CONSERVATOIRE NATIONAL DES ARTS ET METIERS

CENTRE REGIONAL ASSOCIE DE STRASBOURG

ORAL PROBATOIRE

présenté en vue d'obtenir

Le DIPLOME D'INGENIEUR C.N.A.M.

en

INFORMATIQUE

par

PFOTZER Gilles

Méthodes d'authentification vocale d'utilisateurs dans les systèmes informatiques

Soutenu le 27 mai 2000

JURY

PRESIDENT : C. Kaiser

MEMBRES : J. Korczak, A. Napoli, S. Metzger, E. Maurice, J.L. Steffan

CONSERVATOIRE NATIONAL DES ARTS ET MÉTIERS

CONCEPTION DÉVELOPPEMENT INFORMATIQUE

ORAL PROBATOIRE
SESSION DE MAI 2000

Sujet : *Méthodes d'authentification vocale d'utilisateurs dans les systèmes informatiques*

Aujourd'hui, l'authentification d'utilisateurs par les caractéristiques biométriques est de plus en plus appliquée dans les systèmes de haute sécurité. Le but du présent travail est de présenter et d'élaborer une taxonomie des méthodes d'authentification par la voix (séries de Fourier, chaînes de Markov, réseaux neuronaux ...). Il faut également présenter quelques systèmes qui fonctionnent dans un environnement PC/Windows.

Bibliographie :

- **Andrews, R. Diederich, J. and Tickle, A.B.** : *A survey of techniques for extracting rules from trained artificial neuronal network* ; Knowledge-Based Systems, 1995
  (ftp : fit.edu.au::pub:NRC:ps:OUTNRC-95-01-02.ps.Z);

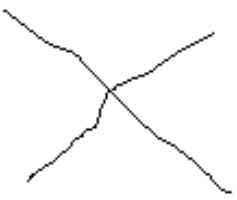
Responsable du sujet : M. Jerzy KORCZAK, Professeur des Universités, Université L. Pasteur
tél. : 03 88 65 55 37 e-mail : jjk@dpt-info.u-strasbg.fr

Plan du probatoire à remettre au plus tard le : *27 avril* (un exemplaire est à remettre au responsable du sujet et un au responsable de filière)

Les rapports sont à remettre au secrétariat au plus tard le : *mardi 16 mai avant 16h*
(prévoyez environ 6 exemplaires, le nombre exact vous sera précisé ultérieurement)

Je soussigné, **Gilles PFOTZER**, reconnaît avoir pris connaissance du présent sujet et reçu en prêt les ouvrages marqués ci-dessus par (*). Ces ouvrages étant à rendre au plus tard le jour de la soutenance.

Fait à Illkirch, le 8 avril 2000



CENTRE RÉGIONAL ASSOCIÉ STRASBOURG

72, ROUTE DU RHIN
67400 ILLKIRCH
GRAFFENSTADEN

TÉLÉPHONE
03 88 67 63 46
TÉLÉCOPIE
03 88 67 63 45

P.S. : la date et l'heure de la soutenance orale ainsi que la composition du jury, vous seront communiquées ultérieurement.

Table de matières

Introduction

La croissance internationale des communications, tant en volume qu'en diversité (déplacements physiques, transactions financières, accès aux services...), implique le besoin de s'assurer de l'identité des individus. En effet, l'importance des enjeux peut motiver les fraudeurs à mettre en échec les systèmes de sécurité existants.

Il existe donc un intérêt grandissant pour les systèmes électroniques d'identification et de reconnaissance. Leur dénominateur commun est le besoin d'un moyen simple, pratique, fiable et peu onéreux de vérifier l'identité d'une personne sans l'assistance d'un tiers. Le marché du contrôle d'accès s'est ouvert avec la prolifération de systèmes, mais aucun ne se révèle efficace contre la fraude, car tous utilisent un identifiant externe tel que: badge/carte, clé, code, ….

Il est fréquent d'oublier un code d'accès. Il existe d'ailleurs de nombreux bureaux où les mots de passe sont notés dans des listes, ce qui représente une dangereuse faille dans la sécurité informatique de l'entreprise puisque toute confidentialité est alors perdue [Vnunet, 2000]. De même, un badge ou une clé peuvent être, volés ou copiés par des personnes mal intentionnées.

Dans le domaine de l'informatique par exemple, le CLUSIF (Club de la Sécurité Informatique) évalue à 7,8 millions de francs les pertes annuelles engendrées par les fraudes, les malveillances et les copies illicites de fichiers informatiques [Biométrie Online].

Le défaut commun à tous les systèmes d'authentification est que l'on identifie un objet (code, carte...) et non la personne elle-même.

Face à la contrainte de l'authentification par «objets », la biométrie apporte simplicité et confort aux utilisateurs.

Cette discipline s'intéresse en effet à l'analyse du comportement ainsi qu'à l'analyse de la morphologie humaine et étudie, par des méthodes mathématiques (statistiques, probabilités), les variations biologiques des personnes.

Les moyens biométriques permettent donc une authentification sûre car ils sont basés sur l'individu lui-même.

Il est alors indispensable de caractériser l'individu par une empreinte afin de le différencier des autres sans aucune ambiguïté ; cette empreinte est une clé codant l'identité d'une personne sans redondance ni variabilité. La plupart des indices biométriques, comme les empreintes digitales ou génétiques, répondent à ces critères [Besacier, 98].

Il en est différemment pour la voix dont la disposition à varier est inscrite dans sa nature même [Rossi, 89]. Si nous ne pouvons pas vraiment parler d'empreinte vocale, la variabilité d'une personne à une autre démontre tout de même les différences du signal de parole en fonction du locuteur.

Cette variabilité, utile pour différencier les locuteurs, est également mélangée à d'autres types de variabilité - variabilité due au contenu linguistique, variabilité intra-locuteur (qui fait que la voix dépend aussi de l'état physique et émotionnel d'un individu), variabilité due aux conditions d'enregistrement du signal de parole (bruit ambiant, microphone utilisé, lignes de transmission) – qui peuvent rendre l'identification du locuteur plus difficile.

Malgré toutes ces difficultés apparentes, la voix reste un moyen biométrique intéressant à exploiter car pratique et disponible via le réseau téléphonique, contrairement à ses concurrents.

Nous vous présenterons donc dans un premier temps les différents moyens d'authentification biométrique, leurs marchés, leurs domaines d'utilisation, leurs principes de fonctionnement ainsi que leurs performances.

Nous étudierons ensuite de façon plus approfondie l'un des moyens d'authentification biométrique : la « voix », en la comparant aux autres moyens puis en présentant les différents niveaux d'authentification par la voix ainsi que leurs performances.

Puis nous aborderons la structure d'un système d'authentification par la voix, les diverses méthodes qui y sont attachées en établissant une taxonomie des méthodes utilisées.

Seront ensuite présentés des produits et prototypes de recherche en la matière, une brève description des sociétés qui les ont établis, le marché et les applications visés, ainsi que leurs caractéristiques. Quelques tests de produits sélectionnés vous seront alors exposés.

Nous conclurons enfin par les problèmes et les limites des systèmes actuels en suggérant quelques solutions.

1. La biométrie

1.1. Généralités

Le mot Anglais « **Biometric** », utilisé pour définir « La mesure des éléments morphologiques des humains », est fréquemment traduit en français par « **Biométrie** ».

La définition de « Biométrie » donnée par le Petit Robert est une « science qui étudie à l'aide de mathématiques (statistiques, probabilités) les variations biologiques à l'intérieur d'un groupe déterminé ».

La biométrie est donc une discipline qui s'intéresse à la mesure de caractéristiques physiques d'êtres vivants et à leur traitement statistique. Lorsqu'il est question de mesurer des organismes humains, le terme anthropométrie est également utilisé.

Les termes "biométrie" et "biométrique" se rapportent donc à des dispositifs destinés à reconnaître des êtres humains à partir de mesures effectuées automatiquement. L'authentification peut concerner le visage, la forme de la main, les empreintes digitales, l'iris, la rétine, la voix, …. Les possibilités sont illimitées.

Il existe 2 catégories de technologies biométriques:

- d'une part les techniques d'analyse du comportement:
  - la dynamique de la signature (la vitesse de déplacement du stylo, les accélérations, la pression exercée, l'inclinaison),
  - la façon d'utiliser un clavier d'ordinateur (la pression exercée, la vitesse de frappe).
- et d'autre part les techniques d'analyse de la morphologie humaine (empreintes digitales, forme de la main, traits du visage, dessin du réseau veineux de l'œil, la voix). Ces éléments ont l'avantage d'être stables dans la vie d'un individu et ne subissent pas autant les effets du stress par exemple, que l'on retrouve dans l'identification comportementale.

Le cadre juridique

Les systèmes biométriques permettent un traitement d'informations nominatives; leur mise en oeuvre est soumise, en France, à la loi n°78-17 du 6 janvier 1978, relative à l'informatique, aux fichiers et aux libertés. Cette mise en oeuvre sur le territoire français est soumise à l'autorisation de la commission nationale de l'informatique et des libertés (CNIL), ce qui garantit au public qu'il n'y a pas atteinte à la vie privée, ou aux libertés individuelles ou publiques.

La biométrie et l'informatique

Dans le passé, le traitement automatique (informatisé) de l'authentification d'empreintes digitales nécessitait l'utilisation d'importants moyens matériels de traitement. Le coût d'élaboration d'un tel système en cantonnait l'usage à des applications spécifiques et à des organismes très motivés qui y mettaient les moyens (système judiciaire, fichier national d'identité, contrôle d'accès haute sécurité).

A présent, les microprocesseurs possèdent la puissance nécessaire à un traitement de ce type et leur coût ne cesse de décroître.

1.2. Le marché de la biométrie

Les ventes de logiciels de sécurité qui représentaient 3,1 milliards de dollars en 1998 ont progressé de 67% de 1996 à 1997 et de plus de 55% en 1998. Le cabinet IDC estime que la croissance devrait se maintenir aux environs de 40% jusqu'en 2002, pour des ventes totales de 7,4 milliards de dollars. Pour le marché français de la sécurité des systèmes d'informations, IDC prévoit un triplement du chiffre d'affaires entre 1998 et 2002, passant de 1,180 milliards de francs à 3,350 milliards de francs.

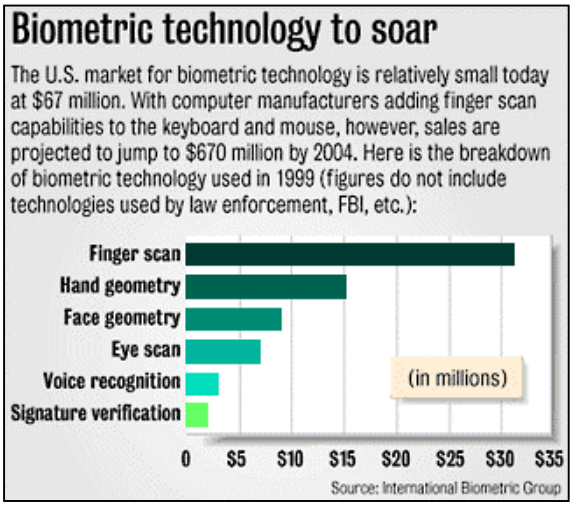


Figure 1 – [Biométrie Online, page Marche]

Aux Etats-Unis, le marché de l'authentification biométrique double chaque année, passant de 25 millions de dollars de chiffre d'affaires en 1997 à 50 millions en 1998, pour atteindre les 100 millions en 1999 [Biométrie Online].

Ces chiffres restent modestes, mais plusieurs facteurs devraient participer au développement prochain de la biométrie, y compris en France.

Le marché de la sécurité informatique est encore atomisé, peu de fournisseurs peuvent prétendre offrir une gamme complète de produits. Toutefois, les spécialistes estiment que ce marché est en pleine croissance et qu'il va également se concentrer.

En effet, Internet et le commerce électronique sont des marchés porteurs pour la sécurité, mais ils ne sont pas les seuls. Le télétravail, la mise à dispositions d'informations aux clients et sous-traitants sont également des facteurs de risque pour les entreprises qui ouvrent leur système d'informations.

Le marché des produits d'authentification individuelle par l'approche biométrique est donc en forte croissance et la technologie dominante actuellement serait celle employant les empreintes digitales. La raison est simple: on a accepté depuis longtemps le fait que les empreintes digitales soient uniques pour un individu donné, et ce fait est supporté par des analyses de probabilité qui affirment que la probabilité théorique de retrouver deux configurations similaires sur les empreintes digitales de deux individus est de l'ordre de 10^{-20} (équivalant à une chance sur des milliards de milliards) [Biométrie Online].

Le GARTNER GROUP a publié en 1998 une liste des dix technologies clés à suivre. Le résultat montre que parmi les systèmes biométriques, les empreintes digitales, l'œil et la forme du visage arrivent en tête de liste ; vient ensuite l'authentification vocale.

Le marché en France

Du coté des utilisateurs ou clients potentiels, il n'y a plus de réticence et la demande est en forte croissance.

Les sollicitations les plus fréquentes à ce jour concernent le remplacement du mot de passe par la biométrie à l'ouverture d'un logiciel ainsi que le contrôle d'accès aux locaux.

Des produits de ce type en provenance des USA existent déjà sur le marché, mais les utilisateurs préféreraient un produit européen pour être certain d'obtenir facilement et rapidement un soutien technique et plus particulièrement si l'utilisateur souhaite intégrer cette technologie au sein de sa propre application.

Les industriels compétents existent en Europe. Ils sont souvent prêts pour la production. En revanche, pour commercialiser ces produits à des prix acceptables sur le marché, il faut que le volume de production soit important dès le début de la commercialisation.

C'est là que se situe le problème européen pour ne pas dire français, il manque tout simplement l'engagement d'investisseurs pour soutenir ces entreprises dans la pénétration de ce marché qui émerge. Un marché qui n'est même plus à risque puisque la demande existe. Bien entendu, cette difficulté n'existe pas aux USA.

1.3. Les domaines d'utilisation de la biométrie

La liste des applications pouvant utiliser la biométrie pour contrôler un accès (physique ou logique) peut être très longue. La taille de cette liste n'est limitée que par l'imagination de chacun dans son domaine d'activité [Biométrie Online].

- contrôle d'accès aux locaux,
  - salle informatique,
  - site sensible (service de recherche, site nucléaire).
- systèmes d'informations,
  - lancement du système d'exploitation,
  - accès au réseau,
  - commerce électronique,
  - transaction (financière pour les banques, de données entre entreprises),
  - tous les logiciels utilisant un mot de passe.
- équipements de communication,
  - terminaux d'accès à Internet,
  - téléphones portables.
- machines et équipements divers,
  - coffre-fort avec serrure électronique,
  - distributeur automatique de billets,
  - casier sensible (club de tir, police),
  - cantine d'entreprise (pour éviter l'utilisation d'un badge par une personne extérieure),
  - casier de piscine (plus d'objet à porter sur soi),
  - contrôle des adhérents dans un club, carte de fidélité,
  - contrôle des temps de présence,
  - voiture (anti-démarrage).
- état / administration,
  - fichier judiciaire,
  - titres d'identité (carte nationale d'identité, passeport, permis de conduire),
  - services sociaux (sécurisation des règlements),
  - services municipaux (sécurisation des accès aux écoles),
  - système de vote électronique.

1.4. Principe de fonctionnement et performance

On peut décomposer le fonctionnement de l'authentification biométrique de la façon suivante : (*figure 2*)

1. capture de l'information à analyser,
2. traitement de l'information et création d'un fichier "*signature*", puis mise en mémoire de ce fichier de référence sur un support (disque dur, carte à puce, code à barres),
3. phase de vérification, l'on procède ici comme pour la création du fichier "*signature*" de référence, puis l'on compare les deux fichiers pour déterminer leur taux de similitude et prendre la décision qui s'impose.

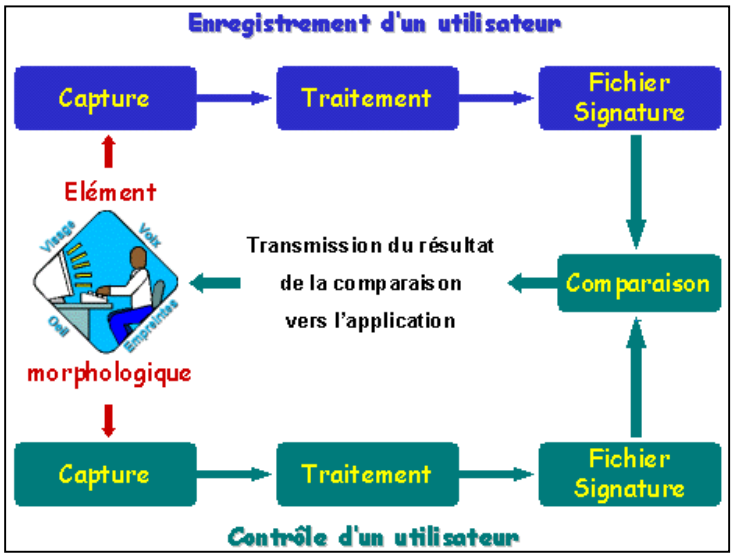


Figure 2 – [Biométrie online, page Techno.htm]

Les moyens biométriques

Voici quelques moyens actuellement à l'étude:

- o les empreintes digitales,
- o la forme de la main,
- o le visage,
- o **la voix,**
- o l'iris,
- o la rétine,
- o la signature dynamique,
- o la thermographie,
- o le code génétique (ADN).

Les performances biométriques

Il est impossible d'obtenir une coïncidence absolue (100% de similitude) entre le fichier "signature" créé lors de l'inscription et le fichier "signature" créé lors de la vérification. Les éléments d'origine (une image, un son...) utilisés pour les traitements informatiques ne pouvant jamais être reproduits à l'identique.

Plutôt que de comparer les performances de ces systèmes, il faut surtout tenir compte de l'environnement de leur usage. Chaque technologie possédant des avantages et des inconvénients, acceptables ou inacceptables suivant les applications. Ces solutions ne sont pas concurrentes, elles n'offrent ni les mêmes niveaux de sécurité ni les mêmes facilités d'emploi.

En ce qui concerne « la voix », les avantages relevés sont les suivants:

- c’est, à ce jour, le seul moyen pour authentifier un interlocuteur via une liaison téléphonique,
- par rapport aux autres technologies, il est plus facile de protéger le récepteur ; on pourra en effet aisément protéger un micro derrière une grille anti-vandalisme.

Mais « la voix », possède aussi des inconvénients:

- elles est sensible à l’étatphysique de l’individu,
- elle est sensible aux bruits ambiants,
- il sera possible de frauder par un enregistrement,
- les taux de faux rejets et fausses acceptations sont élevés.

1.5. L’authentification par « la voix »

Le domaine de recherche concernant l’authentification vocale d’utilisateurs dans les systèmes informatiques est l’Authentification Automatique du Locuteur (AAL). Ce domaine pluridisciplinaire rassemble phonéticiens (production et perception de la parole) et ingénieurs (traitement du signal, informatique, théorie de la décision) [Besacier, 98].

L'identification de la voix est considérée par les utilisateurs comme une des formes les plus normales de la technologie biométrique, car elle n'est pas intrusive et n'exige aucun contact physique avec le récepteur du système.

Les systèmes d'identification de la voix se concentrent sur les seules caractéristiques de la voix qui sont uniques à la configuration de la parole d'un individu. Ces configurations de la parole sont constituées par une combinaison de facteurs comportementaux et physiologiques.

Les sons se caractérisent par une fréquence, par une intensité et par une tonalité. Le traitement informatique tient compte des distorsions liées au matériel utilisé, et sait analyser un son de mauvaise qualité tel qu'une transmission téléphonique ou radiophonique.

Chaque personne possède donc une voix propre que l'on peut analyser à l’aide d’un micro.

La plupart des systèmes d'identification de la voix utilisent l'affichage d'un texte, des mots spécifiques doivent être lus puis répétés afin de vérifier que la personne à authentifier est bien présente et qu'il ne s'agit pas d'un enregistrement.

Les imitateurs essayent habituellement de reproduire les caractéristiques vocales qui sont les plus évidentes au système auditif humain et ne recréent pas les caractéristiques moins accessibles. Il n'est donc pas possible d'imiter la voix d'une personne inscrite dans une base de données [Besacier, 98].

Toutefois, la fatigue, le stress ou un rhume peuvent provoquer des variations de la voix et générer des perturbations.

La fraude est également possible en enregistrant, à son insu, la voix d'une personne autorisée.

La prochaine partie de notre étude vous présentera les différents niveaux d’authentification par la voix dans un certain nombre de contextes (application, identification, vérification, dépendance du texte) puis les critères d’évaluation de leurs performances.

2. L'Authentification Automatique du Locuteur (AAL)

Il s'agit de reconnaître automatiquement l'identité d'une personne prononçant une ou plusieurs phrases, comme un auditeur humain identifie son interlocuteur au cours d'une conversation. Nous distinguerons :

1. les applications «sur site » : serrures vocales pour contrôle d'accès, cabines bancaires en libre service,
2. les applications liées aux télécommunications: ces applications concernent l'identification du locuteur à travers le réseau téléphonique pour accéder à un service de transactions bancaires à distance ou pour interroger des bases de données en accès privé,
3. les applications judiciaires: recherche de suspects, orientations d'enquêtes, preuves lors d'un jugement [Hollien, 90] [Künzel, 94].

La difficulté de la tâche d'authentification n'est pas la même d'une application à une autre. Dans le cas des applications «sur site », l'environnement de prononciation de la phrase ou du mot de passe est plus facilement contrôlé que dans le cas des applications via le réseau téléphonique (distorsions dues au canal, différences entre les combinés téléphoniques, bande passante limitée). Les applications judiciaires présentent quant à elles des difficultés d'un autre ordre (locuteurs non-coopératifs, enregistrements de mauvaise qualité).

2.1. Terminologie

On distingue deux tâches différentes en Authentification Automatique du Locuteur (AAL): l'identification du locuteur et la vérification du locuteur [Atal, 76] [Doddington, 85] [O'Shaughnessy, 86] [Furui, 94] [Eagles, 95].

L'identification du locuteur consiste à reconnaître ce locuteur parmi une population (ou base) composée de N locuteurs connus. L'entrée du système est l'échantillon de parole d'un locuteur inconnu. La sortie du système correspond à l'identité du locuteur de la base de référence qui est la plus "proche" du signal de parole inconnu. Dans cette tâche, on fait l'hypothèse que le signal de parole à identifier est prononcé par l'un des locuteurs de la base de référence (identification en ensemble fermé).

La vérification du locuteur [Naik, 90] [Naik, 94b] [Rosenberg, 76] consiste à déterminer si un locuteur est bien celui qu'il prétend être. Le système dispose en entrée d'un échantillon de parole et d'une identité proclamée. Une mesure de ressemblance est calculée entre l'échantillon et la référence du locuteur correspondant à l'identité proclamée. Si cette mesure est en dessous d'un certain seuil, le système accepte le locuteur ; dans le cas contraire, le locuteur est considéré comme un imposteur et rejeté.

Il est à noter que pour une identification en ensemble ouvert, la combinaison des deux tâches précédentes est nécessaire :

1. identification du locuteur le plus probable parmi les locuteurs de la base,
2. puis vérification que l'échantillon inconnu a bien été prononcé par le locuteur choisi dans l'étape d'identification.

2.2. Dépendance et Indépendance au texte

La distinction est faite entre les systèmes dépendants et indépendants du texte. En mode dépendant du texte, le texte prononcé par le locuteur (pour être reconnu du système) est le même que celui qu'il a prononcé lors de l'apprentissage de sa voix. En mode indépendant du texte, le locuteur peut prononcer n'importe quelle phrase pour être reconnu.

Néanmoins, il existe plusieurs niveaux de dépendance au texte suivant les applications (listés selon le degré croissant de dépendance au texte) [Bimbot, 93] [Bimbot, 94] :

- systèmes à texte libre (ou free-text) : le locuteur prononce ce qu'il veut,
- systèmes à texte suggéré (ou text-prompted) : un texte, différent à chaque session et pour chaque personne, est imposé au locuteur et affiché à l'écran par la machine,
- systèmes dépendants de traits phonétiques (ou speech event dependent) : certains traits phonétiques spécifiques sont imposés dans le texte que le locuteur doit prononcer,
- systèmes dépendants du vocabulaire (ou vocabulary dependent) : le locuteur prononce une séquence de mots issus d'un vocabulaire limité (ex.: séquence de digits),
- systèmes personnalisés dépendants du texte (ou user-specific text dependent) : chaque locuteur a son propre mot de passe.

Les systèmes dépendants du texte donnent généralement de meilleures performances d'authentification que les systèmes indépendants du texte car la variabilité due au contenu linguistique de la phrase prononcée est alors neutralisée.

2.3. Evaluation des performances en AAL

Les performances d'identification du locuteur en ensemble fermé sont données par le taux d'erreur d'identification (pourcentage des cas où le système ne reconnaît pas le bon locuteur).

Dans le cas d'un système de vérification du locuteur, on distingue le taux de fausse acceptation (pourcentage des cas où le système accepte le locuteur alors que celui-ci n'est pas la personne qu'il prétend être) ; et le taux de faux rejet (situation où le système rejette le locuteur alors qu'il est vraiment la personne qu'il prétend être).

L'évaluation des performances d'un système d'AAL n'est cependant pas un problème commun et on ne peut comparer deux systèmes à partir de ces seuls taux d'erreur qui dépendent de multiples facteurs. Ainsi, les éléments suivants doivent également être pris en compte :

- qualité de la parole : enregistrements en studio ou via le canal téléphonique ; environnement calme ou bruyant ; type de réseau téléphonique,

- quantité de parole : durée de parole pour l'apprentissage des références de chaque locuteur ; durée de parole des sessions de test,

- variabilité intra-locuteur : la voix d'un locuteur dépend de son état physique et émotionnel ; de plus, le comportement d'un locuteur se modifie lorsque celui-ci s'habitue à un système,

- population de la base de locuteurs : en identification du locuteur, la taille de la population a une influence directe sur les performances ; la qualité de la population (proportion hommes/femmes, bonne répartition géographique des locuteurs parlant une même langue) est égalementun facteur à intégrer,

- intention des locuteurs : la distinction est faite entre les locuteurs coopératifs (qui veulent être reconnus par le système) et les locuteurs non-coopératifs qui modifient leur voix pour ne pas être reconnus (cas de certaines applications judiciaires par exemple). Enfin, certains locuteurs imitent la voix d'une autre personne pour être reconnus à sa place : ce sont des imposteurs. A ce propos, lors de l'évaluation d'un système, les imposteurs sont en général d'autres locuteurs de la base de référence ce qui n'est pas très réaliste. En effet, en pratique, un imposteur réel qui tentera d'imiter la voix du locuteur pour lequel il voudra être reconnu, n'existera pas forcément dans la base de référence.

Les problèmes d'évaluation sont largement discutés dans le cadre du projet européen EAGLES [Chollet, 97] qui a pour but d'uniformiser les procédures d'évaluation. Des campagnes d'évaluation en AAL ont également été lancées (campagnes NIST (National Institute of Standards and Technology)) pour comparer les performances des systèmes sur une même base de données ("bench-mark programmes") et dans des conditions identiques pour tous. On trouvera aussi un bon exemple sur le problème de l'évaluation des performances dans [Oglesby, 95].

3. Systèmes d'AAL

Dans cette section, sont présentés la structure générale et les différents modules d'un système d'AAL. Une revue critique des méthodes existantes est ensuite proposée en soulignant les atouts et défauts respectifs de chaque méthode. A ce propos, il est à noter que le taux d'erreur d'identification (ou les taux d'acceptation / faux rejet) d'un système d'AAL n'est pas le seul critère de sa qualité. Sont à ajouter :

- la rapidité de l'apprentissage des modèles et de la phase d'authentification,
- la quantité de données nécessaire pour l'apprentissage des modèles de locuteurs,
- la modularité, c'est-à-dire la possibilité d'ajouter ou de supprimer un locuteur de la base sans modifier complètement l'architecture du système [Artières, 95],
- la robustesse aux variations intra-locuteurs ou aux conditions d'enregistrement.

3.1. Structure d'un système d'AAL

La tâche d'authentification automatique du locuteur peut se subdiviser en trois étapes:

- la paramétrisation,
- la classification,
- la décision.

Un premier module de traitement du signal réalise **l'analyse acoustique du signal de parole**. A l'issue de cette étape, le signal est représenté par des vecteurs de coefficients, ce qui permet de réduire l'information en quantité et en redondance. Ces vecteurs sont éventuellement représentés par un modèle mathématique; on parle alors de méthodes paramétriques. Dans la phase de **classification**, les vecteurs du signal de test (ou leur modèle) sont comparés aux vecteurs des locuteurs de référence (ou à leurs modèles). La phase de **décision** désigne le locuteur finalement reconnu.

La structure d'un système d'identification du locuteur en ensemble fermé est représentée sur la *figure 3*.

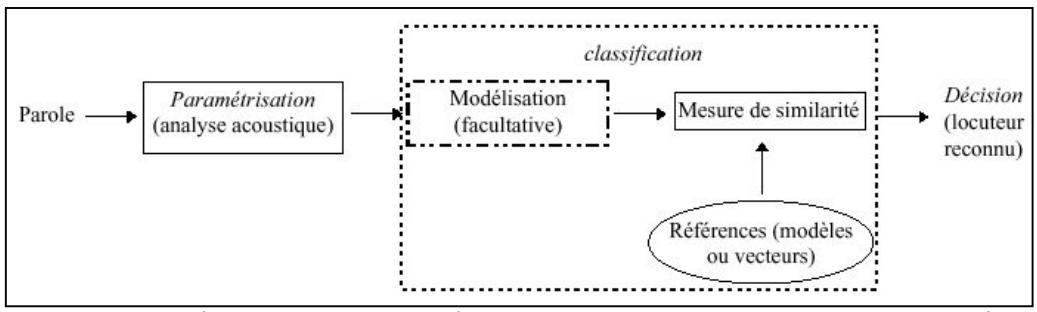


Figure 3 - Schéma modulaire d'un système d'identification du locuteur en ensemble fermé [Besacier, 98, page 9]

Les différents systèmes d'AAL existants se distinguent, d'une part suivant les paramètres qu'ils utilisent, et d'autre part suivant les différents classificateurs qui prennent la décision finale.

3.2. La paramétrisation

Dans son article sur le choix de paramètres efficients pour l'authentification du locuteur, [Wolf, 72] décrit les attributs nécessaires des «bons paramètres» pour l'AAL. Idéalement, les paramètres (ou traits acoustiques) doivent :

- être fréquents,
- être facilement mesurables,
- ne pas être trop sensibles à la variabilité intra-locuteur,
- ne pas être affectés par le bruit ambiant ou les variations dues au canal de transmission,
- être robustes face aux imitateurs.

En pratique, il est très difficile de réunir tous ces atributs en même temps. La sélection de traits acoustiques pertinents pour l'AAL est donc un sujet largement traité: sélection de paramètres séparant les locuteurs en terme de F-ratio (ou ses variantes) [Sambur, 75] [Bonastre, 92] ; sélection par programmation dynamique [Cheung, 78] ; sélection suivant les taux d'identification [Atal, 74]. Finalement, il ressort que les seuls types de paramètres vraiment pertinents et utilisables efficacement sont les paramètres de l'analyse spectrale et éventuellement les paramètres prosodiques. Nous pouvons noter qu'ils sont respectivement corrélés à la forme du conduit vocal et à la source de l'appareil de production de la parole.

3.2.1. Paramètres de l'analyse spectrale

Les principaux paramètres de l'analyse spectrale utilisés en AAL sont les coefficients de prédiction linéaire et leurs différentes transformations (LPC (Linear Predictive Coefficients), LPCC (Linear Predictive Cepstral Coefficients), ...), ainsi que les coefficients issus de l'analyse en banc de filtres et leurs différentes transformations (coefficients banc de filtres, MFCC (Mel Frequency Cepstral Coefficients), ...).

Pour les coefficients de prédiction linéaire, on se référera par exemple aux thèses de [Grenier, 77] et [Homayounpour, 95]. Les articles suivants proposent quand à eux une bonne synthèse sur le choix de paramètres spectraux : [Reynolds, 94a] [Homayounpour, 94] [Ong, 94] [Charlet, 97]. Sans oublier l'utilisation des versions numériques : la TDF (Transformée Discrète de Fourier), un algorithme de calcul rapide : FFT (Fast Fourier Transform). Toutefois, la TDF ne peut opérer sur des séquences trop courtes de signal [Haton, 91].

3.2.2. Paramètres prosodiques

Le terme "paramètres prosodiques" réunit l'énergie, la durée [Van den Heuvel, 94] et la fréquence fondamentale (ou pitch) [Atal, 72]. Ces paramètres s'avèrent cependant fragiles en pratique et ne permettent pas, à eux seuls, de discriminer les locuteurs. En conséquence, ils sont souvent associés aux paramètres de l'analyse spectrale (surtout l'énergie). C'est aussi le cas pour la durée dans [Forsyth, 93] et pour la fréquence fondamentale dans [Matsui, 90] et [Dubreucq, 94].

3.2.3. Paramètres exploitant la dynamique du signal de parole

La prise en compte d'une information de type dynamique peut être un facteur d'amélioration des performances d'identification du locuteur.

3.2.4. Nouvelles paramétrisations

Aujourd'hui, les paramètres utilisés sont pratiquement les mêmes pour la plupart des systèmes d'AAL. Il existe cependant quelques exceptions comme [Thevenaz, 95] et [Hayakawa, 97] qui proposent d'utiliser le résidu de l'analyse par prédiction linéaire, combiné avec les coefficients LPC. [Wenndt, 97] utilise des paramètres issus d'un bispectre (statistiques d'ordre supérieur), plus robustes aux dégradations en milieu bruité. Enfin, AEGIR SYSTEMS qui a participé à la campagne d'évaluation NIST 97 [Nist, 97], utilise des coefficients issus d'une transformée en paquets d'ondelettes. La transformée en ondelettes ainsi que les autres transformées permettant une analyse multi-résolution du signal [Cohen, 95] sont très peu utilisées en traitement de la parole, malgré leur présence dans de nombreux autres domaines. On trouvera cependant quelques références sur le sujet dans [Navarro-Mesa, 92] [Wassner, 96] et [Bernstein, 97].

3.3. La classification

Cette étape consiste à comparer les vecteurs du signal du locuteur testé aux vecteurs des locuteurs de référence de la base de données. Il existe différentes techniques de classification utilisées lors de l'identification du locuteur indépendante du texte; quelques-unes vous sont présentées ci-après :

3.3.1. Méthodes algébriques

- spectres moyens :

[Pruzansky, 63] fut une des premières à utiliser les paramètres du spectre moyen à long terme pour l'AAL. Elle obtint un taux d'identification de 90 % sur une base de dix personnes.

- méthodes statistiques du second ordre :

Des mesures entre matrices de covariance ont été proposées par [Grenier, 77] et [Gish, 90]. Elles sont faciles à implémenter et donnent de très bons résultats [Gish, 94] [Bimbot, 95] avec des durées de test relativement courtes (moins de 3s).

- Modèles Auto Régressif Vectoriels (MARV) :

Ces modèles ont pour vocation de prendre en compte la dynamique du signal de parole. On trouvera notamment l'application des MARV pour l'AAL dans [Montacié, 92a] et [Montacié, 92b]. Une étude sur le choix de l'ordre des modèles (i.e. le nombre de trames utilisées pour la prédiction) est proposée dans [Griffin, 94]. Cependant, un ordre élevé des modèles engendre une complexité de calcul difficile à contrôler.

3.3.2. Méthodes connexionnistes

L'utilisation des réseaux de neurones en AAL est relativement récente [Oglesby, 90] [Bennani, 90]. On trouvera cependant un bon exemple sur le sujet dans [Bennani, 95].

- réseaux de neurones et discrimination:

Les réseaux multicouches (MLP (Multi Layer Perceptron)) utilisés au départ ont rapidement présenté des problèmes lors de l'apprentissage qui devient long et complexe quand le nombre de locuteurs est grand [Rudasi, 91]. Pour éviter ce problème, la tâche de classification est divisée en plusieurs sous-tâches de complexité moindre pour chaque paire de locuteurs. Un apprentissage plus rapide peut également être obtenu en remplaçant les réseaux multicouches par des réseaux RBF (Radial Basis Function) [Oglesby, 91] [Frederickson, 94] [Furlanello, 95]. Les réseaux TDNN (Time Delay Neural Networks) permettent quand à eux de prendre en compte l'information dynamique en réalisant la classification sur des segments de plusieurs trames concaténées [Bennani, 92]. Enfin, l'approche LVQ (Learning Vector Quantization) [Driancourt, 92] [Bennani, 95] est une méthode de type quantification vectorielle avec apprentissage discriminant des vecteurs de référence à l'aide d'un réseau de neurones.

- réseaux de neurones et modélisation:

Un défaut majeur des réseaux de neurones en classification est le problème de modularité [Artières, 95]. En effet, dans le cas d'un apprentissage discriminant, les modèles de tous les locuteurs doivent être re-appris quand une nouvelle personne est ajoutée dans la base. Les modèles prédictifs permettent de modéliser un locuteur indépendamment de tous les autres.

3.3.3. Modélisation multi-classes

- approches par segmentation explicite:

Dans cette approche, le signal de parole segmenté est utilisé pour entraîner des modèles de classes acoustiques dépendants du locuteur. Dans [Bonastre, 94a] et [Bonastre, 94b], un score d'authentification est calculé pour chaque phonème du signal de parole préalablement segmenté, puis ces scores sont combinés afin de prendre une décision finale. [Olsen, 97] propose un système de vérification du locuteur en deux phases: une première phase de Décodage Acoustico-Phonétique (DAP) utilisant des HMM (Hidden Markov Model), puis une phase d'authentification du locuteur basé sur des réseaux RBF dépendants des phonèmes. On trouve également ce type d'approche dans [Savic, 90] et [Matsui, 91] qui obtiennent de bonnes performances avec des durées de test courtes. Il est intéressant de noter qu'avec ces systèmes, les taux d'erreur sont pratiquement les mêmes en mode dépendant ou en mode indépendant du texte.

- approches par segmentation implicite:

Une première possibilité, introduite par [Soong, 85] consiste à regrouper les vecteurs acoustiques en classes. La méthode de quantification vectorielle (VQ (Vector Quantization)) [Soong, 86] est la plus souvent utilisée. L'emploi de la quantification vectorielle en AAL est notamment proposé dans [Matsui, 91] [Matsui, 92] [He, 97]. Une prise en compte de la nature séquentielle des événements phonétiques, associée à la quantification vectorielle, a également été proposée par [Higgins, 86].

La seconde possibilité consiste à utiliser des modèles probabilistes. [Poritz, 82] propose un HMM à 5 états pour classer les vecteurs de paramètres du signal d'un locuteur en 5 catégories correspondant chacune à un état du HMM. [Tishby, 91] propose une extension de ces modèles en décrivant un état comme une combinaison linéaire (mixture) de gaussiennes. Cependant, une expérience de [Matsui, 92] comparant les approches VQ aux HMM en mode indépendant

du texte n'a pas montré une différence de performance significative entre les deux techniques. Ces modèles à base de mixtures de gaussiennes (GMM (Gaussian Mixture Model)) sont désormais largement utilisés en AAL [Reynolds, 94b] [Gish, 94] [Reynolds, 95] [Markov, 96] [Lamel, 97] [Schmidt, 97] et fournissent les meilleurs résultats actuels. Les GMM semblent également être un peu plus robustes quand les environnements d'apprentissage et de tests diffèrent [Van Vuuren, 96].

3.3.4. Un point sur les performances actuelles

L'institut américain NIST organise chaque année une campagne d'évaluation des systèmes d'identification du locuteur. En 1997, la campagne portait sur la tâche de vérification du locuteur indépendante du texte [Nist, 97]. Neuf compétiteurs ont participé à cette campagne : Aegir, BBN, Dragon, ENST, IDIAP, ITT, MIT, OGI et SRI. Le classement final s'est fait sur :

- un apprentissage sur environ 1 minute de parole correspondant à un mélange de 2 conversations enregistrées sur 2 combinés téléphoniques différents,
- une mesure de performances réalisée à partir d'un segment de test de 30 secondes environ.

Les performances sont évaluées séparément pour les portions de test utilisant un combiné téléphonique déjà présent dans la base d'apprentissage et pour les portions de test utilisant un combiné inconnu de la base d'apprentissage. Le score, qui permet le classement final des systèmes, est donné par une fonction de coût égale à la somme pondérée des probabilités de faux rejet et de fausse acceptation. Sur cette évaluation, huit laboratoires ont été classés [Besacier, 98].

Le classement final, ainsi que les méthodes utilisées par les laboratoires sont rassemblés dans le *tableau 1*.

Tableau 1 - Classement final de la campagne d'évaluation NIST 97. – [Besacier, 98, page 18]

Laboratoire	Dragon 2	MIT1	BBN1	Dragon 1	OGI	ITT	IDIAP - ENST	SRI
Classement	1	2	3	4	5	6	7	8
Méthode	GMM	GMM	GMM	LVCSR	GMM	VQ	Hybrid HMM / MLP	GMM / LVCSR

Il en ressort que la méthode de classification GMM est la plus performante.

3.4. La décision

La phase de décision désigne le locuteur finalement reconnu. Le procédé de cette phase dépendra fortement de la phase de classification choisie. Dans cette phase de décision, le locuteur sera accepté, reconnu ou rejeté suivant un seuil de décision, car on ne pourra jamais avoir 100% de similitude entre le signal du locuteur testé et le signal des locuteurs de la base de référence.

4. Analyse de produits commercialisés et de prototypes de recherche

Une certain nombre de sociétés ainsi que leurs produits vous sont présentés ci-après. Vous trouverez une brève description de chaque société, une énumération et description de leurs produits d'AAL, les caractéristiques techniques de chaque produit, ainsi que le marché et les applications visés.

7 sociétés/produits sont présentés ci-après :

4.1. CONFIGATE

E-mail : info@configate.com
Web : http://www.configate.com
Téléphone : +972-8-9316701
Fax : +972-8-9316702
Adresse postale : Configate, Unit 231 2 Professor Bergman St., Rabin Science Park, Rehovot 76705, Israel

CONFIGATE est une compagnie spécialisées en biométrie et plus précisément dans la vérification de la voix sous forme d'authentification pour l'accès à distance sur Internet et la téléphonie. Fondée en 1999, la philosophie de cette société est d'identifier et d'évaluer les caractéristiques naissantes de la vérification de la voix [Configate].

Produit :

CONFIGATE propose un produit appelé « Verimote ». Son utilisation se décompose en 2 phases :

1. l'inscription : l'utilisateur doit exprimer oralement un mot de passe composé de trois mots de son choix. Le profil de la voix de l'utilisateur, appelé un « voiceprint », est enregistré dans une base de données,

2. la vérification : l'utilisateur est amené à réitérer la prononciation de son mot de passe. La vérification se fait immédiatement et donne droit ou non à l'accès.

Marché et Application :

Entre autres :
- les services financiers,
- le commerce électronique,
- la formation en ligne,
- le télétravail,
- la sécurité de l'information,
- le contrôle d'accès physique.

Caractéristiques Techniques :

- longueur du mot de passe : 2 à 4 secondes,
- temps d'inscription : environ 30 secondes,
- taille d'un « voiceprint » : approximativement 10 Ko comprimé,

- temps de vérification: environ 2 secondes (dépend de l'encombrement du réseau),
- système d'exploitation du serveur: Windows NT, Sun Solaris,
- système d'exploitation du client: Windows 95/98, Windows NT/2000, Macintosh.

4.2. BUYTEL Ltd et ITT INDUSTRIES

E-mail :	speakerkey@buytel.com voicekey@buytel.com	E-mail :	speakerkey@itt.com voicekey@itt.com
Web :	http://www.buytel.com http://www.voicekey.com	Web :	http://www.ittind.com http://www.voicekey.com
Téléphone :	+353 (0)1 603 9500	Téléphone :	(219) 451-6321
Fax :	+353 (0)1 478 4478	Fax :	(219) 451-6126
Adresse postale:	BuytelTM (Ire) Limited, Clonmel House, 17 Harcourt Street, Dublin 2, Ireland.	Adresse postale :	ITT Industries, Frank Smead, Attn : SpeakerKey – M/S 668, 1919 W.Cook Road, PO Box 3700, Ft. Wayne, IN 46801 USA

La société ITT est une compagnie de fabrication industrielle. Elle est le principal fournisseur des systèmes militaires sophistiqués de la défense ainsi que des d'éléments électroniques pour les téléphones cellulaires et cartes PC pour les ordinateurs portables. L'associé d'ITT pour les services de « SpeakerKey » est BUYTEL. BUYTEL est l'un des principaux fournisseurs de services téléphoniques.
Les services de « SpeakerKey » fournissent de nouvelles normes d'exécution et de convenance pour l'accès et l'identification d'utilisateurs [Buytel], [ITT and Buytel].

Produits :

- PhoneKey : authentification vocale via le téléphone,
- NetKey : via un réseau local,
- WebKey : via Internet.

Marché et Application :

Entre autres :
- commerce électronique,
- services financiers en ligne,
- téléphone,
- formation en ligne,
- assurance chômage en ligne,
- avantages d'assistance sociale en ligne.

Correspondance

A la suite d'échange d'e-mails (annexes 1 et 2a) avec ces deux sociétés, j'ai constaté qu'il n'était pas possible de connaître précisément les méthodes utilisées pour l'élaboration de leur produit d'AAL. La société ITT m'a tout de même indiqué qu'elle utilisait la méthode HMM ainsi que d'autres méthodes non divulguées. En effet, ces informations sont de propriété industrielle donc strictement confidentielles.

4.3. MOTOROLA

E-mail
P14215@email.mot.com
Web
http://www.motorola.com
Téléphone
+602-441-5009

MOTOROLA est l'un des principaux fournisseurs mondiaux des transmissions sans fil, des semi-conducteurs ainsi que des systèmes, des composants et des services électroniques avancés.

Produit :

Le kit[tm] de logiciel de vérification du locuteur «CipherVox » est le dernier né de la famille de produits de sécurité de l'information de MOTOROLA.

Marché et Application :

C'est un système que les développeurs pourront insérer dans leurs programmes. « CipherVox » se prête à beaucoup de plate-forme et à une variété d'applications.

Caractéristiques Techniques :

- temps d'inscription : environ 1 minute,
- taille d'un « voiceprint » : très petit,
- temps de vérification : moins d'1 seconde.

4.4. OTTAWA TELEPHONY GROUP

E-mail
info@otg.ca
Web
http://www.otg.ca
Téléphone
(613) 745-4441 ou (800) 684-5121
Fax
(613) 745-7473
Adresse Postale
OTG (the Ottawa Telephony Group inc.)
1900 City Park Drive
Suite 204
Gloucester
Ontario K1J 1A3 Canada

Fournisseur international de produits de sécurité, il développe également des technologies d'identité biométrique de la voix [OTG].

Produits :

- « Help Yourself » est une solution prenant en charge une partie du travail du service de maintenance informatique dans une société. La prise en charge porte sur la maintenance des mots de passes, codes PIN (Personal Identification Number), identifiant de profils,

etc. oubliés par les utilisateurs. Par exemple, l'utilisateur en panne pourra obtenir son mot de passe oublié grâce à sa propre voix en se reliant au système «Help Yourself »,

- « SecurPBX » est un produit de téléphonie qui fournit la sécurité d'accès à distanceaux sociétés telle que l'audio-messagerie, les données, etc..

Marché et Application :

- toute entreprise ayant un service d'aide pour le dépannage des mots de passe oubliés ou autres codes pour les systèmes informatiques, pourra réduire les coûts de cette maintenance en utilisant «Help Yourself »,

- vérification du locuteur à distance à travers le réseau téléphonique.

Caractéristiques Techniques :

- configuration : PC, 400 Mhz, 5Go de disque dur, 64 Mo de RAM,
- système d'exploitation : Windows NT,
- taille d'un « voiceprint » : 40 Ko.

4.5. T-NETIX

E-mail : Pat.Flannery@T-NETIX.com
Web : http://www.T-NETIX.com
Téléphone : (303) 705-5525
Fax : (303) 790-9540
Adresse postale : T-NETIX, Inc., 67 Inverness Drive East, Englewood, Colorado 80112, USA

T-NETIX fournit des services de sécurité pour les télécommunications. Elle développe des technologies de pointe d'authentification du locuteur appelées «SpeakEZ » [T-Netix].

Produits :

- VoicEntry I : remplace votre mot de passe écrit de votre économiseur d'écran de Windows 95/NT, par un mot de passe vocal, en utilisant la technologie d'AAL dépendante du texte,

- VoicEntry II : remplace tous les mots de passe rencontrés sous Windows 95/98/NT (ouverture de session, accès au réseau, économiseur d'écran, etc.), par un mot de passe vocal,

- VeriNet WEB : spécifique aux serveurs Web, garantit une vérification de l'utilisateur par contrôle vocal,

- Software Development Kit : kit permettant d'intégrer la technologie d'AAL dans les programmes développés par l'entreprise.

Caractéristiques Techniques :

VoicEntry I et VoicEntry II :

- PC 150 Mhz,
- carte son,
- microphone,
- 5 Mo d'espace disponible sur le disque dur,
- 24 Mo de RAM pour Windows NT et 16 Mo pour Windows 95/98.

VeriNet WEB :

- IIS (Internet Information Server) 4,
- Internet Explorer 4.x ou Navigator 4.x au minimum.

SoftWare Development Kit :

- taille du code pour l'inscription : 90 Ko,
- taille du code pour la vérification : 150 Ko,
- longueur d'un mot de passe : 1 à 2 secondes,
- durée d'inscription : 30 secondes,
- durée d'apprentissage : 5 secondes,
- taille du « VoicePrint » : de 16 à 25 Ko,
- durée de vérification : 0,2 secondes,
- configuration : Pentium 100 Mhz,
- RAM : 32 Mo,
- disque dur : 40 Mo,
- système d'exploitation : Windows NT 3.51, 4.0, Windows 95/98.

Partenaires :

- BioNetrix (http://www.bionetrix.com),
- Envox (http://www.envox.com),
- IBM (http://www.ibm.com),
- Lucent Technologies (http://www.lucent.com),
- Nortel Networks (http://www.nortelnetworks.com),
- OTG (http://www.securpbx.com),
- Periphonics (http://www.peri.com),
- Sentry Systems (http://www.sentry-systems.com),
- Stratus (http://www.stratus.com),
- Visionics Corporation (http://www.faceit.com).

Correspondance

Après un échange d'e-mails (annexes 1 et 2c), T-NETIX m'a fourni un fichier joint illustré de figures décrivant globalement les méthodes utilisées (*figure 4.a*), la robustesse, les phases d'enregistrement (*figure 4.b*) et de vérification. Ce fichier qui est à votre disposition sur mon site Internet à l'adresse http://www.chez.com/gipp/oraux/aal/SpeakEZ_Voice_Print_Overview.doc (255 Ko) m'a apporté un certain nombre d'informations que vous trouverez ci-après :

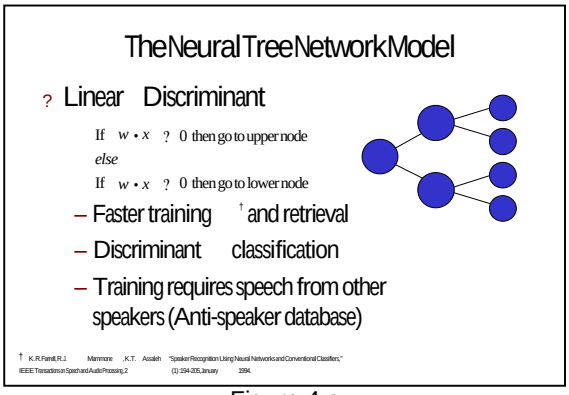


Figure 4.a

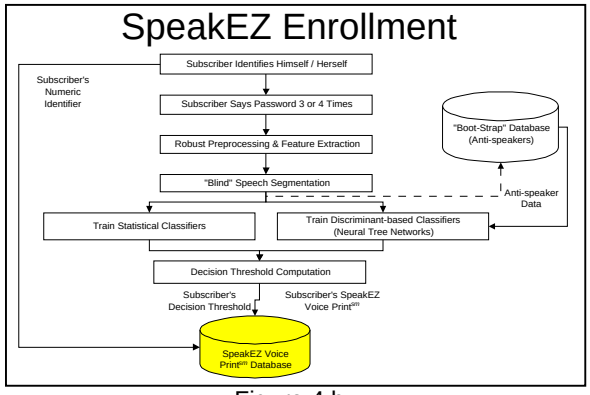


Figure 4.b

SpeakEZ utilise une combinaison de deux classificateurs:

1. l'un basé sur la discrimination (Discriminant-based),
2. l'autre basé sur les statistiques (Statistical-based).

Un réseau de neurone NTN (Neural Tree Network) est un arbre de neurones composé de nœuds (neurones) discriminants. Pendant le procédé de vérification dans le classificateur NTN, chaque nœud doit décider si les caractéristiques acoustiques sont identiques à ceux de la personne proclamée ou identique à ceux des autres personnes enregistrées dans la base de données. La technologie NTN permet une prise de décision rapide(plus rapide que les autres technologies aux dires d'T-NETIX). Des décisions sont prises à chaque nœud de l'arbre et une conclusion est faite après avoir parcouru 5 ou 6 branches de l'arbre.
Ce produit était initialement développé pour l'utilisation lors de combats (guerres) c'est-à-dire dans des conditions d'enregistrement difficile (environnement de bruit, de chaos et des canaux de transmission faibles).

4.6. VERIVOICE

E-mail : info@verivoice.com
Web : http://www.verivoice.com
Téléphone : (609) 452-9220
Adresse postale : VeriVoice, Inc., 5 Vaughn Dr, Princeton, NJ 08540

Fondée en 1995, « VeriVoice » est une société spécialisée dans les solutions biométriques de sécurité pour l'accès aux systèmes informatiques et aux réseaux à distance. La compagnie fournit un moteur de vérification qui peut être intégré dans les applications.

Produit :

- VeriVoice.

Correspondance

A la suite d'échange d'e-mails (annexes 1 et 2c), j'ai pu obtenir une version bêta d'un économiseur d'écran. La seule information qui m'ai été donnée est que les méthodes mentionnées dans mon e-mail (annexe 1), à savoir HMM, FFT, réseau de neurones, ont été utilisées pour la conception de leur produit.

4.7. LSIIT, ULP-CNRS

E-mail : jjk@dpt-info.u-strasbg.fr
Web : http://lsiit.u-strasbg.fr/
Téléphone : 03 88 65 55 00
Fax : 03 88 65 55 01
Adresse postale : LSIIT (Laboratoire des Sciences de l'Image, de l'Information et de la Télédétection), Pôle API, Boulevard Sébastien Brant, 67400 ILLKIRCH GRAFFENSTADEN CEDEX

Le LSIIT (Laboratoire des Sciences de l'Image, de l'Information et de la Télédétection) est une unité mixte de recherche de l'ULP (Université Louis Pasteur) et du CNRS (Centre National de la Recherche Scientifique). C'est un laboratoire interdisciplinaire fédéré par l'imagerie. Les grandes disciplines qui y sont représentées sont l'Informatique, le Traitement du signal, l'Automatique, la Télédétection.

Produit :

- un projet d'authentification par le visage et par la voix est en cours de réalisation.

Correspondance

Ayant la possibilité de recueillir directement des informations auprès du LSIIT, j'ai appris que ce projet, dans sa partie concernant l'authentification par la voix, utilise la méthode de transformation de l'ondelette pour la phase d'analyse (ou paramétrisation) et un réseau de neurones dans la phase d'apprentissage (ou classificateur). C'est le seul projet où les informations et les détails sur les méthodes utilisées m'ont été donnés à profusion (par M. POH N. - stagiaire travaillant sur le projet – e-mail : *normanpoh@i.am*). Ce projet associera deux méthodes biométriques. La première méthode est l'authentification par le visage, sous-projet réalisé en 1999 et présenté dans [Metzger, 99]. La deuxième méthode étant l'authentification par la voix, sous-projet en cours de réalisation (2000) (*figure 5*).

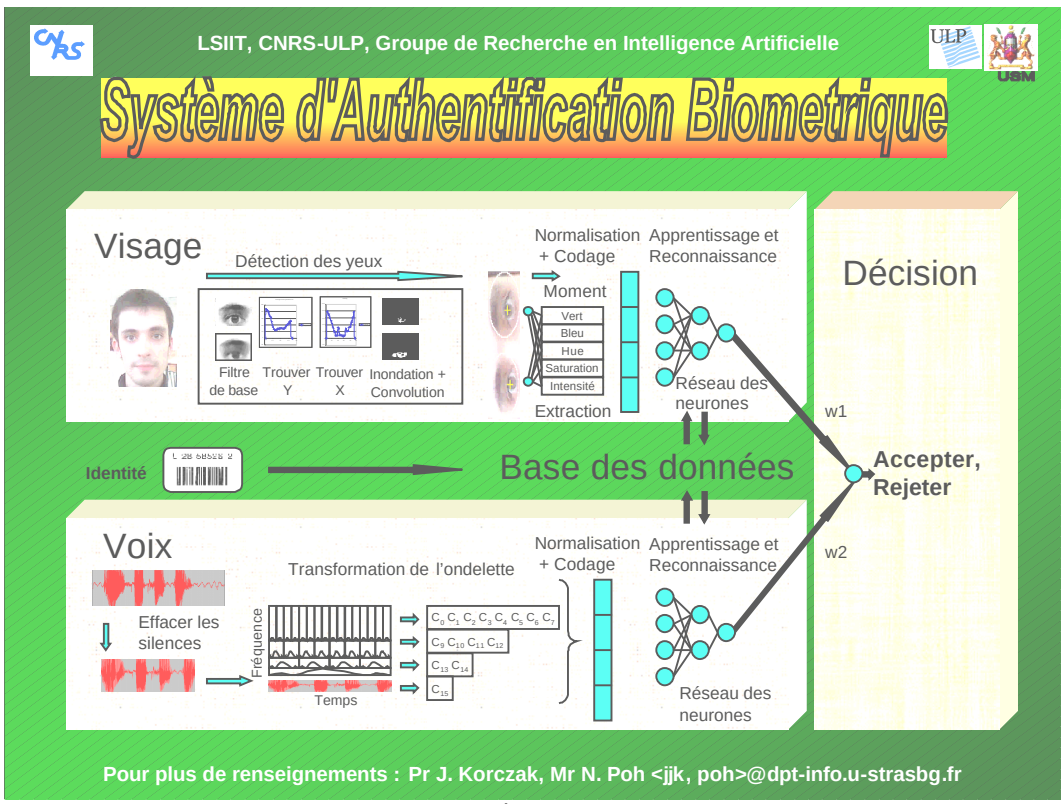


Figure 5

4.8. Tests des produits sélectionnés

SpeakerKey de BUYTEL Ltd et ITT INDUSTRIES :

D'après ces deux sociétés, une démonstration en ligne est proposée par téléphone au 00 1 (880) 775-7515, mais, malheureusement, après plusieurs tentatives, nous avons constaté qu'aucun serveur ou service ne répondait à l'appel.

Une autre démonstration est proposée directement via Internet à l'adresse http://www.buytel.com/WebKey/index.asp ; elle simule un accès en ligne à un compte bancaire. Cette démonstration nous demande au préalable de télécharger et d'installer un « Plug-in » pour son bon fonctionnement.

Il nous est demandé, dans un premier temps, de saisir notre nom (*figure 6.a*) puis ensuite, de répéter oralement 12 paires de nombres aléatoires compris entre 40 et 99 (« 97-46 » par exemple) (*figures 6.b et 6.c*). Le nom est alors associé au modèle d'apprentissage de la voix.

Pour accéder ultérieurement au compte bancaire simulé, il faudra saisir notre nom et répéter oralement 2 à 4 paires de nombres aléatoires. Si le système affirme nous reconnaître, il nous présentera une page simulant notre compte bancaire (*figure 6.d*), le cas échéant, nous serons rejetés.

L'inscription effectuée restera stockée dans la base de données générale pour tout accès ultérieur au compte bancaire. Pour toutes tentatives d'accès par un autre ordinateur, il sera nécessaire de ré-installer le « plug-in » utile à cette démonstration.

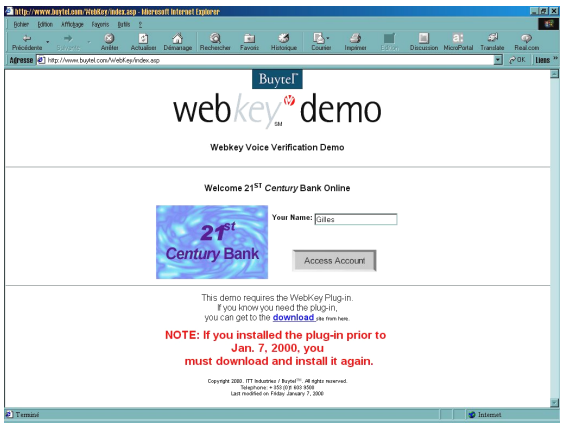


Figure 6.a

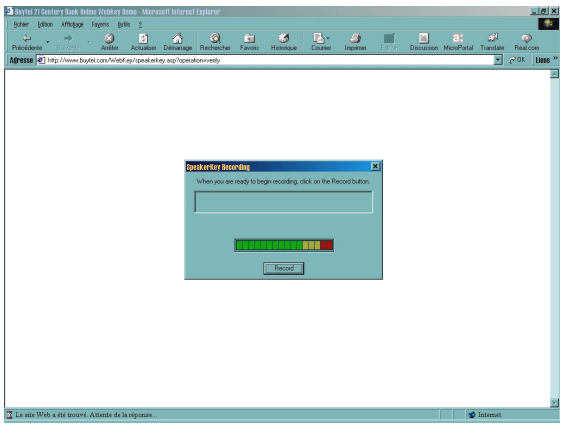


Figure 6.b

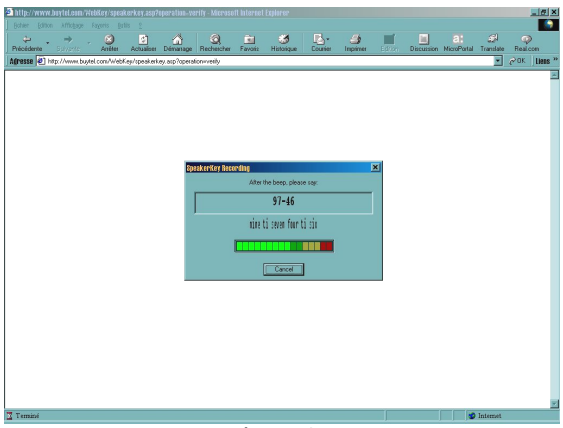


Figure 6.c

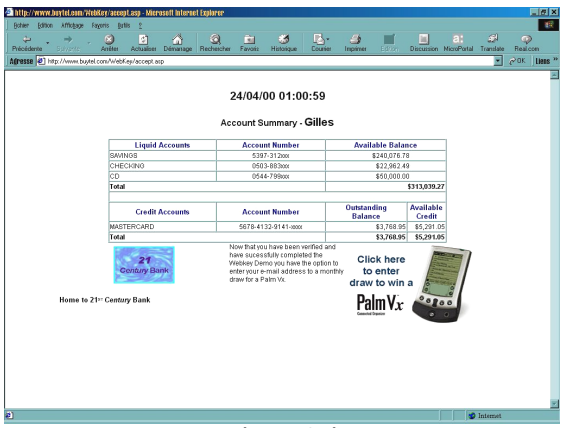


Figure 6.d

Mes propres essais ont été très concluants. En effet, avec l'aide de plusieurs locuteurs, nous avons tenté quelques essais d'intrusions, mais sans réussite. Il me semble que le fait que cette version de démonstration soit dépendante des traits phonétiques, la rende performante, même en déformant volontairement sa voix.

Une autre version d'évaluation est également téléchargeable à l'adresse http://www.voicekey.com/le/html/le_keygen1.html. Mais jusqu'à présent ce téléchargement n'a jamais fonctionné !

VoicEntry de T-NETIX :

Une version de démonstration est téléchargeable à l'adresse Internet http://www.t-netix.net/speakez/download.html ou directement à http://www.t-netix.net/speakez/ve1.exe . Cette version s'installe sous Windows et propose un économiseur d'écran qui est désactivé par l'identification vocale du locuteur.

Après son installation, le programme, peut être paramétré par le biais des «propriétés d'affichages» de Windows ; on y trouvera des réglages possible de sensibilité et delongueur de mot de passe (*figure 7.a*).

Une première étape consiste à enregistrer sa voix, en prononçant 4 fois le même mot de passe de son choix (*figures 7.b et 7.c*). Suit, une étape d'apprentissage de la voix enregistrée qui dure environ 8 secondes (sur un Pentium 350) (*figure 7.d*).

La troisième étape est la vérification. Lorsque l'économiseur d'écran est en fonction, il suffit de déplacer la souris ou de presser une touche du clavier pour être invité à prononcer son mot de passe (*figure 7.e*). Si le locuteur est identifié, l'économiseur cessera.

Une solution de secours est toutefois proposée après plusieurs échec vocaux. En effet, un mot de passe écrit (paramétré dès le début de l'apprentissage) est demandé pour désactiver l'économiseur d'écran.

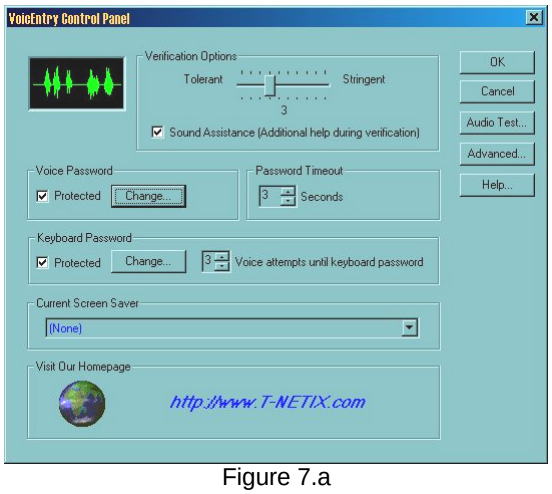


Figure 7.a

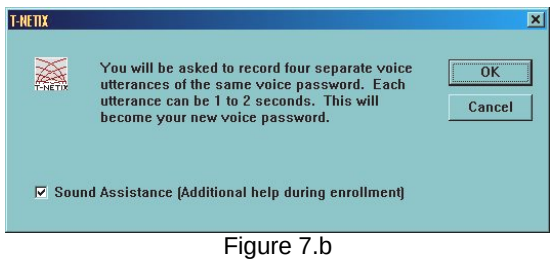


Figure 7.b

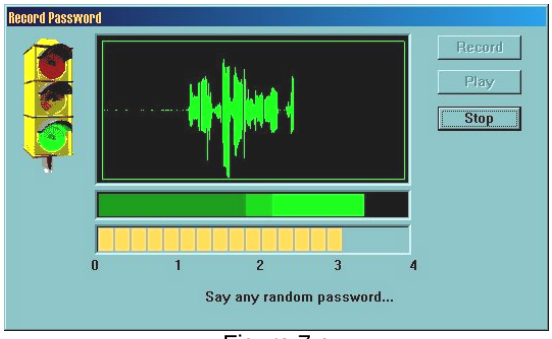


Figure 7.c

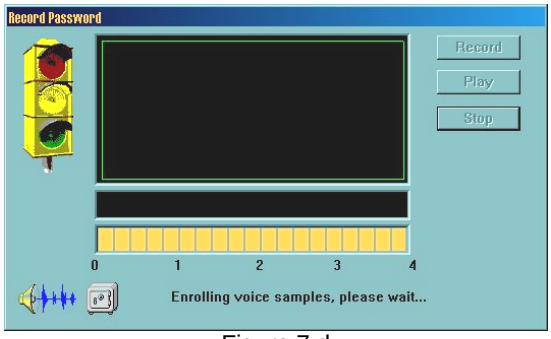


Figure 7.d

Figure 7.e

Après avoir effectué de nombreux tests auprès plusieurs locuteurs, il ressort les points suivants :

- il s'agit d'une AAL dépendante du texte,
- en déformant volontairement sa propre voix, l'identification semble quand même fonctionner,
- en tentant de réaliser quelques intrusions avec environ 5 locuteurs, l'un d'entre eux a tout de même réussi à tromper le programme !

VeriVoice :

Suite à un échange d'e-mails, j'ai pu obtenir une version de démonstration similaire à VoicEntry. Un économiseur d'écran pour Windows, mais dans ce cas précis, possibilité est donnée d'inscrire plusieurs «VoicePrint » en leur donnant des noms (*figure 8.a*). Dans cette version, la phase d'apprentissage (inscription) demande la prononciation de 16 séquences de 5 chiffres (exemple: 13052) (*figures 8.b, 8.c, 8.d, 8.e et 8.f*). Pour la vérification, il suffit de prononcer 1 seule séquence de 5 chiffres pour désactiver l'économiseur d'écran (*figure 8.g*).

Cette version de démonstration peut être téléchargée sur mon site à l'adresse Internet suivante : http://www.chez.com/gipp/oraux/aal/VVSS20Apr00.exe (1 Mo).

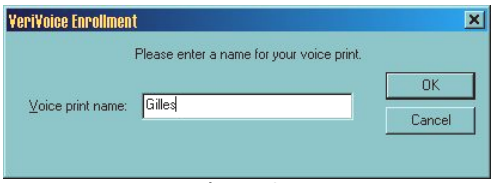


Figure 8.a

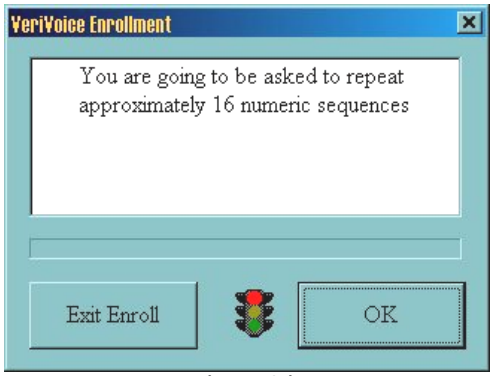


Figure 8.b

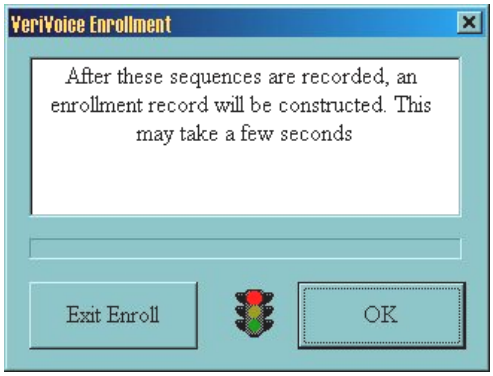


Figure 8.c

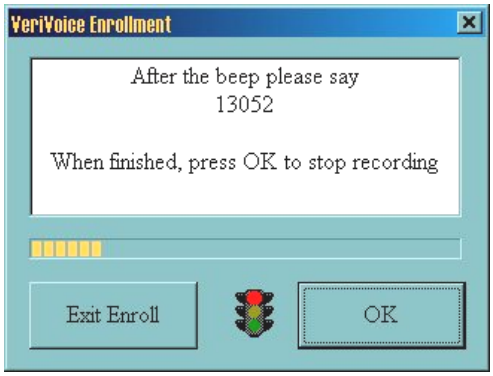


Figure 8.d

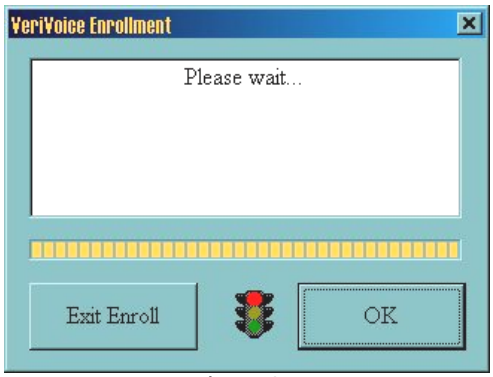


Figure 8.e

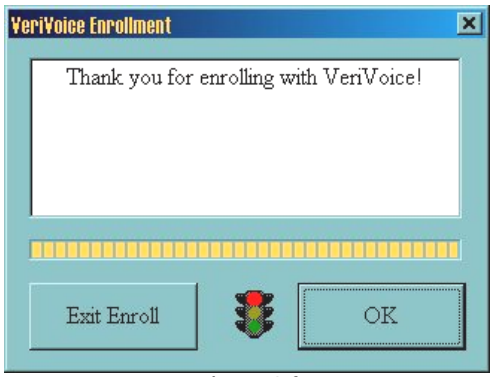


Figure 8.f

Figure 8.g

Après avoir effectué de nombreux tests auprès de plusieurs locuteurs, les points suivantsont été constatés :

- il s'agit d'une AAL dépendante du vocabulaire,
- ce système n'apparaît pas probant car une voix masculine a pu être remplacée par une voix féminine lors des essais,
- les chiffres à prononcer lors de l'apprentissage et lors de la vérification peuvent être énoncés dans n'importe quelle langue,
- en déformant volontairement sa propre voix, l'identification semble quand même fonctionner.

LSIIT, ULP-CNRS :

J'ai eu l'opportunité de voir fonctionner un projet d'authentification par le visage et par la voix réalisé à LSIIT. Le programme résultant de ce projet (que j'ai d'ailleurs pu tester à cette occasion) nous demande de nous enregistrer par « le visage » et par « la voix » (*figure 9.a*). Dans la première phase d'apprentissage, le programme prend un certain nombre de clichés à l'aide d'une caméra et localise immédiatement la position des yeux du visage filmé (*figure 9.b*). Dans la deuxième phase, nous sommes invités àrépéter oralement plusieurs fois un mot de passe d'une longueur maximum de 2 secondes (*figure 9.c*). En ce qui concerne la vérification, il est demandé de répéter ce même mot de passe devant la caméra. Ce programme est donc dépendant du texte. Ses performances ne peuvent malheureusement pas encore être évaluées puisque le projet n'en est qu'à ses débuts.

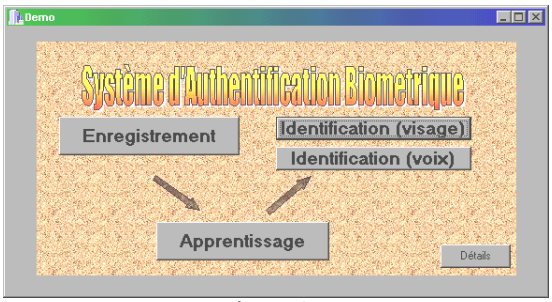


Figure 9.a

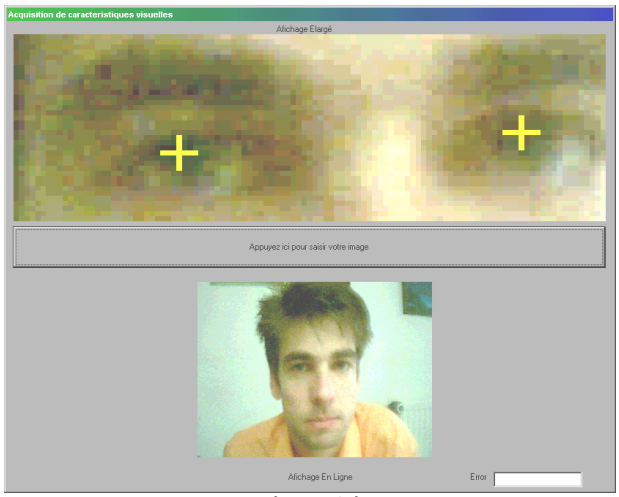


Figure 9.b

Figure 9.c

5. Problèmes et limites des systèmes actuels

Dans un tutorial sur les avancées récentes en authentification du locuteur, [Furui, 97a] propose 16 questions ouvertes concernant les interrogations et les problèmes restés sans solution à ce jour. De nombreux problèmes sont liés à la variabilité: variabilité due au locuteur et variabilité due aux conditions d'enregistrement.

5.1. Variabilité due au locuteur

Une dégradation croissante des performances a été observée au fur et à mesure que le temps qui sépare la session d'apprentissage de la session de test augmente [Furui, 72] [Furui, 74] [Rosenberg, 76]. De plus, le comportement des locuteurs se modifie lorsque ceux-ci s'habituent au système. Les modèles des locuteurs doivent donc être régulièrement mis à jour avec les nouvelles données d'exploitation du système [Setlur, 95]. Les altérations de la voix dues à l'état physique (fatigue, rhume) ou émotionnel (stress) mettent aussi en échec l'efficacité des systèmes [Homayounpour, 94].

5.2. Variabilité due aux conditions d'enregistrement et de transmission

La parole téléphonique est sujette à des dégradations parmi lesquelles on peut citer la limitation de la bande utile et les distorsions dues au combiné ou au canal de transmission [Reynolds, 92].

Une diminution des performances pour de la parole téléphonique est systématiquement observée [Hunt, 83] [Gish, 85] [Gish, 86]. [Reynolds, 94b] observe une dégradation des performances d'identification qui passent de 99.7 % sur TIMIT (Texas Instruments Massachusetts Institute of Technology) à 76.2 % sur NTIMIT (Network TIMIT) pour 168 locuteurs. Plus récemment, [Van Vuuren, 96] a fait le point sur les problèmes dus aux différences entre les environnements téléphoniques. Ainsi, dans le cas où les données d'apprentissage et les données de test ne viennent pas du même environnement téléphonique, la dégradation des performances d'identification du locuteur est très importante. [Reynolds, 96] a montré que la plus grande part de ces dégradations est due aux différences de combinés téléphoniques entre l'apprentissage et le test. Une détection préalable du combiné téléphonique semble donc nécessaire. Ce point précis était d'ailleurs l'un des enjeux essentiels lors de la campagne d'évaluation NIST 97.

5.3. Autres problèmes

Récemment, [Kuitert, 97] a étudié l'effet du codage de la parole utilisé dans le réseau téléphonique mobile GSM sur les performances de vérification du locuteur.

Peu d'articles traitent du problème de la robustesse des systèmes confrontés à de la parole enregistrée dans un environnement bruité. La robustesse au bruit ambiant est pourtant une condition nécessaire au succès des systèmes d'AAL dans des applications en conditions réelles.
Enfin, une autre condition est la robustesse vis à vis des imitateurs occasionnels ou professionnels [Homayounpour, 94]

6. Quelques solutions aux problèmes de robustesse

Nous avons vu dans la section précédente que la plupart des problèmes rencontrés en AAL sont dus à une inégalité entre les conditions d'apprentissage et les conditions de test: variabilité due au locuteur, au canal de transmission ou aux conditions d'enregistrement.
Les méthodes traitant de la réduction des écarts dus aux variations du signal de parole peuvent être regroupées en deux niveaux [Furui, 97b] :

- niveau des paramètres,
- niveau des modèles.

6.1. Paramétrisations robustes

Le problème de la robustesse des paramètres pour l'AAL a notamment été abordé par [Assaleh, 94] [Naik, 94a] et [Reynolds, 94a]. Les paramètres peuvent également être retraités après l'analyse acoustique: égalisation de canal [Furui, 81] [Wang, 93], filtrage RASTA (RelAtive SpecTraAl) [Hermansky, 94] [Hermansky, 97], masquage du bruit par addition d'un offset aux paramètres spectraux [Openshaw, 94]. Le défaut de l'égalisation de canal est qu'elle supprime en même temps une partie de l'information spécifique du locuteur [Furui, 97a].

6.2. Ré-estimation ou adaptation des modèles

Comme la voix des locuteurs évolue au cours du temps, il est nécessaire de mettre à jour les modèles des locuteurs pour éviter leur vieillissement. Pour des raisons pratiques, les modèles doivent être mis à jour en utilisant les données d'exploitation. On peut soit réestimer les modèles des locuteurs en utilisant les données d'apprentissage initiales et les nouvelles données d'exploitation, soit adapter le modèle initial du locuteur avec les données d'exploitation. Cette deuxième alternative ne nécessite aucun stockage des données de sessions précédentes puisque l'adaptation se fait 'en ligne' [Matsui, 96].

L'adaptation des modèles est également nécessaire sur de la parole téléphonique pour prendre en compte les différentes conditions d'appel (combiné, canal, ...). Une première solution consiste à créer le modèle d'un locuteur à partir de différents environnements d'appel [Gauvain, 95]. [Heck, 97] propose quant à lui d'entraîner différents modèles dépendants du combiné téléphonique pour normaliser le score d'un locuteur.

6.3. Modèles parallèles

Récemment, de nouvelles techniques sont apparues en vue d'augmenter la robustesse des systèmes d'authentification: leur caractéristique commune est l'utilisation de plusieursreconnaisseurs (travaillant en parallèle) qui sont recombinés pour prendre une décision finale [Besacier, 98].

Conclusion

Nous avons pu constater que le domaine de la biométrie est une véritable alternative aux mots de passe qui permet de vérifier que l'utilisateur soit bien la personne qu'il prétend être. Cette technologie est en pleine croissance et l'authentification par la voix, ainsi que les autres moyens d'authentification, tendent à s'associer à court terme, aux technologies actuelles comme la carte à puce, le badge, la clé, etc..

La fabrication des produits d'authentification est en pleine augmentation, dû en l'occurrence à la nécessité croissante du besoin de sécurité de chacun (tant dans le domaine privé que dans le domaine professionnel).

Les méthodes actuelles utilisées pour la réalisation de l'apprentissage et de la vérification sont nombreuses et en pleine mutation. Elle sont toutefois la propriété des fabricants et centres de recherche qui travaillent souvent de façon autonome et sans aucune corrélation. Aussi, ne serait-il pas judicieux d'associer toutes les connaissances sur le sujet (quelque soit le pays et le domaine de recherche), afin de pouvoir obtenir des outils probants et fiables?

Pour autant, le coût prohibitif de ces technologies a longtemps freiné leur développement. Aujourd'hui, les entreprises entrevoient les économies qu'elles réaliseraient à long terme en les utilisant (ex : temps perdu par les services informatiques pour retrouver les mots de passe oubliés ou perdus).

Notons que la société MICROSOFT se lance également dans la biométrie puisqu'elle a annoncé son intention d'offrir le support de technologie biométrique aux utilisateurs de Windows 2000 avant de l'incorporer pleinement à la sécurité Windows à l'avenir.

L'industrie biométrique représentera un milliard de dollars en 2000 et elle devrait connaître un boom fantastique dès lors qu'elle aura converti le secteur privé (pour le moment les plus gros utilisateurs sont les prisons, les services de police, etc.); toutefois, il devrait se passer un peu de temps avant que ces technologies soient largement adoptées [Vnunet, 2000].

L'on a également pu constater que l'authentification par la voix est un moyen biométrique difficile à réaliser. Elle semble fonctionner, mais les programmes de démonstration ont été testés avec des locuteurs dont le but premier n'était pas la fraude ! De plus, les produits testés dans ce présent rapport, étaient pour certains en cours de réalisation et pour d'autres des versions de démonstration. Aussi, pour une meilleure évaluation de ces produits, il serait à l'évidence nécessaire d'attendre la version finale pour les tester au mieux.

Enfin, pour qu'un système d'authentification soit robuste et paré à toute épreuve l'on peut penser qu'il serait préférable d'associer simultanément plusieurs méthodes d'authentification biométrique en combinant, par exemple, «la voix » et « les traits du visage », comme l'a d'ailleurs suggéré le LSIIT.

Toutefois, « la voix » reste à ce jour le seul moyen d'authentification par téléphone…

Annexes

Annexe 1 : E-mail envoyé aux adresses suivantes : sksupport@acdca.itt.com ; speakerkey@itt.com ; voicekey@itt.com ; voicekey@buytel.com ; info@configate.com ; info@otg.ca ; P14215@email.mot.com ; SDKSales@T-NETIX.com ; info@verivoice.com ; speakerkey@buytel.com ; Information@T-NETIX.com ; Ray.Reid@T-NETIX.com ; Ron.Beyner@T-NETIX.com ; UK_office@T-NETIX.com

De: Gilles PFOTZER <gpfotzer@ifrance.com>
Objet: Vocal Recognition
Date : samedi 29 avril 2000 10:15

**** ENGLISH ****

Hello,

I am currently student in CNAM in Strasbourg and I prepare my oral examination of which the subject is " Methods of vocal recognition of users in the information processing systems ".

This subject is in direct relation with your products about which I want to speak.

Therefore I would be grateful to you if you could inform me about mathematics methods which you used during the development of your product (HMM, FFT, Network Neural, etc...).

Would it be also possible to forward me a booklet presenting your product ?

By addressing all my thanks for the attention you will carry well to my request,

Sincerely yours

Gilles PFPOTZER
gpfotzer@ifrance.com

**** FRANCAIS ****

Bonjour,

Je suis actuellement étudiant en CNAM à Strasbourg et je prépare mon oral probatoire dont le sujet est « Méthodes d'authentification vocale d'utilisateurs dans les systèmes informatiques ».

Ce sujet est en relation directe avec vos produits dont je souhaiterais parler.

Aussi je vous serais reconnaissant si vous pouviez me renseigner sur la/les méthode(s) mathématique(s) que vous avez utilisées lors de l'élaboration de votre produit (HMM, FFT, Réseau Neuronaux, etc.).

Serait-il également possible de me faire parvenir une brochure présentant votre produit ?

En vous adressant tous mes remerciements pour l'attention que vous voudrez bien porter à ma demande,

Je vous prie de croire à mes salutations les meilleures.

Gilles PFOTZER
gpfotzer@ifrance.com

Annexe 2a : Réponses de BUYTEL et ITT

De: Vance Harris <Vance@buytel.com>
À: 'Gilles PFOTZER' <gpfotzer@ifrance.com>
Objet: RE: Vocal Recognition
Date : samedi 29 avril 2000 13:41

Hi Giles.

The algorithms inherent in the VoiceVault product are proprietary and I am therefore sorry to say that I am unable to provide you with any material.

For more general information I would recommend www.bio1.com <http://www.bio1.com> as a great source of white papers and other material.

Hope the examination goes well.

Best regards

Vance

De: Smead, Frank <Frank.Smead@itt.com>
À: 'Gilles PFOTZER' <gpfotzer@ifrance.com>
Objet: RE: Vocal Recognition
Date : lundi 1 mai 2000 15:18

The only readily available information is what we post on our website: www.speakerkey.com <http://www.speakerkey.com>. We use HMM's in the product, but we do not have anything we could send you on the algorithms. They are proprietary.
Good luck in your orals.

Annexe 2b : Réponses de T-NETIX

De: Ray Reid <Ray.Reid@T-Netix.Com>
À : 'Gilles PFOTZER' <gpfotzer@ifrance.com>
Objet : RE : Vocal Recognition
Date : lundi 1 mai 2000 15 :53

I sent your request to our Denver office. The person is Carol Godfrey,
Marketing Manager, 303-705-5534.

De: Bill Mistretta <Bill.Mistretta@t-netix.com>
À: <gpfotzer@ifrance.com>
Cc: Bill Mistretta <Bill.Mistretta@t-netix.com>
Objet: RE: Vocal Recognition
Date : mercredi 10 mai 2000 19:49

Hello Mr. Pfotzer,
I am sending you a document that provides an overview of our SpeakEZ Voice Print product. This is what we make publicly available to people requesting information about our speaker verification technology. I hope this may be what you are looking for.

Sincerely,

William Mistretta

Annexe 2c : Réponse de VERIVOICE

De : Joseph A. Mannino <jmannino@verivoice.com>

À : ‘Gilles PFOTZER’ <gpfotzer@ifrance.com>
Objet : RE : Vocal Recognition
Date : lundi 1 mai 2000 15 :15

Gilles,

Thank you for your inquiry.

I have taken the liberty of attaching a beta copy (self-extracting file) of a Screensaver incorporating the VeriVoice technology (please be sure to read the installation instructions also attached).
This is a demonstration application, put together to facilitate demonstrating how voice can work.
The user interface can change to meet the application requirements.
It takes about 5 minutes to install and about 2 minutes to enroll a voice on a Windows 95/98 multimedia computer. (this is an English challenge phrase version, the technology can work with French or any other language if desired).

Potential e-commerce applications abound. It can replace or supplement passwords and PINs adding a significantly higher level of security.
It can be implemented client or client sever, work over the internet, private networks, telephones, cell phones, is installable on Smartcards.

Most of our information is on our website http://www.verivoice.

Our product uses several of the mathematical techniques you mention.
What we do however is obviously proprietary to our technology otherwise everyone else would be able to implement it.

I hope this will be helpful. You can help us by giving us your feedback.

Regards,

Joe

Joseph A. Mannino
President & CEO
VeriVoice, Inc.

http://www.verivoice.com

Tel: 609 452 9220
Fax: 609 452 2700

Table des abréviations

AAL	Authentification Automatique du Locuteur
DAP	Décodage Acoustico-Phonétique
GMM	Gaussian Mixture Model
HMM	Hidden Markov Model
LPC	Linear Predictive Coefficients
LPCC	Linear Predictive Cepstral Coefficients
LVCSR	Large Vocabulary Continuous Speech Recognition
LVQ	Learning Vector Quantization (Algorithm)
MARV	Modèle Auto Régressif Vectoriel
MFCC	Mel Frequency Cepstral Coefficients
MLP	Multi Layer Perceptron
NIST	National Institute of Standards and Technology
NTIMIT	(Telephone) Network TIMIT
NTN	Neural Tree Network
PIN	Personal Identification Number
RASTA	RelAtive SpecTrAl (Methodology)
RBF	Radial Basis Function
TDNN	Time Delay Neural Network
TIMIT	Texas Instruments Massachusetts Institute of Technology
VQ	Vector Quantization

Bibliographie

[Artières, 95] Méthodes prédictives neuronales : applications à l'identification du locuteur. Thèse de l'Université de Paris XI Orsay. 1995.

[Assaleh, 94] **ASSALEH K.T. MAMMONE R.J.**, Robust cepstral features for speaker identification. In Proc. ICASSP 94, Adélaide, Australia, pp 129-132. 1994.

[Atal, 72] **ATAL B.S.**, Automatic speaker recognition based on pitch contours. The Journal of the Acoustical Society of America, n° 52, pp 1687-1697. 1972.

[Atal, 74] **ATAL B.**, Effectiveness of linear prediction characteristics of speech wave of automatic speaker identification and verification. JASA, vol. 55, pp 1304-1312. June 1974.

[Atal, 76] **ATAL B.S.**, Automatic recognition of speakers from their voices. Proc. IEEE, n° 64(4), pp 470-475. 1976.

[Bennani, 90] **BENNANI Y., SOULIE F.F., GALLINARI P.**, A connectionist approach for automatic speaker identification. In Proc. ICASSP 90, pp 265-268. April 1990.

[Bennani, 92] **BENNANI Y.**, Speaker Identification through a modular connectionist architecture. Evaluation on the TIMIT database. In Proceedings ICSLP 92, pp 607 -610. Banff (Canada). October 1992.

[Bennani, 95] **BENNANI Y., GALLINARI P.**, Neural networks for discrimination and modelization of speakers. Speech Communication, n° 17(1-2), pp 159-176. August 1995.

[Bernstein, 97] **BERNSTEIN E., EVANS W.**, Wavelet based noise reduction for speech recognition, In ESCA-NATO Workshop on Robust speech recognition for unknown communication channels. Pont-à-Mousson, France, pp 111-114. 17-18 Avril 1997.

[Besacier, 98] **BESACIER Laurent - PhD**, PhD Thesis. Adresse Internet : http://herakles.imag.fr/besacier/, Un modèle Parallèle pour la Reconnaissance Automatique du Locuteur. 1998.

[Bimbot, 93] **BIMBOT F., PAOLONI A., CHOLLET G.**, Assessment Methodology for Speaker Identification and Verification Systems. Technical report – Task 2500 – Report 19, SAM-A ESPRIT Project 6819. 1993.

[Bimbot, 94] **BIMBOT F., CHOLLET G., PAOLONI A.**, Assessment Methodology for Speaker Identification and Verification Systems. In Workshop on Automatic Speaker Recognition and Verification, pp 75-82. Martigny (Switzerland). April 1994.

[Bimbot, 95] **BIMBOT F., MAGRIN-CHAGNOLLEAU T., MATHAN L.**, Second-order statistical methods for text-independent speaker identification. Speech Communication, n° 17(1-2). August 1995.

[Biométrie Online] **BIOMETRIE ONLINE**, Biométrie Online. Adresse Internet : http://biometrie.online.fr/

[Bonastre, 92] **BONASTRE J-F., MELONI H.**, A study of spectral variability for speaker characterisation. In 19èmes Journées d'Etudes sur la Parole, p 555. Juin 1992.

[Bonastre, 94a] **BONASTRE J-F.**, Stratégie analytique orientée connaissances pour la caractérisation et l'identification du locuteur. Thèse de Doctorat : Université d'Avignon. 1994.

[Bonastre, 94b] **BONASTRE J-F., MELONI H.**, Inter and Intra-speaker variability of French phonemes. Advantages of an explicit knowledge based approach. In Workshop on Automatic Speaker Recognition and Verification, pp 157-160. Martigny (Switzerland). April 1994.

[Buytel] **BUYTEL™**, Buytel. Adresse Internet : http://www.buytel.com/

[Charlet, 97] **CHARLET D., JOUVET D.**, Optimising feature set for speaker verification, In Proc. AVBPA Spriner LNCS, Bigün, et al., Eds.. 1997.

[Cheung, 78] **CHEUNG R.S., EISENSTEIN B.A.**, Feature selection via dynamic programming for text-independent speaker identification. In IEEE Transactions on Speech and Audio Processing, vol 26, n° 5, pp 397-403. October 1978.

[Chollet, 97] **CHOLLET G., BIMBOT F.**, Assessment of speaker verification systems. In Handbook of standards and resources for spoken language systems. Mouton de Gruyter. 1997.

[Cohen, 95] **COHEN L.**, Time –Frequency Analysis, Prentice-Hall, Englewood Cliffs. 1995.

[Configate] **CONFIGATE**, Configate. Adresse Internet : http://www.configate.com/

[Doddington, 85] **DODDINGTON G.R.**, Speaker recognition. Identify people by their voices. Proceeding IEEE, n° 73(11), pp 1651-1664. November 1985.

[Driancourt, 92] **DRIANCOURT X., GALLINARI P.**, A speech recogniser optimally combining learning vector quantization, dynamic programming and multi-layer perceptron. In Proc. ICASSP 92, San Francisco, USA. 1992.

[Dubreucq, 94] **DUBREUCQ V., VLOEBERGHS C.**, The use of the pitch to improve an HMM based speaker recognition method, In Workshop on Automatic Speaker Recognition and Verification, Martigny (Switzerland), pp 15-18. 1994.

[Eagles, 95] Assessment of speaker verification systems, In EAGLES Spoken Language Systems, Eagles Document EAG-SLWG-Handbook Phase 2. February 1995.

[Forsyth, 93] **FORSYTH M.E., JACK M.A.**, Duration modelling and multiple codebooks in semi-continuous HMM for speaker verification. Eurospeech 93, Berlin, Germany, pp 319-322. 1993.

[Frederickson, 94] **FREDERICKSON S.E., TARASSENKO L.**, Radial basis functions for speaker identification. In Workshop on Automatic Speaker Recognition and Verification, pp 107-110. Martigny (Switzerland). April 1994.

[Furlanello, 95] **FURLANELLO C., GIULANI D., TRENTIN E., FALAVIGNA D.**, Applications of generalised radial basis functions in speaker normalisation and identification. In Proceedings of IEEE International Symposium on Circuit and Systems, pp 1704-1707. Seattle (USA). April-May 1995.

[Furui, 72] **FURUI S., ITAKURA F., SAITO S.**, Talker recognition by long time averaged speech spectrum. Elect. Commun, Japan 55-A(10) pp 54-61. 1972.

[Furui, 74] **FURUI S.**, An analysis of long term variation of feature parameters of speech and its application to talker recognition, In Trans. IECE, 57-A, vol. 12, pp 880-887. 1974.

[Furui, 81] **FURUI S.**, Cepstral analysis technique for automatic speaker verification. In IEEE Trans. Acoust. Speech Signal Processing, vol. 19, n° 2, pp 254-272. 1981.

[Furui, 94] **FURUI S.**, An overview of speaker recognition technology. In Workshop on Automatic Speaker Recognition and Verification, pp 1-9. Martigny (Switzerland). April 1994.

[Furui, 97a] **FURUI S.**, Recent advances in speaker recognition. In Proc. AVBPA, Springer LNCS, Bigün, et al., Eds, pp 237-252. 1997.

[Furui, 97b] **FURUI S.**, Recent advances in robust speech recognition. In ESCA-NATO Workshop on Robust speech recognition for unknown communication channels. Pont-à-Mousson, France, pp 11-20. 17-18 Avril 1997.

[Gauvain, 95] **GAUVAIN J-L., LAMEL L-F, PROUTS B.**, Experiments with speaker verification over the telephone. Eurospeech 95. Madrid, Spain, pp 651-654. 1995.

[Gish, 86] **GISH H., KRASNER M., RUSSEL W., WOLF J.**, Methods and experiments for text-independent speaker recognition over telephone channels. In Proc0 ICASSP 86, pp 865-868. 1986.

[Gish, 90] **GISH H.**, Robust discrimination in automatic speaker identification. In Proc. ICASSP 90, vol. 1, pp 289-292. 1990.

[Gish, 94] **GISH H., SCHMIDT M.**, Text independent speaker identification. IEEE Signal Processing Magazine, p 18. October 1994.

[Grenier, 77] **GRENIER Y.**, Identification de locuteur et adaptation au locuteur d'un système de reconnaissance phonétique. Thèse de Docteur Ingénieur : E.N.S.T. Pari s. 1977.

[Griffin, 94] **GRIFFIN C., MATSUI T., FURUI S.**, Distance measures for Text-independent speaker recognition based on MAR Model. In Proceedings ICASSP. Adelaïde (Australia). 1994.

[Grish, 85] **GRISH H., KARNOFSKY K., KRASNER M., ROUCOS S., SCHWARZ R., WOLF J.**, Investigation of text-independent speaker identification over telephone channels. IN Proc. ICASSP 85, pp 379-382. 1985.

[Haton, 91] **HATON J.P., PIERREL J.M., PERENNOU G.**, Reconnaissance automatique de la parole. 1991.

[Hayakawa, 97] **HAYAKAWA S., TAKEDA K., ITAKURA F.**, Speaker Identification using harmonic structure of LP-residual spectrum. In Proc. AVBPA, Springer LNCS, Bigün, et al., Eds, pp 253-260. 1997.

[He, 97] **HE J., LIU L., PALM G.**, A new codebook training algorithm for VQ-based speaker recognition. In Proc. ICASSP 97, Munich, Germany, pp 1091-1094. 1997.

[Heck, 97] **HECK L-P., WEINTRAUB M.**, Handset dependent background models for robust text-independent speaker recognition. In Proc. ICASSP 97, Munich, Germany, pp 1071-1074. 1997.

[Hermansky, 94] **HERMANSKY H., MORGAN N.**, RASTA Processing of Speech. IEEE Trans. On Speech and Audio Processing, vol. 2, n° 4, pp 578-589. 1994.

[Hermansky, 97] **HERMANSKY H.**, Should recognisers have ears ? In ESCA-NATO Workshop on Robust speech recognition for unknown communication channels. Pont-à-Mousson, France. 17-18 Avril 1997.

[Higgins, 86] **HIGGINS A.L., WOHLFORD R.E.**, A new method of text-independent speaker recognition. In Proc. ICASSP 86, pp 869-872. 1986.

[Hollien, 90] **HOLLIEN H.**, The acoustics of crime. Applied Psycholinguistics and Communication Disorders 1990. Plenum Press : New-York & London. p 370. 1990.

[Homayounpour, 94] **HOMAYOUNPOUR M.M., CHOLLET G.**, A comparison of some relevant parametric representations for speaker verification. In Workshop on Automatic Speaker Recognition and Verification, pp 185-188. Martigny (Switzerland). April 1994.

[Homayounpour, 95] **HOMAYOUNPOUR M.M.**, Vérification vocale d'identité : dépendante et indépendante du texte. Thèse de Doctorat : Université de Paris Sud. 1995.

[Hunt, 83] **HUNT M.**, Further experiments in text-independent speaker recognition over communication channels. In Proc. ICASSP 83, pp 563-566. 1983.

[ITT and Buytel] **ITT INDUSTRIES et BUYTEL Ltd.**, SpeakerKey© Services. http://www.voicekey.com/

[Kuitert, 97] **KUITERT M., BOVES L.**, Speaker verification with GSM coded telephone speech. In Proc. Eurospeech 97, Rhodes, Greece. September 1997.

[Künzel, 94] **KUNZEL H.J.**, Current approaches to forensic speaker recognition. In Workshop on Automatic Speaker Recognition and Verification, pp 135-141. Martigny (Switzerland). April 1994.

[Lamel, 97] **LAMEL L., GAUVAIN J-L.**, Speaker recognition with the switchboard corpus. In Proc. ICASSP 97, Munich, Germany, pp 1067-1070. 1997.

[Markov, 96] **MARKOV K., NAKAGAWA S.**, Frame level likelihood normalisation for text independent speaker identification using gaussian mixture models. In Proc. ICSLP 96, Philadelphia, USA, pp 1764-1767. 1996.

[Matsui, 90] **MATSUI T., FURUI S.**, Text-independent speaker recognition using vocal tract and pitch information. In Proceedings ICSLP 90, pp 137-140, 1990.

[Matsui, 91] **MATSUI T., FURUI S.**, A text-independent recognition method robust against utterance variation. In Proc. ICASSP 91, pp 377-380. 1991.

[Matsui, 92] **MATSUI T., FURUI S.**, Comparison of text-independent speaker recognition methods using VQ-distortion and discrete continuous HMMs. In Proc. ICASSP 92, San Francisco, USA, pp 157-160. 1992.

[Matsui, 96] **MATSUI T., FURUI S.**, Robust methods of updating model and a priori threshold in speaker verification. In Proc. ICASSP 96, Atlanta, EU, pp 97-100. 1996.

[Metzger, 99] **METZGER Serge**, Mémoire : Authentification de visages : Une approche neuronale. 1999.

[Montacié, 92a] **MONTACIE C., LE FLOCH J.L.**, AR-Vector models for free-text speaker recognition. In Proceedings ICSLP 92, pp611-614. Banff (Canada). October 1992.

[Montacié, 92b] **MONTACIE C., DELEGLISE P., BIMBOT F. CARATY M.J.** , Cinematic techniques for speech processing : temporal decomposition linear prediction. In Proc. ICASSP 92, vol.1, pp 153-156. San-Francisco, USA. 1992.

[Naïk, 90] **NAIK J.M.**, Speaker verification : a tutorial. IEEE Communications Magazine, pp 42-48. January 1990.

[Naïk, 94a] **NAIK D., ASSALEH K., MAMMONE R.**, Robust speaker identification using pole filtering. In Workshop on Automatic Speaker Recognition and Verification, pp 225-230. Martigny (Switzerland). April 1994.

[Naïk, 94b] **NAIK J.**, Speaker verification over the telephone network : databases, algorithms and performance assessment. In Workshop on Automatic Speaker Recognition and Verification, pp 31-38. Martigny (Switzerland). April 1994.

[Navarro-Mesa, 92] **NAVARRO-MESA J.L.**, Optimum Window Length in Speech Signals– Time-Frequency Analysis, IEEE Trans. On Signal Processing, Vol.40, n°2. February 1992.

[Nist 97] **MARTIN A., PRZYBOCKI M.**, 1997 speaker recognition evaluation NIST workshop, Maritime Institute Linthicum, Maryland. 25-26 June 1997.

[O'Shaughnessy, 86] **O'SHAUGHNESSY D.**, Speaker recognition. IEEE ASSP Magazine, pp4-17. October 1986.

[Oglesby, 90] **OGLESBY J., MASON J.S.**, Optimisation of neural models for speaker identification. In Proc. ICASSP 90, pp261-264. 1990.

[Oglesby, 91] **OGLESBY J., MASON J.**, Radial basis function networks for speaker recognition. In Proc. ICASSP 91, 99 393-396. May 1991.

[Oglesby, 95] **OGLESBY J.**, What's in a number ? Moving beyond the equal error rate. Speech Communication, n°17(1-2). August 1995.

[Olsen 97] **OLSEN J.O.**, A two stage procedure for phone based speaker verification. In Proc. AVBPA, Springer LNGS, Bigün, et al. Eds., pp 219-226. 1997.

[Ong 94] **ONG S., MOODY M.P., SRIDHARAN S.** , Confidence Analysis for speaker identification: the effectiveness of various features. In Workshop on Automatic Speaker Recognition and Verification, pp 91-94. Martigny (Switzerland). April 1994.

[Openshaw, 94] **OPENSHAW J.P., MASON J.S.**, Optimal noise-masking of cepstral features for robust speaker identification. In Workshop on Automatic Speaker Recognition and Verification. Pp 231-234. Martigny (Switzerland). April 1994.

[OTG] **The OTTAWA TELEPHONY GROUP Inc**, OTG. Avril 2000. Adresse Internet : http://www.otg.ca/

[Poritz, 82] **PORITZ A.B.**, Linear predictive HMMs and the speech signal. In Proc. ICASSP 82, Paris, France, pp 1291-1294. 1982.

[Pruzansky, 63] **PRUZANSKY S.**, Pattern matching procedure of automatic talker recognition. JASA, vol.35, pp 354-358. March 1963.

[Reynolds, 92] **REYNOLDS D.A.**, A gaussian mixture modelling approach to text independent speaker identification. A thesis, Georgia Institute of Technology. August 1992.

[Reynolds, 94a] **REYNOLDS D.A.**, Experimental evaluation of features for robust speaker identification. IEEE Transactions on ASSP, n° 2 (4), pp 639-643. 1994.

[Reynolds, 94b] **REYNOLDS D.A.**, Speaker Identification and Verification using gaussian mixture models In Workshop on Automatic Speaker Recognition and Verification, pp 27-30. Martigny (Switzerland). April 1994.

[Reynolds, 95] **REYNOLDS D.A.**, Speaker Identification and Verification using gaussian mixture speaker models. Speech Communication, n° 17(1-2), pp 91-108. August 1995.

[Reynolds, 96] **REYNOLDS D.A.**, The Effects of handset variability on speaker recognition performance: experiments on the switchboard corpus. In Proc. ICASSP 96, Philadelphia, USA. May 1996.

[Rosenberg, 76] **ROSENBERG A.E.**, Automatic speaker verification, a review. Proc. IEEE, n° 64(4), pp 475-487. 1976.

[Rossi, 89] **ROSSI M.**, De la quiddité des variables. In Variabilité et spécificité du locuteur: études et applications, pp 78-86. Marseille Luminy, France.Juin 1989.

[Rudasi, 91] **RUDASI L., ZAHORIAN S.A.**, Text independant talker identification with neural networks. In Proceedings ICASSP 91, pp 389-392. Toronto (Canada). 1991.

[Sambur, 75] **SAMBUR M.R.**, Selection of acoustic features for speaker identification. IEEE Transactions on ASSP, n° 23(2), pp 176-182. April 1975.

[Savic, 90] **SAVIC M., GUMPTA S.K.**, Variable parameter speaker verification system based on Hidden Markov Modelling. In Proceedings ICASSP, pp 281-284. New-Mexico (USA). 1990.

[Schmidt, 97] **SCHMIDT M., GOLDEN J., GIS H H.**, GMM sample statistic log-likelihood for Text independent speaker recognition. In Proc. Eurospeech 97, Rhodes, Greece. September 1997.

[Setlur, 95] **SETLUR A., JACOBS T.**, Results of a speaker verification service trial using HMM Models. In Eurospeech 95, pp 639-642. Madrid (Spain). September 1995.

[Soong, 85] **SOONG F., ROSENBERG A., RABINER L., JUANG B.**, A vector quantization approach to speaker recognition. In Proc. ICASSP 85, pp 387-390. 1985.

[Soong, 86] **SOONG F., ROSENBERG A.**, On the use of instantaneous and transitional spectral information in speaker recognition. In Proc. ICASSP 86, pp 877-880. 1986.

[Thevenaz, 95] **THEVENAZ P., HUGLI H.**, Usefulness of the LPC-residue in text-independent speaker verification. Speech Communication, n° 17(1-2), pp 145-158. August 1995.

[Tishby, 91] **TISHBY N.Z.**, On the application of mixture AR HMMs to text-independent speaker recognition. In IEEE Transactions on Signal Processing, vol. 39, pp 563-570. March 1991.

[T-Netix] **T-NETIX Inc.**, T-NETIX. Adresse Internet : http://www.t-netix.com/

[Van den Heuvel, 94] **VAN DEN HEUVEL H., RIETVELD T., CRANEN B.**, Methodological aspects of segment and speaker-related variability. A study of segmental durations in Dutch. Journal of Phonetics, n° 22, pp 389-406. 1994.

[Van Vuuren, 96] **VAN VUUREN S.**, Comparison of Text independent speaker methods on telephone speech with acoustic mismatch. In Proc. ICSLP, Philadelphia, USA, pp 1788-1791. 1996.

[Vnunet, 2000] **VNUNET.FR**, La Connexion Informatique, Adresse Internet : http://www.vnunet.fr/VNU2/actualite/page_article1.htm?numero=4764&date=2000-05-03&FULL=1 . 3 Mai 2000.

[Wang, 93] **WANG H.C., CHEN M.S., YANG T.**, A novel approach to speaker identification over telephone networks. In Proc. ICASSP 93, Minneapolis, USA, pp 407-410. 1993.

[Wassner, 96] **WASSNER H.**, Etude sur la paramétrisation du signal en Traitement Automatique de la Parole, Mémoire Industriel ESIEA effectué à l'IDIAP. Février 1996.

[Wenndt, 97] **WENNDT S., SHAMSUNDER S.**, Bispectrum features for robust speaker identification. In Proc. ICASSP 97, Munich, Germany, pp 1095-1098. 1997.

[Wolf, 92] **WOLF J.**, Efficient acoustic parameters for speaker recognition. The Journal of the Acoustical Society of America, pp 2044-2056, n° 51(6). 1972.