

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
ECOLE NATIONALE POLYTECHNIQUE



DÉPARTEMENT D'ÉLECTRONIQUE
Complément de Mémoire

En vue de l'obtention du Master M2 en Electronique

Thème :

**Application d'un système d'Identification
Automatique du Locuteur pour les
Transactions Bancaires**

Réalisé par :

Mlle S.Siouane

Devant le jury :

D.Berkani	Professeur	ENP	Président
M.Guerti	Professeur	ENP	Rapporteur
B.Bousseksou	Chargé de cours	ENP	Examineur

Promotion : Juin 2013

REMERCIEMENTS

En premier lieu nous remercions DIEU le tout puissant de nous avoir donné le courage et la force pour réaliser ce travail.

Notre profonde gratitude et sincère reconnaissance vont tout d'abord à Mme M. Guerti qui a bien voulu nous encadrer. Nous la remercions pour sa disponibilité, son aide, les précieux conseils qu'elle nous a prodigués, ses critiques constructives, ses explications et suggestions pertinentes.

Nos remerciements vont également à tous les enseignants de l'Ecole Nationale Polytechnique qui ont contribué à notre formation, qu'ils trouvent ici l'expression de notre profond respect et notre grande considération.

Nous remercions respectivement Mrs M. Kabache de l'ISMAS et F. Ykhlef du CDTA, pour le temps qu'ils nous ont consacré.

DEDICACES

*A mes très chers parents qui ont toujours été là pour moi,
et qui m'ont donnée un magnifique modèle de labeur et de persévérance*

A ma chère binôme Faïza Bendimerad

A ma sœur Nadia et son mari

A mes frères Ilyes, Midou et Raouf

qui m'ont soutenue et encouragée tout au long de ce projet

A mon chère Amine

A tous ceux que j'aime.

Sofia

TABLE DES MATIERES

	Page
1.1. INTRODUCTION	1
1.2. IDENTIFICATION AUTOMATIQUE DU LOCUTEUR.....	2
1.3. CONSULTER UN COMPTE BANCAIRE A DISTANCE	3
1.4. QUELS SONT LES AVANTAGES DE SERVICES DE BANQUE PAR MOBILE ?	4
1.5. PRINCIPE DE FONCTIONNEMENT	4
1.6. DESCRIPTION DE LA BASE DE DONNEES	7
1.7. ANALYSE CEPSTRALE DU SIGNAL DE PAROLE.....	7
1.8. DESCRIPTION DE L'APPLICATION	8
1.9. EXEMPLE D'EXECUTION.....	9
1.10. CONCLUSION GENERALE	10

LISTE DES ABREVIATIONS

FFT	F ast F ourier T ransform
IAL	I dentification A utomatique du L ocuteur
MFCC	M el F requency C epstral C oefficients
RAL	R econnaissance A utomatique du L ocuteur

ملخص:

يعتبر المكبر منخفض الضوضاء عنصرا رئيسيا في جميع اجهزة الاستقبال الراديوية ونظم الاتصالات، لأنه يقوم بتكبير مستوى الإشارة من دون إضافة الكثير من الضوضاء من اجل معالجة صحيحة فيما بعد. الهدف من هذه الأطروحة هو اجراء دراسة وتقديم استراتيجية لتصميم مكبر منخفض الضوضاء. كلمات مفتاحية: مكبر منخفض الضوضاء، ترددات راديوية، استقرار، التغير الخطي.

RESUME

L'objectif de ce master est l'application d'un système d'Identification Automatique du Locuteur (**IAL**) en mode dépendant du texte dans le domaine des transactions bancaires. Pour atteindre ce but, nous avons partagé l'application en phases d'apprentissage et test. Dans cette dernière, l'analyse acoustique des mots isolés en Français prononcées par les locuteurs et l'identification ont été effectuée en utilisant les **MFCC** (Mel Frequency Cepstral Coefficients). Les résultats obtenus donnent un taux de reconnaissance qui atteint les 80%.

Mots-clés : Identification Automatique du Locuteur, MFCC, Transactions Bancaires, Taux de Reconnaissance **TR**.

ABSTRACT

The goal of our master is to develop an Automatic Speaker Identification system text dependent in transactions bancaires.

To achieve this goal, we have constructed a corpus of isolated words in French. The latter was recorded by four speakers (two women and two men). Acoustic analysis of the database was performed using the **MFCC** (Mel Frequency Cepstral Coefficients). The results show a recognition rate reaching 80%.

Keywords : Automatic Speaker Identification, MFCC, recognition rate.

1.1.Introduction générale

Depuis plus de 15 ans nous assistons à l'explosion du nombre de documents numériques accessibles au public à travers des médias multiples comme le Web, le satellite et les réseaux. Ces documents prennent une place croissante dans la vie quotidienne et sont devenus une ressource essentielle.

Le **T**raitement **A**utomatique de la **P**arole (**TAP**) est un domaine de recherche actif, au croisement du traitement du signal numérique et du traitement symbolique du langage.

Depuis les années 60, il bénéficie d'efforts de recherche très importants, liés au développement des moyens et techniques de télécommunications. Un thème important de la recherche actuelle dans le domaine du TAP, est la réalisation de véritables systèmes de dialogue oral entre l'Homme et la Machine dont la **R**econnaissance **A**utomatique de la **P**arole (**RAP**) joue un rôle très important en particulier l'**I**dentification **A**utomatique du **L**ocuteur **IAL**. C'est dans cette optique que se situe notre travail.

Dans ce travail, nous nous intéressons à l'extraction automatique d'informations contenues dans des documents audio, et plus particulièrement aux informations relatives aux locuteurs. Nous cherchons à identifier chaque locuteur d'un document par son code [1].

1.2. Identification Automatique du Locuteur IAL

Il convient de comparer un message vocal avec un ensemble de références acoustiques correspondant à plusieurs personnes, et de déterminer par cet examen quelle est la personne qui a parlé [4].

D'un point de vue schématique (figure 1.1), une séquence de parole est donnée en entrée du système d'IAL. Pour chaque locuteur connu du système, la séquence de parole est comparée à une référence caractéristique du locuteur : identité du locuteur dont la référence est la plus proche de la séquence de parole est donnée en sortie du système d'IAL.

Deux modes sont proposés en IAL, l'identification en ensemble :

- fermé pour lequel on suppose que la séquence de parole est effectivement prononcée par un locuteur connu du système ;
- ouvert pour lequel le locuteur peut ne pas être connu.

En mode "ensemble ouvert", le système d'IAL doit décider de la fiabilité de son jugement en acceptant ou rejetant l'identité qu'il a trouvée.

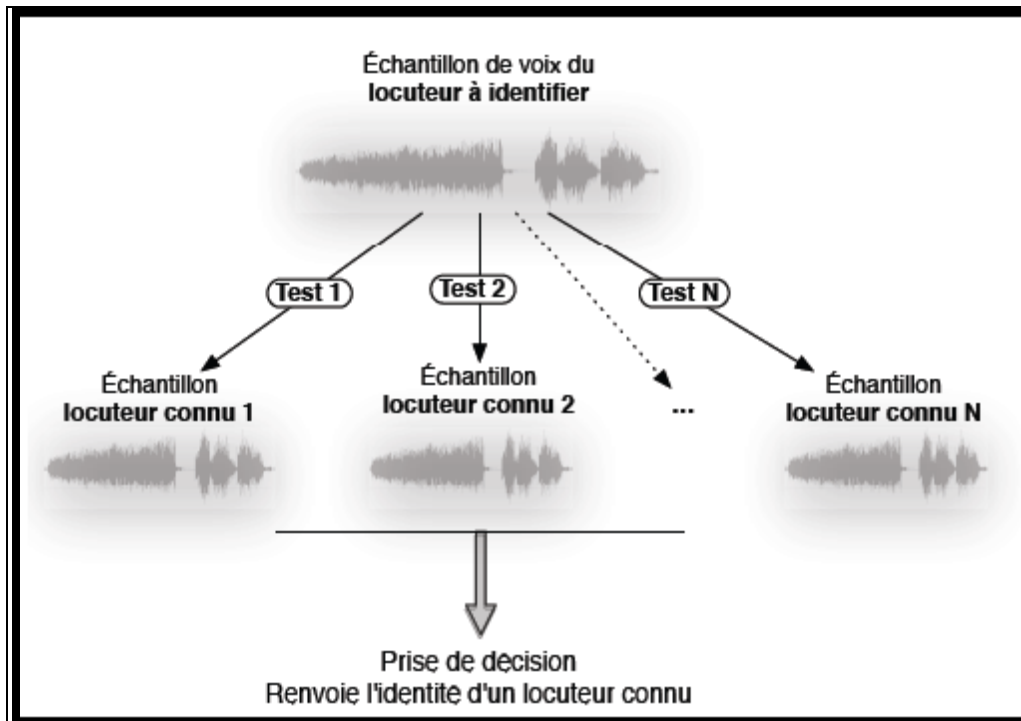


Figure 1.1 : principe de l'identification automatique du locuteur

Les systèmes d'IAL, souvent utilisés pour déterminer l'identité d'un locuteur à son insu, à partir d'une conversation téléphonique par exemple, l'indépendance de la méthode au texte prononcé peut être une nécessité. En résumé :

- en mode indépendant du texte, le système de reconnaissance n'a aucune connaissance sur le message linguistique prononcé par la personne ;

en mode dépendant du texte, la reconnaissance d'une personne est réalisée sur la base d'un message dont le contenu linguistique (mot de passe, phrase...) est connu du système, une terminologie plus fine peut être donnée à un système suivant l'application visée, systèmes à :

- messages fixés : la personne est contrainte de prononcer un message, qu'elle aurait au préalable (mots de passe personnalisés) ou qui sera imposé par le système ;

- messages prompts : un message, différent à chaque nouvelle session de reconnaissance, est imposé par le système sous forme visuelle ou auditive. Ces systèmes ont pour première motivation de se protéger des attaques de personnes malveillantes (imposteurs) qui disposeraient d'un enregistrement de la voix d'une personne ;

- unités segmentales : la personne doit prononcer un message comportant soit une séquence de mots (séquence de chiffres), soit des traits phonétiques (séquence de phonèmes) connus du système.

1.3. Consulter un compte bancaire à distance

Les moyens actuels permettent dans presque toutes les banques de consulter un compte bancaire de manière sécurisée et d'y suivre les opérations au jour le jour en utilisant l'un des canaux suivants :

- Internet : une connexion à distance sur un site déterminé permet d'avoir une vision complète de comptes et de crédits en cours dans une banque, jour par jour, ce qui permet de pointer facilement votre compte ;
- Les Guichets Automatiques de Banque (GAB) : en utilisant une carte et un code confidentiel, on aura l'accès à la consultation du compte (solde et dernières opérations) ;
- Le serveur vocal : ce service de consultation du compte, généralement payant à la communication téléphonique, est parfois disponible seulement sur abonnement. Vous devez appeler le numéro et taper le code confidentiel propres à ce service. Il est disponible 24h/24, il permet notamment de connaître le solde et les dernières opérations apportées aux comptes [2].

1.4. Quels sont les avantages de services de banque par mobile ?

De manière générale, les services de banque par mobile permettent de :

- s'informer : accéder au détail des offres, effectuer des simulations, contacter un conseiller pour avoir une réponse personnalisée ;
- consulter tous vos comptes : pour en connaître le solde, vérifier l'arrivée d'un virement ou le débit d'un chèque, télécharger vos dernières opérations ;
- agir sur vos comptes : réaliser des virements internes ou externes, passer des ordres de bourse ;
- demander : un chéquier, un RIB (relevé d'identité bancaire) [3].

1.5. Principe de fonctionnement :

Lorsqu'un locuteur effectue un appel téléphonique afin de consulter son compte bancaire, et selon sa voix il peut être accepté ou rejeté par le système.



Fig. 1.2 : Fonctionnement d'un système de transaction bancaire [4]

1.6. Analyse cepstrale du signal de parole

Les coefficients cepstraux les plus répandus sont les MFCC (Mel Frequency Cepstral Coefficients). Ils présentent l'avantage d'être faiblement corrélés entre eux, et qu'on peut donc approximer leur matrice de covariance par une matrice diagonale.

Les MFCC sont une extension des cepstres qui sont utilisés pour mieux représenter les modèles de l'audition humaine. Le principe de calcul des MFCC est issu des recherches psychoacoustiques sur la tonie et la perception des bandes de fréquences par l'oreille humaine.

Les MFCC assurent une séparation entre les deux composantes:

- la fonction d'excitation glottique qui est caractérisée par le pitch ou la contribution de l'excitation se localise dans les quéfrences élevées ;
- la fonction de transfert du conduit vocal ou la contribution se retrouve dans les faibles quéfrences (premiers coefficients cepstraux) .

Le cepstre du signal de parole est défini comme étant la Transformée de Fourier Inverse du logarithme de la densité spectrale de puissance. Pour ce signal, la source d'excitation glottique est convoluée avec la réponse impulsionnelle du conduit vocal:

$$S(t) = e(t) * h(t) \quad (1.1)$$

Où $s(t)$ est le signal de parole, $e(t)$ est la source d'excitation glottique et $h(t)$ est la réponse impulsionnelle du conduit vocal.

L'application du logarithme sur le module de la Transformée de Fourier dans l'équation donne :

$$\text{Log } |S(f)| = \text{Log } |E(f)| + \text{Log } |H(f)| \quad (1.2)$$

Par une transformée de Fourier inverse on obtient :

$$S'(cef) = e'(cef) + h'(cef) \quad (1.3)$$

La dimension du nouveau domaine est homogène à un temps et s'appelle la *quéfrence* (cef), le nouveau domaine s'appelle donc le domaine *quéfrentiel*. Un filtrage dans ce domaine s'appelle *liffrage*.

Ce domaine est intéressant pour faire la séparation des contributions du conduit vocal et de la source d'excitation dans le signal de parole. En effet, si les contributions relevant du conduit vocal et les contributions de la source d'excitation évoluent avec des vitesses différentes dans le temps, alors il est possible de les séparer par l'application d'un simple fenêtrage dans le domaine quéfrentiel (liffrage passe-bas) pour le conduit vocal.

Pour simuler le fonctionnement du système auditif humain, les fréquences centrales du banc de filtres sont réparties uniformément sur une échelle perceptive. Plus la fréquence centrale d'un filtre est élevée, plus sa bande passante est large. Cela permet d'augmenter la résolution dans les basses fréquences, zone qui contient le plus d'informations utiles dans le signal de parole. Les échelles perceptives les plus utilisées sont :

- l'échelle Mel qui est linéaire en basses fréquences et logarithmique en hautes fréquences;

$$Mel(f) = 2595 \log \left(1 + \frac{f}{700} \right) \quad (1.4)$$

- l'échelle Bark

$$Bark(f) = 6 \operatorname{arcsinh} \left(\frac{f}{1000} \right) \quad (1.5)$$

f représente la fréquence (Hz).

La procédure de calcul des MFCC est illustrée comme suit (figure 3.4) :

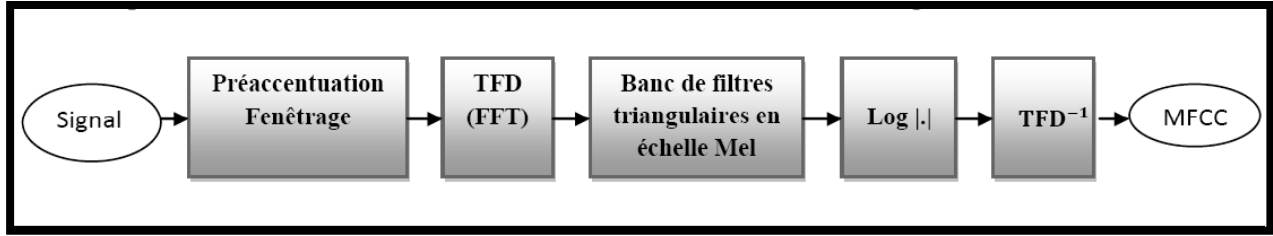


Fig. 1.3 : Schéma d'extraction des MFCC

Soit un signal discret $s(n)$ avec, $0 \leq n \ll N - 1$, N est le nombre d'échantillons d'une fenêtre d'analyse, F_e est la fréquence d'échantillonnage, la Transformée de Fourier Discrète à court terme $S(k)$ est obtenue avec la formule :

$$S(k) = \sum_{n=0}^{N-1} S(n) \exp\left(\frac{-j2\pi nk}{N}\right) \quad , \quad 0 \leq k \leq N-1 \quad (1.6)$$

Le spectre du signal est filtré par un banc de filtres triangulaires, dont les bandes passantes sont de même largeur dans le domaine des fréquences Mel (figure 1.4). Les points de frontières B_m des filtres en échelle de fréquence Mel sont calculés à partir de la formule :

$$B_m = B_b + m \frac{B_h - B_b}{M+1} \quad , \quad 0 \leq m \leq M+1 \quad (1.7)$$

M : Le nombre de filtres.

B_h : La fréquence la plus haute du signal.

B_b : La fréquence la plus basse du signal.

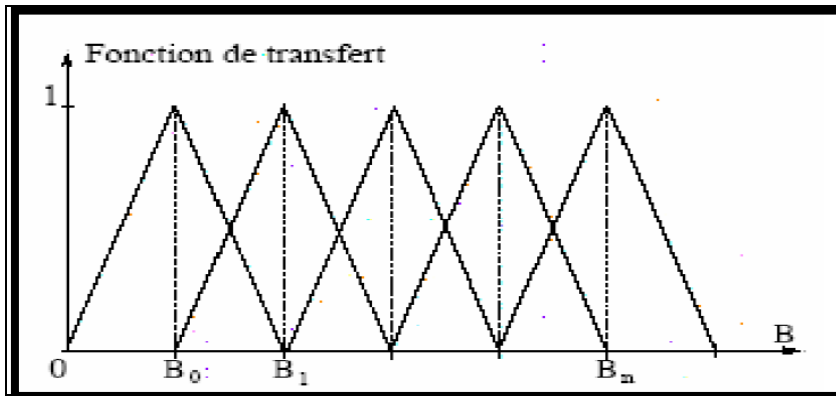


Fig. 1.4: Banc de filtre sur l'échelle Mel

Les coefficients cepstraux peuvent être calculés directement à partir du logarithme des énergies E_i issues d'un banc de M filtres par la transformée en cosinus discrète inverse définie par :

$$C_k = \sum_{i=1}^M \log E_i \cos \left[\frac{\pi k}{M} \left(i - \frac{1}{2} \right) \right] \quad , \quad 1 \leq k \leq d \quad (1.8)$$

et qui permet d'obtenir des coefficients peu corrélés.

Le coefficient C_0 qui est la somme des énergies n'est pas utilisé ; il est éventuellement remplacé par le logarithme de l'énergie totale E calculée dans le domaine temporel et normalisée [19].

$$E = \text{Log} \sum_{n=1}^N x_n^2 \quad (1.9)$$

$$E_{\text{NOR}} = E - E_{\text{max}} \quad (1.10)$$

1.7. DESCRIPTION DE LA BASE DE DONNEES

La **Base de Donnée BD** utilisée dans ce travail est constituée par des enregistrements d'un corpus prononcés par des locuteurs en temps réel (tableau 1.1).

Chaque locuteur a prononcé les différents mots des corpus, constituant la base d'apprentissage. Pour la phase test nous avons choisi d'enregistrer la voix du locuteur en temps réel où chaque locuteur doit prononcer son code choisi.

Tableau 1.1 : Corpus des mots isolés enregistrés

Corpus enregistrés	Un
	Deux
	Trois
	Quatre
	Cinq
	Six
	Sept
	Huit
	Neuf
	dix

1.8.DESCRPTION DE L'APPLICATION

Nous avons utilisé 2 fonctions réparties en 2 fichiers : training.m et testing.m réalisées avec l'outil MATLAB (version 7.8.0.347 :R2009a) ;

- le fichier training.m : est une fonction d'apprentissage dans laquelle on :
 - charge le fichier sonore du locuteur en utilisant la fonction wavrecord ;
 - paramétrise le signal parole de ce locuteur, en initialisant préalablement les paramètres de ce dernier pour le calcul des MFCC ;
- le fichier testing.m : est une fonction qui teste le système d'Identification Automatique du Locuteur. Elle est organisée comme suit :
 - chargement du fichier sonore du locuteur inconnu a l'aide de la fonction wavrecord ;
 - paramétrisation du signal parole du locuteur, en initialisant préalablement les paramètres de ce dernier pour le calcul des MFCC ;

- calcul de la distance pour la mesure de ressemblance entre les paramètres du locuteur inconnu et celle de la partie d'apprentissage ;

le résultat du test est déterminé selon la moyenne de la racine d'erreur (**Mean Squared Error MSE**) :

$$MSE = (\text{sum}((dx - dy).^2)) / 20 \quad (1.11)$$

si $MSE < 1.5$ alors on a une identification du locuteur et donc il est accepté, si non , il est rejeté.

1.9. EXEMPLE D4EXECUTION

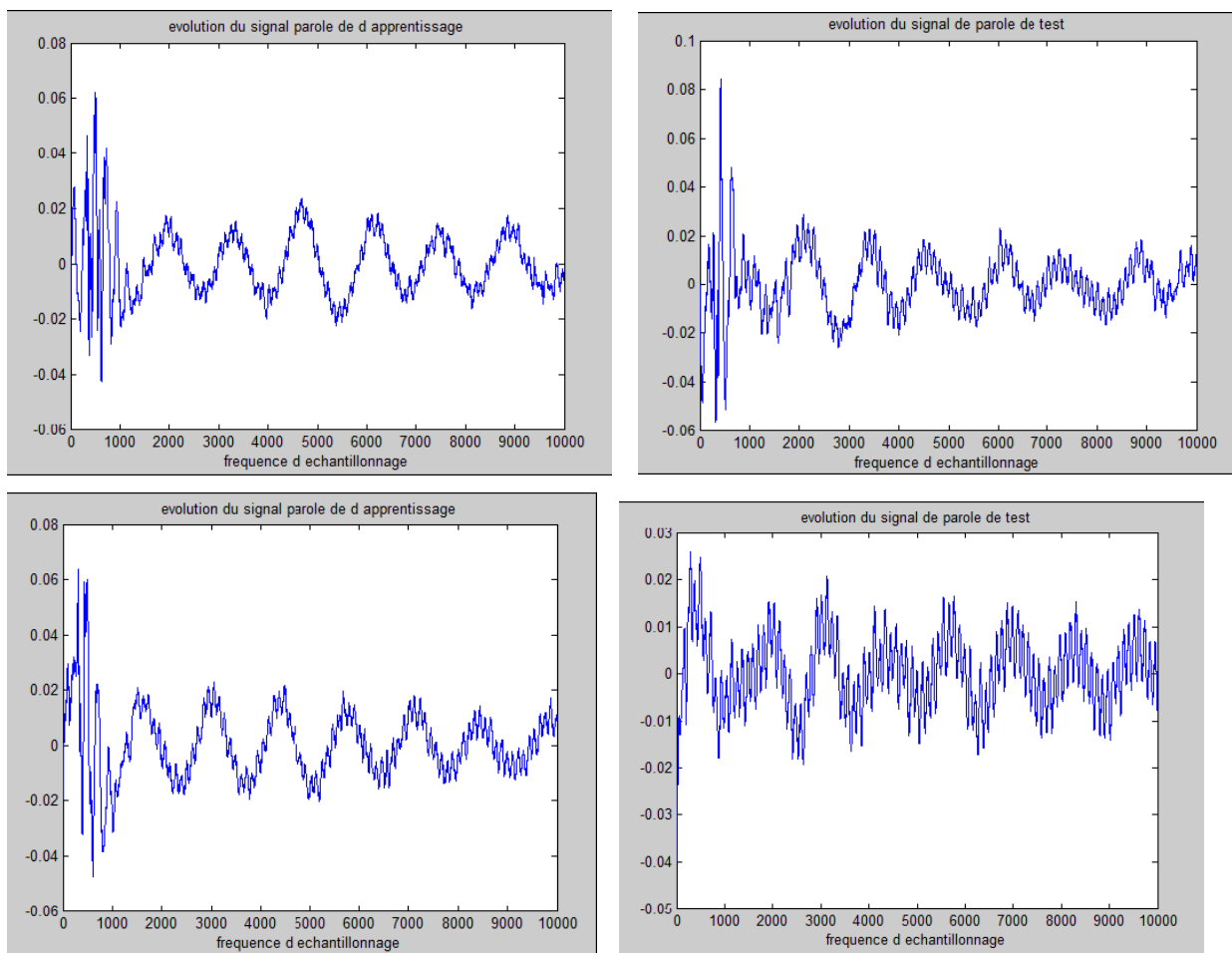


Fig 1.5 : exemple d'exécution de l'application

1.10. CONCLUSION GENERALE

Avec le nombre accru de propriétaires de téléphones intelligents, la plupart des banques offrent actuellement des services bancaires mobiles et des applications qui permettent aux consommateurs d'effectuer un ensemble de transactions quotidiennes à partir de leur cellulaire.

REFERENCES BIBLIOGRAPHIQUES

- [1] V. Jousse, Identification nommée du locuteur : exploitation conjointe du signal sonore et de sa transcription, thèse de doctorat, Université du Maine/France.
- [2] <file:///C:/Users/SOFIA/Desktop/master%20sofi/6W9M92.htm>
- [3] <file:///C:/Users/SOFIA/Desktop/master%20sofi/imgres.htm>
- [4] <file:///C:/Users/SOFIA/Desktop/master%20sofi/imgres2.htm>