الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieur
et de la Recherche Scientifique
Université Akli Mohand Oulhadj - Bouira -
Tasdawit Akli Muḥend Ulḥağ - Tubirett -

وزارة التعليم العالي والبحث العلمي
جامعة أكلي محند أولحاج
-البويرة-

Centre de Développement des Technologies Avancées

مركز تنمية التكنولوجيات المتطورة

Faculté des Sciences et des Sciences Appliquées

كلية العلوم والعلوم التطبيقية

Référence : ...../MM/2021

المرجع:...../م م/ 2021

Mémoire de Master

Présenté au

Département : Génie Électrique
Domaine : Sciences et Technologies
Filière : Télécommunications
Spécialité : Systèmes des Télécommunications

Réalisé par :

BENGHERABI Ayoub

Et

BELABBACI Elouanas

Thème

Descripteurs audio-visuel pour la reconnaissance des marques de téléphones mobiles

Soutenu le: **14/10/2021**

Devant la commission composée de :

Dr :	Ali BOUHADDA	M.A.A	Univ. Bouira	Président
Dr :	Meriem FEDILA	M.R.B	CDTA. Alger	Rapportrice
Dr :	Abdennour ALIMOHAD	M.C.B	Univ. Bouira	Rapporteur
Dr :	Mohammed SAIDI	M.A.A	Univ. Bouira	Examinateur

Année Universitaire: 2020-2021

Remerciements

Tout d'abord, nous remercions Allah le tout puissant de nous avoir donné le courage et la patience nécessaires à mener ce travail à son terme.

Nous tenons à remercier nos deux promoteurs **M. ALIMOHAD Abdennour** et **Mme. FEDILA Meriem** pour l'aide compétente qu'ils nous ont apportée, pour leur patience et leur encouragement. leurs œils critiques nous ont été très précieux pour structurer le travail et pour améliorer la qualité des différentes sections.

Nous tenons aussi à adresser nos plus sincères remerciements à **M. BENGHERABI Messaoud**, chef d'équipe BIOSMC (Biométrie, Sécurité MultiMedia et Criminalistique) pour nous avoir offert l'opportunité d'intégrer son équipe durant notre stage de fin d'études chez CDTA. Nous aimerions également lui remercier pour sa disponibilité permanente, ses nombreux conseil, sa réflexion pertinente et sa discussion adéquate a su guider notre travail.

Que les membres de jury trouvent, ici, l'expression de nos sincères remerciements pour l'honneur qu'ils nous ont fait en prenant le temps de lire et d'évaluer ce travail.

Nous souhaitons aussi remercier l'équipe pédagogique et administrative de l'UAMOB pour leurs efforts dans le but de nos offrir une excellente formation.

Pour finir, nous souhaiton remercier toute personne ayant contribué de prés ou de loin à la réalisation de ce travail.

Dédicace

“

Je dédie ce modeste travail à ceux qui, quels que soient les termes embrassés, je n'arriverais jamais à leur exprimer mon amour sincère.

À l'homme, mon précieux offre du dieu, qui doit ma vie, ma réussite et tout mon respect : mon cher père amar,

À ma maman qui m'a soutenu et encouragé durant ces années d'études. Qu'elle trouve ici le témoignage de ma profonde reconnaissance,

À mes frères, mes grands parents et Ceux qui ont partagé avec moi tous les moments d'émotion lors de la réalisation de ce travail.

À ma famille, mes proches et à ceux qui me donnent de l'amour et de la vivacité

Sans oublier mon binôme Elouanas pour son soutien moral, sa patience et sa compréhension tout au long de ce projet

”

- ayoub

“

Je dédie ce projet :

À ma chère mère,

À mon cher père,

À mon binôme Ayoub,

Qui n'ont jamais cessé, de formuler des prières à mon égard, de me soutenir et de m'épauler pour que je puisse atteindre mes objectifs,

À mes frères ,

Pour leur soutiens moral et leurs conseils précieux tout au long de mes études,

À mon cher oncle boubaker

À qui je souhaite une bonne santé

À mon cher ami abderazek

Pour son aide et support dans les moments difficiles À tous ceux que j'aime et ceux qui m'aiment

”

- elouanas

Résumé

De nos jours, la reconnaissance des téléphones portables est devenue une réelle nécessité en raison de leur immense utilisation pour communiquer. En particulier, dans le cas où c'est le seul moyen pour des applications multiples. Ce travail s'inscrit dans le cadre global de la reconnaissance automatique des téléphones mobiles, qui est un domaine très fertile.

Dans le cadre de ce projet, nous avons utilisé deux bases de données MOBIPHONE et LOCALE pour la mise en place de notre système et l'évaluation des performances de nos algorithmes. Comme tous les systèmes de reconnaissance, notre système est essentiellement composé d'une phase d'apprentissage et d'une phase de test. Le concept ici est d'extraire les caractéristiques acoustiques à l'aide des coefficients MFCC et les caractéristiques visuelles via l'algorithme LPQ. Chaque type de téléphone est ensuite modélisé en utilisant les approches GMM et I-VECTOR. Enfin, les deux scores obtenus à partir de l'extraction acoustique et visuelle sont fusionnés pour obtenir un score de similarité entre les téléphones mobiles. L'évaluation est effectuée à travers le calcul du taux de reconnaissance correct qui a atteint après plusieurs expériences 100 %.

Mots clés : la reconnaissance des téléphones portables , les caractéristiques acoustiques , les caractéristiques visuelles , MFCC, LPQ ,GMM , I-VECTEUR , fusion.

Abstract

Nowadays, the cellphone recognition has become a real necessity because of their immense use to communicate. In particular, in case this is the only way for multiple applications. This work is part of the overall framework of automatic recognition of mobile phones, which is a very fertile field.

As part of this project, we used two databases MOBIPHONE and LOCALE for the implementation of our system and the evaluation of the performance of our algorithms. Like all recognition systems, our system is essentially composed of a learning phase and a test phase. The concept here is to extract the acoustic characteristics using the MFCC coefficients and the visual characteristics via the LPQ algorithm. Each type of phone is then modelled using the GMM and I-VECTOR approaches.Finally, the two scores obtained from the acoustic and visual extraction are merged to obtain a score of similarity between mobile phones. The evaluation is carried out through the calculation of the correct recognition rate which reached after several experiments 100 %.

Keywords : cellphone recognition , acoustic characteristics , visual characteristics , MFCC, LPQ ,GMM , I-VECTOR , fusion.

ملخص

في الوقت الحاضر ، أصبح التعرف على الهواتف النقالة ضرورة حقيقية بسبب استخدامها الهائل في الاتصال. خاصة ، في الحالات التي يكون فيها ذلك هو الوسيلة الوحيدة في تطبيقات متعددة. وهذا العمل يمثل جزء من الإطار العام للتعرف الآلي على الهواتف النقالة ، وهو مجال خصب جدا.

كجزء من هذا المشروع ، استخدمنا قاعدتي بيانات MOBIPHONE و LOCALE لتنفيذ نظامنا وتقييم أداء الخوارزميات لدينا. ومثل جميع نظم التعرف ، يتألف نظامنا أساسا من مرحلة التعلم ومرحلة الاختبار. المفهوم هنا هو استخلاص الخصائص الصوتية باستخدام معاملات MFCC والخصائص البصرية من خلال خوارزمية .LPQ ثم يتم وضع نموذج لكل نوع من الهواتف باستخدام نهج GMM و .I-VECTOR وأخيرا ، يتم دمج الدرجتين اللتين تم الحصول عليهما من الاستخراج الصوتي والبصري للحصول على درجة التشابه بين الهواتف النقالة. ويتم التقييم من خلال حساب معدل التعرف الصحيح الذي وصل بعد عدة تجارب الى 100 في المائة.

كلمات مفتاحية :

التعرف على الهواتف النقالة , الخصائص الصوتية , الخصائص االبصرية , MFCC , LPQ , GMM , I_VECTEUR , الدمج .

Table des matières

Table des figures

Liste des tableaux

Liste des acronymes

AR	*Auto regressif*
BIOSMC	*Biométrie, Sécurité MultiMedia et Criminalistique*
CDTA	*Centre de Développement des Technologies Avancées*
CSF	*Communication Sans Fils*
DCT	*Discrete Cosine Transform*
EER	*Equal Error Rate*
EM	*Expectation Maximization*
FAR	*False Acceptance Rate*
FRR	*False Rejection Rate*
GPLDA	*Gaussian Probabilistic Linear Discrimination Analysis*
GMM	*Gaussian Mixture Model*
GUI	*Graphical User Interface*
HMM	*Hidden Markov Model*
I-VECTEUR	*Vecteur d'Identité*
JFA	*Joint Factor Analysis*
LBP	*Local Binery Patterns*
LDA	*Linear Discrimination Analysis*

LLR *Log-Likelihood Ratio*

LPC *Linear Predective Coding*

LPCC *Linear Predective Cepectral Coefficient*

LPQ *Local Phase Quantization*

MAP *Maximum A Posteriori*

MFCC *Mel Frequency Cepstral Coefficient*

MLE *Maximum Likelihood Estimation*

PCA *Principal Component Analysis*

PLDA *Probabilistic Linear Discriminant Analysis*

PLP *Perceptual Linear Prediction*

PSTE *Public Scientific and Technological Establishment*

VQ *vector Quantization*

RA *reconnaissances automatique*

RATP *reconnaissances automatique de téléphone portable*

SIA *Systèmes d'Information Avances*

TIMIT *Texas Instruments-Massachusetts Institute of Technology*

UBM *Universal Background Model*

WCCN *Within-Class Covariance Normalization*

Introduction générale

Présentation de l'organigramme d'accueil

Le Centre de Développement des Technologies Avancées (CDTA) est un établissement public à caractère scientifique et technologique (EPST). Il a pour mission de mener des actions de recherche scientifique, d'innovation technologique, de valorisation et de formation dans les domaines des sciences et des technologies de l'information, de la vision par ordinateur pour l'interaction homme-machine, de la biométrie, le traitement du signal biomédical et les systèmes intelligents, des technologies industrielles et de la robotique, des dépôts et des traitements des matériaux, des applications et des technologies des lasers. À travers ses missions, le CDTA contribue activement au développement du savoir, à sa transformation en savoir-faire et en produits nécessaires au développement économique et sociétal.

Le CDTA compte cinq division : Architecture des Systèmes et MultiMedia, Microélectronique et Nanotechnologie, Productique et Robotique, Milieux Ionises et Lasers, TELECOM.

La Division Telecom a été créée en 2014. Elle a pour mission de mener des activités RD sur les grands domaines de recherche pertinents tels que : la technologie de l'information, traitement du signal et des systèmes biométriques et multimédia, les télécommunications, les antennes et la propagation des ondes, le smart-X, la sante, la sécurité multimédia et la vidéo surveillance, la biométrie et la criminalistique, L'ingénierie des connaissances, elle comprend quatre équipes de recherche réparties comme suit :

- Equipe CSF (Communication sans fils)
- Equipe Antennes
- Equipe BIOSMC (Biométrie, Sécurité MultiMedia et Criminalistique)
- Equipe SIA (Systèmes d'Information Avances)

Notre stage s'est déroulé au sein de l'équipe de recherche BIOSMC.

Contexte

Le traitement de la parole est l'une des plus importantes branches dans la recherche en traitement du signal. Les chercheurs ont toujours accordé une attention particulière aux signaux vocaux, car la parole est la moyen de communication le plus simple et le plus efficace pour les êtres humains. En raison du développement des technologies de l'information et de la communication, le rêve de communiquer avec les machines devient de plus en plus réalisable. Les recherches actuelles fournissent de nombreux systèmes de reconnaissance automatique, dont des progrès considérables ont été réalisés.

Dans ce projet, nous nous intéressons à la reconnaissance des téléphones portables à partir de leurs caractéristiques vocales. Il s'agit donc, d'identifier à quelle marque ou modèle des téléphones appartiennenent les signaux vocaux enregistrés. Il est à noter que la marque du téléphone mobile fait référence au fabricant, et le terme modèle de téléphone mobile est utilisé pour désigner un type de produit du même fabricant.

Problématique

La reconnaissance des téléphones portables trouve son intérêt dans le domaine de l'analyse multimédia en criminalistique. La détermination de l'authenticité d'un enregistrement, identifier la source de téléphone avec laquelle un son à été enregistré, et le tatouage audio sont quelques exemples d'application dans ce domaine. De nos jours, lorsque les conditions sont sous contrôle, les systèmes de reconnaissance automatique en général et celui du téléphone mobile en particulier donnent d'excellentes performances pour distinguer les voix enregistrées. Cependant, lorsque l'environnement est bruité ou bien les enregistrements sont courts on constate une réduction considérable des performances du système de reconnaissance automatique.

Contribution

Parmi les modalités les plus utilisées dans le système de la reconnaissance automatique, la "reconnaissance automatique basé sur la parole" par ce qu'elle est permanente et unique. Les chercheurs essayent toujours de développer les systèmes de reconnaissance à travers des outils mathématiques habituellement complexes pour faire la discrimination entre les classes. L'objectif suivi dans ce mémoire propose une démarche qui consiste à améliorer la performance du système de reconnaissance de téléphone mobile par l'utilisation de plusieurs méthodes avec un ensemble d'opérations. Pour cela, nous avons fait la combinaison entre différentes méthodes d'extraction des caractéristiques acoustique et celles visuelles, ce qui nous a permis d'en obtenir une meilleure adaptation pour la réali-

sation d'un système de reconnaissance automatique du téléphone mobile et l'amélioration de sa robustesse .

Organisation du mémoire

Ce mémoire final fournit une introduction générale dans laquelle nous aborderons la problématique de la reconnaissance automatique des téléphones portables, nous avons également souligné notre modeste contribution dans ce vaste domaine qui a motivé de nombreux chercheurs ces dix dernières années.

Trois chapitres principaux se présentent dans ce travail comme suit :

- Le premier chapitre est consacré à la présentation du systéme de reconnaissance automatique de téléphone mobile et de ses deux tâches pionnières, à savoir la vérification et l'identification, puis donne une description détaillée de ses différents blocs et des différentes techniques qui opèrent à leurs niveaux.
- Dans le deuxième chapitre, nous retrouvons le détail des techniques qui opèrent dans les trois blocs, de pré-traitement, d'extraction de paramètres, à savoir les coefficients cepstraux à échelle Mel (Mel Frequency Cepstral Coefficient : MFCC) et la quantification de la phase locale (Local Phase Quantization : LPQ), et de modélisation, par les modèles de mélange gaussiennes (Gaussian Mixture Model-Universal Background Model : GMM_UBM) et par le Vecteur d'Identité (I_Vecteur), choisies pour notre étude.
- Le dernier chapitre présente les bases de données MOBIPHONE et LOCALE sur lesquelles nos tests sont effectués, l'environnement de travail, les résultats obtenus ainsi que les interprétations correspondantes ; pour finir avec une implémentation d'une méthode de fusion des scores pour plus de robustesse.
- Le mémoire se termine par une conclusion générale qui présente un bilan du travail réalisé dans ce mémoire et expose les perspectives et les travaux futurs pour améliorer et compléter le travail présenté. Notre interface graphique créée à l'aide du Graphical User Interface (GUI) sous Matlab est donnée en Annexe A.

Chapitre 1

Reconnaissance Automatique du téléphone portable

1.1 Introduction

Les progrès de la technologie numérique ont conduit au développement d'outils portables à faible coût tels que les caméras de poche, les enregistreurs vocaux, les téléphones mobiles et les smartphones, qui font partie intégrante de notre vie quotidienne. Ces outils sont utilisés pour enregistrer et transmettre des données multimédias, qui jouent un rôle de plus en plus important en tant que preuve dans les enquêtes médico-légales.

En conséquence, il existe un besoin croissant d'une analyse et d'une classification approfondies des données multimédias (Hanilçi et al. 2014).

La détermination de l'intégrité et de l'authenticité d'une image, l'identification de l'appareil photo source avec lequel la photo a été prise sont des problèmes dans le domaine multimédia criminalistique. L'identification de l'appareil source avec lequel la photo a été prise est possible car les imperfections des appareils d'acquisition laissent leurs traces spécifiques sur les images. Une méthode similaire est utilisée pour identifier la source de téléphone portable à partir de l'image numérisée. Lorsque les échantillons vocaux enregistrés sont présentés comme preuves médico-légales, il est souvent nécessaire de suivre l'équipement ou l'environnement d'enregistrement (Hanilçi et al. 2014).

Dans ce chapitre, nous commencerons par un aperçu des principes généraux des systèmes de reconnaissance de téléphone portable. Nous décrivons ensuite ces étapes à savoir la production de la parole, le prétraitement de la parole, le paramétrage et la modélisation, et enfin les structures de ces tâches pionnières.

1.2 Traitement automatique de la parole

1.2.1 Production de la parole

Le signal vocal appartient à la classe des signaux acoustiques produits par vibrations des couches d'air. Les changements de ce signal reflètent les fluctuations de pression de l'air. Le processus de production de la parole est un mécanisme très complexe qui est basé sur l'interaction entre les systèmes neurologique et physiologique.

La parole commence par une activité neurologique (Leman 2011), et c'est le cerveau qui dirige les opérations liées à la mobilisation des organes phonatoire. Le fonctionnement de ces organes, est physiologique (Ouni 2001). De nombreux organes et muscles sont impliqués dans la production de la parole. La base du fonctionnement du système de parole humain interaction entre trois unités : poumons, larynx et voies vocales (Voir figure 1.1).

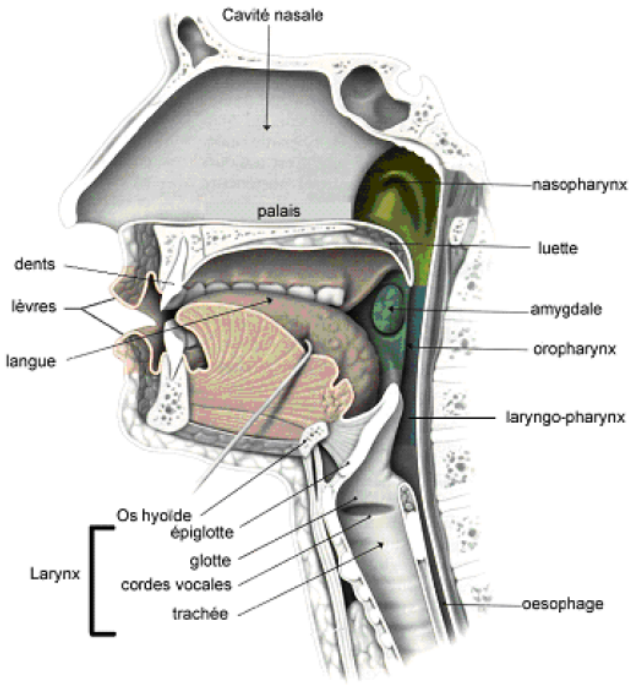

FIG. 1.1 : Organes de production de la parole (LARCHER 2009).

Le larynx est une structure cartilagineuse qui a une fonction de régulation particulière flux d'air par le mouvement des cordes vocales. Le conduit vocal s'étend des cordes vocales aux lèvres dans la partie buccale et les narines dans la région nasale (FUCHS 2007). L'air dans les poumons est comprimé par l'action du diaphragme. Cet air comprimé atteint alors les cordes vocales. Si les cordes sont dépliées, l'air circule librement et il vous permet de générer du bruit. S'ils sont fermés, la pression peut les faire vibrer et il en résulte un son quasi-périodique dont la fréquence fondamentale correspond essentiellement à la hauteur de la voix reçue. L'air, qu'il soit vibré ou non, traverse le tractus vocal puis voyage dans l'atmosphère.

La forme de ce canal, déterminée selon la position des articulateurs tels que la langue, la mâchoire, la bouche ou le palais mou détermine la couleur de divers sons de la parole. Ainsi, le chemin de la voix est considéré comme filtre pour diverses sources de production vocale telles que vibration des cordes vocales ou turbulence causée par le mouvement de l'air à travers spasmes des voies vocales (OUNI 2001) .

Le son résultant peut être classé comme exprimé ou non selon que l'air émis a été produit. faire vibrer les cordes vocales ou non. Dans le cas de sons vocaux, la fréquence de vibration les cordes vocales, appelées fréquence ou hauteur fondamentale, notées F0, s'étendent généralement entre 40 - 140 Hz pour les hommes, 180 à 300 Hz pour les femmes et 300 à 600 Hz pour enfants (LEMAN 2011) (FUCHS 2007).

1.2.2 Description acoustique du signal de parole

Le signal de parole véhicule des informations spécifiques à la personne qui l'émis, telles que le timbre, la manière de parler, l'état émotionnel ou pathologique, etc. Ces informations caractéristiques sur le locuteur peuvent être divisées en deux catégories distinctes :

- Des informations statiques telles que les paramètres spectraux caractérisant la voix et les voies nasales, la moyenne et les variations de la fréquence fondamentale.

- Informations à caractère dynamique, reflétant les phénomènes de coarticulation, les trajectoires d'entraînement et les informations temporelles (vitesse de parole, répartition des pauses) Ici, nous allons parler des caractéristiques statiques du signal vocal. Ce dernier peut être défini à l'aide de 4 paramètres principaux (Debbeche 2008) :

- **Intensité :** L'intensité du son correspond à l'amplitude de la vibration acoustique, elle caractérise le volume et permet de distinguer les sons forts des sons faibles.L'intensité de la voix humaine change principalement en fonction de la pression sous la glotte.
- **Timbre :** Le timbre peut distinguer deux sons de même hauteur et de même amplitude. Il se compose d'un ensemble de fréquences appelées spectre. La richesse du spectre indiquera que le son est riche, brillant, profond, etc. Le timbre dépend des trois critères suivants : l'état d'attache des cordes vocales, l'épaisseur des cordes vocales, et enfin les caractéristiques anatomiques de la cavité de résonance (pharynx, cavité buccale et cavité nasale).
- **Hauteur :** La hauteur dépend de la fréquence du changement de pression acoustique correspondant au son. Cela dépend de la périodicité du mouvement des lèvres de la glotte, c'est-à-dire en pratique du nombre d'ouvertures de la glotte par seconde. La hauteur dépend aussi de la taille du larynx : plus les cordes vocales sont longues, plus la voix est basse.
- **Fréquence :** Elle représente le nombre de fois que l'air vibre en une seconde.

1.3 Système de reconnaissance automatique de téléphone portable (RATP)

1.3.1 Tâches de système de RATP

Le système de reconnaissance de la marque du téléphone se base essentiellement dans son fonctionnement sur les caractéristiques spécifiques du signal vocal. Cette discipline

rentre dans le cadre général de la reconnaissance de formes. C'est un terme général qui regroupe des problématiques liées à l'identification ou à la vérification de la marque de téléphone en utilisant les informations contenues dans le signal sonore. Le champ d'application de ce système est très large, allant du simple contrôle d'accès, aux applications militaires, puis aux applications judiciaires.

1.3.1.1 Identification

L'identification est le processus qui consiste à déterminer, parmi un ensemble de modèles connus, celui qui génère un message donné. D'un point de vue schématique (voir figure 1.2), une séquence de parole est donnée en entrée du système, Pour chaque modèle connu du système, la séquence de parole est comparée à une référence caractéristique du modèle : identité dont la référence est la plus proche de la séquence de parole est donnée en sortie du système d'indentification. Deux modes sont proposés dans ce système :

- l'identification en ensemble fermé pour lequel on suppose que la séquence de parole est effectivement prononcée par un le modèle connu du système.
- l'identification en ensemble ouvert pour lequel le modèle peut ne pas être connu.

En mode « ensemble ouvert », le système d'identification doit décider de la fiabilité de son jugement en acceptant ou en rejetant l'identité qu'il a trouvé. De par son principe déterminait une identité parmi les identités potentielles.

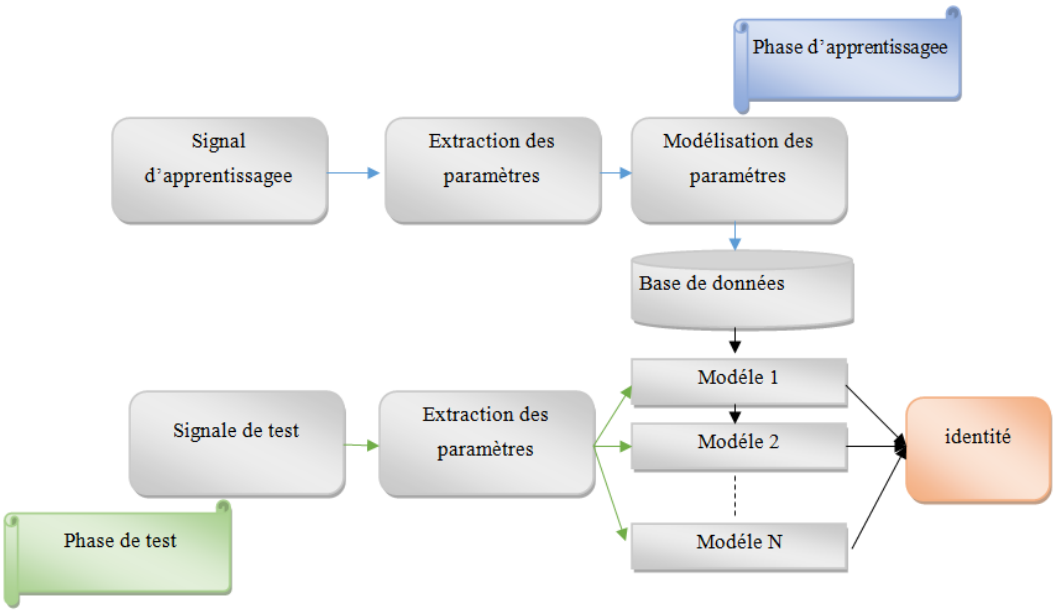

FIG. 1.2 : Structure du système d'identification automatique.

1.3.1.2 Vérification

La vérification est un processus décisionnel qui permet de déterminer la véracité de l'identité déclarée par le modèle au moyen d'un message vocal. L'identité et le message vocal sont les deux entrées du système . L'identité, nécessairement connue du système, détermine automatiquement une référence de modèle caractéristique. Une mesure de similarité est calculée entre cette référence et le message vocal puis comparée au seuil de décision. Lorsque la mesure de similarité est supérieure au seuil, le modèle est accepté. Dans le cas contraire, le modèle sera considéré comme une fraude et rejeté.

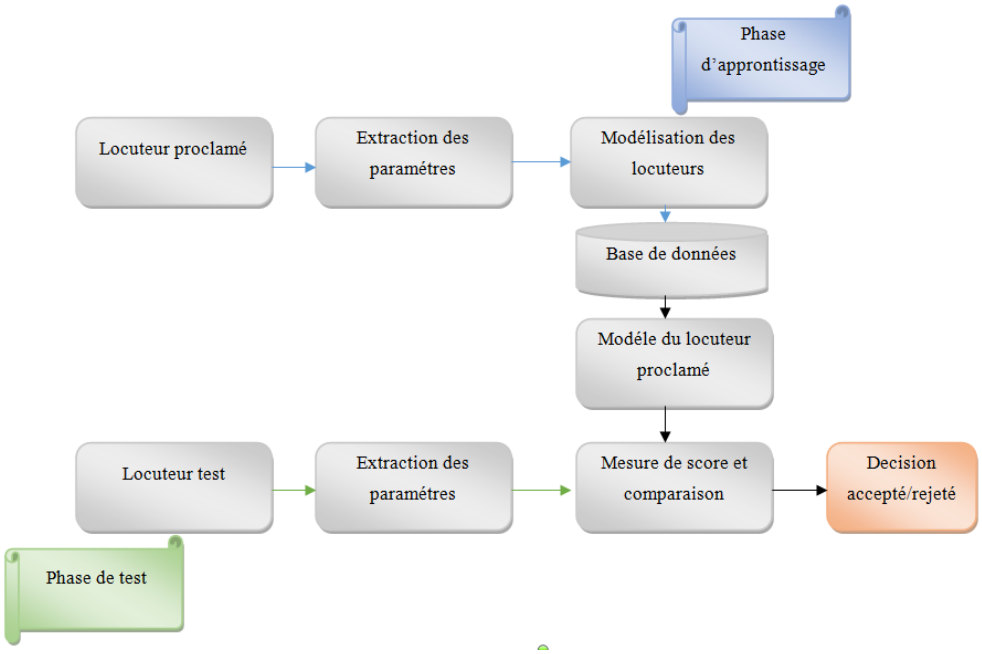

FIG. 1.3 : Structure du système de vérification automatique.

1.3.2 Phases de système de RATP

Comme tout système de reconnaissance de formes, un système de reconnaissance de la marque du téléphone portable travaille en deux modes : l'apprentissage et le test.

1.3.2.1 Phase d'apprentissage

A l'issue de cette étape, le signal est représenté par un vecteur coefficients appropriés pour la modélisation des paramétres.La modélisation est définie comme un processus de description des proprieteés des paramétre . Le modèle résultant de cette opération doit fournir des moyens pour sa comparaison avec un énoncé inconnu.

1.3.2.2 Phase de test

la mesure des similitudes sont calculées entre les paramètres acoustiques du signal de test parlé et les modèles d'enceintes présents dans la base de données. Enfin, un module sur une stratégie de décision donnée fournit la réponse du système.

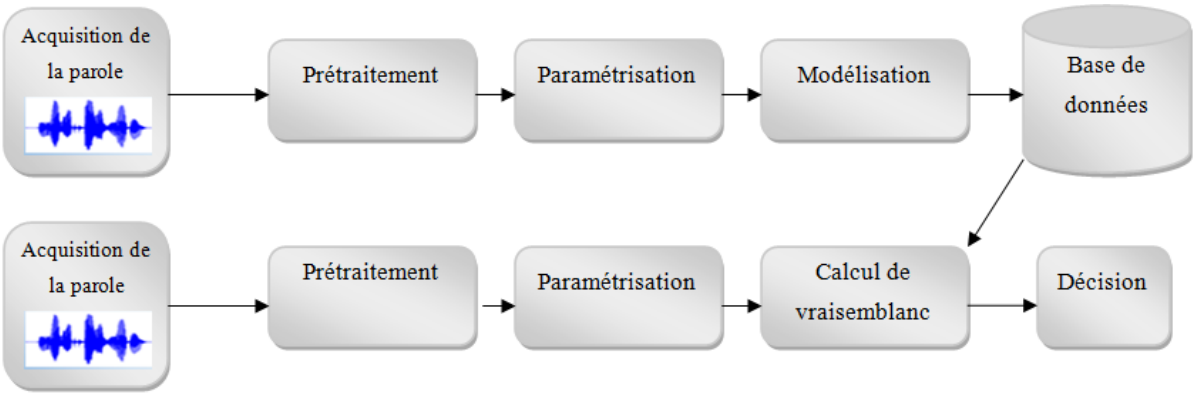

FIG. 1.4 : Étapes générales du SRA basé sur la parole.

1.3.3 Applications de système de RATP

Un intérêt croissant est accordé aux technologies basées sur la reconnaissance vocale dans les domaines public et industriel. En effet, le systéme de RATP intervient de nos jours dans un grand nombre d'applications dans le domaine criminalistique. Dans le cas où l'enregistrement de la parole est parfois la seule preuve accessible, l'authentification de cet enregistrement et la reconnaissance de la source de la voix enregistrée font l'objets de recherches intenses.

1.4 Architecture générale du système de RATP

Différents modules sont présents dans le système de RATP (figure (1.5)). Tout d'abord, le message vocal, capté par un microphone, est converti en signal numérique pour être, par la suite, analysé dans un étage d'analyse acoustique. Á l'issue de cette étape, le signal est représenté par un vecteur de coefficients pertinents pour la modélisation du locuteur. A la reconnaissance (phase de test) ; une mesure de similarité est calculée entre les paramètres acoustiques du signal de test prononcé et les modèles de locuteurs présents dans la base. En dernier lieu, un module de décision, basé sur une stratégie de décision donnée, fournit la réponse du système.

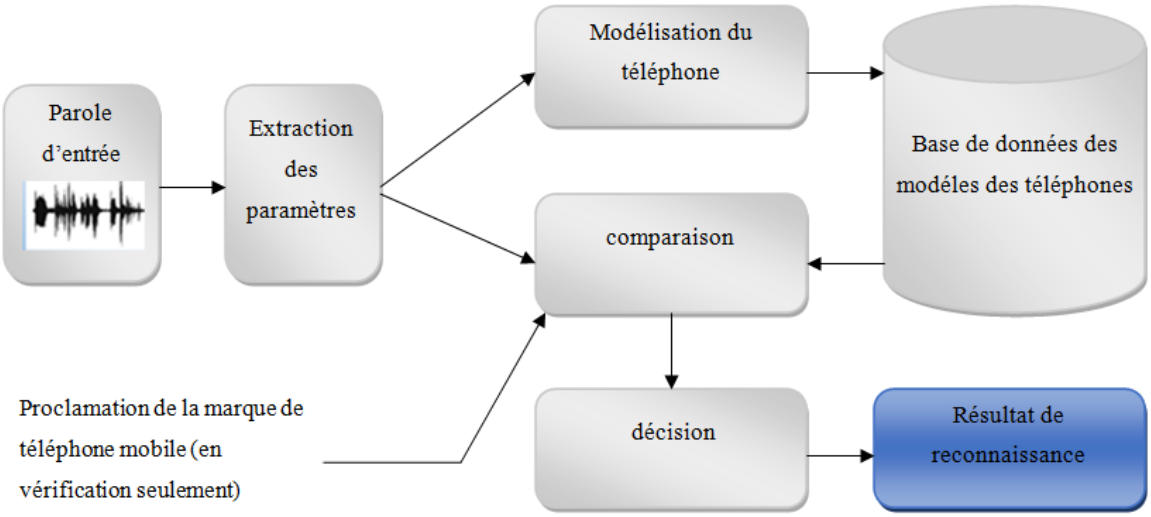

FIG. 1.5 : Système de reconnaissance de téléphone cellulaire.

1.4.1 Extraction des paramètres

1.4.1.1 Prétraitement

En général, avant de procéder à d'autres étapes d'extraction des paramètres, Le signal de parole subit des prétraitements qui incluent : la détection de la parole en retirantle silence avant des traitements ultérieurs, la segmentation qui vise à découper le signal de parole en petits segments (chaque segment dure 10 à 30 millisecondes), où il peut être considéré localement comme quasi-stationnaire.

Après une étape de numérisation du signal vocal, une préaccentuation (annuler l'effet des lèvres qui atténuer le signal sorti donc utiles pour garder les haut fréquence) peut être utilisé. Elle consiste à faire passer le signal d'entrée à travers un filtre passe-haut numérique de premier ordre à réponse impulsionnelle finie (FIR), donné comme suit (HARRINGTON et al. 1999) :

$$H(z) = 1 - az^{-1} \; avec \; 0.9 \leq a \leq 1 \tag{1.1}$$

1.4.1.2 Paramétrisation

Les caractéristiques vocales d'un individu peuvent ne pas être faciles à distinguer, mais ils sont uniques ; À titre d'exemple, prenons le cas des jumeaux monozygotes qui diffèrent par la voix mais la recherche montre qu'ils ont une forme de tractus vocal et des propriétés acoustiques similaires, et sont difficiles à distinguer perceptuellement. Ainsi, que la reconnaissance soit effectuée par des humains (par exemple un expert) ou par des machines, il y a certains aspects mesurables et prédéfinis de la parole qui doivent être pris

en compte afin de faire des comparaisons significatives entre les voix (Lippmann 1997) En général, nous nous référons à ces aspects en tant que paramètres de la caractéristique qui sont évoqués dans (Nolan 1982), ces paramètres doivent être :

- Difficiles à imiter par des imposteurs.
- Relativement faciles à extraire et à mesurer.
- Très fréquents dans le signal vocal.
- Robustes aux différents bruits et variations inter-sessions, ayant une grande variabilité inter-locuteurs et une faible variabilité intra-locuteur pour les considérer idéaux.

  Ces paramètres peuvent être :

- Spectraux à court terme extraits sur des petites fenêtres et reflètent les caractéristiques du conduit vocal.
- Prosodiques extraits sur des grands segments (de dizaines à des milliers de millisecondes),des paramètres représentatifs de la source vocale.
- De haut niveau (caractérisant la sémantique, l'accent, la prononciation, etc) des locuteurs .

1.4.1.3 Segmentation de la parole

Lorsque l'on écoute de la parole, on a l'impression qu'elle est composée d'un enchaînement de sons distincts. La séparation des mots à partir du flux de la parole est indispensable à la compréhension du discours et à l'accès lexical. La plupart des systèmes d'écriture possèdent des espaces blancs entre les mots. Dans le signal de parole, il n'existe pas d'indices clairs et univoques qui permettent de marquer le début et la fin des mots. La parole est dite « continue » : la segmentation est donc une étape majeure dans la reconnaissance des mots parlés. On peut définir plusieurs types de segmentation (organisés du segment le plus courtau segment le plus long) :

- la segmentation en sons voisés ou non voisés.
- la segmentation en phonèmes.
- la segmentation en syllabes.
- la segmentation en mots.
- la segmentation en locuteurs.

La tâche de segmentation de la parole est indispensable pour l'apprentissage des modèles acoustiques d'un système de reconnaissance de la parole et de synthèse vocale. Il existe deux formes de segmentation : La segmentation manuelle et la segmentation automatique (Laleye 2016).

1.4.2 Modélisation

C'est la tâche qui suit l'extraction des paramètres, qui se définit comme le processus de description des propriétés de ces derniers. Le modèle résultant de cette opération doit fournir un moyen de comparaison avec l'assertion inconnue.

En effet, les propriétés non idéales des caractéristiques extraites nécessitent différentes techniques de compensation lors de la phase de modélisation afin que l'impact des changements nuisibles observés dans le signal soit minimisé lors du processus de vérification.

La plupart des techniques de modélisation font des hypothèses mathématiques différentes sur les caractéristiques (par exemple, la distribution gaussienne). Si ces hypothèses ne sont pas vérifiées par les données, des imperfections sont introduites au stade de la modélisation .

1.4.3 Décision

Le SRA est conçus de telle manière que la comparaison entre le modèle et le signal test donne un résultat (valeur scalaire) indiquant si les deux entrées correspondent au même modèle. Dans le cas de l'identification la phase de décision détermine l'identité du modèle reconnu. En cas de vérification, cette décision est binaire et consiste à confirmer ou infirmer la conformité de la session de test avec l'identité revendiquée. Si ce résultat est supérieur (ou inférieur) au seuil prédéfini, le système accepte (ou rejette) ce qui a été saisi comme test.

1.5 Conclusion

Dans ce chapitre, nous avons présenté la structure générale d'un système de Reconnaissance Automatique du téléphone portable à savoir, le prétraitement, l'extraction, la modélisation et la décision et ses composants modulaires. Le chapitre 2 sera consacré à une description détaillée des deux étapes d'extraction et de modélisation.

Chapitre 2

De la paramétrisation à la modélisation des systèmes de reconnaissance portable

2.1 Introduction

La variation de la nature du signal acoustique rend le traitement des données brutes issues de ce dernier très difficile. En effet, ces données contiennent des informations complexes, souvent redondantes et mélangées au bruit. Le module de paramétrisation, qui traite le signal acoustique reçu, doit remplir plusieurs objectifs : séparer le signal du bruit ; extraire l'information utile à la reconnaissance et convertir les données brutes à un format directement exploitable par le système. Chacune de ces taches pose des problèmes complexes et influe fortement sur les résultats des systèmes automatiques de reconnaissance. Dans le présent chapitre, nous présentons quelques paramètres utilisés en reconnaissance automatique du téléphone portable. On se penchera ensuite sur le domaine audiovisuel qui se base sur des algorithmes de traitement d'image et on donnera une description de quelques paramètres visuels à savoir LBP et LPQ. par la suite nous allons présenter un module important de la reconnaissance automatique du locuteur. comme nous l'avons déjà mentionné dans le chapitre précédent, les systèmes RATP nécessitent une étape de modélisation des paramètres acoustiques. En effet, plusieurs méthodes de modélisation ont été utilisées et chacune d'elle présente des avantages et des inconvénients. Dans ce chapitre, nous nous focalisons davantage sur deux méthodes de la classification à savoir les modèles de Mélanges de Gaussiennes GMM et le I-Vecteur. Ces méthodes sont adoptées dans l'implémentation des systèmes de l'état de l'art actuel dans le domaine de reconnaissance.

Dans ce chapitre on va explorer les procédures nécessaires pour accomplir la tâche de reconnaissance de la marque du téléphone portable dans le sens large, comme le prétraitement du signal parole, les techniques d'extraction de paramètres acoustique et visuelle et les différents modèles utilisés (QV, GMM, HMM, I-VECTEUR). On se penchera ensuite dans ce travail sur le domaine audiovisuel qui se base sur des algorithmes de traitement d'image.

2.2 Extraction des paramètres acoustique

La variation de la nature du signal acoustique rend le traitement des données brutes issues de ce dernier très difficile. En effet, ces données contiennent des informations complexes, souvent redondantes et mélangées au bruit. Le module de paramétrisation, qui traite le signal acoustique reçu, doit remplir plusieurs objectifs : séparer le signal du bruit ; extraire l'information utile à la reconnaissance et convertir les données brutes à un format directement exploitable par le système. Chacune de ces taches pose des problèmes complexes et influe fortement sur les résultats des systèmes automatiques de reconnaissance. Dans le présent chapitre, nous présentons quelques paramètres utilisés en reconnaissance automatique du téléohone mobile.

2.2.1 Coefficients cepstraux en fréquence Mel (Mel-Frequency Cepstral Coefficients MFCC)

Ces paramètres sont les coefficients cepstraux calculés par la transformée en cosinus discrète DCT appliquée au coefficient d'énergie. L'analyse du banc de filtres à l'échelle Mel convertit le spectre de puissance du signal en coefficients d'énergie par bande de fréquence suivi par une compression logarithmique appliquée à ces coefficients.

Les étapes de cette extraction qui sont montrées dans la figure 2.1 sont détaillées juste après :

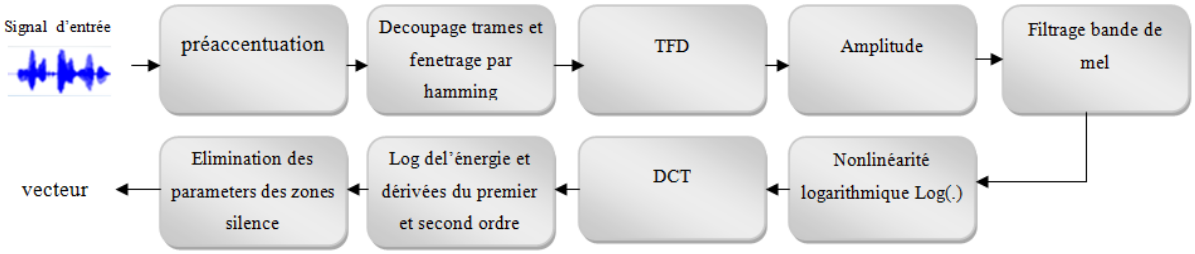

FIG. 2.1 : Extraction des paramètres MFCC.

La préaccentuation permet d'amplifier la partie haute de la fréquence.Pour des raisons citées précédemment (paragraphe 1.4.1.1) et dans le but de diminuer la distorsion spectrale, une multiplication par une fenêtre de Hamming (figure 2.2) (la plus adaptée pour la voix) est effectuée.

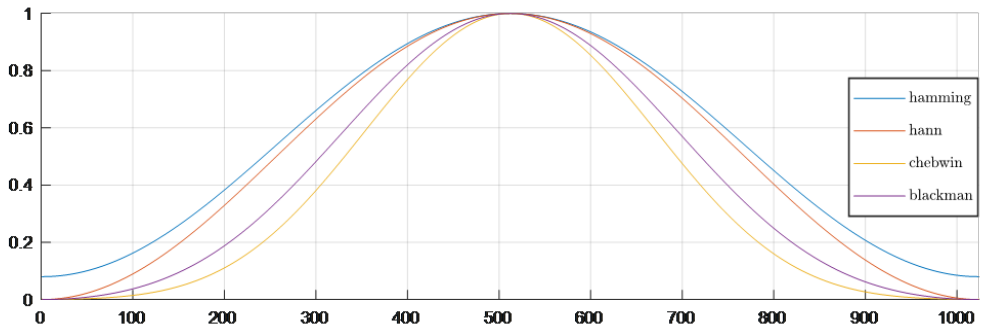

FIG. 2.2 : des examples des fenêtres.

Le signal parole est alors segmenté en trames tel que :

$$y(n,t) = x(n,t)\, w(n), n = 0,1,2,\ldots N-1\ ; t = 0,1,2,\ldots,T-1 \tag{2.1}$$

Avec $x(n,t)$: le signal parole original, N : nombre d'échantillons, T : nombre de trames,$w(n)$: est la fenêtre de Hamming donnée par :

$$w(n) = 0.54 - 0.46cos\left(\frac{2\pi}{N-1}n\right),\ 0 \leq n \leq N \tag{2.2}$$

- Entre les trames consécutives, un chevauchement, de 10ms dans notre travail, est appliqué afin de préserver le maximum des paramètres et éviter la discontinuité entre les trames.

- La deuxième étape consiste au passage du domaine temporel au domaine fréquentiel par la transformée de fourier discrète tel que :

$$Y(k,t) = \frac{1}{n}\sum_{n=0}^{M-1} y(n,t)\exp(\frac{-2\,\mathrm{jkn}}{N}), k = 0,1,\ldots,M-1\,; t = 0,1,\ldots,T-1. \quad (2.3)$$

  Dans le cas complexe, seulement la valeur absolue est considérée. Où la gamme de fréquence $0 \leq f \leq f_e$ Correspond à $0 \leq k \leq \frac{M}{2} - 1$ et la gamme de $-\frac{f_e}{2} \leq f \leq 0$ correspond à $\frac{M}{2} + 1 \leq k \leq M - 1$.

- L'application du module produit le spectre de puissance de chaque trame, et constitue alors la troisième étape du processus.

- Comme l'échelle de perception de fréquence de l'oreille humaine n'est pas linéaire, l'étape suivante,de traitement,consiste à filtrer le spectre de puissance du signal vocal à travers un ensemblede k filtres positionnés selon l'échelle de Mel (Figure 2.3)

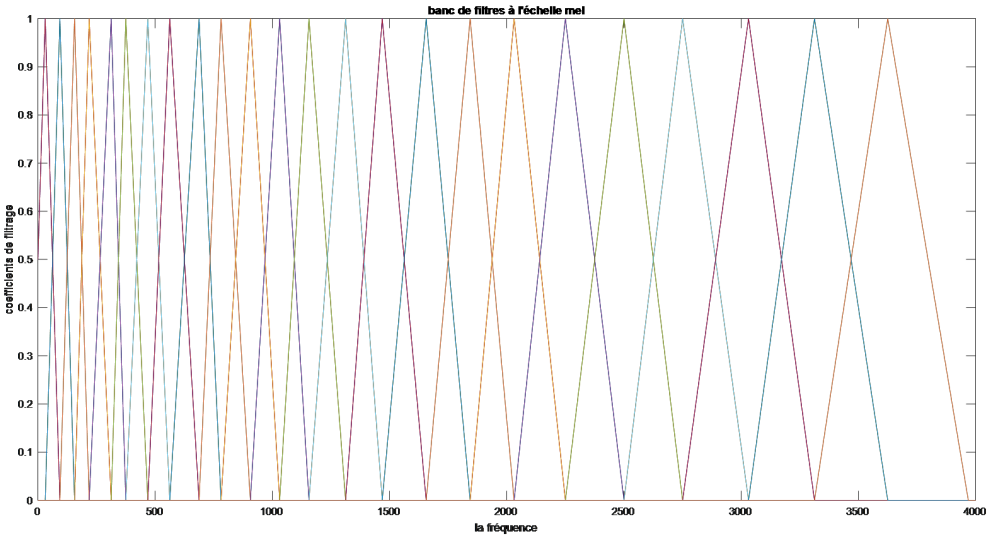

FIG. 2.3 : Filtre Mel.

L'échelle se compose d'une série de filtres passe-bande triangulaires pour un positionnement linéaire de basse fréquence (<1000 Hz) et un positionnement logarithmique haute fréquence. Le premier filtre indique que la quantité d'énergie présente au voisinage de 0 Hz est très étroite, et cette largeur est augmentée jusqu'aux hautes

fréquences. Cela simule le fonctionnement de l'oreille humaine, qui est insensible à ces fréquences et ne peut distinguer deux fréquences très proches. A l'issue de cette étape, un vecteur donnant une indication sur la quantitéd'énergie contenue dans chaque filtre résulte (Klautau 2005). Le passage à cette échelle se fait parl'équation suivante (Ponraj et al. 2016) :

$$m = \ln(1 + \frac{f}{700})\frac{1000}{\ln(1 + \frac{1000}{700})} \tag{2.4}$$

Le problème de la reconnaissance est la convolution de la source et du canal (conduit vocal). Afin d'effectuer une déconvolution pour éliminer l'influence de la source, l'analyse cepstre convertit la multiplication en une sommation en appliquant un logarithme dans le domaine fréquentiel (Klautau 2005).

- La dernière étape de ce processus est l'application de la transformée en cosinus discrète (DCT) sur le vecteur E_W d'énergies spectrales logarithmiques pour une éventuelle compression de l'information (Klautau 2005). L'ensemble de ces coefficients cepstraux produits est appelé vecteur acoustique. L'équation de la DCT est la suivante :

$$C_m = \sum_{w=1}^{k} \cos\left(m\,(w - 0.5)\,\frac{\pi}{k}\right) E_w \ , m = 1, 2, \ldots, L \tag{2.5}$$

Avec k le nombre de filtres passe-bande triangulaires, L le nombre de coefficients cepstrauxàl'échelle de Mel.

2.2.2 Analyse par prédiction linéaire perceptuelle (Perceptuel Linear Prediction PLP)

Cette méthode prend en compte la perception humaine de la parole. Elle utilise des connaissances issues de la psychoacoustique lors de l'estimation d'un modèle autoregressif (AR), à savoir, une résolution non linéaire en fréquence à l'aide de bandes critiques sur une échelle de Bark, une préaccentuation du signal selon une courbe d'isotonie, et une compression en racine cubique pour simuler la loi de perception humaine en puissance sonore. Le spectre résultant est sujet finalement à une modélisation Autorégressive (Hermansky 1990).

La méthode PLP est inspirée du principe de prédiction linéaire (LPC). Elle combine ce principe à une représentation du signal qui suit l'échelle humaine de l'audition. La figure 2.4 résume ce principe. Le spectre est ensuite modifié par une interpolation et une transformée de Fourier inverse, enfin le signal obtenu étant passé dans un filtre pour réduire la dimension du spectre et augmenter la résolution fréquentielle (AJGOU 2016).

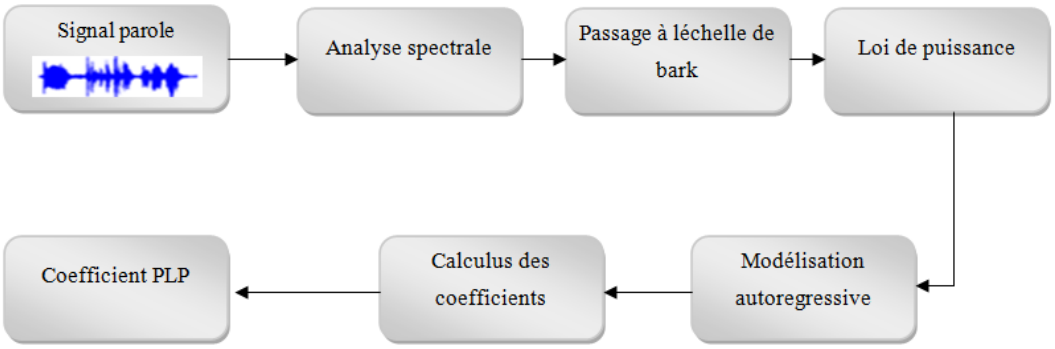

FIG. 2.4 : Méthode de calcul des coefficients PLP.

2.2.3 Coefficients cepstraux Q constants (constant Q cepstral coefficients CQCC)

Les caractéristiques des CQCC sont dérivées à l'aide de la tranformée CQT (Constant Q Transform), qui est un outil d'analyse fréquence-temps piloté par la perception et qui est une alternative à la transformée de Fourier à court terme (Short Time Fourier Transform STFT). Alors que la STFT fonctionne avec une résolution temporelle spectrale fixe, la résolution temporelle spectrale de la CQT est variable.La résolution de fréquence est plus élevée aux fréquences les plus basses et la résolution temporelle est plus élevée aux fréquences les plus élevées. Comme pour les coefficients MFCC traditionnels, l'extraction CQCC est effectuée avec un ensemble de filtres, où le facteur Q est une mesure de la sélectivité de chaque filtre, et définie comme la fréquence centrale du filtre et sa bande passante(DELGADO et al. 2018).

2.2.3.1 Étapes de traitement du signal vocal pour le processus d'extraction de caractéristiques (CQCC)

Le processus d'extraction de paramètres par CQCC comprend une série d'étapes de traitement du signal, y compris la conversion de signaux vocaux, variant dans le temps,au domaine fréquentiel, le ré-échantillonnage pour changer les cellules espacées géométriquement dans les cellules espacées linéairement (MITTAL et al. 2021).

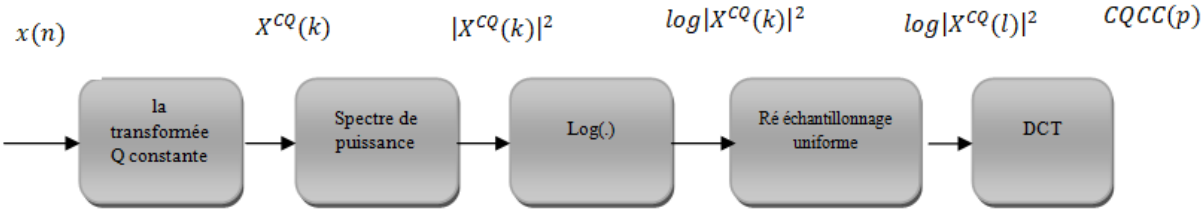

FIG. 2.5 : Extraction des paramètres CQCC.

2.2.3.1.1 Calcul de la transformée Q constante (CQT) Le domaine temporel du signal vocal peut être converti en domaine fréquentiel pour une utilisation plus facile. Tout d'abord, la transformée à Q constant (CQT) est appliquée au signal vocal initial. CQT peut convertir le domaine temporel en domaine fréquentiel tout en maintenant le facteur Q constant tout au long du signal (Mittal et al. 2021).

Le CQT $X^{\mathrm{CQ}}(k,n)$ d'un signal de domaine temporel discret $x(n)$ est défini par :

$$X^{\mathrm{CQ}}(k,n) = \sum_{j=n-\left\lfloor \frac{N_k}{2} \right\rfloor}^{n+\left\lfloor \frac{N_k}{2} \right\rfloor} x(j)\, a_k^*(j-n+\frac{N_k}{2}) \tag{2.6}$$

où $k = 1,\ 2,\ ...,K$ est l'indice des barres de fréquence (Bins frequency), $a_k^*(n)$ est le conjugué complexe de $a_k(n)$ et N_k sont des longueurs variables des fenêtres. La notation$\lfloor \bullet \rfloor$ représente l'arrondissement vers l'entier inférieur le plus proche. Les fonctions de base $a_k(n)$ sont des atomes de fréquence-temps à valeur complexe, définis selon (Todisco et al. 2016) :

$$a_k(n) = \frac{1}{C}\left(\frac{n}{N_k}\right) \exp\left[i\left(2\pi n \frac{f_k}{f_s} + \Phi_k\right)\right] \tag{2.7}$$

où f_k est la fréquence centrale de la barre k, f_s est le taux d'échantillonnage, et $w(t)$ est une fonction de fenêtre (par ex. fenêtre Hann). Φ_kEst un décalage de phase. Le facteur d'échelle C est donné par (Todisco et al. 2016) :

$$C = \sum_{l=-\left\lfloor \frac{N_k}{2} \right\rfloor}^{\left\lfloor \frac{N_k}{2} \right\rfloor} w\left(\frac{l-\frac{N_k}{2}}{N_k}\right) \tag{2.8}$$

Comme un espacement de barre correspondant à l'échelle de revenu égal est souhaité, les fréquences centrales f_k obéissentà (Todisco et al. 2016) :

$$f_k = f_1 2^{\frac{k-1}{B}} \tag{2.9}$$

oùf_1 est la fréquence centrale de la cellule de fréquence la plus basse et B détermine le nombre de cellules par octave. En pratique, B détermine le compromis de résolution de fréquence-temps. Lefacteur Q est alors donné par (Todisco et al. 2016) :

$$Q = \frac{f_k}{f_{k+1}-f_k} = (2^{\frac{1}{B}}-1)^{-1} \tag{2.10}$$

Les longueurs de fenêtre $N_k \in \mathbb{R}$ dans les équations 2.9 et 2.10 sont réallouées et inversement proportionnelles à f_k afin que Q soit constant pour toutes les barres de fréquence$k$,c'est-à-dire (Todisco et al. 2016) :

$$N_k = \frac{f_s}{f_k} Q \tag{2.11}$$

Les travaux ont introduit un paramètre supplémentaire γ qui diminue progressivement les facteurs Q pour les cellules de basse fréquence en accord avec les filtres du système auditif humain. En particulier, lorsque $\gamma = 228.7 * (2^{(\frac{1}{B})} - 2^{(\frac{-1}{B})})$ les largeurs de bande correspondent à une fraction constante de la bande passante critique de l'ERB (Todisco et al. 2016) **.**

2.2.3.1.2 Squaring et Log Opération sur Spectrum Après l'application, la puissance CQT du spectre est calculée qui est suivie par le fonctionnement logarithmique du spectre. Avant de trouver, le logarithme absolu des valeurs au carré est pris (Mittal et al. 2021).

2.2.3.1.3 Ré-échantillonnage Le ré-échantillonnage est une étape importante dans la technologie d'extraction de fonction CQCC, car CQT fournit des unités de fréquence géométriquement espacées, qui sont converties en unités linéairement espacées en appliquant le ré-échantillonnage (Mittal et al. 2021).

Étant donné que les k barres sont espacées géométriquement, la reconstruction du signal peut être considérée comme une opération de sous-échantillonnage sur les premiers$k$ barres (basse fréquence) et comme opération de suréchantillonnage pour les k barres restantes (haute fréquence). Nous définissons la distance entre $f_k$et $f_1 = f_{\min}$comme suit (Todisco et al. 2016) :

$$f^{k \leftrightarrow 1} = f_k - f_1 = f_1 \left(2^{\frac{k-1}{B}} - 1\right) \tag{2.12}$$

où $k = 1,\ 2,\ ..., K$ est l'indice de tranche de fréquence. La distance $f^{k \leftrightarrow 1}$ augmente en fonction de k. On cherche maintenant une période T_l pour le ré-échantillonnage linéaire. Cela revient à déterminer une valeur de $k_l \in 1,\ 2,\ ..., K$ telle que (Todisco et al. 2016) :

$$T_l = f^{k_l \leftrightarrow 1} \tag{2.13}$$

Pour résoudre 2.13 ,il suffit de se concentrer sur la première octave ; une fois T_l fixé pour cette octave, les octaves supérieures auront naturellement une résolution deux fois supérieure à celle de l'octave inférieure. Une résolution linéaire est obtenue en divisant la

première octave en d parties égales de période T_l et en résolvant pour k_l (Todisco et al. 2016) :

$$\frac{f_1}{d} = f_1\left(2^{\frac{k_l-1}{B}} - 1\right) \rightarrow k_l = B\ log_2\left(1 + \frac{1}{d}\right) \tag{2.14}$$

Le nouveau taux de fréquence est alors donné par (Todisco et al. 2016) :

$$F_l = \frac{1}{T_l} = \left[f_1\left(2^{\frac{k_l-1}{B}} - 1\right)\right]^{-1} \tag{2.15}$$

Il existe donc des échantillons uniformes d dans la première octave, $2d$ dans la seconde et $2^j d$ dans la $(j-1)$ ème octave. L'algorithme de reconstruction du signal utilise un filtre anti aliasing polyphasique et une méthode d'interpolation « spline » pour ré-échantillonner le signal à la vitesse d'échantillonnage uniforme F_l. Les coefficients Q cepstral constants (CQCCs) peuvent alors être extraits de manière plus ou moins conventionnelle selon Todisco et al. 2016 :

$$\mathrm{CQCC}\,(p) = \sum_{l=1}^{L} \log\left|X^{\mathrm{CQ}}(l)\right|^2 \cos\left[\frac{p\left(l-\frac{1}{2}\right)\pi}{L}\right] \tag{2.16}$$

Où $p = 0,\ 1,\ ...,L-1$ et les l sont les barres de fréquence nouvellement ré-échantillonnées.

2.3 Extraction des Paramètres visuels

Les étapes d'extraction des paramétres visuels sont illustrés dans La figure 2.6 :

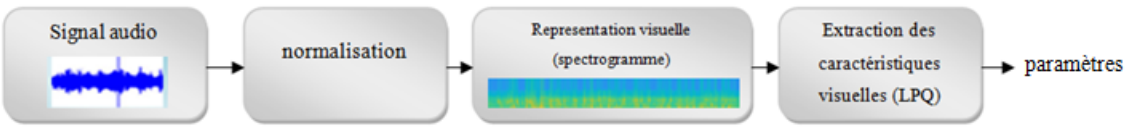

Fig. 2.6 : Extraction des paramètres visuels.

2.3.1 Normalisation

Une technique très simple et très répandue en R vocale, qui consiste à retirer la moyenne (la composante continue) de la distribution de chacun des paramètres, et à ramener l'amplitude à une amplitude unitaire dans intervalle [-1,1] (équation 2.17) (Ji et al. 2021) :

$$s = \frac{s - \mathrm{mean}(s)}{max(\mathrm{abs}\,(s))} \tag{2.17}$$

2.3.2 Spectrogramme

Le spectrogramme montre l'intensité du signal au fil du temps à différentes fréquences de la forme d'onde. Le spectrogramme peut être un graphique à deux dimensions avec une troisième variable représentée par la couleur (voir figure 2.7.a) ou un graphique à trois dimensions avec une quatrième variable de couleur (voir figure 2.7.b) (Kamp 2020).

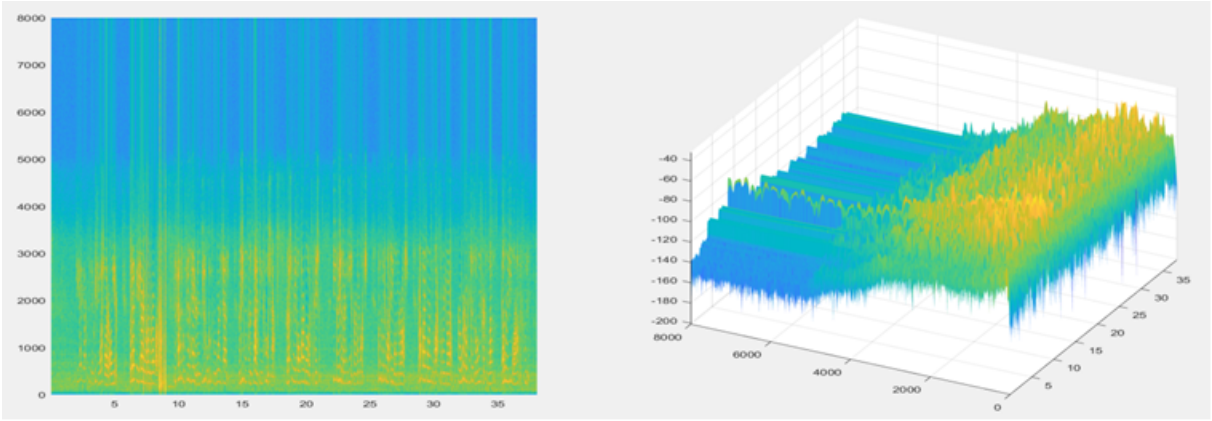

Fig. 2.7 : Spectrogramme en 2D (a) et 3D (b).

2.3.3 Extraction des paramètres visuels (Local Phase Quantization LPQ)

Cette technique est appliquée pour extraire un motif de texture local invariant de flou et d'illumination. La LPQ a été proposée par (Ojansivu et al. 2008) et est basée sur la propriété d'invariance du flou du spectre de phase de Fourier. LPQ prend un voisinage rectangulaire autour de chaque pixel de l'image pour calculer la transformée de Fourier à court terme 2D (STFT) qui donne l'information de phase locale de l'image.

Le flou d'image est une méthode pour réduire le contenu des bords de l'image pour faire une transition en douceur d'une couleur à un autre. Le flou d'image est une fonction spatiale, notée $g(x)$donnée par la convolution entre l'image d'origine $f(x)$ et une fonction d'étalement de points (PSF) $h(x)$. Dans le domaine fréquentiel, cela est représenté par :

$$G\ (u)\ =\ F\ (u).\ H\ (u) \tag{2.18}$$

$G\ (u), F\ (u)$ et $H\ (u)$ sont les transformées de Fourier discrètes (DFT) de l'image floue, de l'image originale et de la PSF respectivement. u désigne l'ensemble des coordonnéesvectorielles $[u,\ v]\ T$. Les composantes de l'ampleur et de la phase peuvent être

Séparés et représentés comme suit :

$$|G\ (u)| = |F\ (u)|.|H\ (u)\ | \quad < G\ (u)\ = < F\ (u) + \ < H\ (u) \tag{2.19}$$

Ici, $< G\ (u)$ représente la phase de $G(u)$.

Lorsque la PSF de la fonction est à symétrie centrale, sa transformée de Fourier H est toujours réelle, c'est-à-dire :

$$< H\,(u) = \begin{cases} 0 \text{ si } H\ (u) \geq 0 \\ \pi \text{ si } H\ (u) < 0 \end{cases} \tag{2.20}$$

L'équation 2.20 montre la propriété d'invariance du flou, c'est-à-dire $< G\ (u)\ = < F\ (u)$ lorsque $H(u)\ =\ 0$. LPQ extrait les informations de phase. En examinant le voisinage local $N_x$de taille M × M à chaque position de pixel x de l'image $f(x)$:

$$F\,(u,x) = \sum_{y \in N_x} f\,(x-y)\, e^{-2\mathrm{j}\, u^T y} = w_u^T f_x \tag{2.21}$$

Ici w_u est le vecteur de base pour la DFT 2-D à la fréquence u et f_x est un vecteur contenant tous les échantillons M^2de N_x. Les coefficients de Fourier locaux sont calculés à quatre points de fréquence $u1\ =\ [a,\ 0]^T, u2\ =\ [0,\ a]^T, u3\ =\ [a,\ a]^T$, et $u4\ =\ [a,\ -a]^T$, où a est la première fréquence en dessous des premiers passages à zéro de $H(u)$ qui satisfait $< G(u)\ =< F\ (u)$ pour tous $H(u) \geq 0$. Pour chaque position de pixel, il en résulte un vecteur :

$$F_x^c = [F\,(u1,x)\,, F\,(u2,x)\,, F\,(u3,x)\,, F\,(u4,x)] \tag{2.22}$$

L'information de phase dans les coefficients de Fourier est enregistrée en observant les signes des parties réelles et de l'imaginaires de chaque composant en F_x^c .Ceci est réalisé en utilisant un quantizer scalaire simple $q_j\ (x)\ =\ 1$, si $g_i\,(x) \geq 0\ et\ 0$ sinon où $g_i\,(x)$ est la composante j^{th}du vecteur $G_x =$ [Re $\{F_x^c\}\,,\ Im\ \{F_x^c\}$] .Le résultat quantifiéles coefficients $q_j\ (x)$ sont représentés sous forme de valeurs entières entre 0 et 255 en utilisant le codage binaire $b = \sum_{j=1}^{8} q_j 2^{j-1}$.

L'histogramme de la valeur entière ci-dessus est utilisé comme vecteur caractéristique.

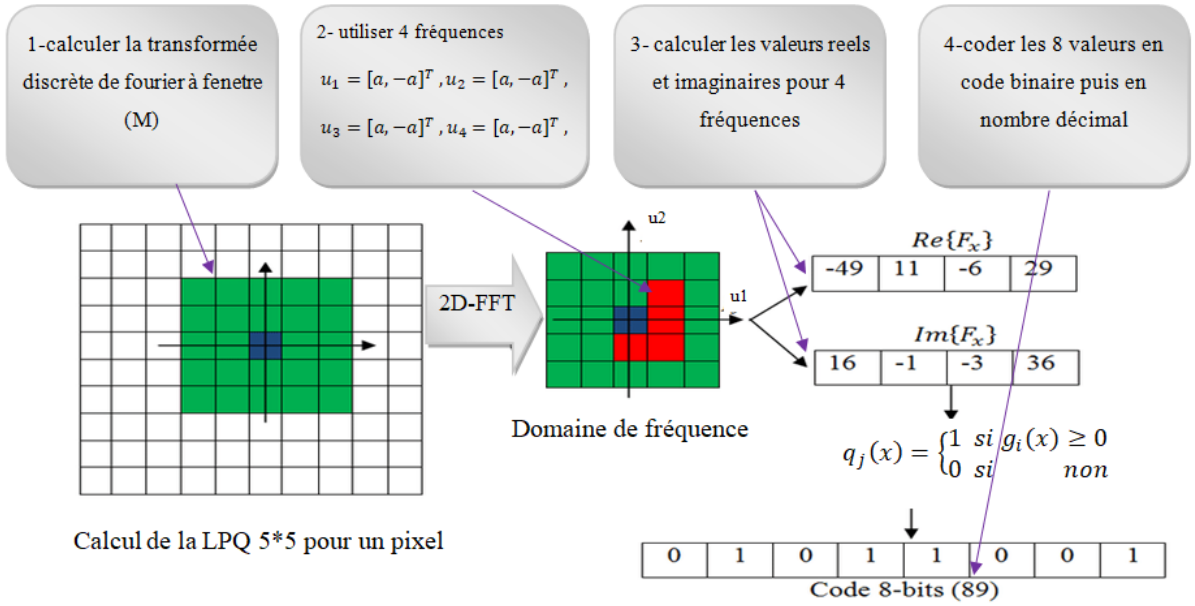

FIG. 2.8 : Organigramme de l'ensemble des étapes nécessaires du descripteur LPQ.

2.3.4 Extraction des paramètres visuels (Local Binery Patterns LBP)

Une autre méthode qui s'appele l'opérateur LBP a été initialement proposé par (OJALA et al. 2002) afin d'exprimer la texture des patchs de l'image.En tant que méthode d'extraction de caractéristiques locales, elle a été largement utilisée dans divers algorithmes de systèmes de reconnaissance faciale (YUAN et al. 2012) .L'opérateur LBP de base attribue un motif binaire à chaque pixel. Le LBP de l'image de pixel est traité en seuillant le voisinage 3×3 du pixel central (si la valeur du pixel central est supérieure à la valeur de ses pixels voisins) et en le transmettant sous forme de code binaire, puis en le convertissant en un nombre décimal. Après cela, l'opérateur est étendu pour utiliser des voisinages R de différents rayons et différents points d'échantillonnage P,de sorte que des caractéristiques d'échelles différentes puissent être extraites (HADID et al. 2014) .

Les valeurs des niveaux de gris de 3×3 pixels et le code LBP est calculé en utilisant la formule suivante :

$$\mathrm{LBP}\,(x_c, y_c) = \sum_{n=0}^{7} S\,(x)\,(i_n - i_c)\,2^n \tag{2.23}$$

$S\,(x)$Est la fonction de seuillage, donnée par :

$$S\,(x) = \begin{cases} 1 \text{ si } (x \geq 0) \\ 0 \text{ si } (x < 0) \end{cases} \tag{2.24}$$

Ici x_c et y_c montrent la position du pixel central, i_n et i_c sont des valeurs des niveaux de gris des pixels environnants et du pixel central respectivement (Hussain et al. 2012)

Le LBP étendu sélectionne les pixels voisins comme un ensemble de points d'échantillonnage répartit uniformément le long d'un cercle avec comme centre le point i_c pixel central) et un rayon R comme représenté dans la figure 2.9.

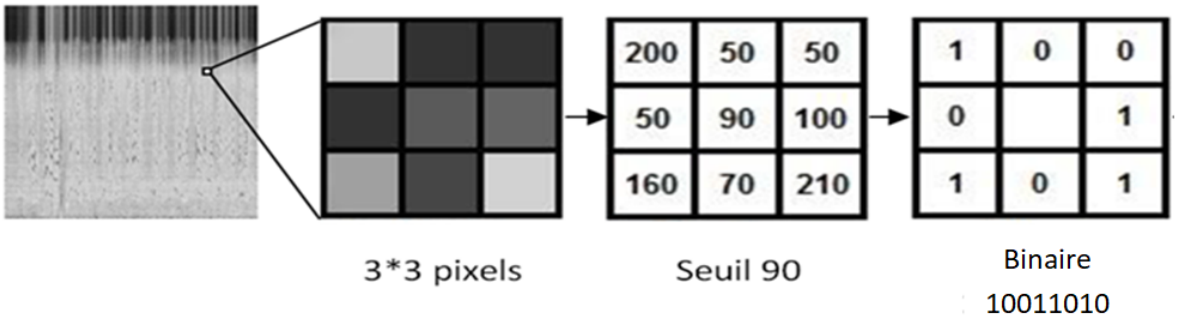

Fig. 2.9 : Illustration de LBP basique.

Dans la littérature la notion LBP est généralement utilisée pour désigner l'opérateur LBP basique, tandis que la notion LBP $P.R$ est utilisée pour représenter l'étendue LBP où, l'indice P représente le nombre des points d'échantillonnage et l'indice R représente le rayon du cercle.

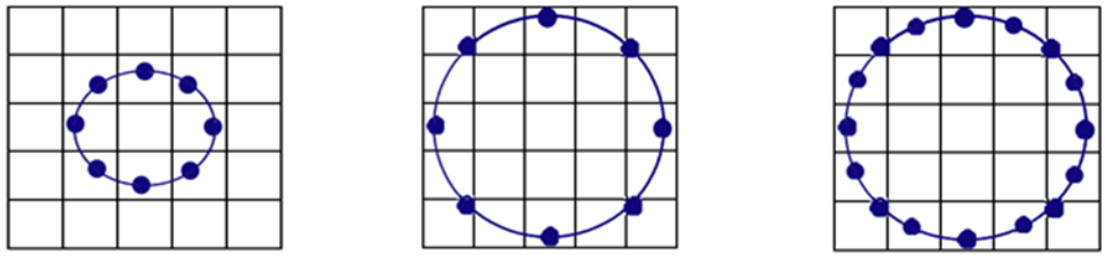

Fig. 2.10 : Exemples de l'operateur LBP P.R

.

Afin d'exprimer les caractéristiques de l'empreinte par la méthode LBP, nous considérons d'abord un voisinage carré, et la valeur du niveau de gris du pixel central est utilisée comme seuil des 8 pixels adjacents. Après avoir scanné tous les pixels de l'image, l'histogramme calculé de l'image est généré. Cet histogramme représente le vecteur des caractéristiques de l'image. Il est à noter, qu'il existe plusieurs variantes de cette méthode. La méthode utilisée dans notre travail est la variante de base (le cas le plus simple) la figure 2.11.

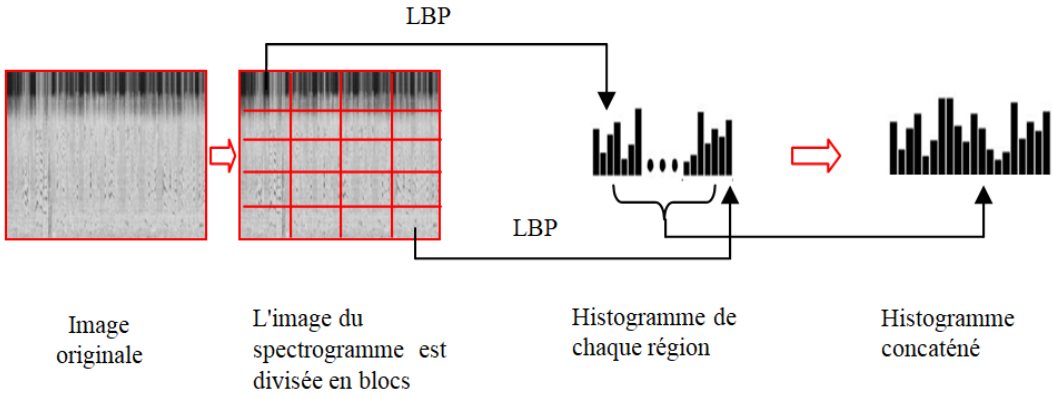

FIG. 2.11 : Histogramme global pour la représentation d'un spectrogramme a base de LBP

.

2.4 Modélisation

La modélisation vise à créer une référence qui représente chaque objet (ici il s'agit de téléphones mobiles). Tous les enregistrements prennent en compte la dépendance temporelle entre les vecteurs de paramètres extraits. Par conséquent, il est possible de prédire l'alignement temporel des séquences vectorielles d'apprentissage et de test car elles doivent contenir la même séquence. Cependant, dans les applications qui ne sont pas liées au texte, seule la distribution des paramètres acoustiques est modélisée. Les techniques de modélisation peuvent être dérivées de diverses méthodes générales telles que les vecteurs, le connexionnisme, les prévisions et les méthodes statistiques (JOURANI 2012). Cette section détaille plusieurs approches fonctionnant dans le bloc de modélisation.

2.4.1 Quantification vectorielle (Vector Quantization VQ)

La quantification vectorielle (VQ) divise l'espace acoustique en sous-espaces. Chaque sous-espace est associé à son vecteur centroïde, c'est-à-dire à un vecteur paramètre qui représente un ensemble de vecteurs qui composent le sous-espace. Dans ces conditions, le modèle de téléphone mobile est constitué d'un ensemble de vecteurs centroïdes, appelé dictionnaire de quantification (codebook). Dans la phase de reconnaissance, la distance entre le vecteur de test et chaque vecteur du centroïde du dictionnaire est calculée. La distance minimale est retenue. La quantification vectorielle est utilisée dans des modes liés au texte ou indépendants (SOONG et al. 1992), (MASON et al. 1989) , (MATSUI et al. 1994) . La vitesse et les performances de cette technique dépendent en grande partie de la taille du dictionnaire : plus la taille du dictionnaire augmente, meilleures sont les performances. Cependant, ce processus devient plus lent.

2.4.1.1 K-moyennes (K-means)

C'est l'un des algorithmes de clustering les plus populaires. Il permet l'analyse d'un ensemble de données caractérisé par un ensemble de descripteurs, afin de réassembler des groupes (cluster) de données "similaires". La similitude entre deux données peut être déduite par la « distance » entre leurs descripteurs ; donc deux données très similaires sont deux données avec des descripteurs très similaires. Cette définition permet de poser le problème du partitionnement des données telles que la recherche de K "données prototypes", autour d'autres données pouvant être regroupées. Ces données prototypes sont appelées centroïdes ; En fait, l'algorithme associe chacun à son centre le plus proche, pour créer des clusters.

D'autre part, les moyens de ses descripteurs de données précisent la position de leur centroïde dans l'espace : c'est l'origine du nom de cette algorithme. Après avoir initialisé ses centres en prenant des données aléatoires dans l'ensemble de données, K-means entrelacera certaines de ces deux étapes pour optimiser les centres et leurs groups :

1. Regrouper chaque objet autour du centroïde le plus proche.

2. Replacer chaque centroïde selon la moyenne des descripteurs de son groupe.

Après quelques itérations, l'algorithme trouve une répartition des données stable : on dit que l'algorithme a convergé.

Comme tout algorithme, K-means présente des avantages et des inconvénients : il est simple, rapide et facile à comprendre ; Cependant il ne permet pas de trouver des groupes ayant des formes complexes (Stéphanie 2021) .

2.4.2 Modèles de Markov caches(Hidden Markov Models HMM)

Le modèle de Markov a été largement utilisé dans la reconnaissance automatique de la parole. Récemment, leur utilisation a été étendue à la reconnaissance automatique du locuteur. Dans ce cas, la modélisation se fait à travers une série d'états avec des probabilités de transition d'un état à un autre. La reconnaissance se fait en calculant la probabilité qu'une série de vecteurs de test provenant d'une chaîne de Markov. L'utilisation de HMM en mode texte fournit d'excellents résultats (Rosenberg 1992). Le modèle HMM est utilisé pour modéliser un processus aléatoire qui change au fil du temps. Pour cela, ils ont combiné les caractéristiques de la distribution de probabilité et de la machine à états. Les vecteurs acoustiques seront utilisés comme observations dans les modèles de Markov cachés HMM. Le but de HMM est de trouver la meilleure séquence de mots du lexique pour définir des mots reconnaissables et la grammaire pour déterminer les séquences de mots correctes ou au moins les plus probables. HMM est une collection d'états et de transitions

entre eux. Le nom du modèle de Markov caché vient du fait que le chemin emprunté par le processus aléatoire, modélisé par HMM, est inconnu car les états qu'il traverse ne sont pas directement observables (Aziza 2013).

2.4.3 Modèles de mélange gaussiennes (Gaussian Mixture Model GMM)

Cette approche consiste à modéliser un téléphone portable par un mélange de gaussiennes qui représente une somme pondérée de M gaussiennes multidimensionnelles. Chaque gaussienne gi est supposée modéliser un ensemble de classes acoustiques. Elle est caractérisée par son poids w_m, un vecteur moyen m de dimension d et une matrice de covariance Σ_m de dimension D x D. La fonction de densité de probabilité s'écrit par :

$$P\,(x|\,\gamma) = \sum_{m=1}^{M} w_m N(x|\,\mu_m, \Sigma_m\,) \tag{2.25}$$

$$N(\,x\,|\mu_m, \Sigma_m\,) = \frac{1}{(2\pi)^{\frac{D}{2}}\left|\sum_m^{\frac{1}{2}}\right|} \exp\left(-\frac{1}{2}(x-\mu_m)^T \sum_m^{-1}(x-\mu_m)\right) \tag{2.26}$$

$$\sum_{m=1}^{M} w_m = 1 \tag{2.27}$$

L'apprentissage du modèle GMM comprend l'utilisation de l'ensemble de données d'apprentissage X = { X1, X2, ..., XT } pour estimer tous les paramètres. Ce type d'apprentissage nécessite généralement la technique d'estimationdu maximum de vraisemblance MLE (Maximum Likelihood Estimation) (Jourani 2012) .

Le principal inconvénient de cette technique est le nombre de signaux d'apprentissage requis pour une bonne estimation des paramètres du modèle.

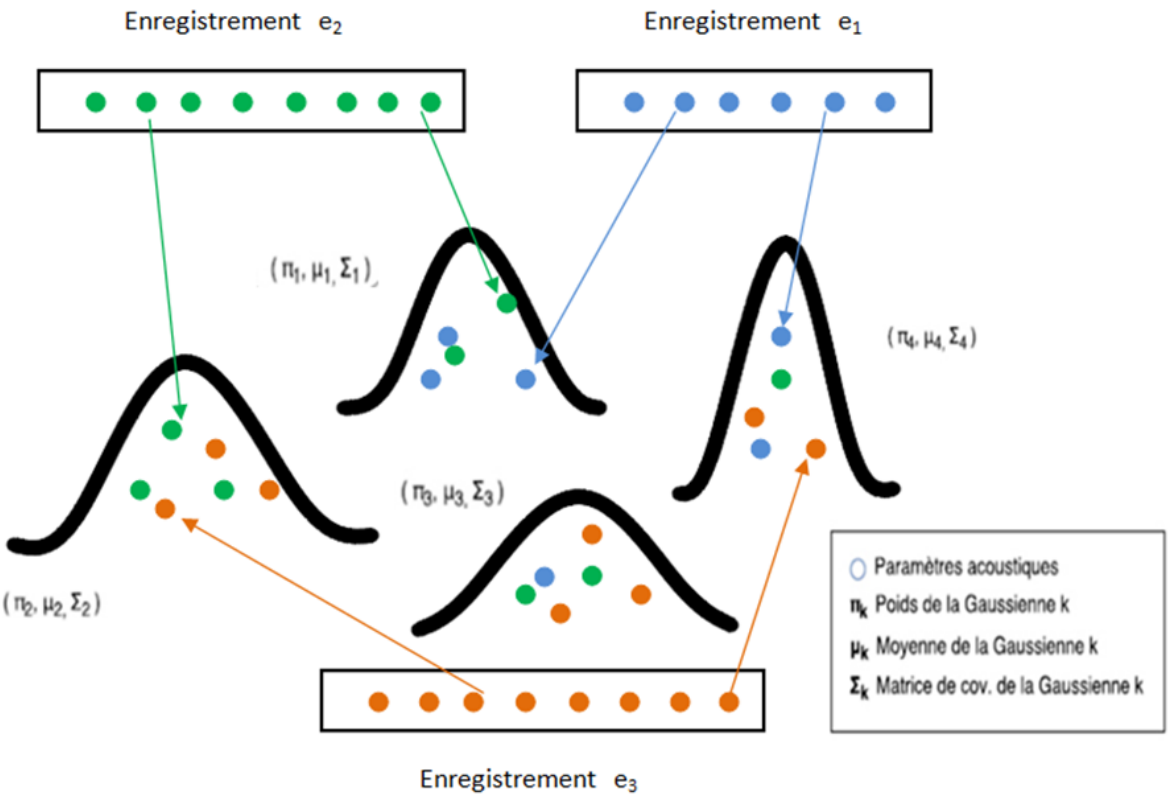

FIG. 2.12 : Mélange de Gaussiennes (GMM) construit en utilisant des paramètres acoustiques issus de plusieurs enregistrements

.

2.4.4 Approche GMM-UBM (Gaussien Mixture Model-Universel Background Model)

Une version améliorée du modèle GMM a été proposée pour surmonter le problème de données insuffisantes. La description et le schéma de principe de l'utilisation de la méthode GMM-UBM pour modéliser les téléphones mobiles sont les suivants (Figure 2.13) :

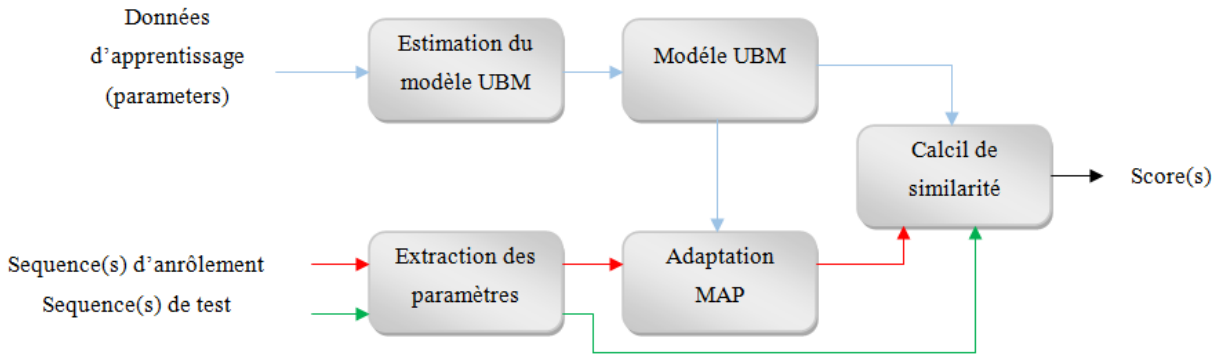

FIG. 2.13 : Architecture du système RA à base de GMM-UBM.

.

2.4.4.1 Estimation du modèle du monde UBM

1. Un seul modèle indépendant des téléphone λ_{UBM} , appelé modèle du monde (Universel Background Model UBM), est utilisé. L'apprentissage du modèle UBM se fait en utilisant un ensemble large de données de parole obtenu par la concaténation d'un large épouvantail detéléphones.

2. Un algorithme itératif appelé Expectation Maximisation (EM) est utilisé pour estimer la vraisemblance maximale du modèle par rapport au vecteur de paramètres d'apprentissage. La moyenne et la matricede covariance de l'ensemble des V vecteurs sont calculées et une valeur de 1 est affectée aupoids. . Pour chaque itération j allant de 1 à $log_2(M)$ (où : M est le nombre de composantes gaussiennes),les étape de l'algorithme EM sont détallées comme suite :

3. Etape de l'algoritme EM

- **Estimation :** L'estimation consiste à calculer l'appartenance de chaque vecteur acoustique $x_t$de la matrice X_{UBM} à chacune des gaussiennes i avec $(1 \leq i \leq M)$ du modèle λ_{UBM}.
- **Maximisation :** La maximisation consiste à mettre à jour les poids, les moyennes et les matrices de covariance obtenus lors de l'estimation.

Poids du mélange (Reynolds et al. 2000) :

$$\overline{w}_i = \frac{1}{V} \sum_{t=1}^{V} P_r(\ i|\, x_t\ , \lambda_{\text{UBM}}) \tag{2.28}$$

Moyennes (Reynolds et al. 2000) :

$$\overline{\mu}_i = \frac{\sum_{t=1}^{V} P_r(\ i|\, x_t\ , \lambda_{\text{UBM}}) x_t}{\sum_{t=1}^{V} P_r(\ i|\, x_t\ , \lambda_{\text{UBM}})} \tag{2.29}$$

Variances (covariances diagonales) (Reynolds et al. 2000) :

$$\overline{\sigma}_i^2 = \frac{\sum_{t=1}^{V} P_r(\ i|\, x_t\ , \lambda_{\text{UBM}}) x_t^2}{\sum_{t=1}^{V} P_r(\ i|\, x_t\ , \lambda_{\text{UBM}})} - \overline{\mu}_i^2 \tag{2.30}$$

2.4.4.2 Estimation des modèles des locuteurs par l'adaptation MAP

La dérivation des marques du téléphone se fait, dans le système GMM-UBM de façon adaptative. Les signaux d'enregistrement d'apprentissage de chaque téléphone serrent à

adapter les paramètres du modèle du monde (UBM) en utilisant l'algorithme d'estimation du maximum a posteriori (MAP) qui se résume à (Alimohad 2015) :

$$n_i = \sum_{t=1}^{V} P_r(\ i|\, x_t\ , \lambda_{\text{UBM}}) \tag{2.31}$$

$$E_i\,(x) = \frac{1}{n_i} \sum_{t=1}^{V} P_r(\ i|\, x_t\ , \lambda_{\text{UBM}}) x_t \tag{2.32}$$

$$E_i\left(x^2\right) = \frac{1}{n_i} \sum_{t=1}^{V} P_r(\ i|\, x_t\ , \lambda_{\text{UBM}}) x_t^2 \tag{2.33}$$

x^2est la notation signifiant $diag(xx')$.

Ces nouvelles statistiques suffisantes sont utilisées pour mettre à jour les paramètres de la gaussienne i (Reynolds et al. 2000) :

$$\widehat{w}_i = \left[\frac{\alpha_i^w n_i}{V + (1 - \alpha_i^w) w_i}\right] \gamma \tag{2.34}$$

$$\widehat{\mu}_i = \alpha_i^m E_i\,(x) + (1 - \alpha_i^m)\mu_i \tag{2.35}$$

$$\widehat{\sigma}_i^2 = \alpha_i^v E_i\left(x^2\right) + (1 - \alpha_i^v)\left(\sigma_i^2 + \mu_i^2\right) - \widehat{\mu}_i^2 \tag{2.36}$$

Les coefficients d'adaptation controlant l'équilibre entre les anciennes et nouvelles estimations sont $\alpha_i^\rho,\ \ \{w, m, v\}$pour le poids, la moyenne et la variance respectivement avec : $\alpha_i^\rho = \frac{n_i}{n_i + r_p}$. Le facteur d'échelle γ.est calculé sur tous les poids du mélange adapté assurer leur somme a l'unité (Reynolds et al. 2000)

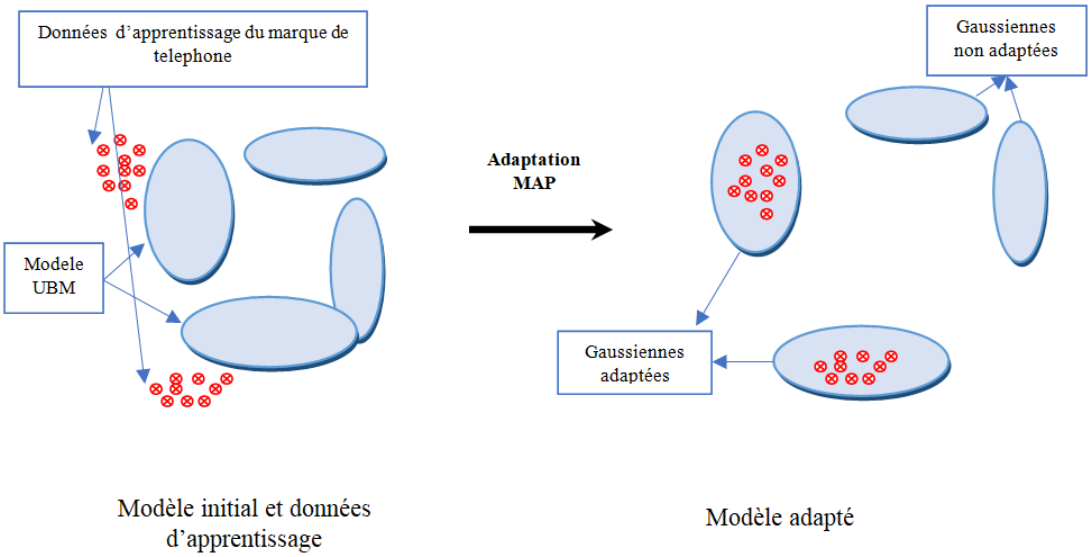

FIG. 2.14 : Adaptation MAP d'un modèle GMM-UBM.

2.4.5 Approche I-VECTOR

Le succès de l'architecture GMM-UBM dans le domaine de la reconnaissance du locuteur et la proposition de supervecteurs ont donné naissance au JFA (Joint Factor Analyse). Ce modèle est une tentative pour résoudre le problème des disproportions causées par les effets de canal dans un très grand espace de supervecteur. À son tour, le modèle JFA a largement ouvert la voie à l'émergence de représentations de la parole utilisant des vecteurs de faible dimension appelés i-vecteurs (DEHAK et al. 2009) (DEHAK et al. 2011). La définition la plus basique du i-vecteur est de le considérer comme une projection d'un supervecteur dans un espace de réduction de dimensionnalité. Cet espace est différent des deux sous-espaces de JFA, et ne fait pas de distinction entre les différents types de variabilité (c'est-à-dire) la variabilité des interlocuteurs et la variabilité induite par le canal), il est donc appelé l'espace de variabilité totale. Mathématiquement, un supervecteur S d'un segment audio dépendant du locuteur et du canal. Il peut s'écrire sous la forme suivante :

$$S = m + Tx \tag{2.37}$$

Le supervecteur de la moyenne m de dimensions C×F est le supervecteur du modèle du monde UBM, et la matrice rectangulaire T de $C \times F \times D$ (où $D \ll CF$ est appelée matrice de variabilité totale . Variabilité de l'interlocuteur et la variabilité causée par le canal) et x est le vecteur caché de dimension D, qui suit la distribution normale standard. Par conséquent, le superviseur S suit la distribution normale du vecteur moyen m et la matrice de covariance égale à TT'. Le modèle génératif proposé dans (2.37) peut être estimé par un modèle d'analyse factorielle ou une analyse probabiliste en composantes principales (Probalistic Principal Component Analysis, PPCA) (DEHAK et al. 2011) (KENNY 2012) (CM 2007). Notez que l'estimation ponctuelle (point estimate) du vecteur caché x nous

fournit le i-vecteur.

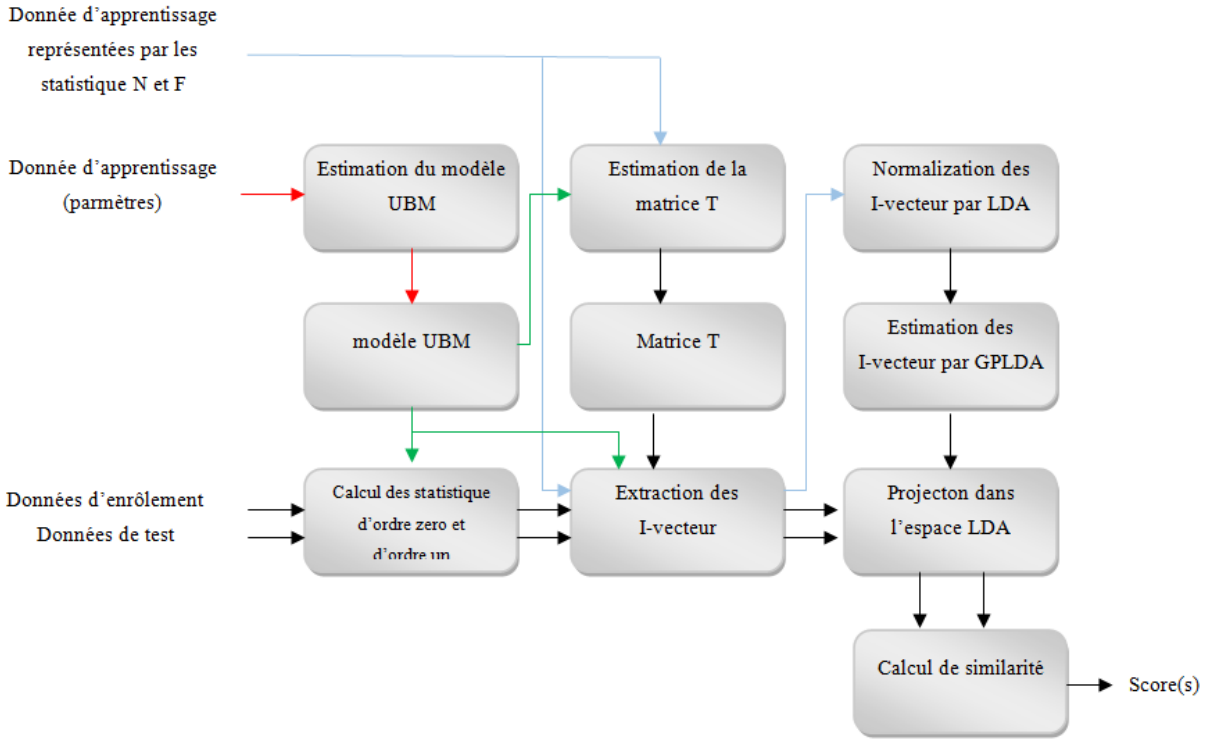

FIG. 2.15 : Architecture du système RA à base de I-Vecteur.

2.4.5.1 Estimation de la matrice T

Dans l'estimation de la matrice T les séquences de parole du même locuteur sont traitéescomme si elles étaient produites par différents locuteurs. Les étapes d'estimation de cettematrice sont données par :

Pour chaque énoncé u :

- Calculer les statistiques d'ordre zéro et d'ordre un $N(u)$ et $F(u)$ respectivement.

$$N(u) = \sum_{t \in u} \gamma_t(c) \tag{2.38}$$

$$F(u) = \sum_{t \in u} \gamma_t(c)\, X(t) \tag{2.39}$$

Avec ; $\gamma_t(c)$est la composante gaussienne c à posteriori pour l'observation t de l'énoncé u, et $X(t)$ est le vecteur acoustique.

- Centrer la statistique de premier ordre

$$\widetilde{F}(u) = F(u) - N(u) \tag{2.40}$$

- Calculer la matrice

$$G(u) = (I + T^T \sum^{-1} N(u) T)^{-1}) \tag{2.41}$$

- Calculer les statistiques du vecteur $w(u)$

$$E_1(u) = E(w) = G(u) T^T \sum^{-1} \widetilde{F}(u) \tag{2.42}$$

$$E_2(u) = E\left(ww^T\right) = E_1(u) E_1^T(u) + G(u) \tag{2.43}$$

- Résoudre le système d'équations

$$\sum_u N(u) \widetilde{T} E_2(u) = \sum_s \widetilde{F}(u) E_1^T(u) \tag{2.44}$$

- Remplacer T par $\widetilde{T}$

2.4.5.2 Estimation des paramètres du téléphone

Dans le système à base de I-Vecteur, un téléphone est représenté par son vecteur d'identité w (Ambikairajah et al. 2012) défini par :

$$w(s) = (I + T^T \sum^{-1} N(s) T)^{-1} T^T \sum^{-1} (F(s) - N(s) M) \tag{2.45}$$

où s représente chaque téléphone.

2.5 Compensation de l'effet de la variabilité

2.5.1 Méthode LDA (Linear Discrimination Analysis)

Afin de compenser l'influence du canal la méthode LDA est utilisée. Cette méthode sert à minimiser la variabilité intra-classe et maximiser la variabilité inter-classes. La matrice de projection (A) de cette méthode est obtenue en résolvant le problème des valeurs propressuivant (Ambikairajah et al. 2012) :

$$S_b v = \lambda S_w \tag{2.46}$$

Où S_b et S_w sont respectivement la matrice de variabilité inter-classes et la matrice de variabilité intra-classe.

$$S_b = \sum_{s=1}^{S} (w_s - \overline{w}) (w_s - \overline{w})^T \tag{2.47}$$

$$S_w = \sum_{s=1}^{S} \frac{1}{n_s} \sum_{i=1}^{n_l} (w_i^s - \overline{w}_s)(w_i^s - \overline{w}_s)^T \tag{2.48}$$

S est le nombre des classes, n_s est le nombre des I-Vecteurs pour chaque classe$s$, $\overline{w}_s$ est la moyenne des I-Vecteurs pour chaque classe$s$ et $\overline{w}_s$est la moyenne de tous les I-Vecteurs (CM 2007) la moyenne sur tous les locuteurs est définie par :

$$\overline{w}_s = \frac{1}{n_s} \sum_{i=1}^{n_s} w_i^s \tag{2.49}$$

$$\overline{w} = \frac{1}{N} \sum_{s=1}^{S} \sum_{i=1}^{n_s} w_i^s \tag{2.50}$$

où N est le nombre total de sessions.

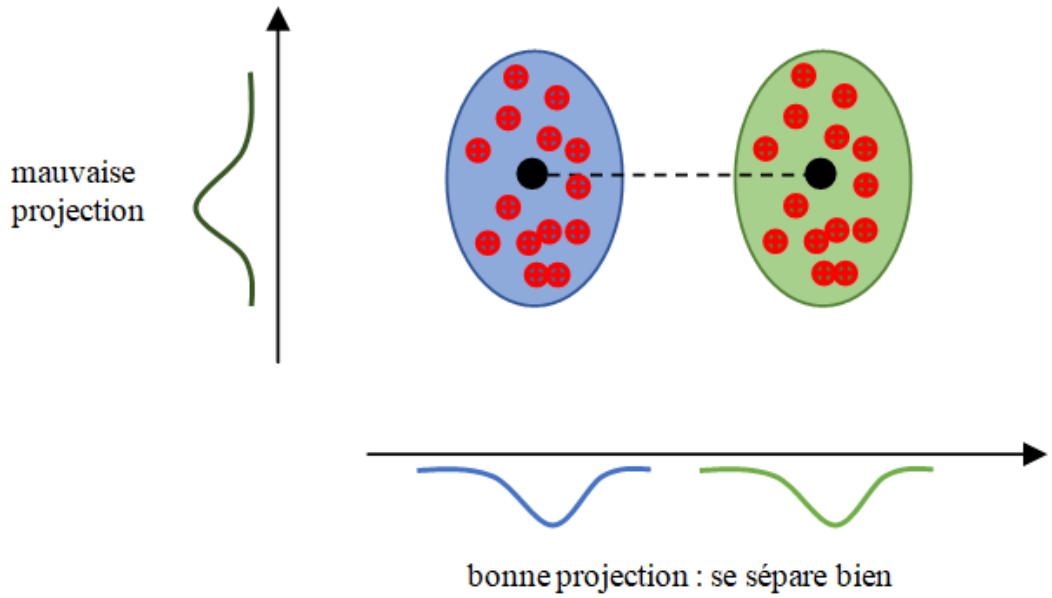

FIG. 2.16 : maximisation des axes des composants pour la séparation des classes.

2.5.2 Méthode WCCN(Within-Class Covariance Normalization)

Cette méthode est utilisée comme technique de compensation de la variabilité de session supplémentaire pour mettre à l'échelle le sous-espace afin de réduire la dimension de la variance intra-classe élevée. La matrice de transformation WCCN (B) est entraînée à l'aide des i-vecteurs projetés par LDA (DEHAK et al. 2011) de la première étape. La matrice WCCN (B) est calculée en utilisant la décomposition de Cholesky suivante :

$$BB^T = W^{-1} \tag{2.51}$$

où la matrice de covariance intra-classe W est calculée en utilisant

$$W = \frac{1}{S}\sum_{s=1}^{S}\sum_{i=1}^{n_s}(A^T(w_i^s - \overline{w}_s))(A^T(w_i^s - \overline{w}_s))^T \tag{2.52}$$

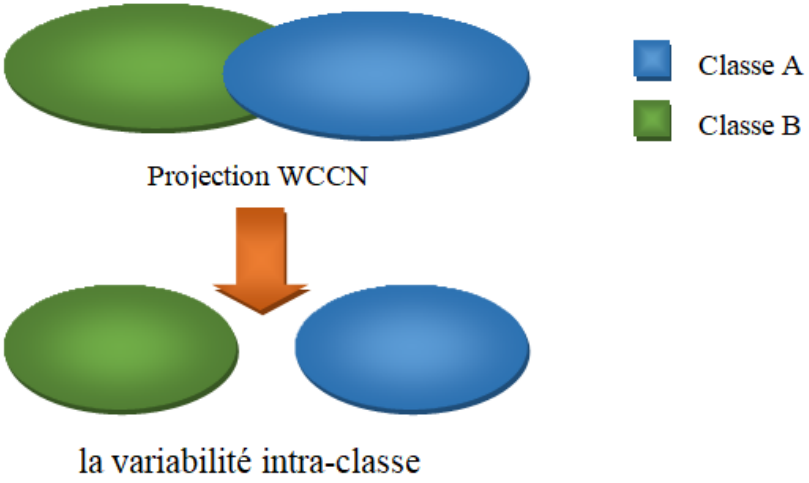

FIG. 2.17 : Normalisation de la covariance intra-classe.

— Le i-vecteur compensé par la variabilité de la session WCCN[LDA] finale peut être calculé comme suit :

$$\widehat{w}_{\text{WCCN[LDA]}} = B^T A^T w \tag{2.53}$$

2.5.3 Méthode G-PLDA (Gaussien Probalistic LDA)

Afin de modéliser directement lavariabilité de la conversation et du locuteur dans l'espace vectoriel I, Kenny a proposé une technique d'analyse discriminante linéaire PLDA. On peut voir que cette méthode est très similaire à la méthode JFA, mais utilise le vecteur I dépendant du locuteur et du canal au lieu du super vecteur GMM comme base pour la modélisation factorielle. Le modèle G-PLDA (KENNY 2010), suppose que chacun de ces I-Vecteurs se décompose comme suit :

$$w = m + U_1 x_1 + U_2 x_2 + \epsilon \tag{2.54}$$

Où U_1 est la matrice des voix propres et U_2 la matrice des canaux propres, x_1 et x_2 sont respectivement les facteurs du locuteur et du canal, et ϵ est le résiduel du locuteur supposé gaussien. La distribution normale est utilisée pour x_1, x_2 et ϵ.

Ce modèle est donc formé de deux parties : $m + U_1 x_1$ La composante qui ne dépend que du locuteur et ne décrit que la variabilité inter-locuteurs, et celle du canal $U_2 x_2 + \epsilon$ qui ne dépend que de ce dernier et ne décrit que la variabilité intra-locuteurs.

Cependant, le G-PLDA intégrée dans notre travail ne considère que la partie dépendante du locuteur et la décomposition de chacun des I-Vecteurs devient comme suit :

$$w = m + U_1 x_1 + \epsilon \tag{2.55}$$

2.5.4 Méthode PCA (Principal Component Analysis)

L'analyse en composantes principales (ACP) peut définir un sous-espace à partir d'un ensemble de données d'apprentissage, ce qui permet de sauvegarder des informations distinctives et de supprimer des informations secondaires (non informatives) en même temps.

Cette méthode consiste à trouver une nouvelle base dans l'espace de données, où tous les vecteurs sont orthogonaux les uns aux autres. Le premier de ces vecteurs correspond à la direction de variance maximale des données d'apprentissage. Les autres composantes sont déterminées par des contraintes orthogonales entre les vecteurs, en tenant compte de la direction de la variance maximale. Dans la méthode ACP, la standardisation de l'éclairage est toujours essentielle.

L'ACP est très utilisée en reconnaissance de formes pour sa rapidité, sa simplicité. C'est la meilleure façon pour reconstruire une base de dimension réduite car les projections de l'ACP sont optimales (M.Benatia 2012) .Elle consiste à trouver les vecteurspropres de la matrice de covariance formée par les différentes images de notre base d'apprentissage par la procédure qui suit :

Etape1 : Sélectionnez data matrix, X^Tmoyenne nulle.

Etape2 : Calculer la moyenne.

$$\Psi = \frac{1}{N} \sum_{i=1}^{N} X_i \tag{2.56}$$

Etape3 : Soustraire la moyenne de la distribution à partir de l'ensemble de données.

$$X_i = X^T - \Psi \tag{2.57}$$

Etape4 : Calculer la matrice de covariance XX^T.

$$C = \sum_{i=1}^{N} X_i X_i^T \tag{2.58}$$

Etape5 : Calculer les valeurs propres et les vecteurs propres V de la matrice de covariance. Où $= 1 \dots N$.

Etape6 : Ordonner les vecteurs propres $V_i(i = 1 \dots N)$ par leurs valeurs propres correspondantes λ_i, par ordre décroissant.

Etape7 : Ne conserver que les vecteurs propres avec les valeurs propres les plus importantes (les composants principaux),$k(k \ll N)$ $\mathrm{X}^k = V^k.X$

Etape8 : Résoudre pour PCA.

$$\lambda V_{x^T} T = C_x V_{x^T} \tag{2.59}$$

Le fonctionnement de l'ACP peut être considéré comme révélateurde la structure interne des données de manière à mieux expliquer la variance dans les données.

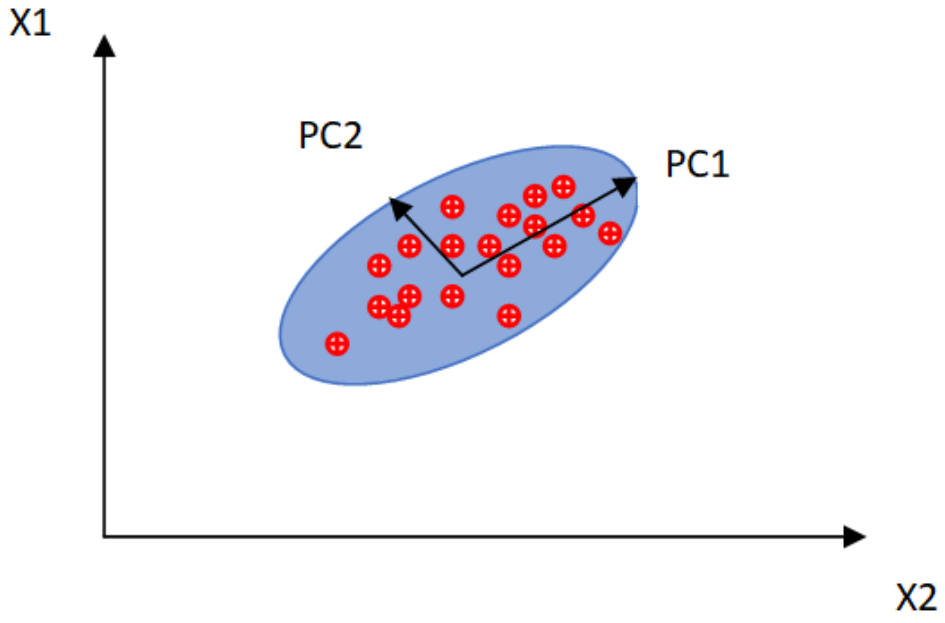

FIG. 2.18 : Analyse en Composantes Principales.

2.6 Calcul des scores

2.6.1 Méthode LLR (Log Likelihood Ratio)

Pour l'approche GMM-UBM, le calcul des scores qui représente la similarité entre la donnée test $X_{\text{Test}} = x_1, x_2, x_3, \dots, x_V$ et le modèle du client λ_c (REYNOLDS et al. 2000).

Il est donnée par :

$$\mathrm{LLR}\left(\ X_{\mathrm{Test}}|\,\lambda_c\right) = \frac{1}{V}\sum_{t=1}^{V}\log(P_r\left(\ x_t|\,\lambda_c\right)) - \log(P_r\left(\ x_t|\,\lambda_{\mathrm{UBM}}\right) \tag{2.60}$$

2.6.2 Méthode GPLDA (Gaussian Probabilistic LDA)

Le calcul des scores pour I-Vecteur basé sur GPLDA utilise le rapport de vraisemblance par lots (Kenny 2010) Pour deux I-Vecteurs w_{client} et w_{test}, le calcul du rapport de vraisemblance par lots est donné par :

$$\text{score } = \ln(\frac{P\left(w_{\mathrm{client}},\ w_{\mathrm{test}}|\,H_1\right)}{P(\ w_{\mathrm{client}}|\,H_0)\ \ P(w_{\mathrm{test}}|\,H_0)}) \tag{2.61}$$

où H_1 : est l'hypothèse que les deux locuteurs sont les mêmes ; alors que H_0 est l'hypothèse que les deux locuteurs sont différents.

2.6.3 Méthode CSS (Cosine Similarity Scoring)

La mesure de similarité en cosinus (Cosine Distance) a été utilisée initialement dans les articles qui sont à l'origine du paradigme de la variabilité totale (Dehak et al. 2011). Puis reprise en reconnaissance faciale. Pour cette mesure, le score entre les paramètres du client w_{clien} et ceux de test w_{test} est calculé en tant que produit scalaire normalisé :

$$\text{score }_{\mathrm{CD}} = (w_{\mathrm{client}}, w_{\mathrm{test}}) = \frac{w_{\mathrm{client}}.w_{\mathrm{test}}}{\|w_{\mathrm{client}}\|\,\|w_{\mathrm{test}}\|} \tag{2.62}$$

Le "." fait référence au produit scalaire entre deux vecteurs.

Malgré sa structure simple, cette approche s'est avérée très efficace dans la RA car elle ne contient aucune information préalable sur l'apprentissage des classes (ici les téléphones mobiles). Cette distance est généralement utilisée dans la littérature en conjonction avec des algorithmes de compensation de variabilité canal/session (tels que la normalisation WCCN et/ou l'analyse discriminante linéaire).

2.7 Conclusion

Ce chapitre présente en détail ce que nous avons mis en œuvre dans les deux blocs d'extraction et de modélisation, visant à améliorer la robustesse du système d'identification automatique du téléphone mobile. Au niveau de l'extraction, plusieurs méthodes sont utilisées, à savoir les descripteurs acoustiques MFCC et CQCC et les descripteurs audio-visuels tels que LPQ et LBP. Au niveau de la modélisation, nous avons étudié la méthode GMM-UBM et la méthode I-Vecteur.Plusieurs algorithmes de réduction de dimension ont été utilisés comme : LDA, PCA, GPLDA, et WCCN. Le chapitre 3 présentera les tests effectués, et les résultats obtenus ainsi que leurs interprétations.

Chapitre 3

Résultats et discussions

3.1 Introduction

À partir de ce qui a été cité dans le chapitre 2 comme méthodes existantes au niveau des deux blocs d'extraction de paramètres et de modélisation du SRA de téléphone mobile, six d'entre elles sont retenues dans le chapitre 3 pour faire l'objet d'expérimentation dans nos travaux, à savoir la méthode d'extraction des paramètres acoustique MFCC la méthode d'extraction des paramètres visuels LPQ, ainsi que les deux méthodes de modélisation à savoir la mèthode GMM-UBM et celle du I-Vecteur.

Dans un premier temps, une description des deux bases de données MOBIPHONE et LOCALE sur lesquelles l'apprentissage et les tests sont effectués, l'environnement de travail et l'organisation expérimentale de ces deux bases de données, avec des interprétations appropriées est donnée.

L'évaluation des performances de notre SRA de téléphone mobile, exprimée par le taux des résultats corrects est enfin calculé à partir des scores de tests obtenus.

Pour optimiser au mieux les performances de l'évaluation dans les deux bases de données avec le problème de variation des locuteurs, une fusion des scores des deux systèmes, à base d'extraction de LPQ et celle des MFCCs avec l'approche I-Vecteur pour la modélisation, est réalisée.

3.2 Description des bases de données

3.2.1 Base de données MOBIPHONE

La base de données MOBIPHONE contient des enregistrements à large bande (16 000 Hz) de 24 locuteurs provenant de la base de données TIMIT. Chaque locuteur est enregistré en lisant 10 phrases phonétiquement riches en anglais, d'environ 3 secondes chacune.

La collection des marques utilisés dans les expériences comprend les marques de téléphone mobiles suivantes : IPHONE, NOKIA, LG, SONY, SAMSUNG, VODAFONE qui sont répertoriés dans le tableau 3.1.

Numéro	Marque de téléphone	Modèles
01	APPEL	I PHONE 5
02	HTC	SENSATION X
03	LG	L3
04	NOKIA	C5
05	SAMSUNG	E 2600
06	SAMSUNG	GALAXY GT-19100 s2
07	SAMSUNG	GT-18190 mini
08	SAMSUNG	GT-N7100
09	SAMSUNG	S 5830i
10	SONY	ERICSON C902
11	VODAFONE	JOYE 845
12	HTC	DESIRE C
13	LG	GS290
14	LG	OPTIMUS L5
15	LG	OPTIMUS L9
16	NOKIA	5530
17	NOKIA	N70
18	SAMSUNG	E1230
19	SAMSUNG	E2121B
20	SAMSUNG	GALAXY NEXUS S
21	SONY	ERICSON C510I

Tab. 3.1 : Les marques et modèles de téléphones portables utiliséss dans la base de donnée MOBIPHONE

3.2.2 Base de données LOCALE

Cette base de données a été créé de la même manière que celle de MOBIPHONE. Par conséquent, on a pris enregistré 24 locuteurs de la base de données TIMIT.

L'ensemble de marques de téléphones mobiles utilisé comprend : CONDOR, HUAWEI, IRIS, LG, OPPO, REDMI, SAMSUNG, URBAN,HONOR,comme indiqué dans le tableau 3.2.

Numéro	Marque de téléphone	Modèles
01	CONDOR	PLUM
02	HUAWEI	P10
03	IRIS	N31
04	LG	STYLO2
05	OPPO	A5S
06	REDMI	NOTE7
07	SAMSUNG	GRAND PRIME
08	URBAN	URBAN1
09	HUAWEI	P10 plus
10	HUAWEI	Y7 PRIME
11	REDMI	NOTE 9
12	REDMI	NOTE 10
13	REDMI	NOTE 10 (2)
14	OPPO	A31
15	SAMSUNG	GALAXY G3 PRO
16	SAMSUNG	GALAXY S9
17	HONOR	X7
18	HUAWEI	Y7
19	OPPO	A5 2020
20	REDMI	NOTE8 PRO

Tab. 3.2 : Les marques et modèles de téléphones portables utiliséss dans la base de données LOCALE

3.3 Environnement de travail

Le langage de programmation utilisé dans ce travail est MATLAB, émulé par l'environnement de programmation du même nom (dans notre cas MATLAB 2017a) et développé par la société The Math Works. Matlab permet une mise en œuvre simple et rapide d'algorithmes, la mise en œuvre de tâches nécessitant une puissance de calcul élevée, la manipulation et l'affichage de courbes, ainsi que la création d'interfaces graphiques. Nous avons utilisé ce langage dans le MSR IdentityToolkit (Mankibi 2021).

Afin de concevoir une interface graphique qui offre une meilleure présentation de notre travail et des résultats obtenus, nous avons utilisé le même environnement de développement qui nous a fourni les outils nécessaires à une bonne conception et un bon développement de notre application.

3.4 Protocole de travail

3.4.1 Base de données MOBIPHONE

- **Modèle du monde UBM :** Le modèle UBM est créé à partir de données d'enregistrements de 24 locuteurs provenant de dix marques de téléphones mobiles. Ces dix marques sont numérotées de 01 à 10 dans le Tableau 3.1.
- **Création des modèles de téléphones :** Chaque modèle est créé avec un seul fichier parmi les vingt-quatre enregistrements de locuteurs pour les onze marques de téléphones mobiles restantes (numérotées de 11 à 21 dans le tableau 3.1.
- **Ensemble des locuteurs sélectionnés pour le test :** Le test est effectué à partir des enregistrements des 23 locuteurs pour les marques de téléphones mobiles restantes. Ceux ci sont numérotées de 11 à 21 dans le tableau 3.1.

3.4.2 Base de données LOCALE

- **Modèle du monde UBM :** Le modèle UBM a été créé à l'aide des données de huit marques de téléphones portables. Dans chaque téléphone on a enregistré 24 locteurs avec le même environnement (Locale Même Environnement LME). Ces huit marques sont numérotées de 01 à 08 dans le tableau 3.2 .Ensuite, nous avons ajouté quatre autres marques avec le même nombre d'enregistrement, mais dans un environnement différent(Locale Différent Environnement LDE). Ces quatre marques sont numérotées de 17 à 20 dans le tableau 3.2.
- **Création des modèles de téléphones :** Chaque modèle est créé avec un seul fichier parmi les vingt-quatre enregistrements de chacune des huit marques de téléphones mobiles restantes (numérotées de 09 à 16) dans le Tableau 3.2.
- **Ensemble des locuteurs sélectionnés pour le test :** Nous utilisons les fichiers restants (23 enregistrements) correspondant aux marques de téléphones numérotées de 09 à 16 dans le tableau 3.2 pour les tests.

3.5 Expérimentations et résultats

3.5.1 Extraction des paramètres acoustiques du téléphone mobile à l'aide des coefficients MFCC

En utilisant le logiciel de programmation MATLAB, nous avons commencé à construire notre système de reconnaissance comme expliqué au chapitre 2. La première étape consiste

à extraire les paramètres du téléphone à l'aide des coefficients MFCC . Ces paramètres sont ensuite utilisés pour créer chaque modèle de téléphone à l'aide des techniques GMM_UBM et I-Vector. Les modèles résultants seront stockés dans la base de données. Ces étapes forment la phase d'apprentissage.

La phase de test comprend l'extraction des paramètres et la mise en correspondance. Cette extraction s'effectue de la même manière qu'en phase d'apprentissage. La mise en correspondance s'obtient par le calcul d'une mesure de similarité avec les modèles créés.

Les résultats obtenus sont évalués sur la base du taux de reconnaissance correct, qui correspond au nombre de fois que le système a fait un test correct sur le nombre total de tests.

3.5.1.1 Modélisation avec la technique GMM-UBM

Fondamentalement, nous nous appuyons sur trois facteurs principaux dans toutes nos simulations, qui sont le nombre de paramètres (N MFCC), le nombre de Gaussiennes (N MIX) et le nombre d'itérations des algorithmes d'estimation GMM (NITER). Aussi, on a noté le paramètre énergie par 'e', et les paramètres dynamiques à savoir la première dérivée par 'd' et la deuxième dérivée par 'dd'.

3.5.1.1.1 Influence du nombre de coefficients MFCC

Dans la première simulation, nous avons fixé deux facteurs qui sont N MIX à 8, et le nombre d'itérations à 10. Les résultats obtenus sont illustrés dans le tableau suivant :

Base de données	N_MFCC	12	14	16	18	20
MOBIPHONE	TC(%)	88.93	89.93	88.93	89.32	88.53
MOBIPHONE	TC(+e) (%)	89.32	**90.51**%	**90.51**	**90.51**	**90.51**
LME	TC (%)	92.94	94.02	94.02	95.65	95.65
LME	TC(+e) (%)	96.73	93.47%	94.57	96.20	**96.73**
LDE	TC(%)	92.39	94.02	95.10	94.56	95.11
LDE	TC(+e) (%)	94.02	95.11%	96.20	**95.65**	**95.65**

Tab. 3.3 : Influence du nombre de coefficients MFCC sur le SRA

D'après les résultats obtenus dans le tableau 3.3 nous constatons que :

- pour la base de données MOBIPHONE, les meilleurs taux de reconnaissance correct sont observés pour les nombres coefficients MFCC 14, 16, 18 et 20.
- pour la base de données LOCALE et dans le même environnement, 20 coefficients MFCC donne le meilleur taux de reconnaissance correct.

- pour la base de donnée LOCALE avec différents environnements, 18 et 20 coefficients MFCC réalisent le meilleur taux de reconnaissance correct.

On constate aussi que l'ajout du paramètre énergie contribue à améliorer les résultats dans les deux base de données.

3.5.1.1.2 Influence du nombre de GMM

Dans cette partie, nous avons fixé le nombre de coefficients MFCC à 18 pour la base de donneés MOBIPHONE, à 20 pour la base de données LOCALE dans le cas du même environnement, et à 16 quand l'environnement est différent. Pour le nombre d'iteration ; il a été fixé à 10 itérations dans tous les cas. Différentes valeurs ont été utilisées pour le nombre de GMM (N MIX) et les résultats obtenus sont illustrés dans le tableau suivant :

base de données	N MIX	8	16	32	64	128	256	512
MOBIPHONE	TC(%)	89.32	89.72	91.30	92.09	91.30	91.30	91.30
MOBIPHONE	TC(+e)(%)	90.51	92.49	94.07	93.67	93.67	94.07	**94.86**
LME	TC(%)	95.65	95.10	95.10	95.65	**97.28**	91.30	91.30
LME	TC(+e)(%)	**96.74**	**96.74**	**96.74**	94.57	95.10	95.10	95.65
LDE	TC(%)	95.10	94.56	94.56	95.65	**96.73**	96.73	96.19
LDE	TC(+e)(%)	96.19	95.10	94.02	93.47	94.65	**95.10**	94.65

Tab. 3.4 : Influence du nombre de GMM sur le SRA dans la base de donnée

Nous avons testé sept valeurs de GMM pour voir quel est le nombre qui convient le mieux à notre système.

- Pour la bese de données MOBIPHONE nous avons constaté que le meilleur résultat est obtenu lorsqu'on utilise 64 GMM sans apport d'énergie et 512 GMM lorsqu'on ajoute l'énergie. L'erreur dans ce cas est égale à **5.14 %**.
- pour la base de données LOCALE et dans le même et différent environnement, le nombre de 128 GMM donne les meilleurs taux de reconnaissance corrects sans apport d'énergie. Les erreurs dans ce cas sont égales consécutivement à **2.72 %** et **3.27 %**.

3.5.1.1.3 Influence des paramètres dynamiques

Dans cette simulation, nous avons fixé trois facteurs qui sont N MFCC, N mix, et le nombre d'itération qui ont donné les meilleurs résultats précédemment. Les résultats obtenus après l'ajout des paramètres dynamiques(« d », « dd ») avec les coefficients MFCC sont présentés dans le tableau suivant :

Base de données	MOBIPHONE	LME	LDE
N MFCC	18	20	16
N MIX	512	128	128
Nombre d'itération	10	10	10
TC(d) (%)	92.49	89.13	90.76
TC (e + d) (%)	**95.25**	93.47	**94.02**
TC (dd) (%)	92.09	89.13	88.67
TC (d + dd) (%)	94.86	**94.02**	91.30
TC (e + dd) (%)	92.49	91.30	**94.02**
TC (e + d + dd) (%)	94.86	92.93	**94.02**

Tab. 3.5 : Influence des paramètres dynamiques sur le SRA

Dans ce cas, nous avons étudié différentes situations pour montrer l'efficacité de notre système. D'après les résultats obtenus dans le tableau 3.5 nous constatons que :

- pour la base de données MOBIPHONE, lorsque nous avons combiné les paramètres MFCC avec le paramètre énergie (e) et la première dérivée (d), nous avons pu identifier la marque de téléphone mobile avec un pourcentage d'erreur de **4.74 %** seulement.
- pour la base de données LOCALE et dans le même environnement, la combinaison entre la première dérivée (d) et la deuxième dérivée (dd) donne le meilleur taux de reconnaissance correct avec un pourcentage d'erreur de **5.98 %**.
- pour la base de données LOCALE avec différents environnements, La combinaison des paramètres MFCC avec le paramètre énergie (e) et les paramètres dynamiques permet de réaliser un meilleur taux de reconnaissance correct. Ce dernier est dans ce cas égal à **5.98 %**.

D'après ces résultats, nous pouvons constater que ces paramètres rajoutent de l'information pertinente au signal parole.

3.5.1.2 Modélisation avec la technique I-Vecteur

Comme pour les simulations basées sur le modèle GMM_UBM, afin de simuler l'approche I_Vecteur nous nous appuyons essentiellement sur plusieurs facteurs majeurs qui sont ; le nombre de paramètres (N MFCC), le nombre de gaussiennes (N MIX), la matrice de la variabilité totale T (TV Dim), nombre de l'analyse LDA (LDA DIM) et le nombre d'itérations des algorithmes d'estimation du modèle GMM, de la matrice T et de l'analyse LDA (NITER).

3.5.1.2.1 Influence du nombre de coefficients MFCC

Dans cette seconde partie, où une nouvelle approche de modélisation est utilisée, nous avons fixé la dimension de l'analyse LDA (LDA DIM) à 9 pour la base de donneés MOBIPHONE, à 7 pour la base de données LOCALE dans le cas du même environnement, et à 11 quand l'environnement est différent. Pour le nombre d'iteration, le nomre de GMM et la dimension de la matrice T ; ils ont été fixé à 10, 128, 40 respectivement. Différentes valeurs ont été utilisées pour le nombre de coeficients MFCC (N MFCC) et les résultats obtenus sont illustrés dans le tableau suivant :

N MFCC	TC(MOBIPHONE)	TC(LDE)	TC(LME) (%)
12	90.90	97.82	98.36
14	**96.62**	98.36	**98.91**
16	86.95	**98.91**	97.28
18	84.58	**98.91**	98.36
20	89.72	96.73	98.36

Tab. 3.6 : Influence du nombre de coefficients MFCC dans l'approche I-Vecteur

D'après les résultats obtenus nous constatons que :

- pour la base de données MOBIPHONE, les meilleurs résultats étaient obtenus pour le coefficient MFCC égal à 14 paramètres. L'erreur dans ce cas est de **3.37** %.
- pour la base de données LOCALE et dans le même environnement, 14 coefficients MFCC donne le plus faible taux d'erreur **1.09** %.
- de même, pour la base de données LOCALE avec différents environnements, 16 et 18 coefficients MFCC réalisent le meilleur taux de reconnaissance correct. L'erreur dans ce cas est de **1.09** %.

3.5.1.2.2 Influence du nombre de GMM

Pour les deux bases de données MOBIPHONE et LOCALE, nous avons fixé tous les facteurs cités précédemment (N MFCC, TV DIM, N ITER, et LDA DIM). Différentes valeurs ont été utilisées pour le nombre de GMM (N MIX) et les résultats sont présentés dans le tableau ci-dessous :

N MIX	TC(MOBIPHONE)(%)	TC(LDE)(%)	TC(LME)(%)
8	90.91	**98.36**	**98.91**
16	90.49	93.47	98.36
32	88.93	96.73	95.65
64	89.72	92.39	96.19
128	**92.92**	94.56	96.19
256	90.92	95.65	96.65
512	90.92	94.56	95.10

TAB. 3.7 : Influence du nombre de GMM dans l'approche I-Vecteur

Nous avons expérimenté sept valeurs de GMM pour voir lequel fonctionnait le mieux avec notre système. et nous avons constaté que :

- pour la bese de données MOBIPHONE, le meilleur résultat est obtenu lorsque ce nombre est égal à 128 .
- pour la base de données LOCALE et dans le même et différents environnement, 8 GMM donne le meilleur taux de reconnaissance correct .Les erreurs dans ce cas sont égales consécutivement à **1.09 %** et **1.64 %**.

3.5.1.2.3 Influence de la dimension de la matrice T (TV DIM)

En utilisant les résultats obtenus précédemments, les meilleures performances ont été réalisées pour le nombre de coefficients MFCC à 18 pour la base de donneés MOBIPHONE, à 20 pour la base de données LOCALE dans le cas du même environnement, et à 16 quand l'environnement est différent. un nombre de GMM fixé à 128, et un nombre d'itérations de 10 dans tout les cas. Les modifications ont été sélectionnées pour la dimension TV DIM.

Afin de voir quelle est la meilleure performance du système par rapport à la dimension de la matrice T, nous avons testé plusieurs valeurs.

les résultats obtenus pour la base de données MOBIPHONE sont présentés dans le tableau 3.8 :

TV DIM	Taux correct GPLDA (%)
40	96.63
50	90.51
60	**96.63**
70	89.33
80	88.93

TAB. 3.8 : Influence de la dimension de la matrice T (MOBIPHONE)

Nous constatons que l'utilisation d'un TV DIM de 60 donne les meilleurs résultats.

Les résultats sur la base de données LOCALE sont présentés dans le tableau 3.9 :

TV_DIM	TC(LDE)(%)	TC(LME)(%)
25	97.28	98.36
30	97.28	97.28
35	**98.36**	**98.36**
40	96.19	97.28
45	98.36	96.19

TAB. 3.9 : Influence de la dimension de la matrice T(LOCALE)

Nous constatons que les meilleurs résultats sont obtenus avec un matrice T de 35 dimension .

3.5.1.2.4 Effet de l'ajout de la projection WCCN au LDA et du calcul de la distance en cosinus (cosine scoring)

Pour chaque simulation, nous nous concentrons sur la recherche de la configuration qui atteint la meilleure précision globale sur l'ensemble de vérification grâce à de multiples modifications des paramètres.

Dans ce cas nous avons ajouté la projection WCCN au LDA (noté par +WCCN), et la méthode de calcul de scores CSS afin d'améliorer les performances de notre système. Les résultats obtenus sont présentés dans le tableau suivant :

Projection	Méthode	TC (MOBIPHONE)(%)	TC(LDE)(%)	TC(LME)(%)
LDA	GPLDA	96.05	96.19	98.91
+WCCN	GPLDA	95.26	97.82	99.45
LDA	CSS	90.12	98.36	98.91
+WCCN	CSS	**97.23**	**98.91**	**99.45**

TAB. 3.10 : Effet de la projection WCCN et du calcul de score CSS

Nous avons constaté que lorsque nous identifions la marque de téléphone mobile :

- pour la base de données MOBIPHONE, le meilleur résultat est obtenu par la dernière configuration (+WCCN et CSS). L'erreur est alors de **2.77 %** .

- pour la base de données LOCALE, dans le même et dans différents environnements, les meilleurs performances sont réalisées avec des taux d'erreurs égaux consécutivement à **0.55 %** et **1.09 %**.

3.5.2 Extraction des paramètres visuels des téléphones mobiles à l'aide de l'algorithme LPQ

Le modèle binaire local représente un descripteur local important, qui donne des résultats efficaces dans certaines applications (telles que la reconnaissance faciale et la détection d'objets). Dans notre système , nous avons utilisé l'algorithme LPQ (Quantification de phase locale) comme descripteur visuel.

La phase d'apprentissage comprend la lecture des images créées par spectrogrammes à partir des bases de données utilisées, l'alignement des images sélectionnées, puis l'utilisation de la méthode de quantification de phase locale (LPQ) pour extraire les paramètres de chaque image.

La phase de test consiste à mesurer la similarité cosinus entre les paramètres extraits de l'image de test et les paramètres uniques caratérisant chaque modèle. Le résultat de similarité permet de faire correspondre ou non ces deux quantités.

L'image du spectrogramme est subdivisée en régions de spectrogramme P sans chevauchement, où les histogrammes de ces blocs rectangulaires sont connectés pour former un vecteur de caractéristiques v avec une taille de $n = P \times 256$, qui représente un descripteur spécifique à une échelle spécifique.

3.5.2.1 Effet du descripteur LPQ

La méthode LPQ peut être résumée en quatre étapes différentes. Tout d'abord, l'opérateur (LPQ) est appliqué à l'image d'entrée pour obtenir l'image étiquetée. Ensuite, divisez l'image résultante en petites zones. Pour chacune d'elle, un histogramme des étiquettes est construit pour obtenir un vecteur de caractéristiques. Une représentation globale (un vecteur de caractéristiques global représentant l'image entière) est obtenue en combinant tous les vecteurs. D'une manière générale, LPQ est une chaîne binaire, obtenue par la concaténation des codes de la partie réelle et la partie imaginaire des huit coefficients de Fourier pour chaque pixel. La figure 3.1 montre des images LPQ avec différentes valeurs du rayon r (r= rayon de l'opérateur).

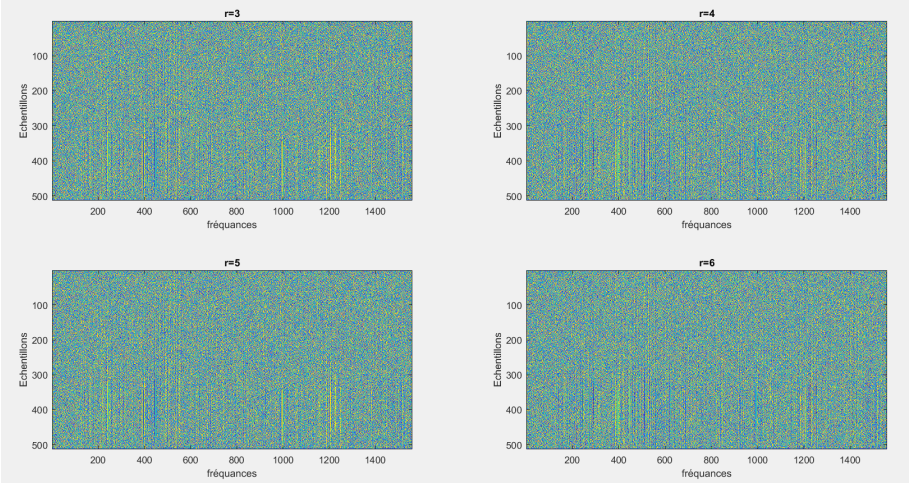

FIG. 3.1 : Représentation d'image du spectrogramme avec le descripteur LPQ.

Cette expérience nous a permis de démontrer l'efficacité du descripteur LPQ utilisé. En effet, plusieurs variations du rayon r ont été testées et comparées. Le tableau 3.11 montre les résultats obtenus :

Base de donneés	Projection	r = 3	r = 4	r = 5	r = 6
MOBIPHONE	LDA	87.75 %	89.72 %	**91.70** %	88.93 %
MOBIPHONE	LDA + WCCN	89.33 %	91.30 %	**92.09 %**	87.75 %
LME	LDA	85.86 %	**88.04** %	85.86 %	88.04 %
LME	LDA + WCCN	88.00 %	**97.28** %	91.30 %	89.13 %
LDE	LDA	87.75 %	89.72 %	**91.70** %	88.93 %
LDE	LDA + WCCN	89.33 %	91.30 %	**92.09 %**	87.75 %

TAB. 3.11 : Effet du descripteur LPQ

D'après ces résultats nous constatons que :

- pour la base de données MOBIPHONE et la base de données LOCALE avec différents environnement, le meilleur résultat est obtenu lorsque la projection (LDA + WCCN) est utilisé avec r égal à 5.
- pour la base de données LOCALE dans le même environnement, un rayon r=4 avec la projection (LDA + WCCN), réalise le meilleur taux de reconnaissance correct.

3.5.2.2 Influence de filtre Mel

Afin de chercher à améliorer les performances des descripteurs LPQ, nous avons ajouté le filtre mel. Dans ce cas, sous différentes tailles de fenêtre r, la représentation de l'image du spectrogramme est illustrée par la figure 3.2. Alors que les résultats obtenus sont présentés dans le tableau 3.12 :

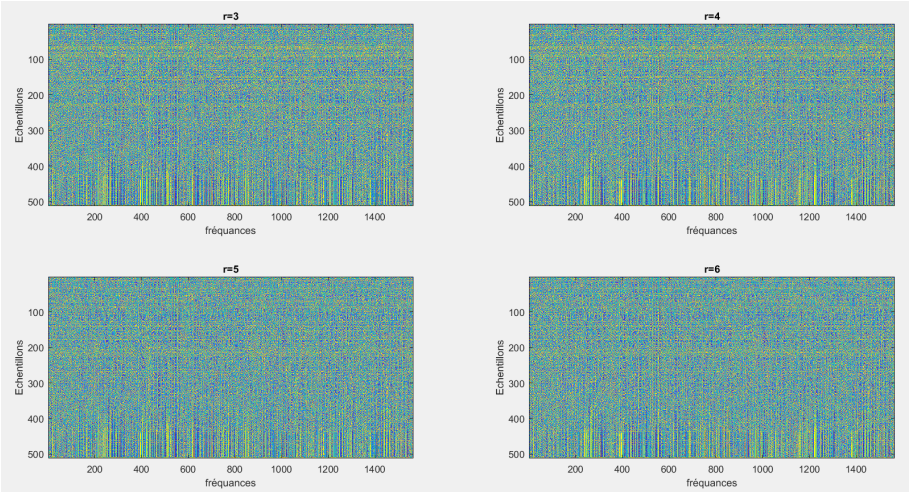

Fig. 3.2 : Représentation d'image du spectrogramme avec le descripteur LPQ.

Base de donneés	Projection	r = 3	r = 4	r = 5	r = 6
MOBIPHONE	LDA	94.07 %	**96.05** %	93.28 %	84.98 %
MOBIPHONE	LDA + WCCN	93.68 %	**96.44 %**	94.07%	88.14 %
LME	LDA	**88.04** %	78.80 %	75.54 %	78.26 %
LME	LDA + WCCN	**88.04** %	82.06 %	78.26 %	79.89 %
LDE	LDA	**96.28** %	94.02 %	90.76 %	92.93 %
LDE	LDA + WCCN	**96.73** %	91.84 %	92.09 %	94.02 %

Tab. 3.12 : Effet du filter MEL et descripteur LPQ

- pour la base de données MOBIPHONE et la base de données LOCALE avec différents environnement, nous avons remarqué que lorsqu'on ajoute le filtre MEL, l'utilisation de la projection LDA+WCCN permet d'améliorer le résultat.

- pour la base de données LOCALE dans le même environnement, nous avons remarqué que le résultat se dégrade avec le filtre MEL.

Il est à noter que l'utilisation de la projection LDA+WCCN nous a toujours permet d'améliorer les résultats.

3.5.2.3 Influence de la concaténation des paramètres LPQ

Dans cette technique nous avons concaténé les paramètres LPQ après l'ajout du filtre MEL pour r=4 et r=5 qui ont donné les meilleurs résultats pour la base données MOBIPHONE. De même, pour la base de données LOCALE, nous avons concaténé les paramètres LPQ avec r=4 et r=3 lorsque l'environnement est le mème puis r=4 et r=6 lorsque l'environnement est différent. Nous avons obtenu ces résultats :

Base de données	Projection	Taux correct (%)
MOBIPHONE	LDA	**98.02**
MOBIPHONE	LDA + WCCN	97.63
LME	LDA	96.73
LME	LDA + WCCN	96.73
LDE	LDA	**100**
LDE	LDA + WCCN	**100**

Tab. 3.13 : Effet de la concaténation des paramètres LPQ.

L'utilisation de cette technique de concaténation nous a permis de trouver de très bons résultats. Ainsi, on a pu identifier les marques des téléphones mobiles avec un pourcentage d'erreur de **1.98 %** dans MOBIPHONE de **00.00 %** pour la base de données LOCALE, et rien que par la technique d'extraction de paramètres visuels.

3.5.3 Fusion des scores

D'après les résultats présentés dans le tableau 3.10 et le tableau 3.13, nous avons constaté que les performances de notre SRA de téléphone mobile utilisant les paramètres acoustiques MFCC et celui utilisant les paramètres visuels LPQ sont similaires. À partir de cette observation et dans le but d'améliorer plus nos résultats, nous avons utilisé un algorithme de fusion des scores de ces deux systèmes. Le détail de cette technique peut se résumer par :

- Calculer le meilleur score (Score1) du premier système utilisant les paramètres MFCC.

- Calculer le meilleur score (Score2) du second système à partir d'une extraction des paramètres LPQ.
- Une combinaison linéaire pondérée des scores relatifs aux paramètres MFCC et ceux relatifs aux paramètres visuels est décrite dans (Alimohad 2015). Le nouveau score du système de reconnaissance aura la forme suivante :

$$Score\_Fusion = aScore1 + (1 - a)Score2 \tag{3.1}$$

Nous avons remarqué que la méthode de fusion des deux types de paramètres dépend du paramètre a. Par conséquent, nous avons expérimenté en modifiant ce paramètre entre les valeurs 0 et 1 pour trouver le meilleur résultat (correspondant à la plus petite erreur). Les résultats trouvés indiquent que la valeur $a = 0{,}2$ permet d'atteindre cet objectif. Ainsi, on réalise le plus grand taux de reconnaissance correct comme illustré dans le tableau 3.14 :

Base de données	Taux correct (%)
MOBIPHONE	**100**
LOCALE	**100**

Tab. 3.14 : Fusion des scores

3.6 Conclusion

Ce chapitre a été dédié à l'implémentation de notre sustème de RATP, basé sur les deux techniques d'extraction, celle des MFCC et des LPQ et les deux approches, GMM-UBM et le I-Vecteur pour la modélisation. Cette dernière (I-Vecteur) réalise la compensation de l'effet du canal. Par conséquent, nous avons présenté les deux bases de données MOBIPHONE et LOCALE qui nous ont permis d'effectuer nos tests afin d'évaluer notre système. Les performances du système de reconnaissance sont données par les valeurs du taux correct. Les différentes implémentations ont montré l'apport obtenu par les paramètres LPQ par rapport aux coefficients MFCC. Un autre constat concerne la grande précision de notre SRA de téléphone mobile apportée par l'approche I-Vecteur par rapport au GMM-UBM. Pour plus de robustesse de notre système, nous avons appliqué une fusion des scores de deux systèmes à base des paramètres MFCC et LPQ. La présentation de notre interface graphique, réalisée à l'aide de GUI sous Matlab, et ses fonctionnalités est donnée dans l'annexe A.

Conclusion et perspectives

Conclusion générale

Le travail réalisé dans ce mémoire consiste à implémenter un système de reconnaissance automatique de téléphone portable. Puis, d'améliorer les performances de celui-ci en utilisant de nouvelles techniques d'extraction de paramètres, de modélisation, et de calcul de scores.

Dans ce genre de systèmes, l'environnement et les différents types de variabilités influent énormément sur ses performances. Le système de base, avec lequel on a commencé, utilise les coefficients MFCC comme descripteur acoustique des enregistrements et l'approche GMM-UBM pour la modélisation. l'évaluation s'est portée sur les bases de données MOBIPHONE et LOCALE. Après plusieurs expériences dans lesquelles on a modifié le nombre de coefficients, ajouté l'énergie du signal, et insérré les paramètres dynamiques, on a réussi à atteindre un taux de reconnaissance correct de **95.25 %** pour la base de données MOBIPHONE et **97.28 %** pour la base de données LOCALE.

Dans le but de pallier le problème de variabilité, nous avons remplacé la modélisation GMM-UBM par l'approche I-Vecteur. Nous avons repris les mêmes expériences qu'auparavant. Nous avons constaté que les résultats se sont amélioré. En terme de taux de reconnaissance correct, le résultat est devenu égal à **96.63 %** pour MOBIPHONE et **98.91 %**pour LOCALE. Dans la même objective, on a utilisé d'autres techniques de compensation d'intra-variabilité (comme WCCN et LDA) et de calcul de score dans diverses expériences. Les résultats obtenus ont atteint un taux de reconnaissance correct de **97.23 %** pour MOBIPHONE et **99.45 %** pour LOCALE.

Une autre méthode d'extraction de paramètres visuels LPQ a été introduite avec un ensemble de configurations pour notre système. Nous avons gardé le même protocole pour faire une comparaison objective avec les paramètres acoustiques. Cette méthode a été appliquée sur les mêmes bases de données MOBIPHONE et LOCALE, les résultats obtenus sont intéressants. En effet, avec la méthode LDA on est arrivé à un taux de reconnaissance correct de **98.02 %** pour MOBIPHONE et **100 %** pour LOCALE. Ce qui rend notre système plus fiable.

Afin de bénéficier des avantages des deux descipteurs (acoustique et visuel) nous avons appliqué une fusion des scores sur le système. En utilisant l'approche I-vecteur et les extracteurs de coefficients MFCCs et LPQ, nous sommes arrivés à un excellent résultat traduit par un taux de reconnaissance correct de **100 %** pour les deux bases de données.

À l'issue de ces travaux, nous estimons avoir réalisé un système répondant à l'objectif que nous nous sommes fixés. Ainsi, l'utilisation des descripteurs audio-visuels en tant que paramètres pour le système de reconnaissance automatique de téléphone mobile permet d'avoir une meilleur robustesse en améliorant les performances de ce système.

Perspectives

Les travaux menés dans le cadre de ce projet représentent un bon début pour plusieurs autres expérimentations futures qui doivent être poursuivi pour parvenir à un système encore plus robuste.

Sur la base de ce projet, nous avons l'intention à l'avenir de poursuivre les recherches sur la reconnaissance automatique des téléphones portables et sur d'autres projets utilisant le système de reconnaissance automatique basé sur la parole.Nous travaillerons fort pour améliorer ce système afin de relever les défis réels qui influencent négativement sur la qualité de l'enregistrement afin d'atteindre l'objectif d'assurer un système performant qui fonctionne sans heurts en toutes circonstances et dans le pire des scénarios.

Bibliographie

AJGOU, Riadh (2016). “Reconnaissance Automatique du Locuteur à Travers les Canaux Digitaux”. Thèse de doct. Université Mohamed Khider-Biskra.

Alimohad, Abdennour (2015). “Contribution a l’inference d’identite en utilisant un systeme de reconnaissance du locuteur GMM-UBM”. Thèse de doct. Université de Bilda.

Ambikairajah, Eliathamby et al. (2012). “PNCC-ivector-SRC based speaker verification”. In : *Proceedings of The 2012 Asia Pacific Signal and Information Processing Association Annual Summit and Conference*. IEEE, p. 1-7.

Aziza, Yassamine (2013). “Modélisation AR et ARMA de la Parole pour une Vérification Robuste du Locuteur dans un Milieu Bruité en Mode Dépendant du Texte”. Mém. de mast. Université Ferhat Abbas –Setif1.

CM, Bishop (2007). “Pattern Recognition and Machine Learning”. In : *Information Science and Statistics* 49, p. 366-366.

Debbeche, Feriel (2008). “Système Acoustico-Anatomique pour l’Identification des Locuteurs par Localisation dans un Espace de Locuteurs de Référence”. Mém. de mast. Université Badji Mokhtar Annaba.

Dehak, Najim et al. (2009). “Support vector machines versus fast scoring in the low-dimensional total variability space for speaker verification”. In : *INTERSPEECH*, p. 1559-1562.

Dehak, Najim et al. (2011). “Front-End Factor Analysis for Speaker Verification”. In : *IEEE Transactions on Audio, Speech, and Language Processing* 19, p. 788-798.

Delgado, Héctor et al. (2018). “ASVspoof 2017 Version 2.0: meta-data analysis and baseline enhancements”. In : *Odyssey 2018 - The Speaker and Language Recognition Workshop*. Les Sables d’Olonne, France.

Fuchs, Guillaume (2007). “Codage audio hiérarchique à faibles débits”. Thèse de doct. Canada : Université de Sherbrooke.

Hadid, Abdenour, Juha Ylioinas et Miguel Bordallo López (2014). “Face and texture analysis using local descriptors : a comparative analysis”. In : *2014 4th International Conference on Image Processing Theory, Tools and Applications (IPTA)*. IEEE, p. 1-4.

Hanilçi, Cemal et Tomi Kinnunen (2014). “Source cell-phone recognition from recorded speech using non-speech segments”. In : *Digital Signal Processing* 35, p. 75-85.

Harrington, Jonathan et Steve Cassidy (1999). *Techniques in speech acoustics*. Springer Science & Business Media.

Hermansky, Hynek (1990). "Perceptual linear predictive (PLP) analysis of speech". In : *the Journal of the Acoustical Society of America* 87.4, p. 1738-1752.

Hussain, Ayyaz et al. (2012). "Survey of various feature extraction and classification techniques for facial expression recognition". In : *Proc. the 11th WSEAS International Conference on Electronics, Hardware, Wireless and Optical Communications, and Proceedings of the 11th WSEAS International Conference on Signal Processing, Robotics and Automation, and Proceedings of the 4th WSEAS International Conference on Nanotechnology*, p. 138-142.

Ji, Xunsheng, Kun Jiang et Jie Xie (2021). "LBP-based bird sound classification using improved feature selection algorithm". In : *International Journal of Speech Technology*, p. 1-13.

Jourani, Reda (2012). "Reconnaissance automatique du locuteur par des GMM à grande marge". Thèse de doct. Université de Toulouse.

Kenny, Patrick (2010). "Bayesian speaker verification with heavy-tailed priors." In : *Odyssey*. T. 14.

— (2012). "A small footprint i-vector extractor". In : *Odyssey 2012-The Speaker and Language Recognition Workshop*.

Klautau, Aldebaro (2005). "The MFCC". In : *Digital Signal Processing*.

Laleye, Frejus Adissa Akintola (2016). "Contributions à l'étude et à la reconnaissance automatique de la parole en Fongbe". Thèse de doct. Université du Littoral Côte d'Opale.

Larcher, Anthony (2009). "Modèles acoustiques à structure temporelle renforcée pour la vérification du locuteur embarquée". Thèse de doct. Université d'Avignon.

Leman, Adrien (2011). "Diagnostic et évaluation automatique de la qualité vocale à partir d'indicateurs hybride (Modèle DESQHI)". Thèse de doct. France : INSA de Lyon.

Lippmann, Richard P (1997). "Speech recognition by machines and humans". In : *Speech communication* 22.1, p. 1-15.

M.Benatia, H.Ouamane et (2012). "Identification de reconnaissance faciale avec des expressions". Thèse de doct. Université Mohamed Khider–Biskra.

Mason, J., J. Oglesby et L. Xu (1989). "Codebooks to optimise speaker recognition". In : *EUROSPEECH*, p. 267-270.

Matsui, Tomoko et Sadaoki Furui (1994). "Comparison of text-independent speaker recognition methods using VQ-distortion and discrete/continuous HMM's". In : *IEEE Transactions on speech and audio processing* 2.3, p. 456-459.

Mittal, Aakshi et Mohit Dua (2021). "Constant Q cepstral coefficients and long short-term memory model-based automatic speaker verification system". In : 87, p. 895-904.

Nolan, FJD (1982). "The phonetic bases of speaker recognition". In : *Phonetics laboratory* 12, p. 85-89.

Ojala, Timo, Matti Pietikainen et Topi Maenpaa (2002). “Multiresolution gray-scale and rotation invariant texture classification with local binary patterns”. In : *IEEE Transactions on pattern analysis and machine intelligence* 24.7, p. 971-987.

Ojansivu, Ville et Janne Heikkilä (2008). “Blur insensitive texture classification using local phase quantization”. In : *International conference on image and signal processing*. Springer, p. 236-243.

Ouni, Slim (2001). “Modélisation de l’espace articulatoire par un codebook hypercubique pour l’inversion acoustico-articulatoire”. Thèse de doct. France : Université Henri Poincaré-Nancy 1.

Ponraj, Narain, Merlin Mercy et al. (2016). “Extraction of speech signal based on Power Normalized Cepstral Coefficient and Mel Frequency Cepstral Coefficient : A comparison”. In : *2016 International Conference on Electrical, Electronics, and Optimization Techniques (ICEEOT)*. IEEE, p. 1843-1846.

Reynolds, Douglas A, Thomas F Quatieri et Robert B Dunn (2000). “Speaker verification using adapted Gaussian mixture models”. In : *Digital signal processing* 10.1-3, p. 19-41.

Rosenberg, Aaron E (1992). “Recent research in automatic speaker recognition”. In : *Advances in speech signal processing*.

Soong, Frank K et al. (1992). “Report : A vector quantization approach to speaker recognition”. In : *International Conference on Acoustics Speech, and Signal Processing (ICASSP), Tampa (USA)* 66.2, p. 387-390.

Todisco, Massimiliano, Héctor Delgado et Nicholas WD Evans (2016). “A New Feature for Automatic Speaker Verification Anti-Spoofing : Constant Q Cepstral Coefficients.” In : *Odyssey* 2016, p. 283-290.

Yuan, Baohua, Honggen Cao et Jiuliang Chu (2012). “Combining local binary pattern and local phase quantization for face recognition”. In : *2012 International Symposium on Biometrics and Security Technologies*. IEEE, p. 51-53.

Webographie

Kamp, Jade Vande (2020). *What is a Spectrogram ?* URL : https://vibrationresearch.com/blog/what-is-a-spectrogram/. (visité le 03/09/2020).

Mankibi, Mohamed El (2021). *MATLAB*. URL : https://www.construction21.org/maroc/products/h/5/matlab-simulink,6.html. (visité le 19/08/2021).

Stéphanie, JULLIEN (2021). *k-means-ou-k-moyennes*. URL : https://dataanalyticspost.com/Lexique/k-means-ou-k-moyennes/. (visité le 12/06/2021).

Annexes

Annexe A

Interface graphique

Dans notre projet, nous avons eu recours à la réalisation d'une interface graphique qui assure une communication aisée entre l'utilisateur et le système de RATP en :

- simplifiant la lecture et la compréhension des résultats en optimisant la manière dont ils sont présentés par le système.
- facilitant à l'utilisateur la configuration de ce système en lui proposant une liste de choix préétablie.

Notre interface est simple et permet d'illustrer les principaux processus du système de RATP.

A.1 Fenêtre d'accueil

La fenêtre principale de notre application (figure A.1) s'affiche lors de son lancement. Elle permet à l'utilisateur l'accès aux différentes fonctionnalités réalisées et lui offre le choix d'être orienté soit vers une fenêtre proposant la configuration de l'extraction a l'aide descripteurs acoustique soit vers celle a l'aide des descripteurs visuels soit a l'aide d'une fusion audio-visuel.

FIG. A.1 : Fenêtre d'accueil.

A.2 Extraction des paramétres acoustiques (audio)

Lorsque l'utilisateur choisit ce procédé, il trouvera deux options pour calculer la modélisation , la première utélésant la méthode de calcul GMM-UBM, et la seconde la méthode I-VECTEUR comme la figure A.2 illustre :

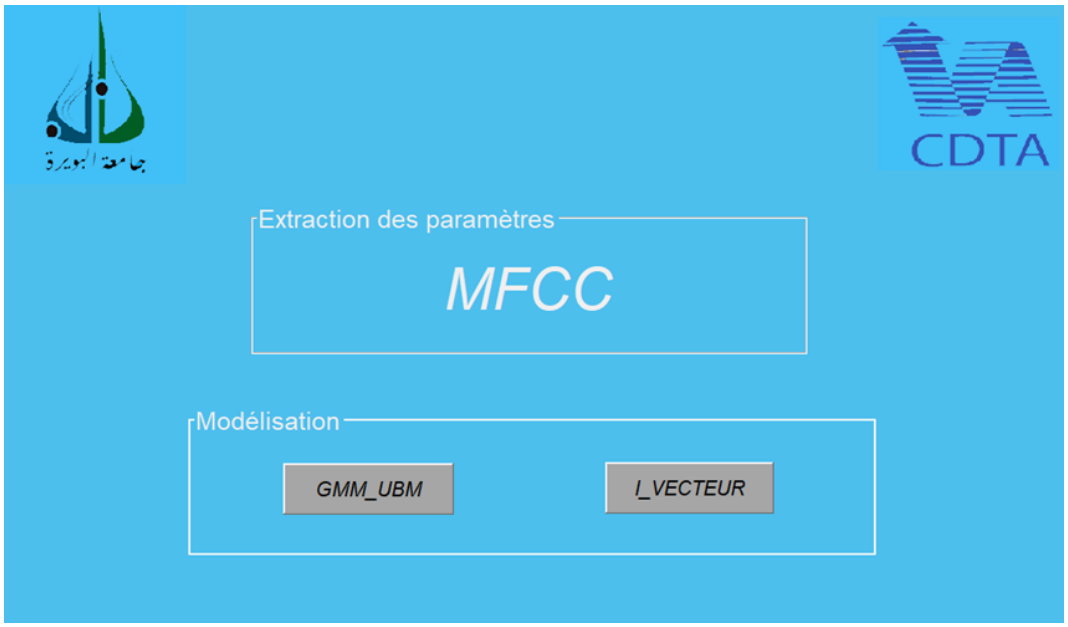

FIG. A.2 : Extraction des paramétres acoustiques.

A.2.1 Modélisation GMM-UBM

Une fois cette modélisation choisie, la fenêtre (figure A.3) apparaît en proposant diverses fonctionnalités.

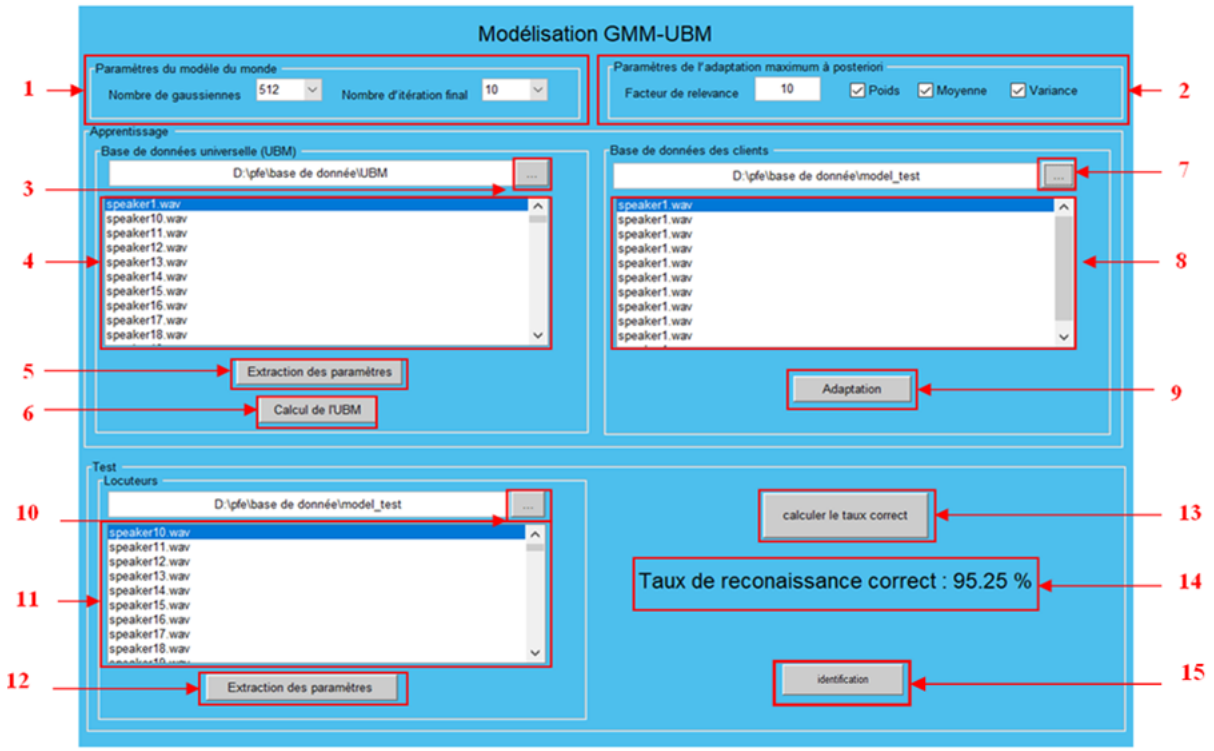

FIG. A.3 : Modélisation GMM-UBM.

- (1) Choix des paramètres de l'UBM .
- (2) Choix des paramètres de l'adaptation MAP.
- (3) Chargement des fichiers souhaités et leur affichage dans (4).
- (5) Extraction des paramètres (figure A.5).
- (6) Calcul de l'UBM une fois le chargement des fichiers correspondants et l'extraction des paramètres de ces derniers sont effectués.
- (7) Chargement des fichiers souhaités et leur affichage dans (8).
- (9) Adaptation MAP une fois le chargement des fichiers son des clients ainsi que l'extraction de leurs paramètres sont effectués.
- (10) Chargement des fichiers souhaités et leur affichage dans (11).
- (12) Extraction des paramètres effectuée après le chargement des fichiers test souhaités.

- (13) Calcul du score.
- (14) Affichage du taux de reconaissance correct.
- (15) Choix d'un test d'identification.

A.2.2 Modélisation I-Vecteur

Si l'utilisateur choisit cette approche, la fenêtre (figure A.4) apparaîtra offrant diverses fonctions.

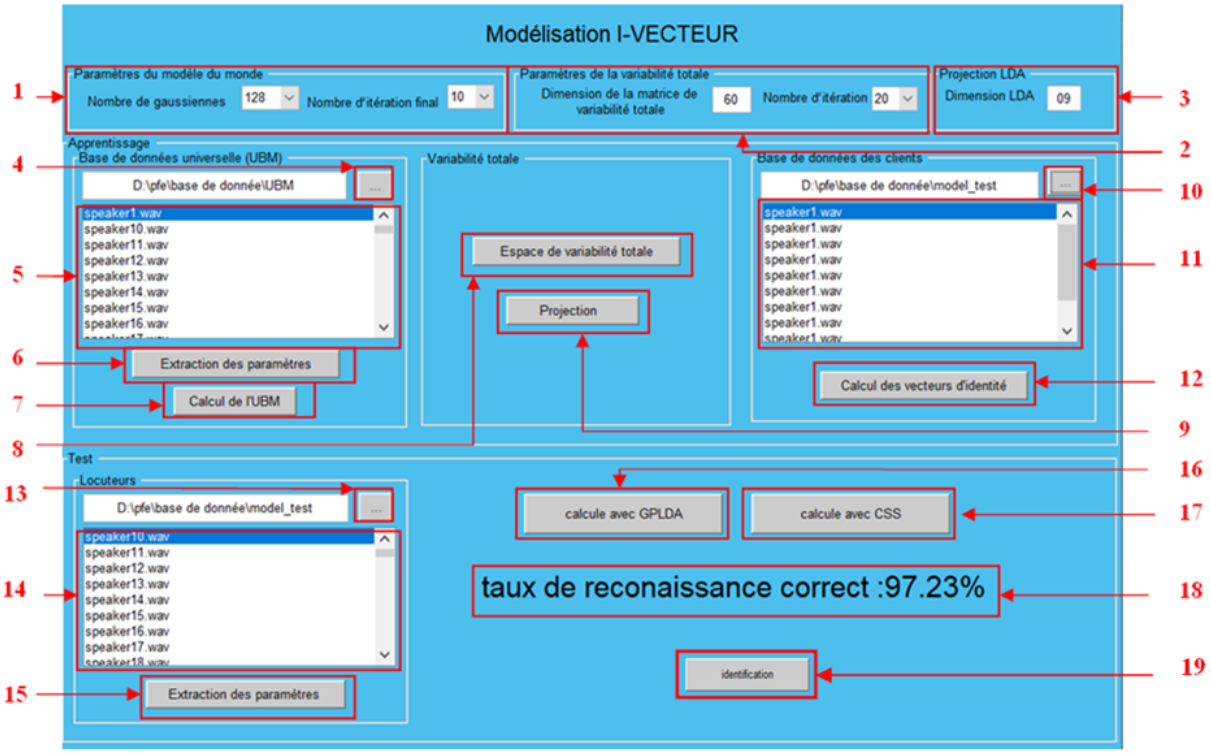

FIG. A.4 : Modélisation I-Vecteur.

- (1) Choix des paramètres de l'UBM.
- (2) Choix des paramètres de la variabilité totale.
- (3) Choix des paramètres de la dimension du LDA.
- (4)Chargement des fichiers souhaités et leur affichage dans (5).
- (6) Extraction des paramètres (figure A.5).
- (7) Calcul de l'UBM une fois le chargement des fichiers correspondants et l'extraction des paramètres de ces derniers sont effectués.

- (8) Calcul de la matrice de variabilité totale une fois le chargement des fichiers correspondants et le calcul de l'UBM sont effectués.
- (9)Projection dans l'espace , une fois le calcul de la matrice de variabilité totale effectué.
- (10)Chargement des fichiers souhaités et leur affichage dans (11).
- (12) Calcul des I-Vecteurs des clients, une fois le chargement des fichiers son des clients effectué.
- (13)Chargement des fichiers souhaités et leur affichage dans (14).
- (15) Extraction des paramètres effectuée après le chargement des fichiers tests souhaités.
- (16) Calcule du score avec la méthode G-PLDA.
- (17) Calcule du score avec la méthode CSS.
- (18) Affichage du taux de reconaissance correct.
- (19) Choix d'un test d'identification.

A.2.3 Extraction des paramètres MFCC

Une étape commune pour les deux types de modélisation, illustrée dans la figure (A.5), qui permet à l'utilisateur le choix des paramètres nécessaires à leur extraction.

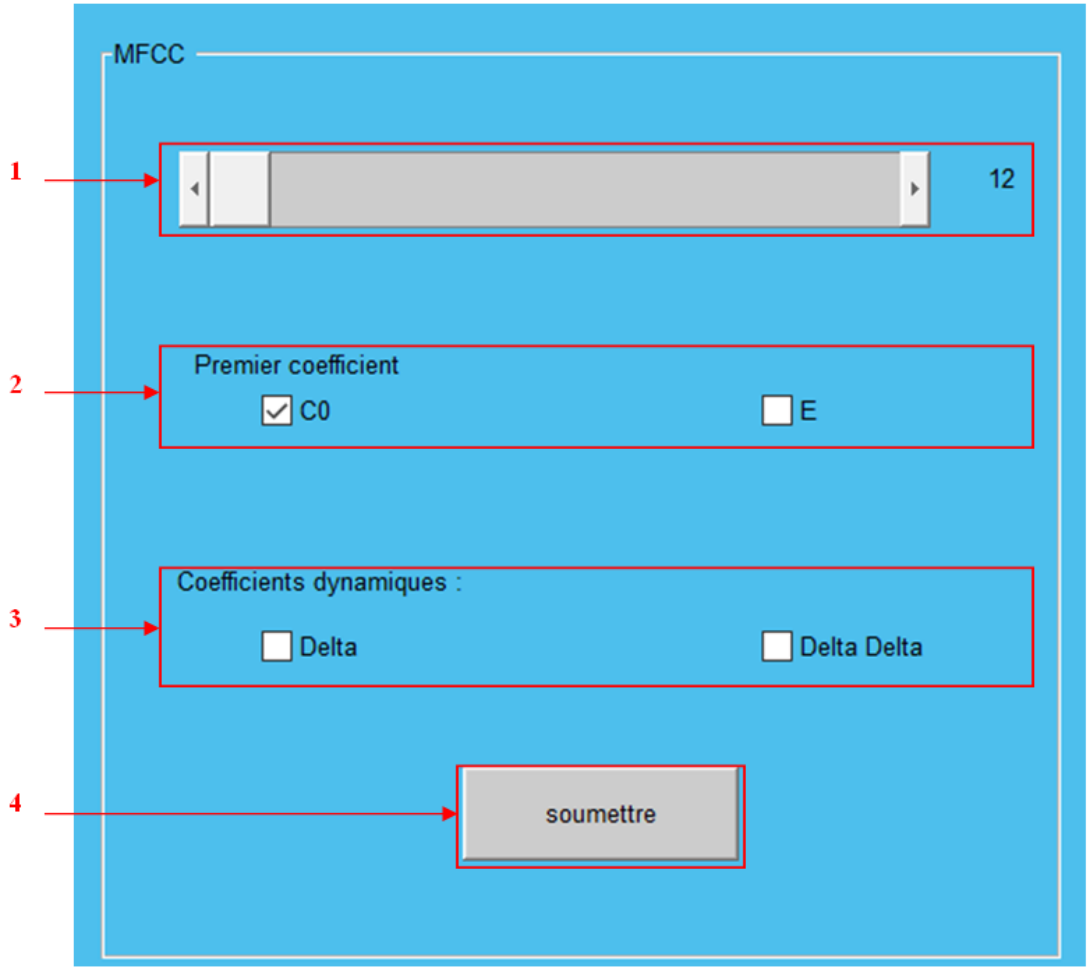

FIG. A.5 : Extraction des paramètres MFCC.

- (1)Choix du nombre des paramètres MFCC.
- (2)Choix du premier coefficient.
- (3)Choix des paramètres dynamiques.
- (4)Lancement de l'extraction .

A.3 Extraction des paramètres visuels

Lorsque l'utilisateur choisit ce procédé, il pourra à partir le descripteur LPQ et la méthode PCA de faire l'extraction des paramètres visuels et puis la classification à la fois.

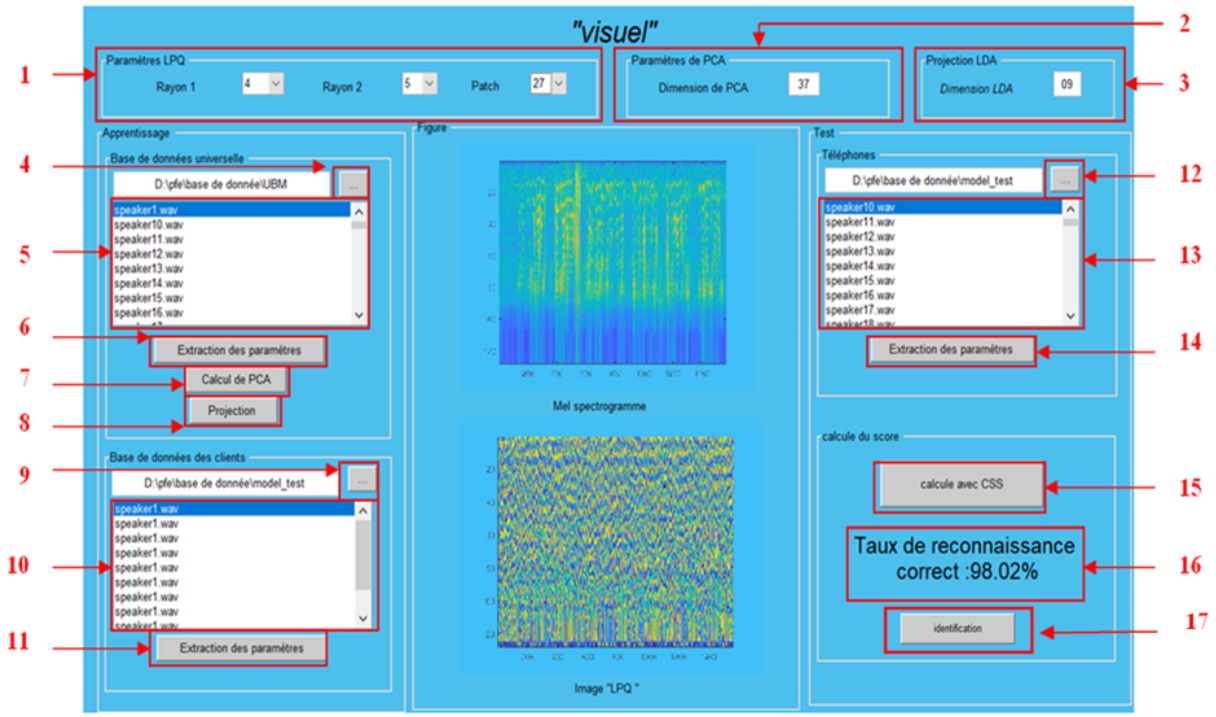

FIG. A.6 : Extraction des paramètres visuels.

- (1)Choix de rayons de l'opérateur.
- (2)Choix de dimension de PCA.
- (3)Choix de dimension de LDA.
- (4)Chargement des fichiers souhaités et leur affichage dans (5).
- (6) Extraction des paramètres.
- (7) Calcul de PCA.
- (8) Projection dans l'espace.
- (9)Chargement des fichiers souhaités et leur affichage dans (10).
- (11) Extraction des paramètres.
- (12)Chargement des fichiers souhaités et leur affichage dans (13).
- (14) Extraction des paramètres effectuée après le chargement des fichiers tests souhaités.
- (15) Calcule du score avec la méthode CSS.
- (16) Affichage du taux de reconaissance correct.
- (17) Choix d'un test d'identification.

A.4 Fusion audio-visuel

Ce procédé permet de fournir aux utilisateurs des résultats plus évidents après une fusion audio_visuel.

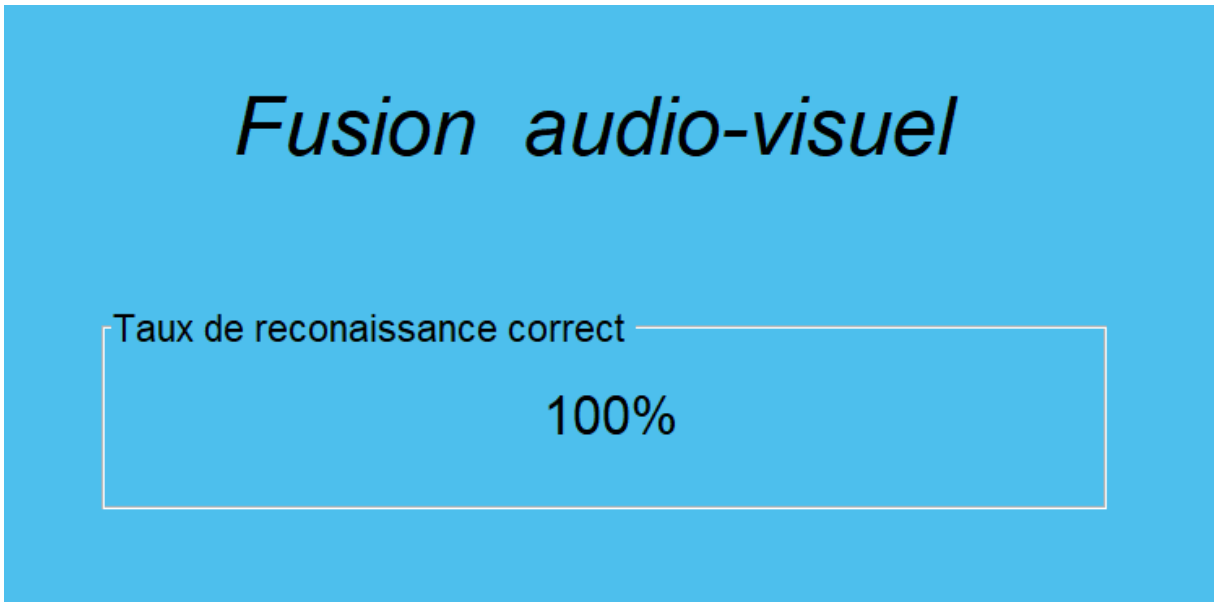

Fig. A.7 : Fusion audio-visuel.

A.5 Test d'identification

L'interface de la figure A.8 a pour le but d'identifier un téléphone mobile de la base de données de ce système en :

- Choisissant le téléphone mobile.
- Chargeant les modèles clients de ce système.
- Prenant le maximum des scores calculés entre cet téléphone mobile et les modèles des clients.
- Affichant l'identité correspondante à ce maximum.

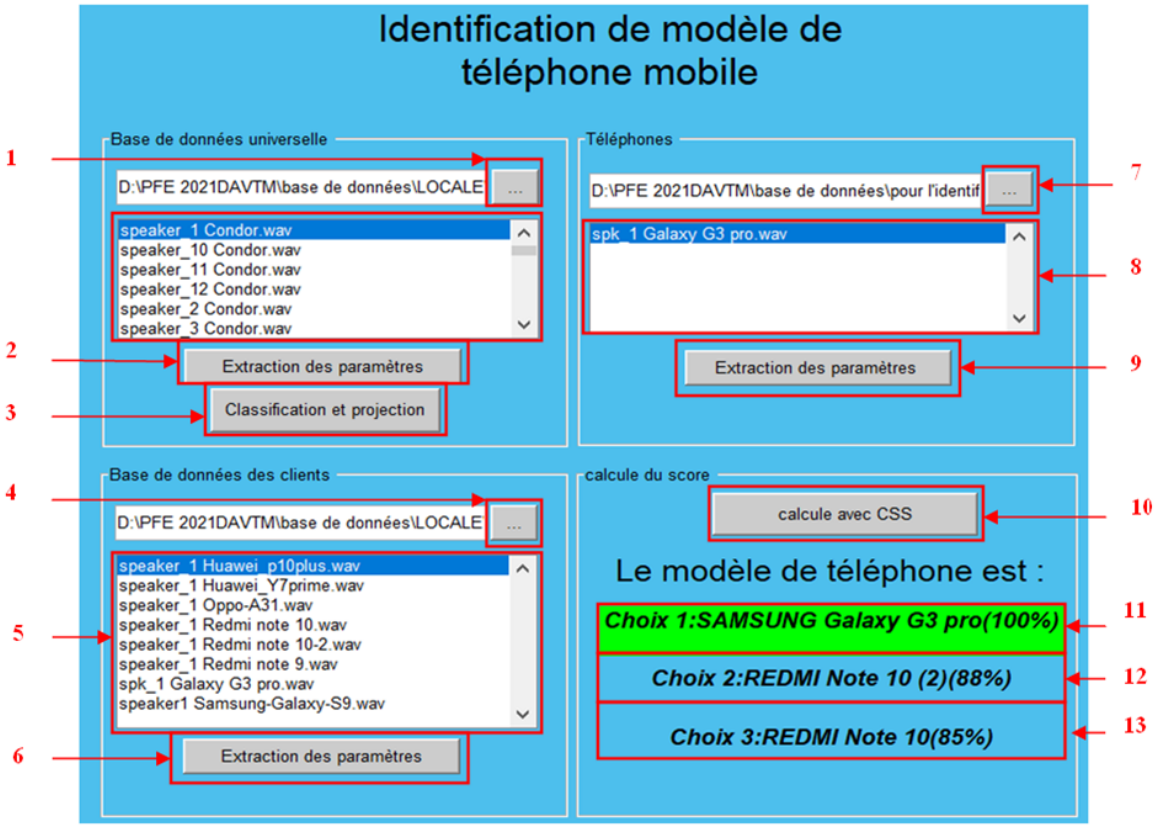

FIG. A.8 : Test d'identification.

- (1)Chargement des fichiers souhaités.
- (2)Extraction des paramètres.
- (3)Projection dans l'espace.
- (4)Chargement des fichiers souhaités et leur affichage dans (5).
- (6)Extraction des paramètres.
- (7) Chargement des fichiers souhaités et leur affichage dans (8).
- (9)Extraction des paramètres.
- (10) Calcule du score avec la méthode CSS.
- (11)Affichage du modèle de téléphone mobile le plus probable avec son taux de probabilité.Les deux probabilités suivantes sont justes affichées en dessous dans (12) et (13).