Contribution à l'évaluation opérationnelle des systèmes biométriques multimodaux

Antoine Cabana

Normandie Université

THÈSE

Pour obtenir le diplôme de doctorat

Spécialité INFORMATIQUE

Préparée au sein de l'Université de Caen Normandie

Contribution à l'évaluation opérationnelle des systèmes biométriques multimodaux

Présentée et soutenue par
Antoine CABANA

Thèse soutenue publiquement le 28/11/2018 devant le jury composé de		
M. HUBERT CARDOT	Professeur des universités, UNIVERSITE TOURS FRANCOIS RABELAIS	Rapporteur du jury
M. AMINE NAIT-ALI	Professeur des universités, université Paris-est créteil (UPEC)	Rapporteur du jury
Mme HÉLÈNE LAURENT	Maître de conférences HDR, Institut National Sciences Appliquées	Membre du jury
M. WILLIAM PUECH	Professeur des universités, UNIVERSITE MONTPELLIER 2 SCIENCES ET TEC	Président du jury
M. CHRISTOPHE ROSENBERGER	Professeur des universités, 14 ENSI de Caen	Membre du jury
M. ABDELHAKIM SAADANE	Maître de conférences HDR, Polytechnique Nantes	Membre du jury
M. CHRISTOPHE CHARRIER	Maître de conférences HDR, UNIVERSITE CAEN NORMANDIE	Directeur de thèse

Thèse dirigée par CHRISTOPHE CHARRIER, Groupe de recherche en informatique, image, automatique et instrumentation

UNIVERSITÉ CAEN NORMANDIE

MIIS

GREYC

Résumé

Le domaine de la biométrie s'intéresse à l'étude de la reconnaissance d'un individu par le biais d'une caractéristique. Les systèmes biométriques permettent de répondre à cette problématique et particulièrement dans le cas des systèmes d'information. Particulièrement, des systèmes biométriques furent intégrés dans différents types de systèmes en tant que solution d'authentification. Une authentification biométrique vise à vérifier que la caractéristique d'un individu revendiquant une identité, correspond bien à celle associée à cette identité. Ainsi, ces systèmes de reconnaissance biométrique sont susceptibles d'être utilisés en tant que solution de contrôle d'accès. En fonction des applications et/ou des dispositifs dans lesquels sont intégrés les systèmes biométriques, leur criticité peut être plus ou moins importante. Dans le cas d'application ayant des exigences sécuritaires particulièrement importante, il peut être nécessaire de réaliser une évaluation afin de déterminer les capacités et les limitations des systèmes évalués.

Ainsi, dans le cadre de la diversification de son activité de laboratoire d'évaluation, Elitt s'est intéressé à l'évaluation des systèmes biométriques afin de l'intégrer à leur offre de services. L'évaluation des systèmes biométriques est un domaine bien développé dans la littérature, qui a donné lieu à une norme internationale. Ce projet a donné lieu à cette thèse Cifre qui vise à intégrer et développer des méthodologies permettant de mettre en place un processus d'évaluation de systèmes biométriques.

Le domaine de l'évaluation des systèmes biométriques a déjà fait l'objet de travaux académiques, ainsi que d'un projet de normalisation. Ainsi, le processus d'évaluation d'un système biométrique permet d'estimer ses performances, celles-ci reflétant les capacités et les limitations du système biométrique évalué. En fonction des limitations et capacités, certains usages du système ne seront pas en adéquation avec celles-ci. La problématique de l'évaluation des systèmes biométriques est un sujet déjà abordé dans la littérature et dans des normes internationales. Les méthodologies issues de la littérature proposent de diviser une évaluation biométrique en deux processus

distincts. En premier lieu, la collecte d'échantillons biométriques permet de constituer une base de données biométriques, en capturant des échantillons biométriques auprès d'individus recrutés et constituant une population de test. Dans un second temps, afin de déterminer les performances du système testé sous la forme de plusieurs taux d'erreurs, de nombreuses comparaisons biométriques sont réalisées en appliquant l'algorithme biométrique sur une référence et un échantillon biométrique.

Néanmoins, les méthodologies proposées postulent que l'évaluateur possède des privilèges lui permettant de manipuler directement l'algorithme biométrique, lui permettant de sélectionner la référence biométrique, ainsi que les échantillons de comparaison, afin de collecter les résultats des comparaisons effectuées. Cependant, les industriels ou les intégrateurs sont généralement réticents à ce qu'une tierce partie ait accès à leurs algorithmes biométriques. Conséquemment lors d'évaluation de systèmes opérationnels, l'évaluateur ne possède pas plus de privilèges que l'utilisateur final. Les options d'administration proposées étant particulièrement limitées, une méthodologie différente doit être mise en place. Cette problématique s'est imposée naturellement compte tenu des objectifs de cette thèse, ainsi une méthodologie d'évaluation des boîtes noires fut définie et éprouvée au cours d'expérimentations et d'évaluations pilotes. La méthodologie définie vise à déterminer les faux positifs et les faux négatifs admis par un système boite noire, néanmoins les traitements respectifs de ces deux types d'erreur furent différenciés. En effet, les cas de faux négatifs sont généralement observables pour un nombre d'essais raisonnable, il est ainsi possible d'estimer un intervalle de confiance en recrutant une population raisonnablement dimensionnée. Le taux de faux positif étant généralement bas, tant le nombre de comparaisons que la taille de la population empêchent la mise en place de son estimation par le biais d'un intervalle de confiance. La méthodologie choisie vise à garantir que ce taux de faux positif est bien inférieur à une valeur choisie lors de la préparation de l'évaluation. La valeur de cette borne supérieure détermine le nombre de comparaisons nécessaires et donc la taille de la population de test.

De plus, en raison des limitations des systèmes biométriques monomodaux, l'usage de la multimodalité commence à se répandre. Les systèmes monomodaux ne se basent que sur une seule source d'information (i.e. une seule caractéristique biométrique), afin de reconnaitre un individu. Alors que les systèmes multimodaux fusionnent quant à eux plusieurs sources d'informations au cours du processus de reconnaissance biométrique. Selon l'implémentation multimodale choisie, les sources d'informations peuvent être issue de différentes modalités, de différents échantillons, de plusieurs instances, ou utiliser plusieurs méthodes d'extractions et représentation sur un même échantillon ... La multimodalité permet de réduire les limitations inhérentes aux

sous-systèmes biométriques composant le système testé, en effet les différentes sources d'informations sont mises en commun par le biais d'un processus de fusion. La fusion est en général un processus entrainé afin de minimiser les probabilités d'occurrence de faux positifs et de faux négatifs. Une des évaluations réalisée en boite noire a testé une évolution de la méthodologie proposée pour les boites noires afin de s'adapter aux systèmes multimodaux. Le système considéré par cette évaluation, implémentait une fusion de deux modalités : la voix et les empreintes digitales.

Les travaux présents dans la littérature permettent de mettre en évidence que les faux négatifs admis par un système biométrique peuvent être corrélés à la qualité de l'échantillon biométrique capturé en vue d'une comparaison. Pour certaines modalités biométriques, des métriques de qualité ont été définies afin de fournir un a priori global sur les performances biométriques d'un échantillon, tout en étant agnostique du ou des systèmes biométriques. Une métrique de qualité particulièrement reconnue et utilisée est NFIQ, qui se propose d'étudier une capture d'empreinte digitale afin d'en déterminer la qualité, et de fournir un a priori pour tout ou majorité de système de reconnaissance d'empreintes digitales. Motivée par l'étude de la multimodalité, cette thèse propose une métrique de qualité dans le cadre de la reconnaissance du locuteur. Celle-ci cherche à corréler certaines caractéristiques extraites du signal de voix avec le score de correspondance biométrique d'un algorithme biométrique. Cette métrique fut entrainée par le biais d'un algorithme génétique, et permit de déterminer une méthode de fusion corrélant de manière significative les caractéristiques extraites avec le score de correspondance de l'algorithme génétique.

Remerciements

« La thèse n'est pas un long fleuve tranquille »

Un thésard anonymisé, à l'insu de son plein gré, par les outrages du temps

Le présent ouvrage s'attarde à présenter les travaux de recherche que j'ai mené pendant ma thèse (une période représentant un peu plus de trois ans), mais néanmoins il n'est pas fait mention des personnes impliquées directement ou indirectement dans la réalisation de cette dernière. D'où la praticité de cette section de remerciement, qui d'une part me permet de leur rendre hommage, et qui d'autre part me permet de changer de ton et d'aborder des sujets beaucoup moins techniques. En effet, une thèse est issue du travail du doctorant, mais aussi de l'apport plus ou moins direct de certaines personnes, je profite donc de cette section afin de leur témoigner mes remerciements.

Ainsi, je souhaite commencer par remercier les personnes directement impliquées dans la réalisation et l'encadrement de cette thèse. En tant que thèse CIFRE, j'ai été amené à travailler au sein de la société Elitt, je tiens donc à remercier d'une manière globale tous mes collègues d'Elitt, et plus particulièrement Alain et Véronique qui ont activement pris part à l'encadrement de ma thèse et aux projets relatifs à la biométrie. Je vous remercie donc de m'avoir apporté leurs connaissances dans le domaine de l'évaluation, et leur vision pragmatique pour la mise en œuvre d'un tel procédé sur des systèmes biométriques. Je tiens aussi à remercier tous mes collègues d'Elitt, qui m'ont permis de m'être en œuvre plusieurs expérimentations, ainsi que les évaluations de plusieurs systèmes biométriques ; et qui l'ont fait dans la bonne humeur et avec la grande patience face à la répétition rébarbative de captures biométriques.

D'autre part, au niveau académique, je tiens à remercier les membres de l'équipe biométrie et monétique au sein de laquelle s'est déroulée cette thèse. Ainsi, je remercie Christophe Charrier de m'avoir encadré au cours de ces trois années, pour ses conseils

et encouragements, ainsi que pour sa confiance qui m'a permis de mener mes travaux avec un grand degré de liberté.

Je tiens aussi à exprimer ma reconnaissance aux rapporteurs de ma thèse, Hubert Cardot et Amine Nait-Ali, pour avoir accepter de rapporter ma thèse et pour le travail que cela a représenté. Je remercie également Hélène Laurent, William Puech, Abdelhakim Saadane et Christophe Rosenberger d'avoir accepté de prendre part à mon jury de thèse. Je tiens à remercier tous les membres de mon jury pour leurs questions et l'attention particulière qu'ils ont porté à mon manuscrit et à ma présentation de thèse. Les remarques et les questions qui ont été faites tant au cours de la séance de question que par le biais des rapports, ont été très intéressantes du point de vue scientifique et m'ont permis d'appréhender mes travaux et mes résultats de manière plus conclusive.

Les bons souvenirs de cette thèse viennent aussi de mes collègues, même si les discussions des pauses-café tournaient rarement autours de la biométrie. Je tiens donc à remercier du coté d'Elitt : Rachid, Pierre, Simon, Arnaud, *et al.* qui m'ont permis de ne pas limiter mes journées au sujet de la biométrie.

Le temps passé au laboratoire trouva une partie non négligeable de son intérêt grâce aux discussions qui eurent lieu au cours des repas au $\mathfrak{RU}$, et des pauses-café. Je tiens ainsi à remercier les autres doctorants : Germain, Mathieu, Tanguy, Xinwei et le grand spécialiste du χ^2-meter : Denis ; les UB-men : Gwen, Peter et Gaétan ; et Kévin pour leur bonne humeur et nos fréquentes discussions. Je tiens à remercier Alexandre, pour m'avoir laissé squatter impunément son bureau, que ce soit pour travailler ou pour discuter de sujets divers et variés. Je remercie Erdal et Estelle qui ont su me guider sur la voix, et m'ont fourni de précieux conseils et informations sur mes travaux d'estimation du degré argentique d'échantillons de parole, ainsi que Jean-Marie et Morgan pour leurs conseils avisés.

Je tiens aussi à exprimer ma reconnaissance pour mes amis qui ont su me soutenir et m'encourager depuis des distances variables, ainsi merci à la récente famille Lucas : Céline, Thomas et Lilou "Dallas" ; Pierre, Aurélie, Coline, Nicolas, les Margiota : Estelle et Adrien ; Jean-Eudes et la famille Gobert au complet. Je remercie aussi ma famille, mes parents et ma sœur qui m'ont soutenu et encouragé tout au long de cette thèse et de mon exil caennais.

Table des matières

Table des figures

Liste des tableaux

Introduction

En ce début de 21ème siècle, une des problématiques les plus prégnantes est celle de l'identité dans le cadre du numérique. L'intégration de la biométrie dans des systèmes d'information permet d'associer une caractéristique issue d'un individu avec l'identité de ce dernier. Les systèmes biométriques autorisent alors deux cas d'usage : l'authentification, qui vise à vérifier que la caractéristique d'un individu revendiquant une identité, correspond bien à celle associée à cette identité ; et d'autre part l'identification, qui ambitionne quant à elle de retrouver l'identité d'un individu en confrontant une caractéristique de ce dernier avec celles présentes dans la base de références. Les principales mises en pratique de l'authentification sont la mise en place d'un contrôle d'accès physique ou logique, ceux-ci permettent donc de réglementer l'accès à une zone ou à des données/applications particulières de manière automatisée, et sans nécessité de supervision humaine. Tandis que les cas d'utilisation de l'identification biométrique relèvent principalement du domaine régalien. L'identification biométrique étant actuellement majoritairement utilisée dans le cadre d'investigations criminelles, ou dans certains cas afin d'éviter des fraudes aux prestations sociales. Ainsi, l'enregistrement des caractéristiques biométriques (se limitant principalement aux empreintes digitales) des criminels et dans certains cas des délinquants ont lieu dans les pays ayant investi dans une infrastructure d'identification à l'échelle nationale. Certains pays ont lancé un projet de recensement biométrique de leur population avec pour objectif de grandement limiter voire supprimer les fraudes, le projet Aadhaar en étant le représentant le plus ambitieux. En effet, ce projet consiste en la création d'une base de données contenant des échantillons biométriques issus de tous les citoyens indiens. Plus proche de nous, le gouvernement français a mis en place le fichier TES [1], liant les identités et plusieurs caractéristiques morphologiques (dont le visage et les empreintes digitales) des détenteurs de passeport et de cartes d'identité.

1. https://www.legifrance.gouv.fr/eli/decret/2016/10/28/INTD1619701D/jo/texte

La mise en place de tels systèmes requérait une infrastructure conséquente, et les limitait donc à des dispositifs fixes. Néanmoins, le perfectionnement et la miniaturisation globaux des composants ont permis aux industriels de mettre au point des dispositifs connectés, et en particulier les smartphones. Ce type d'appareils peut autoriser l'accès à des applications, des données ou des services plus ou moins sensibles (accès aux banques en ligne, consultation de correspondance électronique, achats en ligne ou physiques). En vue du caractère potentiellement sensible de l'usage des systèmes biométriques, et de sa popularisation auprès du grand public par le biais d'intégration dans des produits touchant une grande audience (en particulier les smartphones et les tablettes), la détermination de l'adéquation d'un système biométrique et de ses usages est nécessaire, et peut être réalisée par le biais d'une évaluation.

Le processus d'évaluation d'un système biométrique permet d'estimer ses performances, celles-ci reflétant les capacités et les limitations du système biométrique évalué. En fonction des limitations et capacités, certains usages du système ne seront pas en adéquation avec celles-ci. La problématique de l'évaluation des systèmes biométriques est un sujet déjà abordé dans la littérature et dans des normes internationales. Les méthodologies issues de la littérature proposent de diviser une évaluation biométrique en deux processus distincts. En premier lieu, la collection d'échantillons biométriques permet de constituer une base de données biométriques, en capturant des échantillons biométriques auprès d'individus recrutés et constituant une population de test. Dans un second temps, afin de déterminer les performances du système testé sous la forme de plusieurs taux d'erreurs, de nombreuses comparaisons biométriques sont réalisées en appliquant l'algorithme biométrique sur une référence et un échantillon biométrique.

Néanmoins, les méthodologies proposées postulent que l'évaluateur possède des privilèges lui permettant de manipuler directement l'algorithme biométrique, lui permettant de sélectionner la référence biométrique, ainsi que les échantillons de comparaison, afin de collecter les résultats des comparaisons effectuées. Cependant, les industriels ou les intégrateurs sont généralement réticents à ce qu'une tierce partie ait accès à leurs algorithmes biométriques. Conséquemment lors d'évaluation de systèmes opérationnels, l'évaluateur ne possède pas plus de privilèges que l'utilisateur final. Les options d'administration proposées étant particulièrement limitées, une méthodologie différente doit être mise en place. Cette problématique a été abordée au cours de cette thèse afin de définir une méthodologie d'évaluation de tels systèmes dits « boite noire ». Pour ce faire des expérimentations ont été réalisées, ainsi que des évaluations impliquant une population d'une taille suffisante pour que les résultats

soient significatifs.

De plus, en raison des limitations des systèmes biométriques monomodaux, l'usage de la multimodalité commence à se répandre. Les systèmes monomodaux ne se basent que sur une seule source d'information, afin de reconnaitre un individu. Alors que les systèmes multimodaux fusionnent quant à eux plusieurs sources d'informations au cours du processus de reconnaissance biométrique. Selon l'implémentation multimodale choisie, les sources d'informations peuvent être issues de différentes modalités, de différents échantillons, de plusieurs instances, ou utiliser plusieurs méthodes d'extractions et représentation sur un même échantillon... La multimodalité permet de réduire les limitations inhérentes aux sous-systèmes biométriques composant le système testé. En effet les différentes sources d'informations sont mises en commun par le biais d'un processus de fusion. La fusion est en général un processus entrainé afin de minimiser les probabilités d'occurrence de faux positifs et de faux négatifs. Une des évaluations réalisée en boite noire a testé une évolution de la méthodologie proposée pour les boites noires afin de s'adapter aux systèmes multimodaux. Le système considéré par cette évaluation implémentait une fusion de deux modalités : la voix et les empreintes digitales.

Les travaux présents dans la littérature permettent de mettre en évidence que les faux négatifs admis par un système biométrique peuvent être corrélés à la qualité de l'échantillon biométrique capturé en vue d'une comparaison. Pour certaines modalités biométriques, des métriques de qualité ont été définies afin de fournir un a priori global sur les performances biométriques d'un échantillon, tout en étant agnostique du ou des systèmes biométriques. Une métrique de qualité particulièrement reconnue et utilisée est NFIQ, qui se propose d'étudier une capture d'empreinte digitale afin d'en déterminer la qualité, et de fournir un a priori pour tout ou majorité de système de reconnaissance d'empreintes digitales. Motivée par l'étude de la multimodalité, cette thèse propose une métrique de qualité dans le cadre de la reconnaissance du locuteur.

Chapitre 1

Présentation du domaine et positionnement

« Esse est percipi aut percipere » (Être, c'est être perçu ou percevoir)

George Berkeley (1685 - 1753)

Ce chapitre se propose de présenter le domaine de la biométrie, et d'introduire le sujet de cette thèse et son positionnement au sein de ce domaine. Ce chapitre débutera donc par une définition et un bref historique de la biométrie et de ce domaine, pour se poursuivre avec une présentation des motivations ayant mené à la réalisation de cette thèse, et à la définition de son positionnement.

Sommaire

1.1 Biométrie : un peu d'Histoire

La biométrie est actuellement en train de s'imposer comme une solution d'authentification grand public, en s'intégrant de plus en plus dans le quotidien de ce début de 21ème siècle. Il est ainsi maintenant possible de payer, de passer des frontières, d'accéder à des locaux ou des données par le biais d'un service d'authentification biométrique. Néanmoins, la biométrie est aussi utilisée dans le cadre d'investigations scientifiques, cet usage a été particulièrement popularisé par de nombreuses séries policières modernes. Cet usage vise à détecter de possibles correspondances d'une caractéristique au sein d'une population connue, afin de limiter la liste de possibles suspects dans le cadre d'une enquête policière.

Étymologiquement, la biométrie se compose de deux racines grecques, d'une part *bios* et d'autre part *metron*, qui signifient respectivement « vivant » et « mesure ». Cette étymologie ramène donc à la mesure du vivant, dans le cadre du domaine de l'informatique ce terme désigne un domaine préexistant visant à déterminer l'identité d'un individu à partir de caractéristiques qui lui sont propres. Ce domaine scientifique émergea au cours du 19ème siècle, et continua de s'étoffer au cours du 20ème siècle et particulièrement dans sa seconde moitié, permettant à ce domaine d'obtenir une maturité lui permettant de s'intégrer dans le marché grand public de nos jours ; en particulier par le biais de son intégration en tant que solution d'authentification dans les smartphones. Néanmoins, les études et les utilisations de la biométrie sont antérieures.

Les premières traces d'utilisation de la biométrie furent relevées dans le cadre de recherches archéologiques, en effet des empreintes digitales furent utilisées afin de signer des documents ou des objets d'artisanat, en particulier chez les Babyloniens sur des documents datés du 19ème siècle avant notre ère. En Chine au cours du 3ème siècle av. J. C., les empreintes digitales furent aussi utilisées sur des sceaux en argiles afin d'authentifier des documents scellés[19].

Les premières études scientifiques qui nous sont parvenues datent du 19ème siècle, et en particulier au cours de sa seconde moitié. Ces travaux furent motivés par le manque de moyens pour identifier les individus, et en particulier les récidivistes qui pouvaient alors changer d'identité par le biais d'alias et de faux papiers. Dans le cadre d'investigations, les principaux éléments à charge ou décharge étant des preuves, des témoignages et des aveux ; les décisions de justice se basaient donc sur des éléments à la fiabilité variable[49]. Les méthodes d'investigation d'alors étaient basées uniquement sur des preuves circonstancielles et des témoignages. Ainsi, William J. Hershel étudia les empreintes digitales afin d'authentifier les bénéficiaires d'un service de pension[99], il continua ses études afin de déterminer l'unicité des empreintes

digitales.

En parallèle, Francis Galton étudia les empreintes digitales (ou dermatoglyphes) sous un angle statistique afin d'estimer la probabilité que deux personnes aient des empreintes identiques[92], ou indistinguables ; en 1893, le Home Minister Office reconnut l'unicité des empreintes digitales. Il fournit aussi un système de classification pour les empreintes digitales[32], qui fut par la suite adapté par E. R. Henry pour être utilisé par les forces de l'ordre. Le système de classification de Henry définit ses différentes classes en fonction du motif formé par l'empreinte digitale, l'utilisation de ce système de classification permet d'éviter de comparer des empreintes de classes différentes[38]. La première utilisation d'une empreinte digitale comme preuve dans le cadre d'une enquête criminelle eut lieu en 1891, et fut réalisée par Ivan Vucetic en Argentine[31]. Suite à ces travaux, de nombreux gouvernements perçurent le potentiel de cette technique afin d'identifier leurs citoyens[31].

Concurremment, une autre méthode d'identification des individus fut développée par Alphonse Bertillon, qui mit en place un système dit d'anthropométrie judiciaire en 1888. Ce système permet de reconnaître une personne par le biais de descriptions et de mesures de certains traits (par exemple, la taille de la personne, la couleur des cheveux, des yeux...), accompagné de deux clichés photographiques : un de face, et le second de profil dans des conditions spécifiques. Bertillon propose donc une méthodologie afin de créer des fiches d'identification, néanmoins l'idée d'utiliser des photographies afin d'identifier les personnes lui est antérieure[7, 8]. En effet, l'utilisation de photographie dans le cadre du maintien de l'ordre remonte à 1843 ou 1844 à Liverpool, néanmoins aucune méthodologie rigoureuse ne fut développée avant la méthode de Bertillon (ou Bertillonnage, aussi connue sous le terme de "*mug shot*" dans les pays anglophones). Ce système se répandit alors à travers l'Europe, les États Unis et la Russie. La grande fiabilité du système d'identification par empreintes digitales éclipsa le système mis au point par Bertillon, qui resta en place en France jusqu'aux années dix-neuf cent soixante-dix.

Ces premières applications furent pionnières dans le domaine de la biométrie, mais nécessitaient systématiquement un traitement manuel afin de procéder à l'identification des individus. En dépit d'ingénieuses méthodes de classification mises en place pour les systèmes d'identification, les requêtes d'accès et d'identification devinrent nombreuses ; et ne pouvaient plus être satisfaites en un temps raisonnable avec les méthodologies de comparaisons manuelles. Concomitamment, l'émergence de l'informatique permettait l'automatisation de certains procédés, ainsi des recherches s'intéressèrent à son application au domaine de l'identification biométrique.

Plusieurs gouvernements se lancèrent ainsi dans cette entreprise de recherche

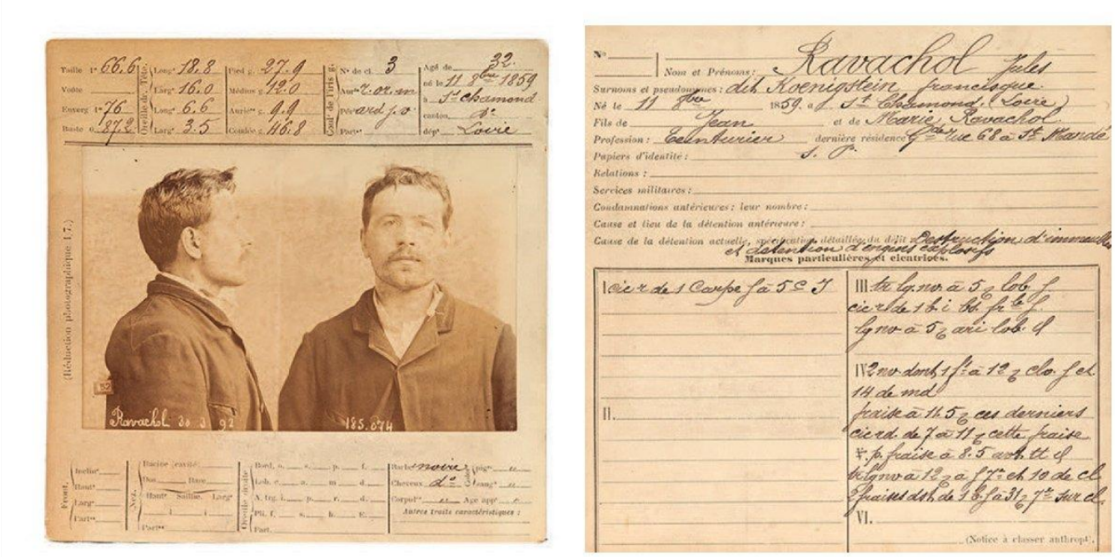

FIGURE 1.1 – Exemple de fiche du système Bertillon. Cette fiche présente une photographie de face et de profil de l'individu, ainsi que de différentes mesures anthropométriques. Parmi celles-ci des mesures de la tête, de la longueur des bras, des jambes et des pieds. La couleur de l'iris est, elle aussi, renseignée.

visant à mettre au point un système automatisé d'identification. Les États Unis se lancèrent dans un projet de système automatique d'identification par empreintes digitales ou AFIS au début des années soixante. Ainsi en 1963, Mitchell Trauring publia la première méthode de comparaison automatique des empreintes digitales se basant sur la détection des minuties[98]. Parallèlement, au Royaume Uni, en France et au Japon, des recherches semblables sont menées, mais concernent particulièrement la confrontation d'une empreinte latente avec une base d'empreintes digitales[39, 52]. En 1972, le FBI reçut le premier prototype d'AFIS, par la suite de nombreux services de police convertirent leurs registres d'empreinte sur carte par le biais des premiers modèles d'AFIS.

Les années soixante-dix virent aussi l'apparition des premières applications permettant l'identification ou l'authentification de manière automatique. Ainsi, le premier système d'authentification biométrique commerciale apparut en 1974, et utilisait la géométrie de la main comme modalité biométrique.

D'autres caractéristiques furent étudiées au cours de ces décennies, ainsi des modèles acoustiques de la voix furent développés dans les années soixante par Gunnar Fant[26], et un système semi-automatique d'identification du visage fut mis au point

par Woodrow Bledsoe[75, 6]. Dans ce dernier, l'opérateur extrayait manuellement les coordonnées de chacune des caractéristiques sur une photographie, et les calculs de distance étaient alors automatisés. La représentation du visage proposée par cette méthode est géométrique, utilisant les distances entre les points caractéristiques désignés par l'opérateur humain. Ainsi, les années soixante-dix virent la progression de l'automatisation de la reconnaissance faciale, néanmoins la solution proposée recourraient toujours à une extraction manuelle des caractéristiques. A la jonction des années quatre-vingts, quatre-vingt dix, une nouvelle technique de modélisation des visages est apparue : Eigenface[102, 101], celle-ci repose sur une approche *holistique* de cette modalité, Ceci permettant de définir un espace des visages, celui-ci étant défini par un ensemble de vecteurs caractéristiques extrait lors d'une phase d'apprentissage. D'autres méthodes *holistiques* furent définies dans la suite de cette dernière au cours des années quatre-vingt dix, mais cette décennie vit aussi apparaître d'autres approches.

En particulier, une approche (Wiskott *et. al.*[105]) sur l'analyse de correspondances au sein de graphes élastiques constitua un travail fondateur pour la reconnaissance faciale basée sur des modèles. Ces méthodes basées sur les modèles visent à dériver une représentation indépendante de la pose, de l'orientation d'un visage par la construction d'un modèle en deux ou trois dimensions. La construction de ces modèles utilise l'extraction de points de repères caractéristiques (extraction des points correspondant à la commissure des lèvres, le coin des yeux, l'extrémité du nez ...). L'usage de modèles en trois-dimensions fut amélioré par le développement de modèles façonnables par Blanz et Vetter[11, 12], qui employèrent en conjonction à la fois des informations issues de la texture du visage et de la représentation en trois dimensions du visage.

Dans un premier temps, le développement du domaine biométrique fut surtout centré sur la création de systèmes de reconnaissance automatique pour la modalité des empreintes digitales. En effet, cette modalité est couramment utilisée par les forces de l'ordre afin d'identifier des suspects.

1.2 Modalités biométriques

La biométrie désigne le domaine d'étude de la reconnaissance des individus par le biais de caractéristiques. Celles-ci sont aussi désignées par le terme de modalités, et doivent satisfaire un certain nombre de propriétés. Ces propriétés ont été définies afin de garantir qu'un système de reconnaissance implémentant une modalité soit fonctionnel, c'est-à-dire qu'il soit en mesure d'être utilisé par la quasi-totalité des

individus, tout en garantissant sa capacité à distinguer deux individus différents. Afin qu'une modalité puisse être utilisée comme caractéristique biométrique, Jain *et. al.*[48] introduisent les conditions suivantes :

- L'universalité : cette propriété spécifie si tous les êtres humains possèdent naturellement ou non cette caractéristique. Ce critère détermine si le système sera en mesure d'être utilisé par toute la population, ou seulement par une portion de celle-ci.
- L'unicité/spécificité qui indique à quel point cette caractéristique est unique/ spécifique à chaque individu. Ce critère impacte la précision théorique du système, et permet de prédire certaines limitations (impossibilité/difficulté à différencier des jumeaux, des frères/sœurs...). La modalité va aussi influencer la spécificité et donc le niveau de sécurité : une spécificité faible dénote, en effet, une tendance aux faux positifs.
- La permanence caractérise quant à elle, la tendance de la modalité à varier dans le temps. La variation de la modalité au cours du temps peut venir soit de la croissance de l'utilisateur, soit de son vieillissement. Une faible permanence peut compromettre la reconnaissance d'un utilisateur, si le temps entre la tentative et l'enregistrement est trop long. Afin de pallier cette faiblesse, plusieurs solutions sont possibles, le renouvellement régulier de la référence, ou un modèle évolutif qui s'actualise au cours des tentatives d'authentification. Une permanence basse garantie un meilleur respect de la vie privée. C'est ce qui motive le choix du contour de la main dans les restaurants scolaires, à un âge où les utilisateurs du système n'ont pas fini leur croissance, les données utilisateur n'étant plus utilisables passé un certain délai.
- La « collectabilité » qualifie la facilité avec laquelle la modalité est collectée. Ce critère peut indiquer la pénibilité de l'utilisation du système pour l'utilisateur, surtout en cas de rejet, ou d'utilisations répétées. La « collectabilité » peut donner un a priori négatif sur le temps nécessaire à la capture et la gêne utilisateur engendrés par le système biométrique, l'ergonomie du système biométrique pour les modalités les moins faciles à capturer. Ce critère peut aussi indiquer si la caractéristique biométrique est facilement capturable par un attaquant à l'insu de l'utilisateur. Un attaquant en possession d'un échantillon de la modalité peut alors tenter une attaque avec un fac-similé, ou attaque par spoofing. Certaines modalités sont particulièrement faciles à capturer à l'insu de leur propriétaire car elles sont susceptibles de laisser des traces latentes sur des objets (en particulier, les empreintes digitales).

Dans le large éventail des modalités biométriques, il est possible de distinguer deux catégories principales : d'une part, les caractéristiques morphologiques ; et d'autre part, les caractéristiques comportementales. Les caractéristiques morphologiques se basent sur des caractéristiques morphologiques, physiques ou biologiques. Parmi ces modalités, on peut citer :

- Les empreintes digitales ou dermatoglyphes
- La forme/le contour de la main
- Le réseau veineux
- L'iris
- La forme de l'oreille
- L'ADN ou acide désoxyribonucléique
- ...

L'implémentation d'une modalité comportementale, utilise l'analyse d'un comportement qui est généralement associé à une action spécifique. Ces analyses étudient généralement la dynamique, la gestuelle d'un utilisateur. Dans cette catégorie, il est possible de citer :

- La dynamique de frappe au clavier,
- La signature et l'écriture,
- La voix,
- La démarche,
- ...

Néanmoins, toutes les modalités ne répondent pas à ces propriétés ou à des degrés divers. Lors de la conception d'un système biométrique, le choix de la modalité consiste en un compromis entre ces différentes propriétés. Par exemple, l'ADN bien qu'extrêmement discriminant (sauf dans le cas des jumeaux monozygotes) est difficilement collectable, tandis que le visage est en revanche facilement capturable mais offre un pouvoir de discrimination moindre (*c.f.* les travaux de Jain *et. al.* [51, 94]). Certaines modalités peuvent même être capturées à l'insu de leur propriétaire, en particulier les empreintes digitales (ou dermatoglyphes) sont susceptibles de laisser des empreintes latentes. Roberts [84] relève la possibilité qu'un attaquant parvienne à capturer une empreinte latente d'une qualité suffisante, lui permettant ainsi d'être en mesure de fabriquer un fac-similé susceptible de tromper certains capteurs et systèmes biométriques, ou encore de réactiver un capteur et d'utiliser une empreinte latente présente sur ce dernier.

1.3 Anatomie d'un système biométrique

Un système biométrique est un système reconnaissant des individus à partir de certaines caractéristiques ; lors de sa conception et dans son fonctionnement, il est possible de distinguer différents blocs fonctionnels dans un système biométrique. Une telle représentation a été proposée par le comité ISO/IEC JTC1 SC37 (voir figure 1.2).

Cette représentation distingue ainsi cinq blocs fonctionnels, qui sont :

- Capture : extraction de données « brutes », ces données contiennent les informations relatives à la modalité biométrique de l'utilisateur. Ce composant prend la forme d'un capteur qui permet de numériser une caractéristique biométrique. Le capteur peut être spécifique à la caractéristique ou au contraire être un périphérique générique utilisé ou détourné pour capturer une modalité biométrique. Parmi les capteurs spécifiques, on peut citer les capteurs d'empreintes, du réseau veineux et dans une moindre mesure la thermographie faciale. Alors que la dynamique de frappe, la reconnaissance faciale utilisent respectivement un clavier et une caméra ou un appareil photographique.
- Pré-traitements : l'extraction des caractéristiques sur la donnée « brute », peut nécessiter des pré-traitements comme par exemple la segmentation de la zone d'intérêt : visage, empreinte.... Le contrôle de qualité de la donnée, utile pour la reconnaissance, nécessaire pour l'enregistrement (enrôlement) fait aussi partie des pré-traitements. L'objectif de l'extraction de caractéristiques est l'obtention de caractéristiques permettant de décrire la modalité biométrique contenue dans les données brutes. L'extraction de caractéristiques est la dernière étape du prétraitement. Le résultat s'appelle le template ou modèle.
- Stockage des données : stockage du/des templates de référence obtenus lors de l'enregistrement. Le stockage peut prendre plusieurs formes :
  - Sur un objet porté par l'utilisateur. Cette solution est utilisée dans le cadre de contrôle d'accès (logique ou physique) par le biais de badge contenant la référence d'un utilisateur
  - D'une base locale intégrée dans le système.
  - D'une base centrale dans le cas d'une application répartie ou sur le web. Ce dernier cas est fortement limité en raison de la législation, la CNIL donnant généralement un avis défavorable à ce genre d'application.
- Comparaison : comparaison des données d'un utilisateur clamant une identité, avec le template de référence de cette identité. A l'issue de cette étape de comparaison un score est obtenu, il représente soit la similarité soit la distance

entre le modèle de référence et le modèle de l'utilisateur. L'accès au score n'est pas disponible dans le cas d'un système biométrique dans ses conditions opérationnelles, néanmoins l'accès aux scores peut faciliter grandement l'obtention et l'amélioration des résultats de l'évaluation d'un système biométrique

- Décision : revient à réaliser un seuillage sur le score obtenu. Le type de seuillage dépend de la nature du score de la comparaison. Si le score est une similarité, on souhaitera que celle-ci soit supérieure à un seuil, afin d'authentifier un individu. Si le score est une distance, on souhaitera que celle-ci soit inférieure à un seuil. La décision est ensuite transmise en sortie du système. En sus des précédentes composantes et non représentés sur la figure 1.2, deux autres blocs fonctionnels d'un système biométrique peuvent être ajoutés :
- Administration : interface d'administration du système biométrique. Elle permet de gérer les références, les effacer, en ajouter... Mais aussi de régler plusieurs paramètres du système, en particulier le ou les seuils utilisés lors de la décision, la qualité minimale des échantillons utilisés lors des comparaisons... Dans le cas de certains systèmes, en particulier ceux embarqués sur smartphones, l'utilisateur n'a accès qu'à très peu d'option d'administration, il n'est pas ainsi capable de régler le niveau voulu de sécurité. Les fonctionnalités sont généralement réduites à la gestion des références (enregistrement et suppression), et éventuellement aux applications pour lesquelles elles sont utilisées. Ce bloc fonctionnel n'est pas représenté sur la figure 1.2.
- Transmission : ensemble composé du transport des informations entre le capteur et le module réalisant la comparaison, du transport de la référence entre la base de données où elle est stockée et le module réalisant l'application de l'algorithme (module de matching) et de la transmission de la décision de l'algorithme. Cette partie doit garantir qu'il n'est pas possible d'injecter des informations dans le système, ou d'en intercepter. Ce bloc fonctionnel n'est pas représenté sur la figure 1.2.

1.3.1 Multimodalité : fusion biométrique

L'implémentation d'un système biométrique repose sur l'extraction d'informations d'une caractéristique biométrique, qui sont comparées à l'aide d'un algorithme de comparaison. Toutefois, ces systèmes admettent des erreurs de classification qui suivant les cas d'usage du système concerné peuvent présenter une gravité plus ou moins importante. L'éventualité d'un usage particulièrement sensible motive l'amélioration des performances, afin d'augmenter la fiabilité globale du système.

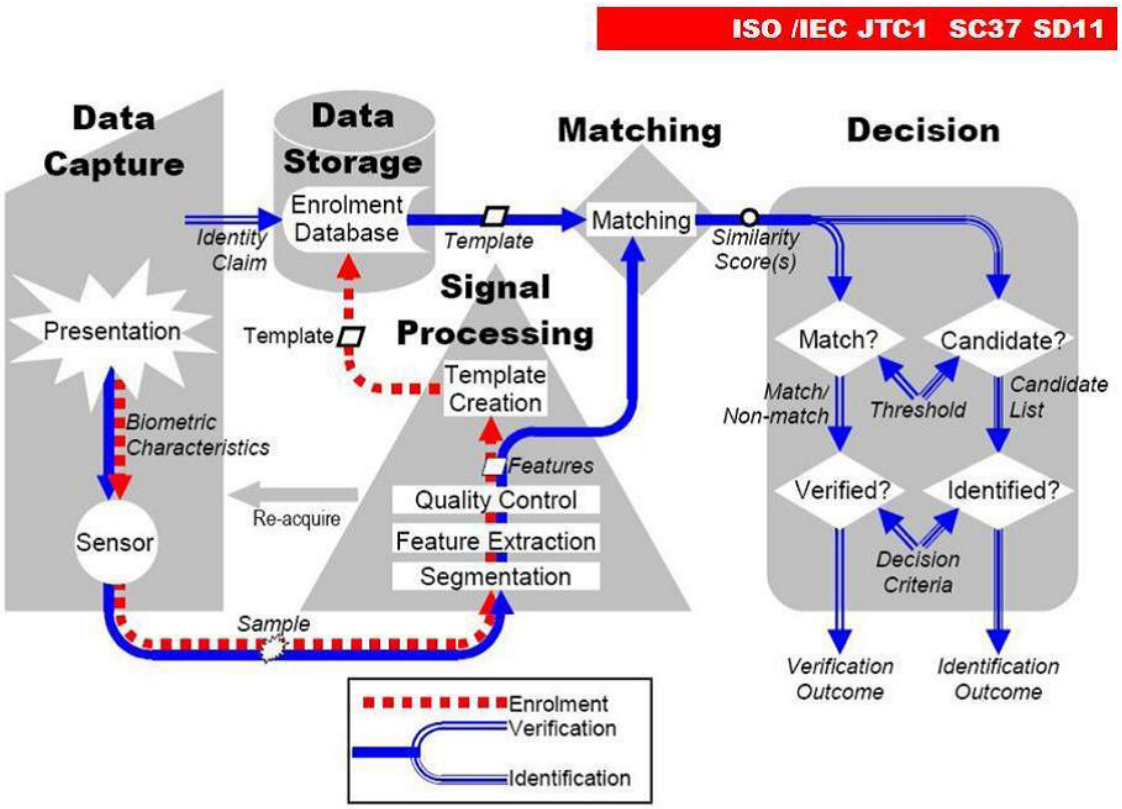

FIGURE 1.2 – Représentation d'un système biométrique selon le modèle proposé par l'ISO/IEC 19795-1[43]. Cette figure permet de distinguer les différents blocs fonctionnels d'un système biométrique et les interactions entre ceux-ci.

La fusion biométrique, aussi désignée sous les termes de multibiométrie ou de biométrie multimodale, implique l'extraction de caractéristiques issues de différentes sources : celles-ci pouvant être des traitements effectués sur une même caractéristique, ou au contraire provenir de différentes caractéristiques biométriques.

Néanmoins, la fusion biométrique ou biométrie multimodale s'accompagne d'un coût issu du procédé de fusion ; en effet cette dernière, suivant la stratégie de fusion, peut nécessiter la présence de plusieurs capteurs, et/ou l'utilisation conjointe de plusieurs algorithmes de comparaison.

1.3.1.1 Présentation des différents types de fusion

Différentes stratégies existent afin de réaliser la fusion de différentes sources d'informations (plusieurs méthodes sont ainsi proposées par Maltoni *et. al.*[65] et Bir Bhanu *et. al.* [9]), ainsi il est possible de distinguer les stratégies suivantes :

- Fusion de plusieurs caractéristiques : cette stratégie de fusion réalise une

reconnaissance biométrique, à l'aide de plusieurs modalités. L'implémentation d'un tel système requiert généralement plusieurs capteurs, à moins que les différentes modalités ne soient susceptibles d'être enregistrées par un seul capteur. Des capteurs spécialisés ont été développés afin de capturer en une seule fois plusieurs caractéristiques. Le principal exemple d'une telle architecture est un capteur dual permettant de capturer en même temps une empreinte digitale et le réseau veineux d'un même doigt.

- Fusion multi capteur : cette stratégie de fusion repose sur l'utilisation conjointe de plusieurs capteurs afin de capturer une même instance d'une modalité biométrique. Cette implémentation peut se trouver être particulièrement contraignante pour l'utilisateur final. En effet, ce dernier doit présenter à plusieurs reprises sa modalité à différent capteurs.
- Fusion de différentes instances : la mise en place de cette stratégie de fusion présuppose que la modalité présente plus d'une unique instance pour la plupart des individus. Par exemple, il n'est pas possible d'utiliser cette stratégie de fusion dans le cas d'un système de reconnaissance faciale ou vocale ; mais celle-ci est particulièrement adaptée pour un système utilisant comme modalité les empreintes digitales ou dans une moindre mesure l'iris.
- Fusion de plusieurs échantillons : cette stratégie d'implémentation de multi biométrie se base sur la capture de plusieurs échantillons à l'aide d'un même capteur. La fusion consiste en la combinaison des différents échantillons, ou des différentes caractéristiques extraites de ceux-ci.
- Fusion multi-algorithmes : cette fusion utilise conjointement plusieurs représentations d'un même échantillon. Ces différents scénarii présentent les différentes sources d'information sur lesquelles peut s'effectuer la fusion biométrique.

Le processus de fusion des informations peut être effectué à différentes étapes du processus de reconnaissance biométrique. Ainsi, selon le premier ajout à la norme ISO/IEC 19795-2[46] et l'article de Ross et Jain[86], les différents niveaux de fusion possibles sont :

- Fusion des échantillons : la fusion prend place juste après la capture des échantillons. Les échantillons sont fusionnés, dans le cas d'échantillons issus de la même modalité et du même capteur, cette étape peut permettre d'améliorer la qualité de l'échantillon final ou de créer un échantillon plus étendu en superposant les parties de recouvrement des différents échantillons.
- Fusion des modèles/représentations biométriques : les modèles sont extraits des différents échantillons (ou dans le cas d'une fusion multi algorithme, d'un même

échantillon selon différentes représentations de la donnée biométrique). Cette méthode permet de consolider le modèle extrait en ne sélectionnant que les caractéristiques les plus pertinentes, ou en permettant d'étendre ce modèle dans le cas d'échantillons ne se recoupant que partiellement. La fusion de template peut aussi permettre de fusionner des modèles issus de différents capteurs, et de les superposer pour corréler la position de plusieurs caractéristiques contenues dans les différents modèles (cette technique est particulièrement pertinente dans le cas où les modèles sont issus d'une même modalité biométrique ayant subi différents traitements, ou encore dans le cas de deux modalités capturées simultanément car fortement liées telles que l'empreinte digitale et le réseau veineux d'un seul et même doigt).

- Fusion des scores : la fusion est réalisée après l'étape de comparaison, en fusionnant les scores. Une étape de normalisation peut être effectuée afin d'harmoniser la portée des valeurs de score. Différentes méthodes peuvent être mises en place afin de réaliser ce type de fusion, par exemple une somme pondérée permet d'intégrer les scores en une valeur unique, et de déterminer une décision en appliquant un seuillage sur cette dernière. La mise en place d'une fusion à ce niveau du système biométrique peut permettre de diminuer à la fois le taux de faux rejets et de fausses acceptations.
- Fusion des décisions : la décision finale est issue des décisions des sous-systèmes de reconnaissance biométrique. Le processus de fusion peut être une simple opération logique parmi le ET et le OU logique. Le cas d'usage d'un ET permettra de diminuer le taux de fausses acceptations, tandis que le OU permet de diminuer le taux de faux rejets.

1.3.1.2 Amélioration des performances

La mise en œuvre et le développement de la multimodalité ont été motivés par l'amélioration des performances biométriques qu'elle induit d'après Hong *et. al.*[41]. En effet, un système biométrique classique ne se base que sur une source d'information afin d'effectuer la reconnaissance d'une personne. Tandis qu'un système multibiométrique se base sur plusieurs sources d'informations qui, fusionnées, permettent de prendre une décision basée sur une plus grande quantité d'informations. Pour les systèmes biométriques les plus largement utilisés (tels que des projets de recensement biométrique d'une population à une échelle nationale), la capacité de discrimination d'une modalité entre deux individus peut être mise à mal. La combinaison de différentes sources d'informations augmente la capacité de discrimination du système entre deux individus.

L'utilisation conjointe de plusieurs modalités permet de limiter l'impact des vulnérabilités d'une des modalités utilisées, ou d'un des sous-systèmes biométriques. En effet, certains utilisateurs ne sont pas en mesure d'utiliser un système biométrique, en raison d'une incapacité liée à la modalité utilisée. Ces incapacités peuvent être temporaires ou permanentes. Afin de palier ce problème, la mise en pratique de la multimodalité permet de diminuer la probabilité qu'un utilisateur potentiel ne soit pas en mesure d'utiliser cette implémentation. La multimodalité peut permettre d'assurer une meilleur aptitude à être utilisée ou utilisabilité, en effet le risque de faux rejet est réduit par la plus grande quantité d'information extraite des échantillons biométriques. Le système est en effet en mesure de distinguer plus aisément différents individus, sans que la qualité ou la quantité d'informations extraites requises n'obligent l'individu à fournir un échantillon d'excellente qualité, et donc potentiellement difficile à produire ; permettant au système d'obtenir des performances comparables à un système biométrique monomodale, pour un taux de faux rejet inférieur.

De même, en fonction des méthodes de fusion mises en place et des paramètres utilisés lors de ce procédé, la multimodalité améliore les performances de reconnaissance. Cette amélioration peut se traduire par une diminution des taux de faux positifs, de faux négatifs ; voire une diminution conjointe de ces deux taux d'erreur.

1.4 Représentation des systèmes biométriques

Comme illustré dans la partie précédente, les données biométriques sont rarement utilisées telles quelles afin de réaliser l'étape de comparaison. En effet, deux captures biométriques issues d'une même instance biométrique peuvent présenter des différences notables. En effet, deux échantillons issus d'une même personne peuvent présenter divers degrés de différence, par exemple des variations d'illumination dans le cas du visage peuvent présenter des difficultés de traitement si les échantillons sont traités tels quels ; pour la reconnaissance du locuteur, dans le cadre d'un système implémentant une reconnaissance indépendante du texte deux échantillons d'une même personne ne présenteront que peu de points communs du point de vue des données brutes ...

Les systèmes biométriques recourent donc de manière quasi systématique à un modèle afin de représenter la donnée biométrique. Les modèles se basent généralement sur l'extraction de certaines caractéristiques qui présentent une variabilité peu élevée entre différents échantillons issus d'une même personne, mais suffisamment entre deux individus différents.

Par exemple dans le cas des empreintes digitales, les caractéristiques les plus

communément utilisées sont des points caractéristiques soit globaux, soit locaux. Les points globaux correspondent à des points particuliers du motif de l'empreinte, les points locaux à des points singuliers des crêtes papillaires. Dans le cas de la voix, plusieurs types de modèles ont été définis et sont utilisés afin de modéliser la voix en tant que modalité biométrique.

1.5 Cas d'usage de la biométrie

Les systèmes biométriques sont principalement utilisés afin de reconnaître un individu par le biais d'une caractéristique biométrique. La possibilité de lier une identité à une personne par une caractéristique est particulièrement utile dans deux cas de figure :

Identifier une personne : une identification vise à déterminer l'identité d'une personne à partir d'une caractéristique, en la confrontant à un ensemble de caractéristiques généralement associées à des identités connues. Ce cas d'usage est particulièrement utilisé dans le cas d'investigation policière afin de déterminer une liste de suspects. Néanmoins, la confiance accordée dans les systèmes biométriques n'est pas suffisante pour que la décision d'un système d'identification puisse être utilisée comme preuve devant un tribunal. Ainsi, chaque décision d'identification est confirmée par un ou plusieurs experts.

Authentifier une personne : l'authentification vise à confirmer l'identité d'un individu par la comparaison de sa caractéristique biométrique avec celle qui est associée à l'identité revendiquée. Ce cas d'usage est particulièrement courant comme solution de contrôle d'accès logique ou physique, et de dans le cadre du domaine régalien avec en particulier le contrôle aux frontières. En raison du perfectionnement des méthodes de reconnaissance biométrique (tant par l'aspect capteur, que traitement logiciel), ainsi que l'augmentation de la puissance de capture, celle-ci est de plus en plus communément embarqué dans des dispositifs connectés, particulièrement les smartphones. La biométrie tend à être intégrée comme solution d'authentification par le domaine bancaire, en particulier dans le cadre de transactions électroniques sécurisées. Dans le cas des empreintes digitales, l'optimisation des algorithmes de comparaison a permis de les intégrer au sein d'éléments sécurisés (ou SE, ces derniers sont particulièrement limités au regard des capacités de calcul). En effet, certaines cartes de paiement embarquent des systèmes de reconnaissance d'empreintes digitales complets.

1.6 Limites de la biométrie

Les systèmes biométriques permettent d'associer une caractéristique avec un individu, néanmoins ceux-ci ne sont pas infaillibles et sont en conséquence susceptibles d'admettre des erreurs. Ces erreurs sont de deux types :

- l'association indue d'une caractéristique enregistrée et d'un individu dont elle ne provient pas,
- ou l'issue négative d'une référence avec la caractéristique correspondante.

Ces deux types d'erreur sont respectivement désignés comme : les faux positifs et les faux négatifs.

1.6.1 Illustration des limites de la biométrie

L'intégration de la biométrie dans des appareils grand public permet d'obtenir une meilleure couverture des limites de cette dernière, tant par les réseaux médiatiques classiques ou spécialisés que par les réseaux sociaux. Cette couverture s'explique principalement par l'effet d'annonce de certains constructeurs présentant les systèmes biométriques comme des solutions d'authentification sûres et équivalentes à d'autres déjà mises en place.

Cette couverture médiatique a permis de mettre en avant les limites de certaines implémentations biométriques, et ce qui est particulièrement utile étant donné l'opacité des constructeurs et des intégrateurs concernant les performances des systèmes commerciaux. En effet, les performances ne sont pas systématiquement communiquées, et quand bien même les informations relatives à ces dernières ne sont généralement pas non plus détaillées (en particulier la méthodologie de test, et la description de la population ayant participé à cette évaluation). Ce type de publication permet aussi de mettre en évidence la fonctionnalité d'usages non prévus pour un système biométrique.

Ainsi pour des systèmes biométriques utilisés en tant que solution d'authentification, des cas d'imposture furent mis en évidence (cas d'imposture sur la reconnaissance faciale d'un iPhone X[16]) ainsi que des limitations concernant les capacités de distinction de certains systèmes pour certaines populations. En particulier dans le cas de systèmes biométriques embarqués sur smartphones, ainsi des implémentations de reconnaissance faciale ont pu se révéler sujettes à être abusées par des attaques de présentation de fac-similé (ici, une photographie de la personne enregistrée pouvait tromper le système d'authentification Face Unlock, mise en place sur Android OS[77]). Sur la constat de la vulnérabilité de ces systèmes à des attaques par fac-similé, des travaux de recherches ont tenté de détecter ce type d'attaque.

De même, l'implémentation de la reconnaissance par empreinte digitale montra une vulnérabilité au attaque par fac-similé, ainsi un regroupement de *hackers*, le *CCC*[1] attaqua le système en reproduisant une empreinte digitale[1]. L'originalité de la méthode proposée est qu'elle modélisait un cas d'attaque plausible, en effet l'acquisition de l'empreinte n'était pas faite par le biais d'un moulage, mais par la photographie d'une empreinte latente mise en évidence.

Un autre aspect moins médiatisé mais néanmoins très présent lors de l'usage d'un système biométrique est l'occurrence de cas de faux rejet. Ce type d'erreur consiste en une non-reconnaissance d'une personne dont la modalité est enregistrée dans le système (indépendamment du type de fonctionnement du système, qu'il s'agisse d'authentification ou d'identification).

1.6.2 Présentation de la problématique

Ces erreurs peuvent mener à des conséquences plus ou moins graves, la gravité d'une erreur étant généralement subordonnée à l'usage et au contexte dans lesquels la solution de reconnaissance est intégrée.

L'estimation des performances d'un système d'authentification biométrique est une forme de test paramétrique au regard des statistiques. Les cas de reconnaissance ou de rejet d'un individu sur la foi d'un échantillon biométrique, constituent alors les hypothèses de ce test paramétrique (ces dernières étant mutuellement exclusives). Les hypothèses H_0 et H_1 du test paramétrique sont alors respectivement, que l'individu réalisant l'authentification est légitime (H_0), illégitime (soit un imposteur) (H_1). La théorie classique des tests, telles qu'exposée par Saporta[88], permet alors de distinguer deux types d'erreur :

- les erreurs de première espèce : erreur qui correspond à choisir l'hypothèse H_1 alors que H_0 est vraie.
- les erreurs de seconde espèce : erreur qui correspond à conserver l'hypothèse H_0 alors que H_1 est vraie.

Ainsi dans le cas de solutions d'authentification, un faux positif amènera le système biométrique à reconnaître indûment un individu non autorisé. La sévérité d'un faux positif dépend donc entièrement du cadre d'usage du système ; ainsi si le système est utilisé comme contrôle d'accès pour un self-service, les conséquences seront mineures. Tandis que si un faux positif peut être admis par un contrôle d'accès à des services bancaires ou de paiement, ou sur des solutions de contrôle aux frontières les conséquences seront plus sévères.

1. ou *Chaos Computer Club*, un regroupement de *hackers* européens

		Positif	Négatif
Vérité terrain	Décision valide	Un utilisateur légitime est normalement reconnu par le système	Un utilisateur non-enregistré est normalement rejeté par le système
	Décision erronée	Un utilisateur non-légitime est anormalement reconnu par le système	Un utilisateur enregistré est anormalement rejeté par le système

TABLE 1.1 – Représentation des différents cas de figure susceptibles d'être admis par un système biométrique

1.6.3 Le contexte d'acquisition

Le contexte d'acquisition est constitué de la présentation de la modalité, de l'environnement dans lequel se déroule la capture et de facteurs influençant la capture de cette modalité. La variabilité intra-individuelle est la variabilité de l'instance d'une modalité observée chez un individu. Ainsi, cette variabilité peut avoir pour cause ses conditions de présentation au capteur.

Ainsi, la distance, l'orientation, la force d'application (dans le cas d'un capteur nécessitant un contact avec la modalité), l'intensité de présentation (par exemple, dans le cas de la voix, le niveau sonore)... influent grandement sur l'échantillon recueilli par le système biométrique. Tandis que l'environnement de présentation représente les conditions environnementales susceptibles de modifier la condition de la modalité, ou d'interférer avec le capteur. Les conditions environnementales sont mesurables, telles que la luminosité, la température, l'humidité ... Le contexte d'acquisition est susceptible de modifier soit la présentation de la modalité, soit la qualité de l'échantillon biométrique. L'orientation peut avoir un impact important en provoquant des occultations ou des pertes d'informations. Les modèles extraits de ces échantillons risquent d'être pauvres en information, ou incomplets. Effectuer des comparaisons avec ces modèles peut mener à des vrais négatifs par manque d'informations pertinentes. Si un échantillon/modèle de mauvaise qualité est utilisé pour l'enregistrement, il sera difficilement utilisable et il sera plus facilement sujet à imposture, en raison du nombre restreint de caractéristiques du modèle.

Certains individus sont susceptibles de présenter des facteurs modifiant la modalité et étant susceptibles d'influencer la qualité des échantillons, et conséquemment les performances du système biométrique. Ces facteurs sont de différentes natures, il peut s'agir de pathologies ou de malformations, ou encore de modifications réalisées sur la modalité. Fernandez *et. al.*[27] proposent une liste de tels facteurs, ceux-ci sont

présentés dans le tableau 1.6.3.

Facteurs	Empreinte digitale	Voix	Iris	Visage	Main	Dynamique de frappe
Fatigue	✓	✓	✓	✓	✓	✓
Distraction	✓	✓	✓	✓	✓	✓
Coopérativité	✓	✓	✓	✓	✓	✓
Motivation	✓	✓	✓	✓	✓	✓
Nervosité	✓	✓	✓	✓	✓	✓
Distance		✓	✓	✓	✓	✓
Occlusion des yeux			✓	✓		
Pression sur le capteur	✓				✓	✓
Expression faciale				✓		
Élément cosmétique				✓		
Vêtements				✓		
Chapeaux				✓		
Bijoux	✓			✓	✓	
Lunettes, lentilles			✓	✓		
Accent, langue		✓				
Mouvements, tremblements	✓		✓	✓	✓	✓

TABLE 1.2 – Ce tableau présente différents facteurs susceptibles d'influer sur la qualité des échantillons biométriques capturés, extrait de [27] .

1.6.4 Extraction de caractéristiques et comparaison

Cependant, les erreurs ne proviennent pas exclusivement de l'étape d'acquisition des échantillons. Des erreurs peuvent être faites à l'étape de comparaisons biométriques. Celles-ci peuvent provenir de l'étape d'extraction des caractéristiques, d'insuffisance du modèle représentant la modalité ou encore de l'algorithme. L'extracteur de caractéristiques est susceptible de produire des approximations, les caractéristiques générées sont alors classées incorrectement, placées de manière approximative...

Lors de l'étape de comparaison, ces erreurs sont susceptibles de créer des vrais négatifs, par le rejet d'une ou plusieurs caractéristiques en raison d'un placement mal

positionné, ou d'une classification incorrecte. De la même manière, des caractéristiques peuvent être indument appariées en raison de ces erreurs ou approximations.

La reconnaissance biométrique passe par l'utilisation de modèles servant à représenter de manière simple une donnée plus complexe et plus riche. Comme lors des processus de compression, des pertes d'informations peuvent se produire. Ces manques peuvent compromettre le processus de reconnaissance. En effet, cette perte d'information peut mener à une reconnaissance de deux modèles très proches en raison des caractéristiques biométriques utilisées par le modèle, alors que la comparaison des données brutes n'aurait pas donné une comparaison positive.

L'algorithme met en place des méthodes permettant de comparer deux modèles, ces derniers peuvent ne pas correspondre directement en les confrontant de manière brutale. Des stratégies peuvent être mises en place afin de minimiser les risques de faux rejets en effectuant des modifications telles que des alignements, des mises à l'échelle, des rotations ...

1.6.5 Les attaques

Suivant les applications dans lesquelles le système est employé, la sécurité peut s'avérer secondaire ou au contraire un enjeu particulièrement important. Un système biométrique peut être sujet à des attaques, certaines spécifiques aux systèmes biométriques, d'autres standards à tout système d'information. En particulier, les attaquants peuvent chercher à être acceptés par le système en lui présentant une reproduction de la modalité d'un utilisateur préalablement enregistré. Une reproduction varie grandement suivant la nature de la modalité, elle peut prendre la forme d'un enregistrement dans le cas de la reconnaissance vocale, d'un moulage en silicone ou gélatine dans le cas d'une empreinte digitale, ou d'une impression, d'une image dans le cas de la reconnaissance de visage ou d'iris.

Ainsi, suivant l'application dans laquelle doit s'intégrer le système, et suivant les exigences en matière d'ergonomie, de précision et/ou de sécurité, il est nécessaire de connaître les limites du système biométrique et pour ce faire, compléter l'étude de ses caractéristiques théoriques par son évaluation.

1.7 Évaluation des systèmes biométriques

Les systèmes biométriques sont susceptibles de réaliser des erreurs de classification, ou d'être abusés par des attaques, ou encore d'être perturbés par des conditions environnementales défavorables. En fonction des usages envisagés, il est nécessaire de procéder à une évaluation biométrique, celle-ci vise à déterminer et quantifier les

vulnérabilités d'un système biométrique sous test. Les résultats obtenus permettent de déterminer l'adéquation d'un système avec l'usage prévu, ou peuvent être utilisés par des industriels ou des acteurs académiques afin d'estimer les capacités du ou des systèmes testés.

1.7.1 Principes et définition

l'objectif d'une évaluation est de déterminer les vulnérabilités et les limitations des systèmes biométriques, néanmoins celles-ci peuvent être de différentes natures. Il est ainsi plus aisé de les séparer et de définir des méthodologies séparées afin d'en évaluer les répercutions. Ainsi, la norme 19795-1[43] et les travaux de Mansfield et Wayman[67, 66] distinguent et définissent trois cadres d'évaluation différents, qui sont :

- Le cadre de performances qui désigne la détermination des taux d'erreurs généraux admis par le système biométrique. Cette phase se déroule dans des conditions normales [2].
- La sécurité désigne l'étude des faiblesses d'un système biométrique face à différents types d'attaques, et de vulnérabilités exploitables par un attaquant.
- Les tests d'environnement visent à déterminer les conditions environnementales susceptibles de modifier les performances observées en conditions normales, et à en quantifier leurs influences.

1.7.2 Cadre d'évaluation

L'évaluation des systèmes biométriques vise à détecter et déterminer la propension d'un système biométrique à admettre de telles vulnérabilités. Afin de déterminer l'inclinaison d'un système pour l'une de ces vulnérabilités, il est nécessaire de mettre en place différents types d'évaluation. Cette section se propose de présenter ces différentes approches plus en détail.

1.7.2.1 Performances

Différents facteurs peuvent influencer l'acquisition d'une caractéristique biométrique, et la qualité intrinsèque de celle-ci. En effet, certaines pathologies ou

2. C'est-à-dire définie en termes de conditions de température, d'humidité de bruit ambiant... en fonction de la modalité et des capteurs utilisés par le système sous test. Ces paramètres sont définis de telle manière à ne pas être considéré comme négatifs, mais comme correspondants à un usage normal du système biométrique

malformations[3] peuvent impacter de manière significative la qualité des échantillons biométriques. De même, certaines modifications (volontaires ou non) apportées à une modalité peuvent aussi influer sur les performances de capture des échantillons biométriques.

1.7.2.2 Sécurité

Les systèmes biométriques sont susceptibles d'être utilisés comme méthodes d'authentification dans des applications ayant des exigences sécuritaires importantes : comme par exemple le contrôle aux frontières avec le passeport biométrique, ou le contrôle d'accès physique ou logique. L'évaluation de cet aspect peut donc s'avérer particulièrement important pour des applications susceptibles d'être la cible d'attaque. Des travaux ont permis d'identifier et de spécifier les points vulnérables d'une implémentation biométrique. En particulier Ratha *et. al.*[80] ont défini huit points de vulnérabilité pour un système biométrique.

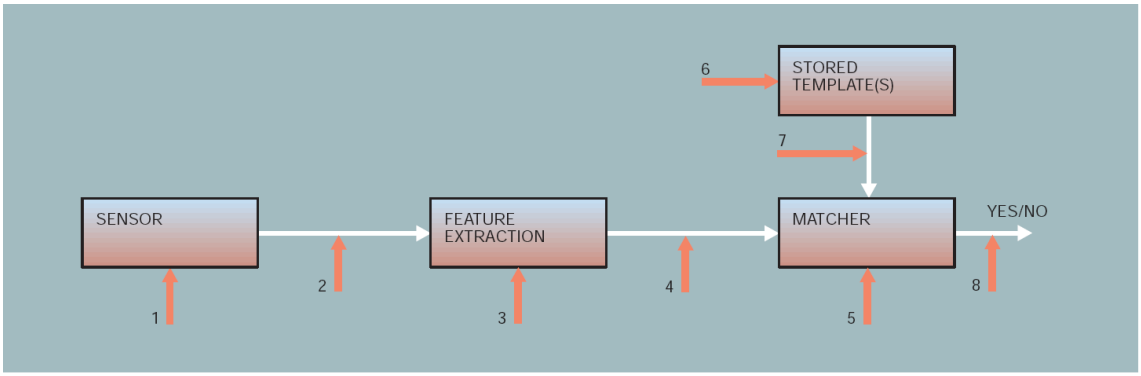

FIGURE 1.3 – Représentation des huit points de vulnérabilité d'un système biométrique selon Ratha *et. al.*[80]. Celle-ci est une version simplifiée du modèle proposé par l'ISO/IEC 19795-1 dans la figure 1.2.

La figure 1.2 présente les points de vulnérabilité suivant :

1. Au niveau du capteur, présentation d'une contrefaçon d'une modalité biométrique : une reproduction d'une modalité biométrique connue du système est présentée en entrée du système. Par exemple dans le cadre d'un système biométrique utilisant comme modalité les empreintes digitales, ce type d'attaque sera la présentation d'un faux doigt.

3. Ces dernières ne sont présentes ou ne touchent généralement qu'une infime partie de la population, ou ne sont que temporaires, sinon le caractère universel de la modalité peut être remis en question.

2. Attaque par rejeu de données biométriques. Cette attaque consiste à injecter juste après le capteur, des données qui ont été précédemment interceptées lors d'une transaction légitime. Par exemple dans le cas d'un système de reconnaissance d'empreintes digitales, l'attaque par rejeu va consister à injecter une image de doigt d'un utilisateur légitime.
3. Réécriture de l'extracteur de caractéristiques, par le biais d'un cheval de Troie, qui renvoie des jeux de caractéristiques choisies par l'attaquant.
4. Falsification des caractéristiques extraites, les caractéristiques envoyées par l'extracteur de caractéristiques sont remplacées par un jeu malveillant choisi par l'attaquant (implique que la méthode d'extraction des caractéristiques soit connue de l'attaquant). Ce point de vulnérabilité peut être extrêmement difficile à exploiter car les modules de comparaison et d'extraction de caractéristiques peuvent ne pas être dissociables.
5. Réécriture du composant de comparaison, qui transmet alors des jeux de scores choisis par l'attaquant.
6. Accès à la base de référence par :

   a) Injection d'un nouveau template correspondant aux caractéristiques de l'attaquant

   b) Suppression de template d'utilisateurs légitimes, rendant le système biométrique non fonctionnel

   Les systèmes d'authentification utilisant un support externe (par exemple : une carte à puce) pour stocker la référence peuvent s'avérer particulièrement vulnérables à ce genre d'attaque, puisque l'attaquant est en possession du support de stockage.
7. Interception du template de référence entre le stockage et le module de comparaison, la donnée peut être interceptée à ce niveau, dans le but de l'enregistrer ou de la modifier.
8. Réécriture de la décision finale.

Les travaux de El-Abed*et. al.*[25] ajoutent deux vulnérabilités qui peuvent aussi être exploitées, en prenant en considération le déroulement du processus biométrique. Les deux vulnérabilités supplémentaires identifiées sont :

9. Exploitation des performances du système, un attaquant peut exploiter des performances "basses", qui peuvent faciliter la mise en pratique de certains types d'attaque. Doddington *et. al.*[21] distinguent différentes catégories d'individus : a) les moutons, b) les agneaux, c) les chèvres et d) les loups. Les moutons

correspondent à des individus facilement reconnaissables par les systèmes biométriques, les agneaux sont des individus qui sont faciles à imiter. Les chèvres sont, quant à elles, des individus qui ont des difficultés à utiliser un système biométrique, en raison d'une tendance élevée à être indûment rejeté. Les loups désignent les individus susceptibles de réaliser facilement des tentatives d'imposture fructueuses. En fonction des performances admises par le système, et en particulier un taux de FAR insuffisamment bas, un *loup* est susceptible de réaliser des impostures.

10. D'autre part, la vérification de la qualité permet de rejeter les échantillons non satisfaisants. Néanmoins, en fonction de la mise en œuvre de la fonction de contrôle de qualité et de la politique d'acceptation des échantillons, l'échantillon de référence peut être de qualité insuffisante et donc impacter de manière significative les performances du système testé. Un échantillon de qualité insuffisante peut ne pas contenir une quantité d'information suffisante pour garantir le caractère discriminant de la décision.

D'autres travaux ont été effectués sur la sécurité des systèmes biométriques, une représentation des différents vecteurs d'attaque a été proposée par Roberts[84] et est présentée dans la figure 1.4. Cette représentation présente des mises en pratique concrètes des différentes vulnérabilités identifiées par Ratha *et. al.*[80].

1.7.2.3 Environnement

Les performances d'un système biométrique sont susceptibles d'être influencées par les conditions environnementales dans lesquelles se déroule la reconnaissance biométrique. En effet, les conditions environnementales sont susceptibles d'interagir avec le capteur biométrique, ou d'influencer la modalité biométrique et/ou le comportement de l'utilisateur lors de la présentation. Par exemple dans le cas de la reconnaissance vocale, des paramètres environnementaux sont susceptibles d'affecter à la fois le capteur et le porteur de modalité. En effet, la présence de bruit ambiant est fortement susceptible de diminuer les performances en interagissant directement avec le capteur, de plus suivant le niveau sonore de ce bruit le comportement de l'utilisateur est modifié, celui-ci pouvant sciemment ou non augmenter son volume de voix.

Le tableau 1.3 présente différents facteurs environnementaux susceptibles d'influencer différents types de modalité.

Lors de la détermination des performances, il est nécessaire de contrôler les paramètres environnementaux susceptibles d'influencer la condition de la modalité, ou le comportement de l'utilisateur. Un environnement de test, dit normal, doit

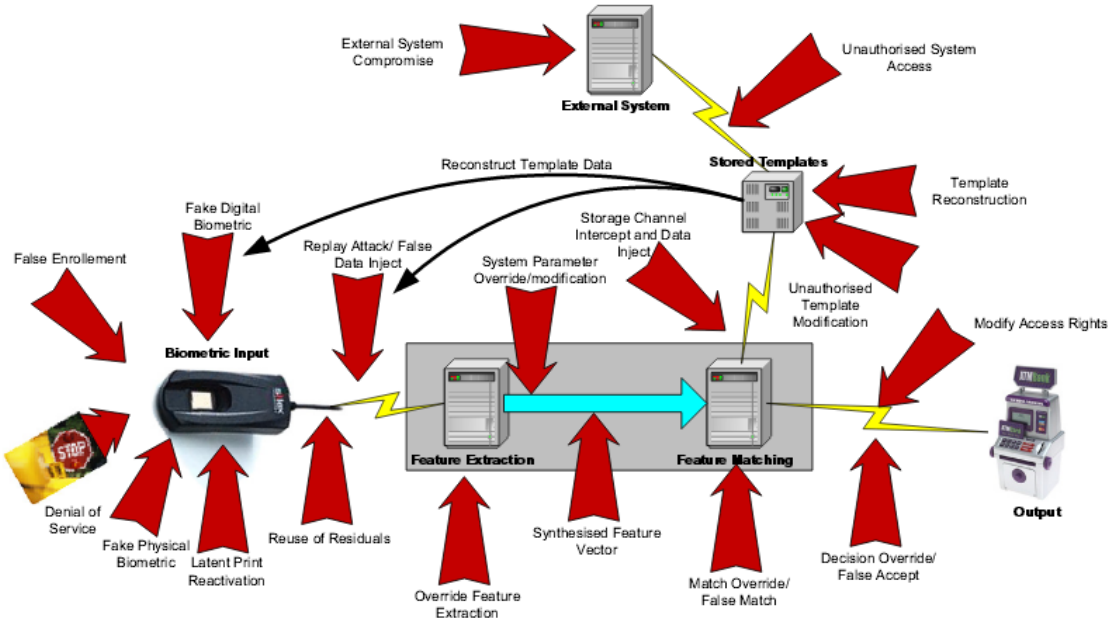

FIGURE 1.4 – Figure présentant les différents vecteurs d'attaque possibles sur un système biométrique tel que proposé par Roberts. Cette représentation illustre les différents vecteurs d'attaque sur une représentation inspirée de celle proposée par l'ISO/IEC 19795-1 (*cf.* figure 1.2)

Facteurs	Empreintes digitales	Iris	Visage	Voix	Dynamique de frappe	Main
Intérieur extérieur	✓	✓	✓	✓	✓	✓
Fond arrière-plan			✓			✓
Température	✓					✓
Humidité	✓					✓
Éclairage	✓	✓	✓			✓
Réflexion de la lumière		✓	✓			✓
Bruits				✓		

TABLE 1.3 – Ce tableau présente les facteurs environnementaux susceptibles d'influencer la qualité de capture d'un échantillon, et par conséquent les performances d'un système biométrique.

être défini en conséquence ; pour chaque condition environnementale, une plage de valeur est définie, de manière à être cohérente avec un usage normal du système testé. L'usage normal est généralement défini par le biais d'un scénario de test, cette méthodologie de test est décrite dans la section 2.4.1

L'évaluation à des fins quantitatives de l'influence de l'environnement requiert la mise en place d'une méthodologie spécifique. La nécessité de mise en place d'un procédé d'évaluation est motivée par la potentielle modification de l'environnement de test par la présence d'équipements de mesure, mais surtout de contrôle des conditions environnementales. La méthodologie proposée par Fernandez-Saavedra[28, 29, 30] propose de comparer les performances d'un système biométrique entre un environnement de référence (la plage de valeurs des conditions environnementales étant cohérente avec un usage normal du dispositif sous test), et d'autre part l'environnement de test ciblé. Néanmoins, la collecte des résultats dans les conditions environnementales normales doit nécessairement être réalisée dans la même configuration que les collectes d'échantillon dans l'environnement cible. En effet, la configuration spatiale lors d'une évaluation d'environnement est susceptible de modifier l'interaction entre l'utilisateur et le système biométrique sous test, en particulier dans le cas de conditions environnementales particulièrement difficiles à maintenir ou à produire. Ainsi, le contrôle des conditions d'humidité et de température induit l'utilisation d'étuves, dans lesquelles sont placés les systèmes biométriques. L'interaction est alors modifiée : l'utilisateur doit présenter la modalité par le biais d'une ouverture et ses possibilités de mouvement, de placement, de mouvement... se retrouvent alors contraintes, et ne correspondent par conséquent pas à la cinématique d'interaction naturelle.

La solution choisie dans cette méthodologie est donc de réaliser ces différentes étapes de collectes au sein des dispositifs de contrôle environnementaux, une session de capture ayant lieu dans des conditions normales, et la seconde dans les conditions ciblées. Les performances obtenues à l'issue des sessions sont comparées afin de déterminer l'influence de l'environnement ciblé.

1.7.2.4 Évaluation d'usage/utilisabilité

L'intégration des systèmes biométriques, en particulier en tant que solution d'authentification, doit répondre à des exigences quant à sa facilité d'utilisation pour les usagers. En effet, dans le cas de l'intégration de l'authentification biométrique dans des systèmes commerciaux (et plus spécifiquement pour les smartphones embarquant un système biométrique), celle-ci est soumise à son acceptation et son adoption par les usagers. Ce type d'évaluation se concentre donc sur l'ergonomie d'un système biométrique, et sur la perception dont en ont les usagers, et est donc particulièrement utile dans le cas de systèmes biométriques

Ergonomie

L'évaluation de l'ergonomie d'un système biométrique implique une étude quantitative et qualitative de l'interaction entre l'utilisateur et le système biométrique. L'étude ergonomique d'un système biométrique vise à en améliorer l'intuitivité ou la clarté d'utilisation. Dans le cadre d'un système biométrique, les principaux écueils à son utilisation sont les échecs d'acquisition qu'il est susceptible d'admettre pour le système ou le capteur. Ces échecs peuvent être dus à une non-détection de l'échantillon biométrique, ou à une qualité insuffisante de l'échantillon capturé. La cause de ces erreurs peut être une cinématique de présentation peu ou mal adaptée, ainsi Theofanos *et. al.*[96] ont mené une étude pour déterminer l'influence du placement d'un capteur biométrique sur les performances observées. Ainsi, si un capteur est placé trop en hauteur, ceci est susceptible de modifier la cinématique de test et donc les performances. De même, Kukula *et. al.*[54] ont étudié l'influence de la pression d'une empreinte digitale lors de sa présentation au capteur, sur les performances observées. Ces dernières sont inversement proportionnelles à la force de présentation, une forte pression déformant les crêtes papillaires empêchant ainsi une extraction précise des minuties.

Une structure d'évaluation a été définie par Kukula *et. al.*[56, 55, 57] à des fins d'investigation de ces échecs d'acquisition. Ce *framework* vise à étoffer le concept d'échec d'acquisition, en prenant en compte les différentes sources pouvant y mener. Ainsi, différents types d'erreur d'acquisition ont été identifiés dans ces travaux :

- interaction déficiente : cette erreur apparait lors d'une présentation incorrecte (ne correspondant pas à la cinématique de présentation), le système traite normalement l'échantillon biométrique.
- interaction faussée : ceci désigne une erreur d'acquisition due à une mauvaise interaction ou à un comportement anormal. Dans le cas d'un capteur d'empreinte digitale par défilement, cette erreur se produit si un utilisateur présente son empreinte de manière trop rapide ou trop lente.
- interaction dérobée : ce type d'erreur qualifie la présentation d'une autre instance de modalité que celle enregistrée dans le système. Par exemple, un utilisateur présente un autre doigt que celui requis pour réaliser sa reconnaissance.
- échec à la détection : cette erreur apparait à la suite d'une présentation correcte de la modalité, c'est-à-dire que l'interaction de l'utilisateur avec le système est conforme à la cinématique de présentation prévue. Néanmoins, le système ne détecte pas qu'une modalité lui est présentée, et par conséquent il ne procède à aucune capture ou extraction.

- échec à l'extraction : ce type d'erreur qualifie une incapacité d'un système à traiter un donnée biométrique, c'est-à-dire que le module de traitement du signal ne parvient pas à segmenter ou extraire de modèle. Le système parvient à capturer un échantillon issu d'une interaction conforme à la cinématique de présentation, mais n'est pas en mesure de réaliser une comparaison à partir de ce dernier.

Ce *framework* permet d'identifier la ou les sources de faux rejets pour une implémentation d'un système biométrique. Ce type d'évaluation permet d'améliorer l'ergonomie du système en conséquence.

Perception d'un système biométrique

La perception d'un système biométrique est une étape importante pour déterminer si ce dernier sera aisément et massivement adopté par les utilisateurs finaux. Différents types de tests permettent de déterminer l'appétence du grand public pour un type de système, ainsi Théofanos *et. al.*[97] proposent une méthodologie dite de *focus group*.

Ce type d'évaluation repose sur la constitution d'un groupe de discussion, l'analyse d'un système par ce biais permet d'estimer l'intérêt d'une ou plusieurs populations ciblées pour son utilisation. La constitution d'un *focus group* débute par le recrutement d'une population de taille restreinte, cette dernière est généralement choisie afin d'être représentative d'une part de la population. Les groupes sont ainsi constitués de manière thématique, par exemple un groupe pourra représenter une population technophile, tandis qu'un autre illustrera une population sceptique voir réfractaire à la biométrie. L'ensemble des groupes est ensuite confronté au système sous test, un scénario d'utilisation intègre l'usage du système afin d'illustrer son utilisation. Le ressenti du groupe est ensuite récolté et analysé. L'utilisation de groupes précautionneusement choisis permet d'estimer l'engouement général que peut susciter une telle solution.

Ce type d'évaluation permet aussi d'identifier les freins à l'adoption de telles solutions, en particulier en terme d'intrusion pour la vie privée, d'acceptabilité, de sécurité ressentie ...

1.8 Positionnement de la thèse

Les usages de la biométrie tendent à se populariser en particulier en tant que solution d'authentification, et donc ceux-ci s'inscrivent de plus en plus dans la vie quotidienne. Comme exposé dans la section 1.6, les systèmes biométriques sont susceptibles d'admettre des erreurs de classification des utilisateurs. Les conséquences

de telles erreurs peuvent être plus ou moins importantes et dépendent principalement des fonctionnalités dont le système biométrique contrôle l'accès, ou encore du contexte d'utilisation.

Étant donné les erreurs que sont susceptibles d'admettre les systèmes biométriques, certains acteurs sont susceptibles d'avoir un besoin de déterminer l'adéquation d'un système biométrique avec l'environnement dans lequel il s'intègre. Ainsi Elitt, dans une politique d'élargissement de son offre de service, s'est intéressé à l'opportunité de se lancer dans l'évaluation des systèmes biométriques. En effet, le domaine d'activité de cette société est connexe à l'évaluation biométrique, en offrant des services d'évaluation en vue d'obtention d'une certification ou d'un agrément dans le domaine des transactions électronique sécurisées ; et l'intégration de la biométrie en tant que solution d'authentification sur différents dispositifs, dont certains permettant des transactions monétiques. Cette thèse fut donc lancée comme une opportunité de prospection d'un nouveau domaine d'activité dans le domaine de l'évaluation.

La première problématique rencontrée au cours de cette thèse fut la mise en application des procédés d'évaluation décrits dans la littérature sur les systèmes biométriques opérationnels. Cette contrainte s'est rapidement imposée avec d'une part la frilosité des industriels qui ne souhaitent pas fournir des systèmes transparents pour protéger leur propriété intellectuelle ; et d'autre part, les référentiels de tests et les organismes de certification qui sont susceptibles d'imposer que le système soit testé sous sa forme finale telle que déployée lors de sa mise sur le marché. Ainsi, cette problématique fut abordée au cours de cette thèse afin de proposer une méthodologie fonctionnelle permettant d'évaluer un système opérationnel (*i.e.* n'autorisant que des interactions limitées avec le système, c'est-à-dire ne permettant pas de manipuler les références et les échantillons en entrée de l'algorithme de comparaison biométrique, voire ne permettant tout simplement pas la capture des échantillons afin d'en étudier la qualité biométrique).

Au cours de l'étude de la problématique précédente, lors de l'évaluation d'un système multimodal, la problématique de la qualité des échantillons s'est imposée après que le dispositif testé ait présenté des performances bien en deçà de celles escomptées. En effet, au cours d'une évaluation précédente suivant un processus d'évaluation différent, les performances au regard des faux négatifs ne sont pas conformes à celles observées lors de l'évaluation utilisant le protocole défini au cours de cette thèse. Cependant après investigation, il s'est avéré que le dispositif sous test introduisait un bruit caractéristique s'ajoutant aux échantillons. Ceci motiva l'étude de la qualité des échantillons vocaux.

1.9 Conclusion

Les premiers travaux scientifiques sur la biométrie sont apparus au cours du 19ème siècle, et sont issus du besoin d'identifier de manière formelle les individus. Ainsi en parallèle du développement de méthodes d'investigation scientifique, la biométrie s'est développée et enrichie tout d'abord par le biais des travaux des pionniers de l'anthropométrie judiciaire et de l'étude des empreintes digitales. Néanmoins, le recours à ces méthodes fut limité et se confina longtemps au domaine policier et régalien, tant que les comparaisons entre les différents échantillons biométriques furent manuels. Le développement des moyens informatiques, que ce soit par le biais de l'augmentation de la puissance de calcul des machines, ou par le développement de capteurs et du traitement de leurs signaux, ont permis d'automatiser le processus de comparaison biométrique.

Les solutions d'authentification biométrique ont ainsi pu être développées et diffusées dans la sphère grand public. Néanmoins, les systèmes de reconnaissance biométriques sont susceptibles d'admettre des erreurs En fonction de l'environnement dans lequel ils s'inscrivent, les conséquences de ces erreurs peuvent être plus ou moins importantes. Ainsi, dans le cadre d'usages réglementés par des normes ou des accréditations, l'estimation de la propension d'un système à admettre des erreurs est nécessaire, afin de déterminer si son intégration est compatible avec les exigences de l'usage envisagé.

Cette thèse étudie l'évaluation des systèmes biométriques afin de les adapter aux usages actuels, et en particulier à leur intégration dans des dispositifs portables (en particulier les *smartphones* et tablettes). Ainsi, une méthodologie d'évaluation de versions opérationnelles de tels systèmes est proposée et présentée par la suite. Au cours d'une évaluation sur un système de reconnaissance du locuteur, l'importance de la qualité des échantillons de voix s'est fait jour, en raison de la présence de parasites dans certains enregistrements. Une méthode d'estimation de qualité de la voix est proposée par la suite.

Chapitre 2

Évaluation des systèmes biométriques

Ce chapitre présente un aperçu global du domaine de l'évaluation biométrique. Ce domaine couvre plusieurs problématiques, chacune couvrant un aspect spécifique de l'évaluation des performances biométriques. Ce chapitre propose donc une présentation des différents facteurs susceptibles d'influer sur les performances biométriques d'un système biométrique, et sur l'évaluation de cette influence.

Sommaire

2.1 Introduction

La biométrie est un ensemble de méthodes permettant de reconnaître un individu par le biais d'une caractéristique qui lui est propre. Le perfectionnement des méthodes de reconnaissance biométrique, le développement et la recherche de nouveaux modèles et représentations, ainsi que l'augmentation de la capacité de

calcul ont permis de rendre le domaine biométrique mature et à même d'être intégré en tant que solution d'authentification. La biométrie se trouve donc être utilisée en tant que solution d'authentification dans des systèmes aux exigences variées. En fonction de l'usage prévu une évaluation biométrique permet de déterminer si un ou plusieurs systèmes biométriques sont en adéquation avec les exigences de cet usage. L'authentification biométrique est le mode de fonctionnement le plus répandu des systèmes biométriques, en particulier dans le cas des systèmes biométriques grand public, et dont la diffusion n'est pas restreinte par les lois protégeant la vie privée et les données personnelles. Ainsi, les travaux d'évaluation ayant eu lieu au cours de cette thèse se sont essentiellement concentrés sur l'authentification biométrique.

Afin d'évaluer l'adéquation d'un système biométrique avec son usage, il est nécessaire de déterminer les performances de celui-ci.

L'évaluation du système biométrique peut se faire en accord avec différentes méthodologies, chacune permettant d'évaluer les performances d'un système biométrique dans des conditions et avec des objectifs différents. Ces différentes méthodologies définissent les conditions dans lesquelles se déroulent l'évaluation, ainsi que la description de la comparaison croisée.

2.2 Définition des performances biométriques

Afin de qualifier et de quantifier les performances d'un système biométrique, différentes valeurs et taux d'erreur ont été définis. Les travaux concernant l'étude et la détermination des performances biométriques ont été motivés par le besoin de déterminer l'adéquation d'un système biométrique avec un usage. En effet, en fonction des limitations d'un système biométrique, certains usages ne seront pas en adéquation avec celles-ci ; afin de mieux définir et cerner ces limitations, différentes métriques et classes ont été définies concernant les types d'erreurs admises par les systèmes biométriques.

Ainsi au début des années 2000, plusieurs travaux ont étudié l'évaluation des systèmes biométriques. Ces travaux se sont attachés à définir tant les différentes métriques permettant de qualifier les capacités d'un système biométrique, que les méthodologies permettant de mesurer ces métriques de performances.

En particulier les travaux de Mansfield [67, 66] ont permis de définir plusieurs taux d'erreurs. L'approche quant à la définition de ces taux d'erreurs, a été de considérer certains modules constitutifs d'un système biométriques indépendamment. La définition de ces différents modules biométriques s'appuie sur une représentation des systèmes biométriques proposée par les auteurs. Cette représentation fut réutilisée

dans l'ISO 19795 [43], afin de représenter les différents composants d'un système biométrique.

2.2.1 Taux d'erreur

Les différents travaux sur l'évaluation biométrique, ainsi que l'ISO 19795 [43] ont permis une définition formelle des performances sous la forme de différents taux d'erreur. Ces différents taux d'erreur peuvent être impartis à une partie ou à la totalité du système biométrique.

Ainsi, l'acquisition des données biométriques est susceptible de donner lieu à des cas d'erreur liée à la capture d'une modalité, ou au pré-traitement et à l'extraction des caractéristiques biométriques. Ces cas d'erreur proviennent donc du capteur, et du module de traitement du signal si l'on se réfère à la représentation des systèmes biométriques fournie par l'ISO 19795. Différents taux d'erreur sont utilisés :

- **FTE ou FTER :** Failure to Enroll (Rate) ou taux d'erreur à l'enrôlement
  Ce taux d'erreur représente la proportion de la population pour laquelle le système biométrique n'est pas en mesure de l'enrôler, c'est-à-dire de créer une référence biométrique viable. Ce cas d'erreur peut être observé lorsqu'une instance de modalité ne permet pas au capteur de créer des échantillons de qualité suffisante, ou porteur de suffisamment d'informations pour créer une référence. En plus de ces problèmes d'acquisition, certains individus peuvent, en raison de malformations ou de séquelles permanentes ou temporaires, ne pas être en mesure d'utiliser un type de système biométrique en fonction de la modalité affectée. Ce taux d'erreur représente dans une certaine mesure, la capacité du système à être utilisé au sein d'une population globale.
- **FTA ou FTAR :** Failure to Acquire (Rate) ou taux d'échec à l'acquisition
  Ce taux d'erreur permet de quantifier la capacité du système à acquérir correctement des échantillons biométriques. Ces échecs permettent de déterminer la proportion de tentatives d'utilisation du système, qui ne permettront pas de donner lieu à une comparaison biométrique. Les raisons de ces échecs sont l'incapacité du système à extraire ou localiser un échantillon viable pour la reconnaissance biométrique. Ainsi, le FTA regroupe des erreurs telles que la non-détection d'une présentation d'une modalité menant par conséquent à une absence de capture biométrique, ou d'une incapacité de l'utilisateur à présenter une instance viable pour une capture d'échantillon ...

De même le module de comparaison est critique dans le cadre de la reconnaissance biométrique. En effet, les erreurs issues de ce module mènent à une fausse

reconnaissance, ou à un faux rejet. Ce qui, en fonction de l'application dans laquelle est intégré le système de reconnaissance biométrique, peut avoir des conséquences plus ou moins importantes.

- **FMR :** False Matching Rate ou taux de fausse correspondance.

  Ce taux d'erreur caractérise la propension de l'algorithme de comparaison à réaliser de fausses correspondances entre échantillons et références biométriques. Dans le cas de l'authentification, ceci permet de déterminer une première mesure naïve de la sécurité d'un système biométrique. En effet une tentative d'imposture peut être considérée comme une attaque à effort nul. Ce type d'erreur apparait lorsque l'algorithme est confronté à une paire d'échantillon et de référence, pour lesquels il n'est pas en mesure de réaliser de discrimination. Cette incapacité peut provenir d'une quantité d'information insuffisante de l'échantillon biométrique ou de la référence.

- **FNMR :** False Non Matching Rate ou taux d'échec à l'acquisition

  Ce type d'erreur représente la tendance d'un système à rejeter anormalement un utilisateur dont la référence est connue par le système. Les causes de ce type d'erreur peuvent provenir d'une lacune d'information dans l'échantillon, ou d'une faible similitude des caractéristiques biométriques extraites de l'échantillon.

Un système de reconnaissance biométrique peut avoir deux finalités : l'authentification, et l'identification. Suivant la finalité implémentée par le système biométrique testé, les types d'erreurs finales seront différents.

- **FAR :** False Acceptance Rate ou taux de fausses acceptations

  Ce taux d'erreur représente la probabilité qu'un individu soit anormalement accepté à l'issue d'une présentation de sa modalité biométrique. Cette métrique représente un aspect des performances pour le système biométrique dans leur globalité.

$$FAR = (1 - FTA) \times FMR \tag{2.1}$$

- **FRR :** False Rejection Rate ou taux de faux rejets

  Ce type d'erreur illustre la probabilité que la tentative d'authentification d'un utilisateur ne soit pas correctement traitée par le système biométrique testé. Ce taux inclut donc les rejets dus à un dysfonctionnement de la capture d'un échantillon, lors de la comparaison réalisée par l'algorithme biométrique.

$$FRR = (1 - FTA) \times FNMR + FTA \tag{2.2}$$

- **EER :** Equal Error Rate ou taux d'erreurs égales.

  Cette mesure indique le taux d'erreurs égales, c'est-à-dire le paramétrage pour lequel le taux de FAR et de FRR admis par le système sont égaux. Cette métrique est utilisée afin de donner un a priori sur les performances globales du système.

- **GFAR et GFRR :** Generalized False Acceptance Rate, et Generalized False Rejection Rate, taux d'erreur de FAR et FRR généralisés.

  La comparaison de systèmes biométriques en confrontant les taux de FAR et de FRR peut présenter un biais, en ne considérant pas le taux d'échec à l'enrôlement des systèmes comparés. La généralisation de ces deux taux d'erreur vise à prendre en considération à la fois les échecs à l'enrôlement, les échecs à l'acquisition et les erreurs de correspondance. Néanmoins en fonction de la méthodologie de test utilisée, l'expression de ces taux d'erreur dépendra de la méthodologie de test.

  La généralisation de ces taux d'erreur doit particulièrement prendre en compte le taux d'échec à l'enrôlement. Lors d'un cas de FTE, quelque soit la méthodologie de test utilisée, toutes les tentatives dans lesquelles la référence ou l'échantillon sont impliquées, sont considérées comme ayant échoué.

Dans le cas de l'identification, la décision du système peut ne pas se définir comme une décision reflétant si l'utilisateur ayant présenté sa modalité est présent ou non dans la base de référence du système testé. Les métriques de performances mesurables pour les systèmes d'identification, sont :

- **IR :** Identification Rate ou taux d'identification

  Ce taux représente la proportion de tentatives d'identification réussies. C'est-à-dire la proportion de transactions d'identification pour lesquelles l'identité réelle de l'individu est retournée dans une liste d'identité candidate. Ce taux est dépendant a) de la taille de la base contenant les références, b) du seuil de décision des scores de comparaison.

- **FPIR :** False Positive Identification Rate ou taux de fausse identification positive

  Ce qui correspond à la proportion de transactions d'identification, pour des utilisateurs non enregistrés dans le système, pour lesquelles le système n'est pas en mesure de déterminer que ceux-ci ne sont pas connus de la base de références, retournant par conséquent une liste de candidats.

$$FPIR = (1 - FTA) \times (1 - (1 - FMR)^N) \tag{2.3}$$

où N désigne le nombre de références dans la base du système d'identification.

- **FNIR :** False Negative Identification Rate ou taux de fausse identification négative

  Cette mesure représente la proportion de transactions d'identification par utilisateurs enregistrés dans le système, pour lesquelles l'utilisateur n'est pas dans la liste retournée ou pour lesquelles la capture a échoué.

$$FNIR = FTA + (1 - FTA) \times FNMR \tag{2.4}$$

2.2.1.1 Courbes de performances

En plus de ces taux d'erreur, des courbes de présentation des résultats peuvent être tracées, elles présentent l'avantage d'une bonne lisibilité des résultats en fonction des seuils choisis. En effet, les taux d'erreurs sont définis pour un seuil donné. Sur un même graphique, il est possible de présenter les taux d'erreurs pour différentes valeurs de seuil et ainsi plus facilement voir l'impact des variations de ces seuils sur les performances. Les résultats sont des courbes paramétriques donnant la valeur de deux taux d'erreurs en fonction du seuil. Les deux principales courbes sont les courbes DET et les courbes ROC. Dans l'ISO/IEC 19795 [43], ces courbes sont définies comme :

- **Courbe ROC :** Receiver Operating Characteristic.

  Cette courbe est le tracé du taux de vrai positif en fonction du taux de faux positif. Cette courbe paramétrique est fonction du seuil d'acceptation fixé pour le système biométrique. Un exemple de courbe ROC est le tracé d'une courbe $(1 - FRR)$ en fonction du FAR. La courbe ROC est une méthode pour résumer les performances d'un système biométrique, et permet de comparer des systèmes entre eux. Il est pertinent de tracer cette courbe sur une échelle logarithmique afin qu'elle soit traçable/observable dans son entier sans pour autant perdre en lisibilité.

- **Courbe DET :** Detection Error Trade-off.

  C'est une courbe ROC modifiée qui consiste à présenter les taux d'erreurs sur les deux axes (i.e. les faux négatifs en fonction des faux positifs). On peut tracer par exemple le FNMR en fonction du FMR pour un seuil d'acceptation du système donné.

- **Courbe CMC :** Cumulative Matching Characteristic.

  Cette courbe est spécifique aux systèmes biométriques d'identification, et représente les résultats d'un processus d'identification en traçant la probabilité d'une

classification correcte en fonction du rang (probabilité d'une classification parmi les k premiers identifiants retournés). Il s'agit d'une probabilité cumulative, c'est-à-dire la probabilité qu'une identification correcte ait eu lieu pour ce rang et pour les rangs inférieurs.

2.3 Qualité des échantillons biométriques

Le principe de la reconnaissance biométrique s'appuie sur une comparaison entre deux vecteurs de caractéristiques biométriques extraits d'échantillons issus d'une seule ou de différentes instances suivant le cas de figure : impostures ou tentatives légitimes. Néanmoins, les vecteurs de caractéristiques extraits d'échantillons d'une même modalité sont sujets à des variations plus ou moins marquées.

Ces variations sont susceptibles d'apparaître au cours de la capture d'un échantillon ; en fonction de la modalité considérée, différents facteurs sont en mesure de conditionner les vecteurs de caractéristiques biométriques. Ainsi l'interaction entre le capteur et la modalité est en mesure de produire des distorsions, des déformations... susceptibles d'induire des variations dans le vecteur de caractéristiques en comparaison avec celui extrait de la référence. Par exemple, dans le cadre des empreintes digitales, la pression du doigt sur le capteur impacte directement l'échantillon, et la quantité d'information qu'il est possible d'extraire (voir les travaux de Kukula *et. al.*[56, 55, 57]) ; tandis que dans le cas de la reconnaissance vocale, le volume de la voix est en mesure d'avoir un impact sur le vecteur de caractéristique extrait d'un échantillon sonore.

L'environnement interfère sur la présentation de la modalité biométrique avec le capteur, les conditions environnementales sont susceptibles de modifier les conditions de capture tant en modifiant les modalités biométriques lors de la présentation, que d'altérer le fonctionnement du capteur biométrique. Dans le cas des empreintes digitales, la température et l'humidité sont susceptibles de modifier la condition des empreintes, en favorisant la sudation, en fonction de la technologie implémentée par le capteur, ces conditions sont aptes à influer sur l'échantillon capturé. L'illumination est aussi capable de perturber directement le fonctionnement d'un capteur d'empreintes digitales implémentant une technologie optique. Dans le cadre de la reconnaissance du locuteur, le bruit ambiant est susceptible d'introduire un biais au cours de la capture d'un échantillon, en effet la superposition d'un bruit avec un signal vocal est en mesure de modifier voire de masquer un certain nombre de caractéristiques spectrales, acoustiques ... utilisables en tant que caractéristiques biométriques. La présence de bruit ambiant peut provoquer une modification du comportement de

l'utilisateur ; en effet placée dans un environnement bruyant, une personne aura tendance à ajuster son volume vocale en fonction du volume de bruit.

Ces variations au sein d'un ensemble d'échantillons influent sur les performances de comparaison, ainsi la définition d'une méthodologie permettant déterminant dans quelle mesure une capture est pertinente et utilisable en tant qu'échantillon biométrique, permet de fournir un a priori sur les performances escomptables pour cet échantillon.

2.3.1 Principes de la qualité biométrique

Les métriques de qualité sont conçues de manière à mesurer la qualité intrinsèque d'un échantillon au regard de la reconnaissance biométrique pour une modalité.

D'après l'ISO/IEC 29794-1 [10], la qualité d'une donnée biométrique est définie selon trois points :

- les caractéristiques de la source représentent la qualité intrinsèque de la modalité biométrique,
- la fidélité caractérise le degré de similarité entre un échantillon biométrique et la référence biométrique d'une même personne,
- l'utilité illustre dans quelle mesure un échantillon biométrique est susceptible d'être exploitable par un système biométrique (présence et facilité d'extraction de caractéristiques biométriques), et par conséquent de prédire ou tout du moins fournir un apriori sur les performances observables.

L'utilité telle que définie ci-dessus, est généralement liée aux métriques de qualité biométrique, qui visent à fournir des informations sur la pertinence d'un échantillon biométrique au regard de l'ensemble des algorithmes biométriques basés sur une modalité particulière. La mesure de la fidélité n'est possible qu'en ayant connaissance de la référence, ce qui ne correspond pas au cas d'usage d'une métrique de qualité où la mesure de qualité est d'une part objective (non supervisée par des moyens humains), et d'autre part sans référence. Quant aux caractéristiques de la source, celles-ci reflètent la qualité intrinsèque d'une modalité biométrique, et sont donc partiellement représentées par l'utilité.

Afin d'aider à la définition d'une telle métrique de qualité, Grother and Tabssi [35] ont préconisé différentes propriétés auxquelles une métrique devrait répondre. Ainsi, une métrique doit être une valeur représentative d'un point de vue statistique, liée aux performances de correspondance biométrique ; cette valeur devrait aussi être quantifiée.

- Une métrique de qualité devrait idéalement se présenter sous la forme d'une valeur statistiquement représentative de la qualité d'un échantillon biométrique. En effet, qu'une métrique se présente sous la forme d'un scalaire en lieu et place d'un vecteur de caractéristiques, permet d'assurer une meilleure lisibilité pour comparer différents échantillons entre eux.

- Une métrique de qualité doit être statistiquement représentative au regard des performances de correspondance biométrique admises par un échantillon. En effet, toujours d'après Grother & Tabassi [35], une métrique de qualité doit présenter une relation avec les performances biométriques observées.

- La quantification conseillée contraint une métrique de qualité à s'exprimer dans l'intervalle $[0, 100]$. Les valeurs tendant vers zéro expriment une qualité extrêmement basse d'un échantillon, tandis qu'un échantillon présentant une qualité tendant vers 100 représente un échantillon d'excellente qualité. Ceci permet de fournir une sémantique à ces valeurs, et de permettre une compréhension aisée.

En plus de leur intérêt dans le cadre de l'évaluation d'un système biométrique afin de déterminer la capacité d'un capteur, et du système dans lequel il s'intègre, à fournir des échantillons de qualité, les métriques de qualité sont susceptibles d'être utilisées dans un cadre opérationnel afin de procéder à une sélection des échantillons utilisés. Ainsi, des exigences de qualité exprimées par le biais d'une métrique peuvent s'appliquer :

- À la sélection du ou des échantillons utilisés au cours de la création d'une référence biométrique, et ce afin de garantir que la référence obtenue contient suffisamment de caractéristiques biométriques. Cette étape est particulièrement importante dans le cas de système ne réalisant l'enrôlement qu'à l'aide d'un seul et unique échantillon biométrique.

- la sélection des échantillons dans le cadre de l'authentification (ou de l'identification) permet d'éviter de solliciter le système avec une tentative de reconnaissance biométrique qui n'aboutira vraisemblablement pas sur un résultat pertinent mais probablement sur une incapacité du système à extraire une quantité suffisante d'informations, ou sur une comparaison biaisée par le manque d'informations concordantes avec celles présentes dans le modèle.

Les métriques de qualité sont particulièrement utiles tant dans le domaine de l'évaluation biométrique que dans l'implémentation de systèmes de reconnaissance biométrique.

2.4 Présentation technique de l'évaluation biométrique

L'évaluation des systèmes biométriques vise à déterminer tout ou partie des taux d'erreurs présentés précédemment. Afin de déterminer les performances d'un système biométrique, il est nécessaire de réaliser de nombreuses tentatives d'utilisation du système biométrique. En fonction de la méthodologie de test optée pour l'évaluation, la réalisation de ces comparaisons peut être amenée à varier.

2.4.1 Présentation des méthodologies de test

La mise en place d'une évaluation biométrique peut se faire selon différentes méthodologies. Cette partie présente les trois méthodologies proposées dans le cadre de la norme ISO 19795 [43]. Celles-ci sont l'évaluation de technologie, de scénario et opérationnelle, chacune adresse un aspect de l'évaluation des systèmes biométriques. Le choix de la méthodologie mise en place repose sur les objectifs de l'évaluation, c'est-à-dire quels types de performances doivent être déterminés et dans quelles conditions. En effet, une évaluation peut être réalisée par un constructeur ou un acteur académique cherchant à déterminer la pertinence de l'algorithme de reconnaissance biométrique implémenté. Dans le cas d'une évaluation réalisée par une tierce partie avec pour objectif l'obtention d'un agrément ou de la certification des performances biométriques, les systèmes sous test peuvent se présenter sous la forme d'une boîte noire. C'est-à-dire que ces systèmes n'autorisent pas d'accès aux données internes de ces derniers. Les trois différents types d'évaluation définis sont :

- Évaluation de technologie : l'évaluation de technologie est généralement utilisée pour déterminer les performances d'un algorithme. Lors de la mise en œuvre de cette méthodologie, les échantillons biométriques proviennent de base de données biométriques. Au sein de ces bases biométriques, les échantillons sont regroupés par instance et par individu. Ce type d'évaluation permet de déterminer les performances d'un algorithme biométrique sous test. Les taux d'erreurs déterminés seront ainsi le FMR et le FNMR.

  Dans le cadre d'une évaluation de technologie, les bases de données biométriques ne sont pas constituées spécifiquement pour une évaluation. Les performances estimées lors de ce type d'évaluation permettent de donner un a priori sur les performances finales de l'algorithme une fois intégré dans le système final. Les évaluations de technologie permettent de tester l'algorithme de comparaison lors de son développement et de son implémentation.
- Évaluation de scénario : l'évaluation de scénario propose une méthode d'évaluation des systèmes biométriques complète. L'objectif de cette évaluation est

de déterminer les performances d'un système pour un scénario donné.

Le scénario permet de représenter un cas d'usage du système sous test, et d'estimer les performances pour ce dernier. La description d'un scénario comporte la politique concernant la collecte des transactions biométriques, les conditions environnementales dans lesquelles se déroulent les comparaisons biométriques. Le scénario conditionne aussi la constitution de la population de test, la description des interactions autorisées avec le système biométrique.

Dans le cadre de ce type d'évaluation, il est possible d'utiliser des bases de données biométriques, il est néanmoins nécessaire que ces dernières respectent le scénario de test, et que le capteur utilisé soit du même modèle que celui du système biométrique complet. En conséquence, l'évaluateur constitue généralement une base de données biométriques spécifiques pour chaque scénario et système testé.

Néanmoins, il est aussi possible de réaliser toutes ou partie des comparaisons manuellement et de stocker les résultats et/ou les échantillons dans une base d'échantillon.

- Évaluation opérationnelle : ce type d'évaluation étudie les performances d'un système biométrique déployé comme solution d'authentification ou d'identification. Le système est dans son environnement opérationnel permettant d'étudier le système en condition réelle, ce type d'évaluation est faiblement supervisé afin de ne pas biaiser les observations. La mise en place de ce type d'évaluation est donc facilement réalisable. Néanmoins, cette faible supervision implique que la classification de certain type d'erreurs sera plus difficile en raison de la difficulté de déterminer la vérité terrain.

2.4.2 Types de transaction biométrique

Selon la terminologie mise en place par la norme ISO 19795 [43], les transactions biométriques sont la confrontation d'un échantillon biométrique avec une ou plusieurs références selon le cas d'usage du système biométrique [1]. La réalisation d'une évaluation biométrique requiert un important nombre de transactions afin d'être en mesure de fournir une estimation fiable des taux d'erreur à l'issue de la comparaison croisée. Ainsi, il est possible de définir trois différentes méthodes permettant de réaliser les transactions nécessaires à l'estimation des performances d'un système biométrique, qui sont :

1. Une référence dans le cas d'un système d'authentification, et plusieurs pour l'identification biométrique.

- tests *online* : les tests *online* correspondent à une confrontation entre une ou plusieurs références et un échantillon capturé pour cette comparaison. Ce type de test ne recourt donc pas à une base de données biométriques, toutes les transactions sont réalisées par un membre de la population de test sur le système biométrique testé.

  En raison du grand nombre de tests nécessaire à l'estimation de certains taux d'erreur particulièrement bas [2], ce type d'évaluation peut être particulièrement couteux en terme de temps et de mobilisation de la population de test.

- tests *offline* : ce type de transaction utilise une base de données biométriques comme source d'échantillons pour réaliser les transactions biométriques. Dans ce type d'évaluation, un échantillon est sélectionné afin d'être comparé avec une ou plusieurs références, elles aussi issues de la base de données biométriques. En fonction du type d'évaluation réalisée (*i.e. scénario ou technologie*), la base peut être une base générique ou une spécialement collectée pour l'évaluation.

- tests *hybrid* : ce type de transaction est un compromis entre les deux méthodologies proposées ci-dessus. Il s'agit de réaliser des transactions *online*, tout en enregistrant l'échantillon afin qu'il puisse être réutilisé lors de transactions de type *offline*. Cette méthodologie propose donc de constituer une base d'échantillon tout en déroulant un certain nombre de tests.

  Ainsi, il est particulièrement intéressant de réaliser la collecte d'échantillons au cours de comparaisons légitimes, en effet l'utilisateur doit dans un premier temps créer une ou plusieurs références, et ensuite réaliser des transactions légitimes ; la réalisation des transactions d'impostures se faisant alors en réutilisant des références et des échantillons issus de différents membres de la population de test.

Ainsi, lors de la mise en place d'une évaluation, l'évaluateur doit choisir la méthode suivant laquelle les transactions se déroulent. Ce choix se fait selon les contraintes imposées par l'évaluation et le système testé.

La mise en place de l'évaluation peut s'avérer un processus particulièrement couteux en terme de temps, de monopolisation de la population de test, de nombres d'instances du système évalué, etc. Les travaux présentés dans la littérature et les normes relatifs à l'évaluation des systèmes biométriques postulent que l'évaluateur possède un accès à un système biométrique en "*boîte blanche*", ou tout du moins avec d'importants privilèges d'accès à ceux-ci. Ainsi afin de simplifier le déroulement de leurs évaluations, celles-ci sont divisées en deux processus majeurs : 1. avec d'une

2. En particulier, l'estimation d'un intervalle de confiance pour le FAR de certains système peut s'avérer particulièrement ardu, si tous les tests sont réalisés selon cette méthodologie.

part la capture des échantillons biométriques afin de constituer une base, 2. et d'autre part la réalisation de la comparaison croisée qui vise à confronter les échantillons capturés précédemment afin d'obtenir les résultats de chaque comparaison.

Néanmoins en fonction du type de collecte de transaction, la séparation de ces deux processus peut ne pas être possible, en particulier dans le cas de transactions *online*. Il est alors nécessaire d'adapter la méthodologie de test afin de pouvoir estimer les performances d'un système biométrique.

2.5 Méthodologies de capture des échantillons

Comme précisé précédemment l'estimation des performances se fait par le biais de comparaison d'échantillons biométrique par le système testé. Il est donc nécessaire de procéder à la capture d'échantillons auprès d'une population de test. Cette partie se propose de présenter différents travaux de la littérature scientifiques et le corpus des normes internationales couvrant ce domaine. Dans un premier temps, les différentes méthodologies de capture et de constitution de base de données biométriques seront présentées, et ensuite les différentes descriptions de la population de test.

2.5.1 Protocoles et politiques de constitution des bases biométriques

A moins de procéder à une évaluation de technologie à l'aide d'une base préalablement constituée, une population de test est nécessaire soit afin de collecter une nouvelle base d'échantillon, soit afin de réaliser des transactions *online*. La constitution d'une base de données biométriques est une étape particulièrement importante lors d'une évaluation biométrique. En effet, certaines erreurs d'archivage sont susceptibles de biaiser les résultats d'une évaluation de manière plus ou moins importante.

Les erreurs de classification peuvent mener à différents types de classification : 1. des identifiants incorrects peuvent avoir été fournis à certains utilisateurs, 2. erreurs de saisie de l'identifiant et 3. présentation d'une mauvaise instance. Ces erreurs sont susceptibles de fausser l'estimation des taux d'erreur en augmentant artificiellement les cas de faux positifs et de faux négatifs. La collecte des échantillons auprès de la population de test doit être un processus supervisé afin de limiter voire d'éviter ce genre d'erreur de classification.

2.5.1.1 Processus de capture

Les bases de données biométriques, librement disponibles ou non, peuvent être accompagnées d'une description des conditions et des méthodes mises en place lors de leurs constitution.

Les différents protocoles spécifient les conditions dans lesquelles les captures ont eu lieu. En effet, certains effets environnementaux sont susceptibles d'influencer le comportement des individus ou du système biométrique. Ces conditions doivent être ainsi reportées et documentées afin de permettre un quasi reproductibilité des résultats obtenus à l'issu d'une évaluation, et dans une autre mesure permettre de qualifier la difficulté d'un corpus d'échantillon.

Ainsi, l'article adossé [5] à la collecte de la base d'échantillons biométriques BANCA [3], propose trois scénarios pour sa capture biométrique :

- un scénario contrôlé,
- un scénario dégradé,
- un scénario défavorable.

Dans l'implémentation proposée par BANCA, le scénario contrôlé correspond à des conditions de capture idéales.

2.5.2 Population de test & base de test

Afin de mettre en place un processus d'évaluation biométrique, il est nécessaire de faire appel à une population d'individus, afin que ces derniers en fonction du mode de fonctionnement de l'évaluation fournissent des échantillons ou des transactions biométriques. Dans le cas des évaluations de technologie, il est possible d'utiliser des bases de données biométriques collectées par des tiers, synthétisées par des logiciels adaptés. Cependant, afin de contextualiser les résultats d'une évaluation biométrique, il est nécessaire, en sus de la description du protocole d'évaluation, de fournir une description de la base de données ou de la population de test.

2.5.2.1 Description de la population de test

Une population de test est nécessaire à la constitution d'une base d'échantillons biométriques. Néanmoins comme exposé dans la section 1.6.3, différents facteurs sont susceptibles d'influencer les performances des systèmes biométriques. En particulier, ces facteurs sont susceptibles de provenir des individus de la population de test, ainsi une partie de ces caractéristiques peut être consignée afin de qualifier la population.

3. La base BANCA est une base biométrique multimodale, en effet celle-ci contient des captures de voix, et du visage des participants.

Dans le cas de certaines évaluations, le scénario d'évaluation est susceptible de viser une population cible, le processus de recrutement devra donc tenir compte de la description de la population cible présente dans le scénario. Suivant les populations visées, les descriptions peuvent être plus ou moins précises. En effet dans le cadre d'une population générique seules quelques caractéristiques sont spécifiées. Ces populations peuvent se définir en termes de catégories d'âge, de types d'occupation, etc.

Lors du recrutement d'une population de test, des biais sont susceptibles d'être introduits en fonction de son mode de recrutement. En effet, lors d'évaluations réalisées par des industriels ou des académiques, la méthodologie de recrutement est généralement *ad hoc* ; c'est-à-dire que l'effort de recruter une population représentative n'est pas mis en œuvre soit sciemment soit par manque de moyens. Il serait nécessaire de mettre en place un échantillonnage aléatoire au sein d'une population afin d'obtenir un échantillon qui pourrait être considéré comme représentatif. Comme défini dans Kruskal & Mosteller [53], un échantillonnage représentatif est entre autre chose défini par une absence de forces sélectives, c'est-à-dire que lors de l'échantillonnage tout élément d'une population doit avoir une probabilité égale de sélection. Étant donné les méthodes de recrutement d'une population de test dans une évaluation biométrique, il est rare que celle-ci puisse se targuer d'être représentative. De plus, dans les articles présentant le recrutement d'une base de données biométriques, une emphase est généralement mise sur la taille de la population et ne précise généralement pas la méthode de recrutement (et donc par la même d'échantillonnage) de la population [5, 78, 23], et ne précise que peu ou pas de caractéristiques de cette dernière. Dans certains cas il est précisé qu'une attention particulière a été portée à une certaine représentativité quant au genre des individus. La base BIOMET [33] précise quant à elle certaines informations de la population utilisée pour constituer sa base. Ainsi la plage d'âge est précisée, ainsi que la proportion d'étudiant la constituant.

La mise en place d'un procédé d'échantillonnage spécifique visant au recrutement d'une population représentative ne semble pas réaliste dans le cadre d'une évaluation biométrique opérationnelle. En effet, cette dernière impose des contraintes fortes en particulier en termes de délai, et de disponibilité de la population de test.

2.5.2.2 Difficulté de la base d'échantillons

Une mesure de la difficulté des bases de données biométriques permet de qualifier le challenge que pose une ou un ensemble de bases biométriques. Ceci permet de contextualiser les résultats observés et d'apporter une confiance supplémentaire dans les revendications de performances. En effet, les conditions de capture des échantillons sont susceptibles d'influer sur la qualité de ces derniers, et par conséquent sur les

scores de comparaison en raison de la corrélation de ceux-ci avec la qualité des échantillons biométriques.

L'ISO/IEC TR 29198 [47] et Li *et. al.* dans [59] et [60], proposent une méthodologie permettant d'évaluer la difficulté d'une base d'empreintes digitales, par le biais d'un niveau de difficulté ou LOD (*Level Of Difficulty*). La mesure de ce niveau de difficulté, proposée par cette méthode, va au delà de la simple mesure de qualité des échantillons biométriques. En effet afin d'estimer la difficulté d'une base, d'autres éléments sont pris en considération. Cette mesure de la difficulté définit le challenge que propose une base d'empreintes digitales dans le cas d'estimation du taux de faux rejet, et étudie ainsi la difficulté que pose une base d'empreintes lors de la phase de détermination du FRR. Le but de la méthode présentée n'est pas de fournir une métrique de difficulté qui se veut absolue, mais plutôt de présenter la difficulté posée par une variation intra-classe des instances de la base.

Cette méthodologie repose sur une détermination de l'aire commune entre toutes les paires d'échantillons issues d'une même instance de modalité. En effet, plus l'aire commune sera importante, plus le système sous test devrait être en mesure de faire correspondre des caractéristiques biométriques issues des échantillons considérés. Ce critère repose sur la détermination d'un point d'alignement commun aux deux échantillons considérés, et qui permettra de servir de base à la détermination de l'aire commune. L'aire commune correspond aux zones qui peuvent être "superposées" l'une à l'autre sans que des différences notables ne puissent être détectées, et correspond donc à des zones de correspondance idéale. Cette mesure de l'aire commune ne comprend donc pas les zones ayant subies des déformations ou distorsions.

Un autre critère considéré par cette méthode est de mesurer la déformation qu'a subie une empreinte digitale au cours de sa capture. Ces déformations/distorsions sont susceptibles d'apparaître en raison du caractère élastique de la peau, qui est en mesure de se déformer sous l'effet de la pression, ou d'une légère rotation par exemple. Ces déformations sont donc particulièrement susceptibles d'apparaître dans le cas de capteur nécessitant un contact avec l'empreinte digitale. La détermination de la déformation repose sur la mesure de la différences des champs d'orientation entre les deux échantillons considérés.

Le troisième critère permettant de définir le niveau de difficulté d'une paire d'échantillons biométriques, est une mesure relative de la qualité de ces échantillons d'empreintes digitales. Dans la méthode proposée par Li *et. al.* [60], cette qualité relative d'une paire d'échantillon est calculée par le biais d'une moyenne géométrique de la qualité des deux échantillons.

Le niveau de chaque paire est ensuite utilisé afin d'obtenir le niveau de difficulté

global de la base d'échantillon. La méthode de fusion des niveaux de difficulté associés aux paires nécessite un entraînement afin de corréler cette mesure de difficulté avec les scores obtenus pour un ensemble de système biométriques de référence.

La mesure de la difficulté d'une base de données biométriques permet d'obtenir un a priori sur les performances d'un système testé quant à son taux de faux rejet. Comme précisé précédemment, cette mesure de difficulté n'inclut pas la détermination du taux de fausse acceptation.

2.6 Estimation des taux d'erreurs

Lors d'une évaluation biométrique, l'estimation des performances est l'étape permettant de transformer une base de résultats "atomiques" (*i.e.* résultat issu d'une transaction biométrique), ou d'échantillons en une ou plusieurs métriques de performance.

2.6.1 Principes

L'estimation des performances est réalisée par le biais de calculs sur des résultats unitaires issus d'une comparaison d'un échantillon biométrique avec une ou plusieurs références en fonction de l'usage du système biométrique. Ces calculs statistiques ont pour objectifs d'estimer un taux d'erreur, et éventuellement des marges d'erreur. Ainsi, les résultats d'une évaluation biométrique sont usuellement présentés sous la forme d'un intervalle de confiance, avec une valeur de confiance donnée.

L'obtention des résultats unitaires nécessite la réalisation d'une étape de comparaisons croisées, qui consiste en la détermination de résultats atomiques à partir d'une base de données biométriques ou d'une population de test. Une méthodologie doit être mise en place afin de spécifier certaines politiques sur les échantillons et les références ; ces dernières sont susceptibles d'influencer l'estimation des taux d'erreur.

2.6.2 Méthodes de comparaison croisée

L'étape de comparaison croisée permet de collecter les résultats des transactions biométriques. En fonction du facteur de forme pris par le système biométrique évalué, l'implémentation de la comparaison croisée peut être différente.

2.6.2.1 Principe

La comparaison croisée consiste en la multiplication de transactions biométriques, les résultats de ces dernières sont ensuite utilisés afin d'estimer les performances d'un

système biométrique.

Afin de réaliser ces transactions, il est nécessaire d'appliquer la comparaison biométrique proposée par le système à un couple composé d'un échantillon biométrique, et d'un ensemble de références biométriques. Dans le cas d'un système d'authentification, cet ensemble ne comprendra qu'une seule et unique référence, tandis que dans le cas d'un système d'identification cet ensemble inclura la totalité ou partie des références connues par le système sous test en fonction de son fonctionnement.

Dans le cas d'un système biométrique d'authentification, deux cas de figure se présentent : l'obtention de transactions légitimes, et celle de tentatives d'imposture. Les transactions légitimes consistent en la comparaison d'une paire constituée d'une référence et d'un échantillon issus d'un même utilisateur. Pour chaque utilisateur, toutes les paires échantillon-référence sont utilisées pour obtenir des résultats de transactions légitimes. Dans le cas de modalités admettant plusieurs instances chez un individu (*e.g.* les empreintes digitales présentent dix instances différentes : pouce, index, majeur, annulaire, auriculaire pour les mains gauche et droite ; l'iris en admet deux instances une pour chaque œil), les paires référence-échantillon ne sont effectives que pour les échantillons et les références issus d'une même instance.

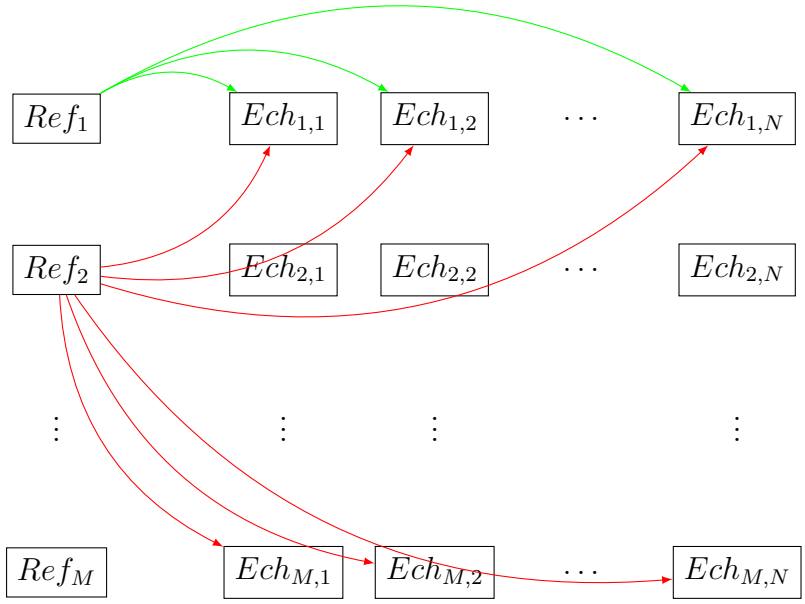

FIGURE 2.1 – Illustration de l'obtention des transactions légitimes et des tentatives d'imposture. Où Ref_i représente la référence du i^{me} utilisateur, et $Ech_{i,j}$ représente le j^{me} échantillon du i^{me} utilisateur. Les relations en vert représente les paires légitimes associant une référence et un échantillon d'une même personne, et les relations rouge illustrent les possibles paires d'imposture pour le premier utilisateur.

Afin d'obtenir l'ensemble des transactions d'imposture, il est nécessaire de compa-

rer les références issues d'un utilisateur avec des échantillons capturés auprès d'autres utilisateurs de la population de test ou présents dans la base de données biométriques. Lors de la réalisation des tentatives d'imposture, dans le cas d'une modalité présentant plusieurs instances, il n'est pas possible de réaliser des comparaisons entre une référence et un échantillon issus de différentes instances d'un même utilisateur.

Si l'évaluateur possède certains privilèges concernant le système biométrique et en particulier l'algorithme de reconnaissance biométrique, celui-ci est en mesure de réaliser des transactions de type *offline* ou *hybrid* (voir la partie 2.4.2). Ainsi, pour les transactions *online* ou *hybrid*, l'évaluateur est en mesure d'utiliser des bases de données biométriques comme entrée de l'algorithme biométrique. Tandis que pour les tests *online*, il est nécessaire de réaliser toutes les comparaisons manuellement, ce qui requiert une méthodologie différente.

Dans le cas des test *offline*, l'évaluateur ne doit pas nécessairement posséder des privilèges d'accès au système biométrique afin de les mettre en place. Néanmoins, l'absence d'accès aux données internes peut être en mesure de conditionner les différents types de résultats présentables à l'issue de l'évaluation.

2.6.2.2 Plateforme d'évaluation

Afin de permettre une automatisation du déroulement des comparaisons croisées, des travaux ont porté sur la définition de plateformes permettant de réaliser l'étape de comparaison croisée de manière automatique. L'objectif de telles plateformes est de permettre d'obtenir un ensemble de résultats de comparaison à partir de bases d'échantillons biométriques, ces plateformes sont donc particulièrement adaptées pour des systèmes sous test permettant de soumettre directement des références et des échantillons à l'algorithme de test.

Ces plateformes de test sont aussi en mesure de calculer les métriques de performances, une fois l'étape de comparaison croisée effectuée. En fonction des données collectables en sortie du système biométrique, ces plateformes peuvent être en mesure de fournir une estimation par le biais d'un intervalle de confiance des taux de FAR et de FRR, des courbes de performances ROC ou DET, et une mesure de l'EER.

Evabio

EvaBio[61] est une plateforme conçue dans l'équipe Monétique et Biométrie du Greyc, afin de réaliser les tests de performances d'un système et de calculer les métriques associées pour l'algorithme sous test. Cette plateforme permet d'interfacer des bases de données biométriques de manière simple : l'ajout d'une base ne présente pas de surcoût. Elle permet aussi de réaliser l'intégration d'un algorithme de recon-

naissance biométrique sans développement supplémentaire pourvu que l'algorithme satisfasse l'exigence de se présenter sous la forme d'une application ou d'un script parfaitement autonome, c'est-à-dire ne nécessitant aucune autre interaction par le biais d'IHM.

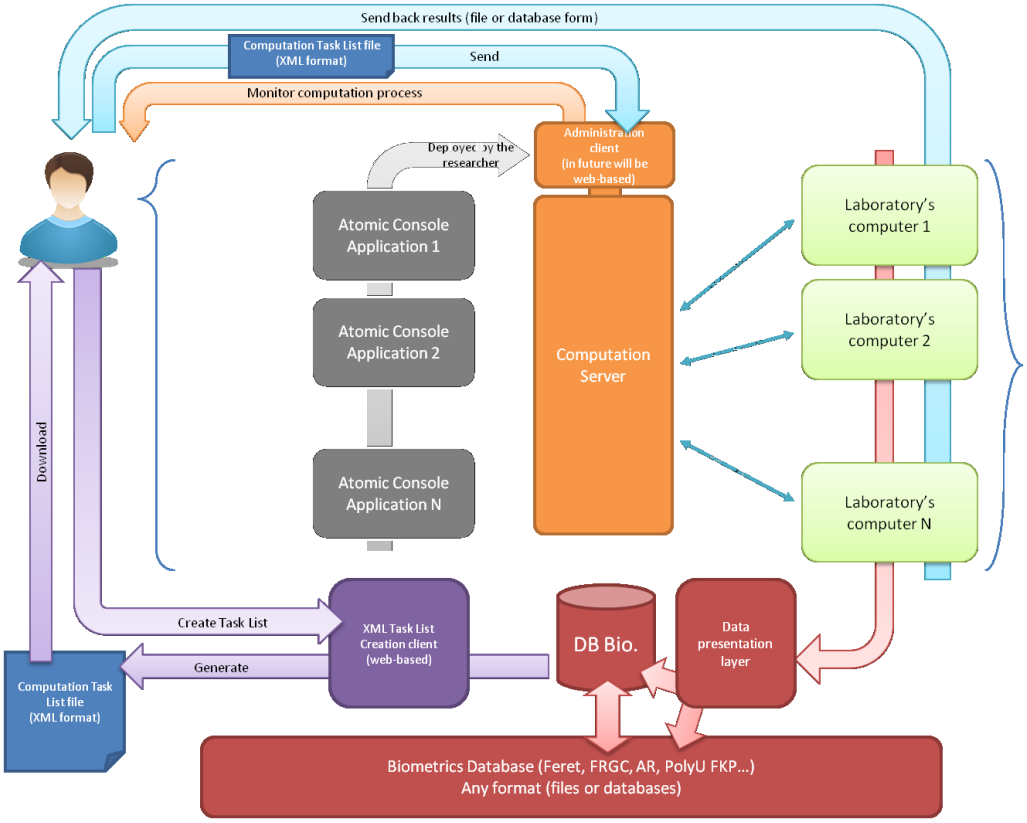

FIGURE 2.2 – Représentation de la plateforme EvaBio, illustrant son fonctionnement [61].

La liste des tâches de calcul ou *computation task list*, telle qu'illustrée dans la figure 2.2, représente le processus global d'évaluation de l'algorithme. Chaque tâche présente dans cette liste est un appel de l'exécutable avec ses éventuels paramètres d'appel. L'utilisation de cette plateforme permet d'exécuter de manière automatisée un processus d'évaluation avec les seules contraintes de disposer de bases biométriques correspondantes à la modalité de l'algorithme que l'on souhaite tester, et de pouvoir exécuter l'algorithme sans qu'aucune interaction extérieure ne soit requise. Cette plateforme a été utilisée pour différents tests. Dans le cadre du banc de test EvaBio, plusieurs capteurs sont placés côte à côte afin que les conditions environnementales soient les mêmes pour chacun d'eux. Une caméra est placée au-dessus des capteurs et l'image filmée est retransmise sur l'écran à gauche. Ceci permet à l'opérateur de test de s'assurer du bon déroulement des opérations et détecter des anomalies (mauvais

placement, ...) qui pourraient impacter les résultats. La lumière pouvant impacter le fonctionnement des capteurs, un luxmètre est placé à côté pour mesurer l'intensité lumineuse tout au long des tests.

FVConGoing : Fingerprint Vendor Competition onGoing

Au cours de plusieurs éditions[62, 63, 64, 13], plusieurs universités ont organisé une compétition entre des algorithmes biométriques de reconnaissance d'empreintes fournis par des acteurs industriels et académiques. Ces compétitions utilisaient plusieurs bases de données biométriques pré-établies, les industriels ou académiques fournissant l'algorithme de reconnaissance à évaluer sous la forme d'un exécutable. Le projet de compétitions ponctuelles a évolué vers une plateforme d'évaluation en ligne [4]. De même que pour les précédentes compétitions, les participants déposent un exécutable de leur algorithme de reconnaissance biométrique sur la plateforme d'évaluation, cette implémentation doit se conformer à un protocole de gestion des entrées/sorties afin d'être exploitable. Pour ce faire, des exemples et des squelettes de fichiers en plusieurs langages sont fournis à destination des acteurs industriels et académiques.

Cette plateforme propose plusieurs types d'évaluation principalement pour les systèmes de reconnaissance par empreintes digitales, néanmoins les modalités faciales et d'empreintes palmaires peuvent être évaluées. FVConGoing propose différentes bases de données biométriques permettant d'évaluer différents aspects d'un système biométrique. Ainsi, cette plateforme possède des bancs d'essai permettant de :

- les performances de reconnaissance biométrique pour les modalités d'empreintes digitales et palmaires pour un algorithme d'authentification.
- Dans le cas d'algorithmes de reconnaissance d'empreintes digitales utilisant des modèles se conformant à la norme ISO 19794-2[44], ou encore des représentations protégées [5].
- D'autre part, un banc de test permet d'évaluer les performances d'algorithmes d'identification face à une base de données adaptées.
- Une base de données permet d'évaluer la précision des algorithmes d'extraction de l'orientation des crêtes papillaires d'échantillons d'empreintes digitales.
- Un des bancs d'essais permet d'évaluer la résistance d'un algorithme de reconnaissance faciale à une altération "*morphing*" (fusion de deux visages différents).

4. https://biolab.csr.unibo.it/FVCOnGoing/UI/Form/Home.aspx
5. Ces représentations des empreintes digitales permettent d'éviter de compromettre une instance en rendant ce modèle non inversible.

- Les algorithmes déterminant la conformité d'une photographie de visage avec la norme ISO 19794-5[45].

BEAT Platform

Le projet européen BEAT [6] a pour objectif de définir et de proposer des outils pour l'évaluation des systèmes biométriques par le biais de la définition d'un référentiel de test, et d'une plateforme d'évaluation. Différents besoins ont été définis au cours de ce projet par Anjos *et. al.*[3] :

- la reproductibilité des résultats,
- la confrontation et la comparaison de ces résultats,
- des résultats recevables dans le cadre d'une certification ou de publications scientifiques.

Ainsi, une plateforme en ligne a été développée au cours de ce projet pour permettre son accès à des industriels et des chercheurs. Une interface de programmation (*API* : *Application Programming Interface*) et une interface utilisateur permettent d'accéder aux ressources de cette plateforme pour évaluer un algorithme de reconnaissance biométrique. La plateforme est évolutive et susceptible de s'adapter aux différents besoins pour une évaluation spécifique. Ainsi, il est possible de spécifier un protocole d'évaluation et de télécharger sur la plateforme les bases de données biométriques. Au cours du développement de cette plateforme, le principe de protection de la vie privée a été intégré en raison de la sensibilité des données biométriques accessible sur la plateforme.

2.6.3 Méthodes d'estimation des taux d'erreur

L'ensemble des résultats obtenus au cours de la collecte des transactions biométriques, correspondent à des résultats expérimentaux et certains peuvent induire un biais dans les résultats observés. Ainsi, il est nécessaire d'avoir recours à des méthodes statistiques afin fournir une estimation réaliste. Lors de l'estimation de la fréquence d'un événement, deux cas de figure sont possibles : 1. le dimensionnement de l'évaluation a permis d'observer des occurrences de l'événement dont on cherche à estimer la fréquence ou la probabilité et 2. un nombre de transactions insuffisant afin d'observer l'événement en question.

6. https://www.beat-eu.org/platform/

2.6.3.1 Calcul des intervalles de confiance

Un intervalle de confiance est un outil permettant de borner l'estimation d'une valeur avec une certaine confiance. Les valeurs standards pour la confiance d'un intervalle sont généralement de 95% et de 90%. À l'issue d'une évaluation biométrique, une estimation d'un taux d'erreur peut s'exprimer différemment en fonction de la méthodologie de comparaison croisée utilisée.

En effet, un intervalle de confiance est centré sur l'estimée d'une variable, et les marges d'erreurs sont calculées à partir de la variance de cette variable. Un intervalle de confiance peut donc s'exprimer comme :

$$IC = \hat{p} \pm z\Big(1 - \frac{\alpha}{2}\Big)\sqrt{\hat{V}(\hat{p})} \tag{2.5}$$

où $\hat{p}$ est l'estimée de la variable, $z()$ est l'inverse de la fonction de distribution cumulative, normale, centrée et réduite ; α est la probabilité d'erreur souhaitée pour l'intervalle de confiance. Et finalement, $\hat{V}(\hat{p})$ est l'estimée de la variance de la variable.

La "*rule of 30*[7]" proposée par Doddington [22] permet d'estimer un intervalle de confiance lorsque au moins trente occurrences d'un événement ont été observées. La "*rule of 30*" affirme ainsi que lors d'une série de tests lorqu'au moins trente erreurs sont observées, le taux d'erreur réel est dans un intervalle de confiance de ±30% autours de la valeur estimée, et ce avec une confiance de 90%.

$$IC = \hat{p} \pm 0,3.\hat{p} \tag{2.6}$$

où $\hat{p}$ est la valeur estimée de la variable considérée.

Ainsi, dans le cadre d'une évaluation biométrique, l'estimation d'un taux d'erreur est le ratio entre le nombre d'erreurs décomptées et le nombre de tentatives d'authentification comptabilisées au cours de l'évaluation. Ainsi, selon le nombre de transactions collectées par utilisateur, les estimations d'un taux d'erreur peuvent s'exprimer comme :

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} a_i \tag{2.7}$$

$$\hat{p} = \frac{1}{m.n}\sum_{i=1}^{n} a_i \tag{2.8}$$

$$\hat{p} = \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} m_i} \tag{2.9}$$

7. ou règle de 30

où $\hat{p}$ est l'estimation du taux d'erreur, n le nombre de participants à l'évaluation, m est un nombre fixe de transactions collectées auprès de chaque participant, m_i est le nombre de transactions collectées pour le i^{me} participant et a_i est le nombre d'erreur observées pour le i^{me} participant.

L'équation (2.7) correspond à l'estimation d'un taux d'erreur pour lequel une seule transaction a été effectuée par chaque participant. L'équation (2.8) permet d'estimer un taux d'erreur pour de multiples transactions pour chaque participant, le nombre de transactions enregistrées étant néanmoins fixe au travers de la population de test. Et finalement, la dernière (2.9) permet d'exprimer un taux d'erreur pour la collecte d'un nombre variable de transactions par personne.

Les calculs de la variance sont, respectivement pour une tentative par utilisateur, un nombre fixe de tentatives par utilisateur, et un nombre variable par utilisateur :

$$\hat{V}(\hat{p}) = \frac{\hat{p}(1-\hat{p})}{n-1} \tag{2.10}$$

$$\hat{V}(\hat{p}) = \frac{\hat{1}}{n-1}\left(\frac{1}{m^2.n}\sum_{i=1}^{n} a_i^2 - \hat{p}^2\right) \tag{2.11}$$

$$\hat{V}(\hat{p}) = \frac{\sum_{i=1}^{n} a_i^2 - 2\hat{p}.\sum_{i=1}^{n} a_i.m_i + \hat{p}^2.\sum_{i=1}^{n} m_i^2}{\frac{n-1}{n}.\left(\sum_{i=1}^{n}\right)} \tag{2.12}$$

où $\hat{p}$ est l'estimation du taux d'erreur, n le nombre de participants à l'évaluation, m est un nombre fixe de transactions collectées auprès de chaque participant, m_i est le nombre de transactions collectées pour le i^{me} participant et a_i est le nombre d'erreurs observées pour le i^{me} participant.

2.6.3.2 Estimation pour un événement non observé

Dans le cas où aucune observation d'un événement n'a pu être faite, il n'est pas possible de calculer un intervalle de confiance autour du taux d'erreur estimé. La *rule of three*[8]présentée dans [67, 66, 43], permet d'estimer que si un événement n'a pu être observé au cours de N tests statistiquement indépendants alors il est possible d'en estimer une limite supérieur p à sa probabilité d'occurrence.

$$p \leq \frac{3}{N} \tag{2.13}$$

Cette borne s'exprime avec un indice de confiance de 95%. La mise en pratique d'une évaluation boîte noire requiert de déterminer une valeur pour cette borne

8. Aucun équivalent en français n'a été trouvé pour le terme de *rule of three*, le terme "règle de trois" étant déjà consacré au calcul de proportionnalité.

supérieure. Une réflexion à ce sujet est développée dans la section 3.2.1.3, et propose une valeur à celle-ci. Les résultats obtenus à l'issu de ce type d'évaluation se présentent donc sous la forme d'un "point de fonctionnement", présentant à la fois les taux *FAR* et *FRR* observés.

2.7 Certification et référentiels de tests pour l'évaluation des systèmes biométriques

En tant que solution d'authentification, les systèmes biométrique sont intégrés dans différents systèmes et applications présentant des exigences quant à leur sécurité et fiabilité. Ces différents dispositifs sont susceptibles d'accorder l'accès à des données, ou des fonctionnalités sensibles (*i.e.* paiements en ligne, application et consultation de comptes bancaires, correspondance électronique ...), ou être utilisé en tant que solution de contrôle d'accès pour des zones sensibles. Afin d'être intégré pour ces cas d'usage, un organisme de certification et/ou le fournisseur de l'application ou de la fonctionnalité sont susceptibles d'imposer la conformité de ces systèmes avec des exigences (fonctionnelles, ou en termes de performances). Ces exigences sont formulées afin de garantir qu'un système est adapté à l'usage prévu tant du point de vue des performances que de la sécurité, mais aussi qu'il est conforme aux spécifications d'interopérabilité.

Dans le cadre d'un schéma de certification, une évaluation est mise en œuvre afin de déterminer la conformité d'un système avec les exigences formulées. Ces évaluations sont régies par un référentiel de test afin de garantir que :

- les résultats obtenus au cours d'une évaluation soient exploitables afin de déterminer la conformité du système sous test avec les exigences formulées dans le référentiel de test.
- D'autre part, les méthodes d'évaluation présentées dans le référentiel de test ont été définies avec un impératif de reproductibilité. En effet, des résultats non reproductibles peuvent mener à des décisions différentes de délivrer ou non la certification pour un même système, ou pour deux systèmes similaires. Concernant l'estimation des performances biométriques, la stricte reproductibilité des résultats nécessiterait d'utiliser une même base d'échantillon pour tous les systèmes testés, ou encore de réaliser un nombre très important de comparaisons biométriques. Les référentiels de test proposent des recommandations et des exigences sur la composition de la population de test afin de limiter le biais apporté par la population de test.

- Et enfin que les résultats soient obtenus dans un environnement normal pour tous les tests non environnementaux. Le référentiel présente aussi un environnement type dans lequel se déroulent tous les tests nécessitant des conditions normales. Afin de quantifier l'influence des conditions environnementales sur les performances du systèmes sous test, plusieurs tests prenant place dans des différents environnements défavorables sont définis.

Les référentiels de test compilent les informations nécessaires à la mise en œuvre d'une évaluation dans le cadre d'une certification, et se structurent au sein de différents documents. Ces corpus de documents s'articulent autour de trois axes :

- Le positionnent du référentiel : Les documents de ce premier axe définissent l'ensemble des aspects couverts au cours de l'évaluation. Dans le cadre du positionnement du référentiel, une représentation des systèmes biométriques est définie, et autour de cette dernière les différents cadres d'évaluation sont définis. Chacun de ces cadres traite un aspect spécifique d'un système biométrique à évaluer au regard du processus de certification (*e.g.* la détermination des performances du système sous test telles que décrites dans la section 2.2.1, ou encore des aspects sécuritaires de ce dernier par l'étude des potentielles vulnérabilités et la résistance à des attaques connues...).
- Les spécifications générales de test : celles-ci présentent pour chaque cadre d'évaluation ses objectifs, les fonctionnalités et/ou les paramètres testés et les relations entre ce cadre d'évaluation et les autres. En plus de cette vision d'ensemble, ces spécifications fournissent aussi une description haut niveau des tests que subiront les systèmes sous test. Ainsi pour chaque test, sont précisés ses objectifs et les différents tests spécifiques le composant.
- La spécification détaillée de test : les spécifications détaillées décrivent en détails chaque test d'un cadre d'évaluation et les méthodologies mises en œuvre dans ces derniers. Dans le cas de tests paramétrables, les différentes configurations sont présentées et décrites (par exemple, dans le cas d'un test d'orientation pour une présentation biométrique, les différents valeurs angulaires à tester sont précisées).

  Pour chaque test, différentes informations sont définies :

  - Les objectifs du test : les objectifs du test sont explicités afin d'interpréter correctement les résultats obtenus à l'issu de celui-ci.
  - Les exigences sur le système sous test : ces informations permettent de déterminer si un test est applicable à un système biométrique. En effet, la mise en œuvre de certains tests peut requérir l'accès à certaines

informations ou fonctionnalités qui peuvent ne pas être disponibles sur le système testé.

- les conditions environnementales sont précisées et décrivent soit un environnement spécifique dans le cas des tests environnementaux, soit rappellent les conditions de l'environnement standard.

La multiplicité des usages biométriques et la propagation à la sphère grand public de l'authentification biométrique ont motivé plusieurs projets de définition de référentiels d'évaluation biométrique. De plus, la définition et la mise en pratique de tels processus d'évaluation ont été rendus possibles grâce à la maturité du domaine biométrique (et en particulier à son utilisation en tant que solution d'authentification pour les smartphones et autres dispositifs portables). La définition de différentes méthodes d'évaluation, à la fois dans la littérature scientifique et technique par le biais de nombreuses publications et de travaux de normalisation, ont aussi contribué à la définition de différents référentiels. En raison de l'ampleur de tels projets, ainsi que pour des impératifs d'impartialité, la définition de référentiels de test et la mise en place d'un schéma de certification se font généralement au sein d'un consortium ou d'une association d'entreprises et d'organismes publics.

- La BAI (Biometric Alliance Initiative) a pour ambition de mettre en place et de promouvoir un schéma de certification des systèmes biométriques. La BAI est un consortium d'acteurs du domaine biométrique et de disciplines connexes, ou amenées à l'être, à celle-ci. La BAI rassemble différents types d'acteurs : des universitaires, des banques, la grande distribution, des laboratoires de test... L'objectif de la BAI est de proposer dans le cadre d'une certification, un référentiel d'évaluation qui répond à des besoins non régaliens, avec pour objectif de promouvoir la biométrie dans ses nouveaux usages auprès du grand public et des décideurs. Dans la définition de son référentiel de test « Framework For The Evaluation Of Biometric Systems », la BAI a adopté une démarche d'évaluation telle que vue par les normes ISO et s'appuie fortement sur ces dernières. Dans un premier temps, le référentiel de la BAI ne définit sa méthodologie d'évaluation que pour les modalités de voix et d'empreintes digitales.
- L'objectif du projet BEAT est de définir et de proposer un cadre standard d'évaluation opérationnelle de la sécurité pour les systèmes biométriques. La mise en place d'un tel projet est motivée par la multiplication des usages de la reconnaissance biométrique, et par la difficulté qu'il y a actuellement à déterminer la fiabilité de tels systèmes. Le référentiel BEAT est publié au fil de l'eau, et plusieurs documents permettant de situer son positionnement sur l'évaluation de la biométrie, sont disponibles. Ils appliquent une approche

globale d'évaluation reposant sur les critères communs. Un des objectif de BEAT est de mettre en place une plateforme de tests en ligne permettant d'évaluer de manière transparente et indépendante un système biométrique face à des bancs de test validés, des protocoles d'analyse de risque sur les systèmes biométriques et un référentiel d'évaluation compatible avec les critères communs.

2.8 Conclusion

Ce chapitre expose les principaux éléments constitutifs du domaine de l'évaluation biométrique : les métriques de performances calculables, et les méthodologies de collecte des résultats et d'estimation des taux d'erreur. Les méthodologies présentées précédemment se concentrent sur la collecte de résultats, et d'échantillons sur des systèmes biométriques accordant d'importants privilèges d'accès. En effet, ces méthodes présuppose généralement qu'un évaluateur a librement accès aux données internes aux systèmes biométriques (qui sont désignés comme systèmes en *boîte blanche*) ; et en particulier aux échantillons et modèles biométriques, aux scores de comparaison, à la référence biométrique, etc.

Néanmoins si les tests sont réalisés par des tierces parties, le système biométrique sous test se présente généralement sous la forme d'une boîte noire (c'est-à-dire qu'un nombre restreint voire aucunes informations internes du système biométrique ne sont accessibles). Afin d'implémenter un procédé d'évaluation des boîtes noire, de nouvelles méthodologies doivent être mises en place afin de permettre l'évaluation de tels systèmes.

Chapitre 3

Évaluation des boîtes noires biométriques

« Insanity is doing the same thing over and over again and expecting different results. »

Citation apocryphe d'Albert Einstein.

Ce chapitre présente les travaux réalisés dans le cadre de l'évaluation des boîtes noires. Une méthodologie d'évaluation a été développée spécifiquement pour les systèmes biométriques opérationnels. Ces derniers limitent les données et les fonctionnalités accessibles à l'évaluateur, ainsi une méthodologie spécifiques aux systèmes biométriques en "boîte noire" a été développée et mise en œuvre au cours de cette thèse.

Sommaire

3.1 Introduction

L'évaluation des systèmes biométriques est un sujet largement étudié, ayant ainsi fait l'objet d'une normalisation. Néanmoins, dans le cadre d'une évaluation biométrique, celle-ci est susceptible d'être réalisée par une tierce partie, et en conséquent

les constructeurs de systèmes biométriques peuvent ne pas être enclins à fournir un système permettant l'accès aux données internes du système, et ce généralement afin de protéger leur propriété intellectuelle. Les méthodologies proposées dans la littérature ne prennent que rarement en compte les limitations imposées par un système en boîte noire. Ce chapitre se propose donc d'étudier et de définir une méthodologie permettant de mettre en place un processus d'évaluation en boîte noire. Dans le cadre de la définition d'une méthodologie d'évaluation concrète, les systèmes biométriques ont été étudiés au sein de la société ELITT, afin de définir les limitations imposées par ce type de système. Les systèmes biométriques sont de plus en plus communément intégrés dans des solutions permettant d'accéder à des données ou des fonctionnalités sensibles, les performances de ces systèmes se doivent donc d'être en adéquation avec ces usages.

3.2 Étude de l'évaluation des systèmes biométriques en boîtes noires

Les méthodologies proposées dans la littérature et les normes présentent des méthodologies séparant les processus de collecte des échantillons biométriques et de comparaison croisée. La séparation de ces deux processus se justifie par une simplification du processus d'évaluation, mais nécessite néanmoins l'accès à un système biométrique en boîte blanche, ou postule que l'évaluateur dispose de privilèges administrateurs sur le système biométrique testé. Les systèmes en boîte blanche désignent des systèmes biométriques dans lesquels l'évaluateur dispose à tout moment d'un accès aux données internes du système. Ainsi dans le cas d'un système biométrique, un boîte blanche fournit à minima à l'évaluateur la capacité d'accès et de manipulation de l'échantillon et de la référence biométrique, ainsi que les moyens de stocker les scores de comparaison et les décisions du système (selon le cas d'usage, la décision peut se présenter sous la forme d'un binaire dénotant la reconnaissance ou non d'un individu pour l'authentification, ou d'une liste de candidats dans le cas d'un système d'identification).

Cependant, les industriels sont soucieux de préserver la propriété intellectuelle de leurs algorithmes et du matériel impliqués dans les processus de reconnaissance biométrique. Ainsi, lorsque une évaluation est réalisée par une tierce partie, celle-ci est susceptible de ne pas avoir accès à un dispositif en boîte blanche, ou de ne pas disposer des privilèges d'administration du système permettant de manipuler le système en tant que boîte blanche. Il est alors nécessaire de mettre en œuvre des procédé d'évaluation différents, mais ceux-ci s'avèrent plus ardus à mettre en

place. Cette section se propose donc de présenter les conclusions et les observations préliminaires effectuées lors de l'étude et de l'utilisation de systèmes biométriques en boîte noire.

3.2.1 Objectifs et définition du protocole d'évaluation en boîte noire

La problématique de définir un protocole d'évaluation en boîte noire est apparue face au besoin d'évaluer des systèmes biométriques embarqués dans des dispositifs portables : les *smartphones*. En effet, les solutions d'authentification intégrées dans ces appareils permettent d'accéder à un certain nombre d'applications sensibles (en particulier avec les solutions de paiements telles qu'ApplePay ou Android Pay, qui proposent toutes deux le paiement par le biais d'une authentification biométrique en lieu et place de la saisie d'un *PIN*), en plus des données personnelles que contiennent les *smartphones*.

3.2.1.1 Forme des résultats

Le challenge posé par ce type d'évaluation est de réaliser une évaluation sur un système opérationnel, ne donnant que peu ou presque pas de privilèges d'administration ou de manipulation sur le système biométrique. Il n'est ainsi pas possible de séparer le processus de collecte des données, de celui de réalisation des comparaisons. De plus étant donné les faibles taux de FAR admis par une majorité de systèmes biométriques, il n'était pas envisageable de répéter des transactions biométriques jusqu'à obtenir une estimation exacte de ce taux d'erreur.

En effet, les limitations d'un système en boîte noire sont que :

- toutes les transactions biométriques requièrent une présence continue et une sollicitation régulière de la population de test ;
- Les capacités d'accès aux données internes sont particulièrement limitées, en effet il n'est pas possible d'accéder aux échantillons biométriques en sortie du capteur, le *template* biométrique extrait et la référence ne sont pas accessibles. De même le score de comparaison n'est pas accessible.
- Les transactions ne peuvent pas être réalisées de manière automatique en raison de l'absence d'accès à l'algorithme biométrique. Il n'est ainsi pas possible d'injecter un échantillon et/ou une référence choisie en entrée de l'algorithme de comparaison, celles-ci devront être réalisées manuellement (transaction *online*).
- le système ne permet l'accès qu'à la décision finale.

En lieu et place d'une estimation pour des taux d'erreur particulièrement bas, l'alternative consistant à déterminer une borne supérieure a été choisie. Les limitations imposées par les systèmes en boîte noire ne permettent donc pas d'obtenir certaines métriques de performances largement utilisées dans le cadre des normes et de la littérature scientifique. En particulier, l'absence d'accès aux scores ne permet pas de connaître la répartition des scores, le tracé des courbes *ROC* ou *DET* ne peut donc pas être fait. De même, le système ne fournit que peu de *feedback* ou retour d'informations quant à la qualité de l'échantillon, l'observation du *FTA* ne peut se faire de manière fiable et automatisée, et requiert l'attention constante du testeur et du superviseur. Étant donné ces contraintes, les résultats se présentent généralement sous la forme d'un intervalle de confiance autours du *FRR*, et d'une borne permettant d'assurer que le taux d'erreur lui est inférieur avec un confiance de 95%.

Cependant, la forme et l'estimation des résultats ne furent pas la seule problématique rencontrée lors de la définition d'un protocole en boîte noire.

3.2.1.2 Déroulement de la collecte des transactions

La limitation d'accès à l'algorithme ne permet pas de séparer les processus de capture des échantillons biométriques et la réalisation des transactions biométriques. Compte tenu de ces limitations, une méthodologie différente de déroulement de l'évaluation est utilisée pour réaliser l'ensemble des transactions biométriques. Une méthode de déroulement de la comparaison croisée est proposée pour le cas des systèmes d'authentification biométriques.

Comme illustré par la figure 3.1, une évaluation en boîte noire est subdivisée en plusieurs tours ou *rounds*. Chacun des rounds correspond à une référence, ceux-ci sont subdivisés en deux parties : 1. la collecte des transactions d'imposture et 2. la réalisation des tentatives de vérification légitime. Au début d'un tour, un ou plusieurs utilisateurs enregistrent des références dans le système sous test ; ils sont désignés en tant qu'utilisateurs légitimes pour le tour courant. Lors de la phase d'enregistrement des références, le ou les utilisateurs effectuent une vérification immédiatement afin de confirmer la fiabilité des références. La référence et l'utilisateur qui la fournit, sont changés à chaque *round* de l'évaluation et tous les utilisateurs de la population sont susceptibles d'être références au cours d'un tour dans le cas des populations de taille modeste.

Lors de la collecte des tentatives d'imposture, chaque utilisateur n'étant pas un utilisateur légitime, présente les instances éligibles pour cette évaluation afin de réaliser des tentatives d'authentification sur le système testé. Les résultats doivent être consignés afin de constituer une base de résultat.

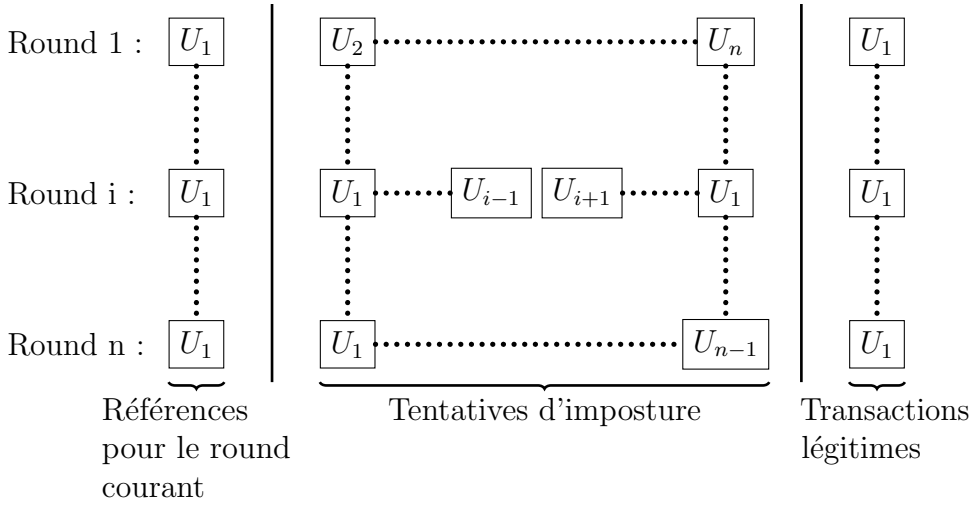

FIGURE 3.1 – Illustration du déroulement d'une comparaison croisée en boîte noire. L'évaluation en boîte noire, telle que définie dans cette thèse, se décompose en plusieurs tours. À chaque tour, les références biométriques sont changées, c'est-à-dire que les nouvelles instances de modalité enregistrées sont différentes de celles utilisées au cours des tours précédents. Dans un premier temps, les références enregistrées sont comparées à des échantillons capturés issus d'autres individus afin de collecter les résultats d'imposture. Ensuite, des échantillons issus des modalités enregistrées sont capturés afin de collecter des tentatives d'authentification légitimes.

La collecte des tentatives légitimes s'effectue après les tentatives d'imposture, ceci permet d'instaurer un délai entre la création de la référence et la vérification par l'utilisateur légitime. Ce délai fut motivé par le processus de création des références de certains systèmes biométriques en boîte noire ; en effet chez ces derniers, le processus de création d'une référence requiert plusieurs captures. En effet, dans la plupart des systèmes biométriques embarqués dans des dispositifs tels que les *smartphones*, le capteur n'est pas d'une taille suffisante pour permettre de capturer la totalité d'une empreinte digitale. Le système met ainsi en place une stratégie de constitution d'une référence par le biais de multiples échantillons, permettant ainsi de limiter l'incomplétude de cette dernière et donc ses conséquences sur les taux d'erreur.

Néanmoins, le nombre de tentatives obtenu par cette méthode peut s'avérer insuffisant afin de réaliser une estimation précise du FRR, d'autres tentatives sont collectées à l'aide d'une méthodologie similaire. Dans ce cas, un temps de latence est toujours observé entre l'enregistrement d'une référence et les tentatives légitimes sur cette dernière. Bien que plus court, ce temps de latence permet d'éviter un effet d'automatisme qu'il est possible d'observer pour des tentatives suivant immédiatement l'enregistrement d'un référence.

3.2.1.3 Principes de l'évaluation en boîte noire

En postulant que l'événement, dont on cherche à estimer une borne supérieure pour le taux d'erreur le caractérisant, est hautement improbable, il devient possible de dimensionner l'évaluation en termes de nombre de transaction, et de nombre d'individus dans la population de test pour une valeur choisie bornant ce taux d'erreur. Ainsi, en partant du contexte des *smartphones* et des applications de paiement, la solution d'authentification la plus implémentée est la saisie du code *PIN*.

Le code *PIN* est un code numérique à quatre chiffres. D'après O'Gorman[73], l'entropie d'un tel code peut être fixée à 13.3 bits. En réalité celle-ci est généralement moindre, en particulier dans le cadre bancaire où certains codes ne sont pas retenus en raison de leur trivialité (*e.g.* 0000 ou 1234), ainsi O'Gorman[73] estime l'entropie de ces codes comme étant de 9.6 bits. En effet, la répétition importante d'un chiffre est généralement proscrite, ainsi le code 0000 n'est pas éligible, ainsi que certains cas trivialement mnémotechniques tels que 1234 ou 4321. De plus, la saisie du code *PIN* autorise trois tentatives, divisant d'autant la probabilité d'être mis en échec en cas d'attaques à effort nul [1]. Cependant si l'on conserve l'approche naïve, il est possible de considérer qu'un code *PIN* présente une probabilité d'être mis en défaut de 10^{-4}.

Motivé par le fait que le domaine de la sécurité informatique considère que la robustesse d'un système complet est égale au degré de robustesse de son élément le moins sécurisé, et que les solutions de paiement proposent conjointement des authentifications par *PIN* et biométrique, la borne choisie a été fixée à 10^{-4}. Le rapprochement entre différents moyens d'authentification a été fait par O'Gorman [73]. En se référant à la *rule of 3*, si 30 000 transactions sont effectuées sans qu'aucun cas d'imposture ne soit observé, le système est en mesure de revendiquer un taux d'erreur correspondant inférieur ou égal à 1.10^{-4}. La détermination du nombre de transactions nécessaires à la confirmation d'une revendication de performance permet d'aider à la définition des dimensions de l'évaluation en terme de population de test.

3.2.2 Évaluation en boîte noire d'un smartphone

La mise en place d'un processus d'évaluation complet sur des systèmes biométriques en boîte noire est coûteuse en terme de temps et d'investissement du point de vue de la population de test (qui doit être recrutée, et disponible pendant l'évaluation). Deux expérimentations ont donc été réalisées au préalable, et ce afin d'estimer la faisabilité d'une évaluation en boîte noire à plus grande échelle, et d'autre part afin d'améliorer les éventuelles faiblesses détectées. Ces deux mini évaluations suivent

1. Une attaque à effort nul est une attaque ne mettant en pratique aucune technique particulière.

la méthodologie présentée dans 3.2.1. Leurs déroulements se sont faits au sein de la société ELITT. En conséquence les populations de test ont été recrutées en interne. La taille de ces dernières est réduite, afin d'évaluer uniquement le déroulement de chaque implémentation de la méthodologie. Les estimations de performances obtenues à l'issue de ces expérimentations ne sont pas considérées comme recevables en raison d'une population de test réduite et donc non suffisamment représentatives.

Ces deux évaluations expérimentales ont été réalisées sur un même système biométrique embarqué dans un *smartphone*. Ainsi pour ce système, le capteur est intégré dans la touche d'accueil du téléphone, les dimensions de ce dernier sont de $10mm$ de long par $4mm$ de large, pour une aire de capture de $40mm^2$.

3.2.2.1 Description des populations impliquées dans les expérimentations

La mise en place de ces expérimentations s'est faite en ayant recours à des populations de test de taille réduite, en effet il ne s'agissait pas d'obtenir des résultats de test exploitables, mais d'observer le déroulement de celles-ci, et d'en déterminer les éventuelles faiblesses et de définir des pistes d'améliorations. Ainsi, des populations de l'ordre d'une dizaine de personnes ont participé à ces évaluations, celles-ci sont présentées dans le tableau 3.1.

TABLE 3.1 – Description des populations impliquées dans les deux expérimentations

Population de test	Première expérimentation	Seconde expérimentation
Nombre total de participants :	12	13
Plage d'âge :	25-60	20-60
Catégories d'âge représentées :		
20-30	1	2
31-40	2	2
41-50	7	7
51-60	2	2

Dans le cadre de ces expérimentations, la population de test n'étant pas dédiée à celles-ci, il est nécessaire de faire des compromis permettant la bonne mise en œuvre des évaluations, tout en s'adaptant à l'emploi du temps des participants.

3.2.2.2 Première expérimentation : protocole nomade

L'objectif de cette première implémentation est de déterminer la faisabilité et d'estimer le temps minimal nécessaire à la mise en pratique d'une évaluation à

une échelle supérieure. Comme précisé dans la section 3.2.2.1, un compromis entre l'emploi du temps des participants à l'expérimentation et le déroulement de cette dernière a été fait. Dans cette première expérimentation, afin de ménager l'emploi du temps des participants, l'évaluation s'est faite selon un processus mobile, afin de diminuer la durée globale de l'évaluation. En effet, cette méthodologie permet d'éviter la mise en place d'une planification détaillée précisant les horaires de passage d'un participant pour chaque tour de l'évaluation. De plus, la mise en place d'un tel planning a été écartée en raison d'une estimation du temps de passage de chaque participant très théorique et approximative.

En conséquence, le dispositif sous test est déplacé dans l'édifice où travaillent les participants, ainsi les transactions de test récoltées ne l'ont pas été systématiquement dans la même pièce. La mise en place de ce processus impose néanmoins certaines contraintes :

- L'évaluation se déroulant dans différentes pièces, les conditions environnementales sont susceptibles de présenter de légères variations entre celles-ci. Néanmoins, les outils de mesure à disposition requièrent un temps de latence avant de fournir une mesure exacte des conditions environnementales. Étant donné l'objectif de cette évaluation visant à estimer une durée minimale d'évaluation, l'introduction d'une latence répétée à chaque changement de bureau [2]est à proscrire. Les conditions environnementales ne sont donc pas renseignées pour cette expérimentation.

- Les résultats de cette évaluation témoin sont enregistrés à l'aide de formulaires. Le superviseur se charge d'effectuer l'enregistrement des résultats et de consigner les observations réalisées. Néanmoins, au cours de l'expérimentation, l'observation des interactions entre le participant et le système sous test s'est révélé particulièrement ardue. D'une part, le superviseur notant le résultat des transactions, et d'autre part les participants réalisant généralement les transactions biométriques à la suite, ceci ne permet pas une observation efficiente.

Cette première expérience permet de déterminer les limites du processus et ses possibles axes d'amélioration, la principale amélioration étant le remplacement des formulaires par un logiciel dédié à l'enregistrement des résultats dans une base de données spécifiquement conçue.

2. En se référant aux descriptions de la population de test disponible dans le tableau 3.1, une estimation haute du nombre de changements de bureaux est de l'ordre de la centaine.

3.2.2.3 Seconde expérimentation : évaluation assistée

La seconde expérimentation implémente un protocole légèrement différent. Étant donné les difficultés à collecter les retours des participants et à observer les interactions utilisateurs-système biométrique, la charge d'enregistrement des résultats se fait maintenant par le biais d'un logiciel dédié, et est effectué par le participant sous la supervision du responsable de l'évaluation.

L'enregistrement des résultats se fait donc par le biais d'un logiciel dédié, permettant de stocker les résultats dans une base de données dont le diagramme est présenté dans la figure 3.2. L'enregistrement d'un résultat est réalisé par le participant à l'issue de chaque transaction biométrique, le logiciel affiche différentes propositions quant aux résultats observés : a) une transaction acceptée, b) une transaction rejetée et c) un cas d'échec à l'acquisition. La gestion de l'enregistrement étant de la responsabilité des participants, le superviseur est en mesure de réaliser des observations, en particulier sur les interactions biométriques afin de consigner des comportements notables ou anormaux, les remarques des utilisateurs ont aussi été consignées, pendant que les participants réalisent et enregistrent leurs interactions avec le système biométrique.

Ce logiciel étant embarqué dans un ordinateur, le protocole mobile implémenté lors de l'expérimentation précédente ne peut donc pas être mis en œuvre. Cette évaluation se déroule ainsi dans une seule pièce, ce qui présente des avantages et des inconvénients. Comme pour l'évaluation test précédente, la mise en place d'un planning prévisionnel de passage s'avère une entreprise difficile et particulièrement problématique en raison d'emplois du temps variables, et d'indisponibilités plus ou moins longues peu prévisibles (telles que des conférences téléphoniques, des rendez-vous ou encore des absences pour cause de maladie, de déplacements ...). D'une part, le déroulement des tests est ralenti, la population de test n'étant pas dédiée, il est nécessaire que chaque membre de la population de test se déplace jusqu'à la pièce où prend place l'évaluation, après y avoir été convié. Ceci suppose généralement que ceux-ci finissent ou interrompent leurs tâches en cours. Néanmoins, la réalisation de l'évaluation dans un local dédié permet d'enregistrer les variations des conditions environnementales tout au long de l'expérimentation.

Présentation de la structure de base de données

Le logiciel d'enregistrement des résultats repose sur une base de données (BDD) spécialement conçue à cet effet. La structure de BDD est inspiré du modèle de données présenté dans Mahier *et. al.* [61] présenté dans la figure 2.2. La structure

implémentée par ce logiciel est présentée dans la figure 3.2, la base de données est une base de données relationnelle.

Les différentes tables présentent dans la base sont :

- Utilisateurs : cette table contient les identités de chaque participant, et n'est utilisée qu'à des fins administratives. L'identifiant de l'utilisateur est utilisé tel quel dans la table contenant les profils utilisateur. Cette table est stockée sous format papier afin de garantir l'anonymisation des données.
- Profil utilisateur : cette table permet de renseigner la catégorie d'âge, le genre et le type d'occupation pour chaque utilisateur. Le type d'occupation renseigne si un participant pratique des activités physiques susceptibles d'altérer certaines modalités biométriques, et en particulier dans le cadre de nos expérimentations les empreintes digitales.
- Instances de modalité : cette table associe à chaque profil utilisateur une ou plusieurs modalités biométriques. La clef composite de cette clef est composée de la clef d'un profil utilisateur et d'une clef issue de la table des modalités.
- Liste de modalité : cette table contient la liste des modalités qui sont susceptibles d'être utilisées au cours d'une évaluation. Dans le cas d'une modalité présentant plusieurs instances, une entrée est présente pour chacune des instances (dans le cas des empreintes digitales, les doigts des deux mains possèdent des entrées dans cette table).
- Références : une entrée dans cette table correspond à une référence enregistrée dans le système biométrique (tant pour les cas d'imposture que pour ceux de tentatives légitimes). Cette table renseigne le cas échéant le nombre de présentations nécessaires à l'enregistrement d'une référence viable, ainsi qu'un champ de description permettant de noter des observations (interactions notables avec le capteur biométrique telles que éventuellement l'orientation du doigt, ou une tendance à ne présenter qu'une zone restreinte de l'empreinte digitale).
- Session de comparaison : cette table correspond à la confrontation d'une instance de modalité avec une référence enregistrée dans le système. La clé de cette table est composée des clés correspondant à l'instance de modalité fournissant les échantillons de comparaison au système d'une part, et d'autre part de l'identifiant de la référence. Si ces deux clés correspondent à une même instance de modalité, il s'agit alors d'une session de transactions légitimes, dans le cas contraire d'une session de tentatives d'imposture.
- Résultats : cette table contient les résultats atomiques de chaque comparaison. Ainsi, une entrée dans cette table renseigne le résultat de la comparaison

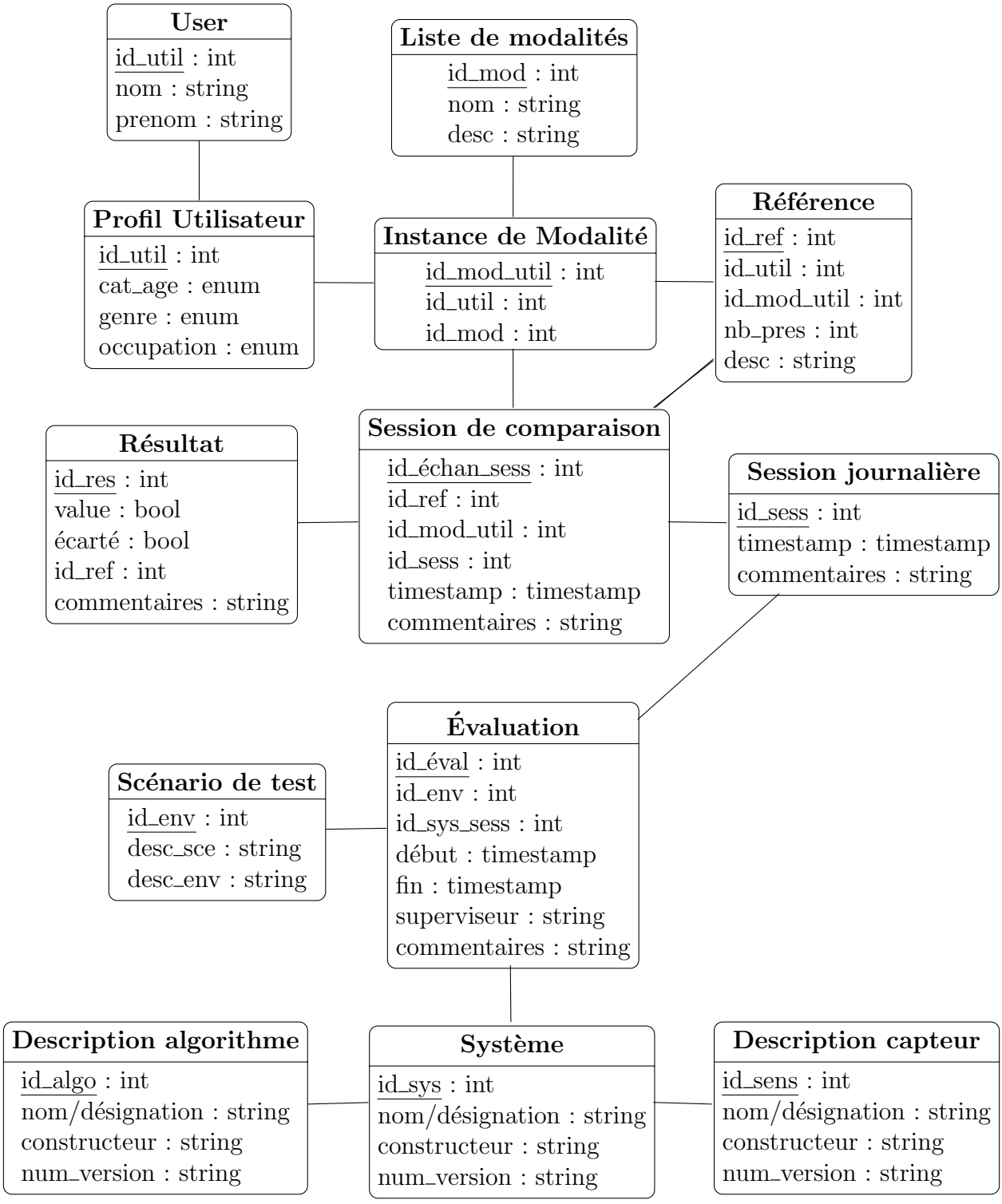

FIGURE 3.2 – Cette figure représente les différentes tables de la base de données relationnelle permettant d'enregistrer les résultats d'une évaluation biométrique d'une boîte noire.

(accepté ou rejeté, la conformité à la vérité terrain est déterminée par la session de comparaison). Un champ permet de déterminer si un résultat doit être écarté ou non lors de la détermination des résultats, ceci fut motivé par la possible présence de présentations incorrectes (par exemple, la substitution d'une instance par une autre, ou une présentation ne répondant à la politique de test).

En plus de ces tables, d'autres sont définies afin de permettre de stocker dans une base unique les résultats de plusieurs évaluations :

- Session : cette table représente une session journalière, et contient donc en conséquence la date de celle-ci.
- Évaluation : Cette table est associée aux descriptions du systèmes biométrique sous test, et à la description de l'environnement ciblé.
- Capteur : cette table permet de stocker les données relatives à un capteur impliqué dans une évaluation, si les informations sur ce dernier ont été fournies ou qu'elles sont librement accessibles.
- Algorithme : de même cette table est amenée à contenir les informations relatives à un algorithme testé, si les informations sur ce dernier ont été fournies ou qu'elles sont librement accessibles.

Conditions environnementales

Au cours de cette évaluation test, le changement de protocole permit de réaliser efficacement l'enregistrement des conditions environnementales. Cependant, en raison des capteurs à disposition, seule la température est renseignée, la période d'échantillonnage est fixée à une minute. Les valeurs minimale, maximale et moyenne observées au cours de l'évaluation sont présentées dans le tableau 3.2.

	1^{re} session	2^{me} session	3^{me} session
Min	$23.3°C$	$22.9°C$	$22.8°C$
Max	$26.7°C$	$25.7°C$	$26.4°C$
Moy	$24.1°C$	$24.5°C$	$24.8°C$

TABLE 3.2 – Ce tableau présente les variations de température au cours de la seconde évaluation test. Ainsi, les valeurs minimale et maximale indiquent la plage de température observée au cours de l'expérimentation, ainsi que la température moyenne.

3.3 Mise en place d'évaluations en boîte noire

Les expériences décrites dans la section 3.2 ont permis de définir un protocole permettant de déterminer les performances biométriques d'un système boîte noire. Ainsi, plusieurs évaluations ont été réalisées à des échelles permettant d'obtenir des résultats pouvant être considérés comme représentatifs des performances réelles du système.

3.3.1 Évaluation pour un système biométrique monomodal

La première évaluation biométrique mise en place a été réalisée afin d'évaluer des systèmes monomodaux, le protocole d'évaluation est une itération de celui implémenté lors de la seconde expérimentation (*cf.* la section 3.2.2.3)

3.3.1.1 Objectif et protocole d'évaluation

Comme précisé précédemment, le protocole mis en place pour cette évaluation est une évolution inspirée de ceux implémentés au cours de ces expérimentations. Les objectifs de cette évaluation sont différents de ceux des évaluations test, ainsi il est nécessaire d'évaluer non pas un unique système, mais bien de paralléliser l'évaluation de plusieurs systèmes biométriques. Tous ces systèmes sont embarqués dans différents modèles de *smarphones*, et reposent sur la modalité des empreintes digitales afin d'authentifier les individus.

Description des systèmes évalués

Au cours de cette évaluation, les phases de collecte des transactions d'imposture et de tentatives légitimes ont été séparées. En effet, cinq systèmes biométriques ont été testés afin de déterminer leurs performance de FAR, et trois pour le FRR. Les manipulations des données relatives à la reconnaissance biométrique sont restreintes pour tous les systèmes testés à la gestion des références (création et suppression). Ces systèmes possèdent une capacité de stockage limité à cinq références, un des systèmes étant néanmoins limité à quatre références en mémoire.

Les différents systèmes sont :

- Système A : sur ce système, le capteur d'empreinte est embarqué dans la touche d'accueil. Les dimensions du capteur sont de $10mm$ de long par $4mm$ de large, pour une surface de capture de $40mm^2$. Ce système fut utilisé afin de réaliser les expérimentations. L'intégration du capteur est visible sur la figure 3.3.

- Système B : le capteur d'empreinte digitale est placé à l'arrière du téléphone. La zone de capture est de $8mm$ de coté, pour une surface de $64mm^2$. L'intégration du capteur est visible sur la figure 3.4.
- Système C : pour ce système, le capteur est embarqué dans la touche d'alimentation latérale. Cette touche est positionnée sur le coté droit de l'appareil. Les dimensions de cette touche sont de $14mm$ par $3.5mm$, pour une surface de capture maximale de $49mm^2$. Ce système présente une différence notable par rapport aux autres systèmes testés quant à la gestion des échecs à l'acquisition. En effet, là où les autres systèmes indiquent les cas de FTA qui ne comptent alors pas comme une transaction biométrique valide, ce système ne fournit pas de retour d'information (*feedback*), et décompte les cas de FTA comme des transactions biométriques valides. L'intégration du capteur est visible sur la figure 3.5.
- Système D : le capteur biométrique est intégré dans une touche située au dos de l'appareil. Cet appareil ne fut utilisé que lors de l'évaluation visant à estimer les taux de FAR. L'intégration du capteur est visible sur la figure 3.6.
- Système E : le capteur de ce système est intégré dans une touche d'accueil placé en bas de la façade. De manière identique au précédent, cet appareil ne fut utilisé qu'au cours de l'évaluation de FAR. L'intégration du capteur est visible sur la figure 3.7.

Étant donné les différentes positions des capteurs et en s'inspirant des travaux de Sanchez-Reillo *et. al.*[87] qui définissaient différents types d'interactions possibles pour un système biométrique, une notion d'éligibilité des doigts quant aux créations de références est ajoutée. Pour chaque positionnement différents doigts sont définis comme éligibles à devenir une référence biométrique, ainsi :

- pour les capteurs situés en façade, qui sont généralement intégrés dans une touche d'accueil, les pouces et les index des deux mains sont considérés comme éligibles. En effet pour ces doigts, les interactions entre l'utilisateur et le capteur sont particulièrement ergonomiques pour la cinématique de présentation.
- Les capteurs situés au dos du *smartphone* sont généralement centrés dans la largeur de l'appareil, et se situent généralement aux deux tiers de la hauteur du téléphone. Ainsi, l'index et le majeur de chaque main sont considérés comme particulièrement adaptés pour l'enregistrement de référence et la réalisation d'authentifications biométriques.
- Les doigts éligibles, pour une position latérale droite du capteur biométrique, sont le pouce et l'index de la main droite ; et l'index et le majeur de la main gauche.

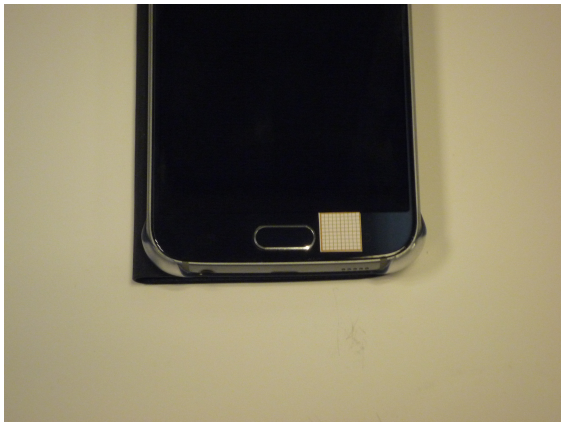

FIGURE 3.3 – Capteur du système A avec un témoin millimétré de 1 cm de coté.

FIGURE 3.4 – Capteur du système B avec un témoin millimétré de 1 cm de coté.

FIGURE 3.5 – Capteur du système C avec un témoin millimétré de 1 cm de coté.

FIGURE 3.6 – Capteur du système D avec un témoin millimétré de 1 cm de coté.

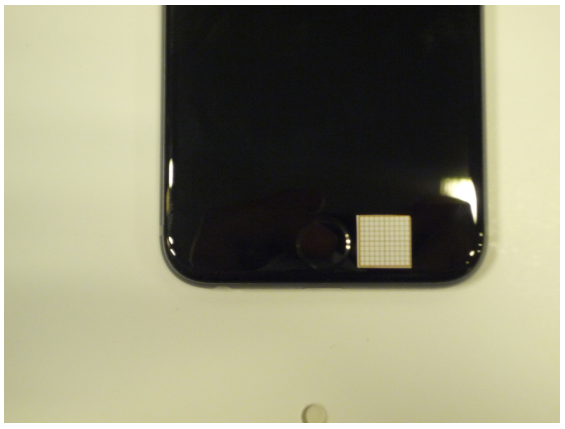

FIGURE 3.7 – Capteur du système E avec un témoin millimétré de 1 cm de coté.

Dans le cas des tentatives d'imposture, l'éligibilité des instances de modalité n'est pas pris en compte ; en effet l'objectif de ces tentatives n'est pas de tester les faux rejets, mais la capacité du système à distinguer deux instances de modalité différentes issues d'individus distincts.

3.3.1.2 Dimensionnement de l'évaluation

Le dimensionnement de la population est réalisé à l'aide du nombre de transactions nécessaires à l'évaluation. L'évaluation s'est faite sur des systèmes de reconnaissance d'empreintes digitales, en effet ceux-ci constituaient la quasi-totalité des solutions d'authentification déployées dans les *smartphones*. Le dimensionnement fût d'abord expérimenté pour le cas des empreintes digitales.

Estimation d'une borne pour le taux de FAR

Afin d'atteindre l'objectif fixé précédemment (voir 3.2.1.3), il est nécessaire de recueillir les résultats de 30 000 transactions. En raison de la collecte des transactions *online* (*cf* section 2.4.2), la taille de la population est issue d'un compromis entre la nécessité d'indépendance des résultats, et les limitations imposées par la méthodologie de collecte des transactions implémentées.

D'autre part, afin de garantir une certaine indépendance statistique des résultats, une importante répétition de transactions biométriques impliquant les même instances biométriques est à proscrire. Idéalement, afin de garantir l'hypothèse de tentatives indépendantes distribuées à l'identique, les tentatives légitimes ne peuvent être réalisées qu'une seule fois par utilisateur, et les tentatives d'imposture ne doivent pas impliquer deux fois les mêmes utilisateurs. Cette méthodologie compromet grandement la faisabilité d'une évaluation en boîte noire, en nécessitant une population de test de grande taille. Néanmoins, les travaux de Wayman [104] ont porté sur la problématique de rejeux de transactions biométriques similaires dans le cadre d'une évaluation, et concluent que la comparaison croisée des références et des échantillons permettent d'obtenir un plus grand nombre de comparaisons biométriques, et que celles-ci, malgré la dépendance introduite entre les tentatives biométriques, permettent d'atteindre des mesures d'incertitude plus réduites que la méthode visant à ne procéder qu'à des tentatives distribuées à l'identique.

La méthodologie d'évaluation en boîte noire proposée ici, repose sur une comparaison croisée ; ainsi plusieurs transactions biométriques impliqueront la même paire de référence-instance de modalité. Comme précisé précédemment, la répétition de ces transactions induit un biais dans les différents résultats collectés, mais permet d'atteindre des mesures d'incertitude réduites, ainsi que des limites de détection plus

fines en raison du nombre plus élevé de transactions biométriques. Néanmoins, un compromis doit être fait entre le nombre de transactions et la taille de la population, afin d'une part de limiter la taille de la population et d'autre part garantir l'indépendance des résultats observés sur les transactions biométriques effectuées. Dans les protocoles de capture proposés par Mansfiel *et. al.* [67] et Cappelli *et. al.* [14], les nombres d'échantillons capturés par instance de modalité sont respectivement 3 et 4. Néanmoins, dans ce cadre afin d'obtenir un nombre de transactions collectées satisfaisant, il a été choisi d'effectuer cinq transactions biométriques par instance de modalité et par référence.

Dans le cadre du protocole expérimental, les doigts éligibles sont : 1. les pouces, 2. les index, 3. les majeurs et 4. les annulaires de chaque main. Huit instances de modalité ont donc été sélectionnées pour chaque individu de la population de test.

Afin d'obtenir les 30 000 transactions d'imposture nécessaires pour atteindre l'objectif fixé, il est nécessaire de réaliser un certain nombre de tours (*cf* section 3.2.1.2) où chaque utilisateur tente de faire cinq transactions biométriques par instances. La taille de la population finale a été fixée à au moins 30, le nombre d'instance de modalité à 8 et le nombre de transaction à 5 par paire référence-instance de modalité.

Le nombre d'impostures permis par une population peut être déterminé à l'aide de l'équation 3.1, et dépend du nombre de sessions de comparaisons admissibles par cette population. L'expression du nombre de sessions de comparaison réalisables en fonction de la taille d'une population s'exprime comme 3.2, et dépend du nombre d'individus dans la population de test, et du nombre d'instances

$$Nb_{impostures} = Nb_{sessions}.Nb_{prsentations} \tag{3.1}$$

$$Nb_{sessions} = Nb_{utilisateurs}.(Nb_{utilisateurs} - 1).Nb_{instances} \tag{3.2}$$

où $Nb_{impostures}$ représente le nombre total d'impostures qui doit être supérieur ou égale au nombre d'impostures qui a été fixé dans les objectifs. Le nombre total de tentatives pour une évaluation de cette dimension est donc de 34 800. Ainsi si aucun cas d'imposture n'est observé, ceci permet de garantir avec une confiance de 95% qu'un système n'admet pas un taux de FAR supérieur à 1.10^{-4} (*cf.* "*rule of 3*" 2.6.3.2).

Estimation du FRR

Dans le cas des systèmes d'authentification commerciaux, le taux de faux rejet est généralement significativement plus haut que le taux de fausse acceptation. Ainsi

lors d'une évaluation, il est probable que des cas de faux rejet puisse être observés, l'estimation du taux d'erreur peut ainsi se faire selon une méthodologie différente de la "*rule of 3*". En effet, étant donné qu'il est possible d'observer des cas de FRR au cours de l'évaluation, l'utilisation de la *rule of 3* ne se justifie pas, il est alors nécessaire de recourir à des méthodes d'estimation par le biais d'intervalles de confiance.

Étant donné que le protocole expérimental mis en place pour la détermination des cas de FRR effectue un nombre fixe de tentatives par utilisateur, l'évaluation des faux rejets nécessite d'enregistrer une référence d'un utilisateur, et que ce dernier effectue des tentatives d'authentification contre cette même référence. Le nombre d'échec du système est alors comptabilisé afin d'estimer le taux de FRR.

Afin d'accélérer le processus d'évaluation, les références sont groupées et chaque groupe comprend des enregistrements issus de différents utilisateurs. Les systèmes sous test sont considérés comme effectuant plusieurs tentatives d'authentification à la suite, et non pas une tentative d'identification contre toutes les références en mémoire.

Ainsi, au cours de cette évaluation, 120 références ont été enregistrées dans les systèmes sous test, et pour chaque référence l'utilisateur légitime correspondant effectue 5 tentatives d'authentification.

3.3.2 Résultats

À l'issue de cette évaluation, le décompte des cas d'erreur ont permis d'estimer les taux d'erreur des systèmes correspondants.

3.3.2.1 Description de la population de test

Afin d'effectuer l'évaluation, une population de testeurs est recrutée au sein de la société Elitt. L'objectif de la constitution d'une population dans le cadre de cette évaluation est de réunir un nombre suffisant de personnes. Au regard du processus d'évaluation, une population de trente à quarante personnes semble être un compromis acceptable entre la nécessité d'un nombre suffisant de personnes et les contraintes sur l'exécution de la méthodologie d'évaluation. La description de cette population est exposée dans le tableau 3.3.

3.3.2.2 Estimation du FAR

L'évaluation de FAR s'est faite sur cinq systèmes biométriques en parallèle, au cours de celle-ci aucune imposture n'a pu être observée sur quelque système que ce soit. Sur un prévision initiale de 34 800 transactions biométriques, 32 040 ont

Population de test	Détermination du FRR	Détermination du FAR
Nombre total de participants :	33	38
Gamme d'âge :	20-60	25-60
Genres :		
Femmes	4	4
Hommes	29	34
Catégories d'âge représentées :		
20-34	9	12
35-49	17	19
50+	7	7

TABLE 3.3 – Description des populations impliquées dans les deux phases de l'évaluation biométrique

été réalisées sur chacun des systèmes testés. Le nombre de transactions capturées est en deçà du nombre initialement prévu principalement à cause de problèmes de disponibilité des participants, d'un rythme de capture plus lent qu'initialement escompté. De plus ces évaluations ont été réalisées en temps contraint, il ne fut ainsi pas possible de disposer d'un délai supplémentaire permettant de remplir l'objectif initial. Néanmoins, ce nombre est suffisant pour que l'on puisse atteindre l'objectif de l'évaluation qui était de vérifier qu'aucun système testé n'admettait de taux de FAR supérieur à 1.10^{-4}.

À l'issue de cette évaluation, il est ainsi possible d'affirmer qu'aucun des systèmes testés n'admet un taux de FAR supérieur à un pour dix mille, et ce avec une confiance de 95%.

3.3.2.3 Estimation des taux de FRR

Lors de l'estimation des taux de FRR, seulement trois systèmes sont testés (les systèmes A,B et C présentés dans la section 3.3.1.1), les résultats observés sont présentés dans le tableau 3.4. Ces trois systèmes ont été choisis suite à une décision interne de n'évaluer que des systèmes implémentant différents positionnements de capteur. Le système A présente un capteur en façade intégré dans le bouton d'accueil, tandis que pour le système B le capteur d'empreinte est situé au dos du téléphone. Le système C quant à lui, intègre son capteur dans un bouton poussoir situé sur la tranche droite du téléphone. Ainsi, cette évaluation a permis de mettre en pratique le protocole d'évaluation pour trois cinématiques de présentation différentes.

Les performances du système C sont notablement passables, en effet comme

Systèmes sous test	Taux de FRR estimé	Marges d'erreur de l'intervalle de confiance	Bornes inférieures	Bornes supérieures
Système A	11.2%	±2.9%	8.2%	14.1%
Système B	7.2%	±3.0%	4.2%	10.1%
Système C	33.7%	±6.6%	27.1%	40.3%

TABLE 3.4 – Ce tableau présente les résultats issus de l'évaluation de FRR, sous la forme de la valeur estimée de FRR pour chacun des systèmes testés, des marges d'erreur admises par l'intervalle de confiance, et par les bornes inférieure et supérieure de celui-ci.

précisé dans sa description en section 3.3.1.1, l'intégration de ce système ne fournit que peu de retour-utilisateurs. Les système A et B fournissent quant à eux, des messages de retour d'information indiquant une interaction ne permettant pas de mener à une capture d'échantillon probante, et ne décompte pas cet essai comme une transaction biométrique. En dépit de ce mode de fonctionnement différent, il fut décidé de garder un seule et unique protocole d'évaluation.

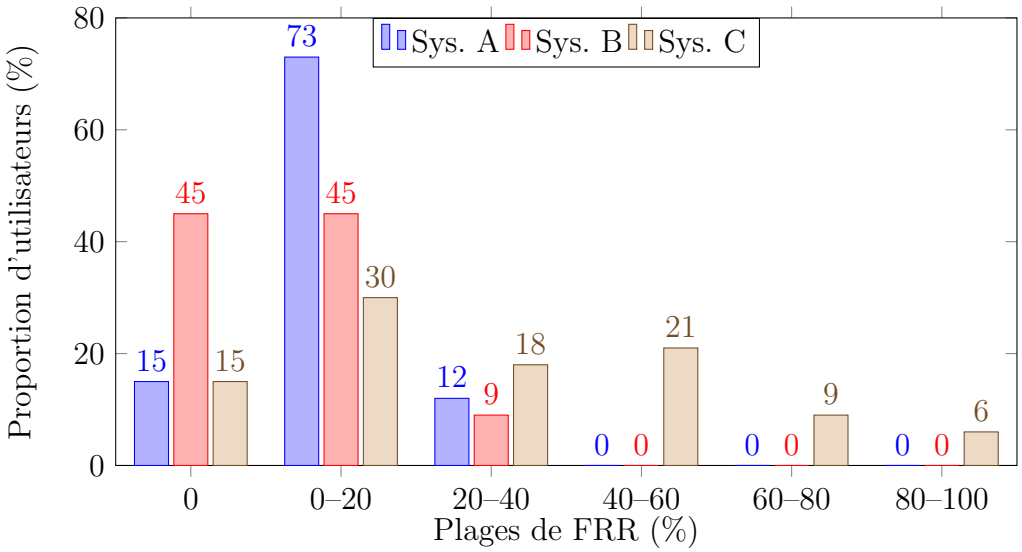

FIGURE 3.8 – Cette figure présente, pour les trois systèmes sous test, le taux de FRR personnel selon plusieurs plages.

Les figures 3.8 et 3.9 présentent les résultats observés au cours de ces évaluations. La figure 3.8 présente la proportion de la population admettant une valeur de FRR comprise dans une plage donnée. Cette figure permet de visualiser la facilité globale

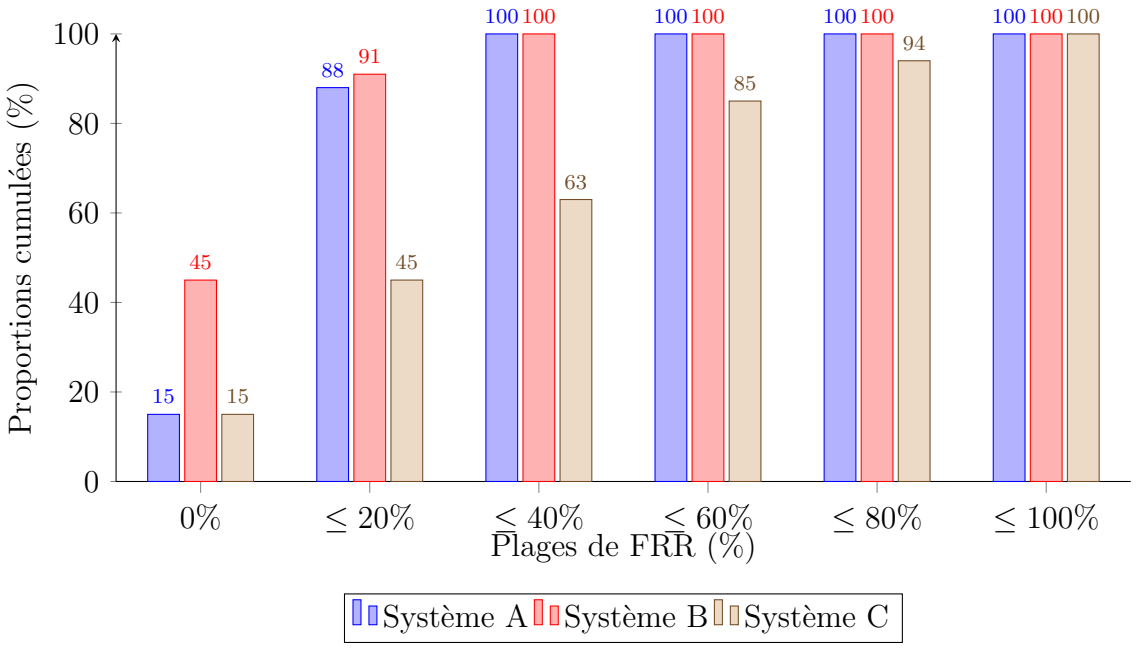

FIGURE 3.9 – Cette figure présente la proportion de population admettant un taux de FRR inférieur à certaines valeurs.

que possède une population à utiliser un système évalué. Il est ainsi possible de visualiser sur cette même figure que les systèmes biométrique A et B n'admettent aucun utilisateur ayant expérimenté une valeur de FRR personnel supérieur à 40%.

La figure 3.9 présente la proportion cumulée des utilisateurs en fonction de la valeur de FRR personnel observé. Ainsi, les systèmes A et B sont particulièrement performants puisqu'aucun utilisateur n'admet de FRR personnel supérieur à 40%. Concernant le système C, 37% de la population de testeurs a expérimenté un FRR personnel supérieur à 40%.

3.3.3 Évaluation pour un système multimodale

L'évaluation en boîte noire précédente (décrite dans la section 3.3.1) visait à déterminer les performances, en termes de FAR et de FRR, sur des systèmes biométriques embarqués dans des dispositifs portables (et plus particulièrement dans des *smartphones*). Pour ce faire une méthodologie particulière a été définie et mise en œuvre, cependant cette évaluation ne portait que sur des systèmes monomodaux utilisant tous les empreintes digitales en tant que modalités biométriques. Néanmoins, afin de permettre l'évaluation en boîte noire d'un système multimodal, des modifications

ont été réalisées sur le protocole de test et sur l'outil d'enregistrement des résultats.

3.3.3.1 Description du système sous test

Le système biométrique sous test au cours de cette évaluation, est un prototype de dispositif portable d'authentification forte du porteur, désigné sous le nom de MBAD (pour *Multi Biometric Authentication Device*, soit dispositif d'authentification biométrique multimodale). Cette authentification se fait auprès de services, ou d'autres dispositifs ; les deux facteurs impliqués dans l'authentification forte sont : d'une part la possession de ce dispositif qui doit jouer le rôle de *token* lors de l'authentification du device, et d'autre part une authentification multimodale du porteur.

L'authentification multimodale se base sur les modalités vocales et d'empreintes digitales. Le système de reconnaissance du locuteur s'appuie sur des caractéristiques cepstrales extraites du signal vocal afin de réaliser l'authentification d'un individu. Tandis que l'authentification par le biais des empreintes digitales est issue d'un module commercial, fonctionnant en boîte noire. Ce module ne propose que des fonctionnalités d'administration restreintes, telles que la gestion des références (enregistrement et suppression), et le choix d'un point de fonctionnement parmi plusieurs proposés (chaque point de fonctionnement étant défini par leur valeur de FAR allant de 1 pour dix mille à 1 pour cinquante mille). En tant que boîte noire, ce système ne permet d'accéder qu'aux décisions d'accepter ou de rejeter une transaction biométrique, les scores de comparaison ne sont donc pas accessibles.

Étant donné le comportement de ce dernier module, le procédé de fusion mis en œuvre ne peut se faire selon tous les types de fusion présentés à la section 1.3.1.1. En effet le module de reconnaissance des empreintes digitales n'autorisant pas l'accès aux scores, ou *templates* ou modèles biométriques, la fusion ne pouvait se faire qu'au niveau des échantillons ou des décisions. Cependant, la fusion des échantillons fut écarté car ce module n'était en mesure que de traiter des échantillons d'empreintes digitales. La fusion s'effectue donc au niveau des décisions.

La stratégie de fusion mise en œuvre se base donc sur la décision du système de reconnaissance d'empreinte digitale afin de moduler le seuil de décision du système de reconnaissance du locuteur.

Le système testé se présente sous la forme d'un boitier permettant de capturer les empreintes et un signal audio, à l'aide respectivement d'un capteur d'empreinte capacitif d'une part, et d'un microphone intégré d'autre part. Néanmoins, compte tenu de l'avancement du projet au moment de l'évaluation, ce boîtier nécessite d'être piloté par un logiciel externe (fonctionnant sur une machine tierce). Ce logiciel de

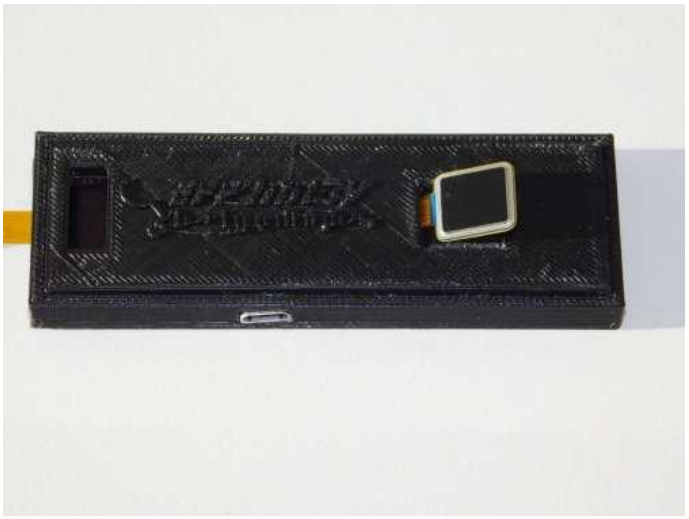

FIGURE 3.10 – Illustration du système sous test, il est possible de distinguer le capteur d'empreinte émergeant à la surface du boitier, un écran est visible sur la gauche de l'appareil. L'ouïe pour le microphone se situe au dos de l'appareil est n'est par conséquent pas visible.

pilotage offre l'opportunité de réaliser à la suite plusieurs comparaisons avec une même paire d'échantillons de voix et d'empreinte digitale. Ainsi, des rejeux de paire d'échantillons sont possibles afin de limiter le nombre de présentations effectuées par chaque participant. L'utilisation de cette stratégie permet d'accélérer l'évaluation et de diminuer l'immobilisation de la population de test.

3.3.3.2 Objectifs de l'évaluation

Les contraintes liées à l'implémentation d'une évaluation en boîte noire et la volonté de limiter la monopolisation des participants, ont mené à revoir la limite de détection du taux de FAR. En effet, mettre en œuvre une évaluation permettant d'affirmer qu'un système n'admet pas un taux de fausse acceptation supérieur à 1.10^{-4} (avec un indice de confiance de 95%), sur un tel système avec une population réduite ne parait pas raisonnable au vu de l'importante répétition de présentation d'une même modalité tout du long de l'évaluation. En effet, chaque participant ne possède qu'une seule modalité de voix.

Ainsi, le seuil de détection est donc rehaussé à une valeur de 1.10^{-3}, le nombre de transactions biométriques nécessaires à la mise ne place de cette évaluation descend donc en conséquence à trois mille.

3.3.3.3 Protocole de déroulement de l'évaluation

Le système sous test permettant de manipuler les références et les échantillons passés en entrée des algorithmes de comparaison de voix et d'empreintes digitales, le déroulement de l'évaluation est modifié afin de tirer parti de cet avantage. En effet, la manipulation des références permet de rejouer des paires d'échantillons de voix et d'empreintes contre des références issues de plusieurs participants. Ainsi, l'évaluation adopte un déroulement différent, et se divise maintenant en deux phases. Ces deux phases ont pris place dans un bureau dédié, dans lequel les conditions environnementales étaient mesurées. Les mesures concernent la température, le taux d'hygrométrie et le niveaux sonore ; en effet les deux premiers sont susceptibles d'influer sur la qualité des échantillons de dermatoglyphes, tandis que le niveau sonore est en mesure de perturber les captures de voix.

Au cours de la première phase, les références biométriques sont créées pour chaque participant à l'évaluation. La création d'une référence pour le système de reconnaissance d'empreinte utilise la fonctionnalité du module correspondant. Celle-ci nécessite plusieurs échantillons. En effet le capteur intégré au système ne fait que $8mm$ de coté pour une surface de capture de $64mm^2$. Concernant les références vocales, celles-ci nécessitent quant à elle un enregistrement d'une quinzaine de secondes, l'enregistrement est réalisé dans un environnement calme.

La seconde phase consiste en la collecte des paires d'échantillons en vue d'une confrontation entre ces dernières et les références précédemment enregistrées. Les références enregistrées sont groupées par trois, une paire d'échantillons est comparée à l'un de ces groupes. De même que pour l'évaluation monomodale en boîte noire, ce groupement de référence s'est fait dans un souci de diminuer le temps nécessaire à cette évaluation, permettant ainsi de diviser le nombre de captures nécessaires par trois. Chaque participant fait cinq transactions biométriques face à chaque groupe de références. À l'issue de ces cinq transactions, le groupe est changé jusqu'à ce que tous les groupes aient été sollicités. À l'issue des comparaisons d'imposture, des tentatives légitimes sont réalisées afin d'estimer le taux de faux rejets admis par le système testé.

L'enregistrement des résultats se fait à l'aide d'un logiciel spécifique, manipulé par le participant ; sous la supervision du responsable de l'évaluation, qui est en charge de manipuler le logiciel de pilotage du prototype, et d'observer les interactions entre le participant et le système sous test.

Évolution de l'outil d'enregistrement des résultats

L'outil d'enregistrement des résultats a été modifié afin de s'adapter aux contraintes liées à cette nouvelle évaluation. Ainsi, la base de données permettant le stockage des résultats a subi des modifications afin de permettre de lier une paire de modalités avec une référence ou un échantillon. Ainsi, une table permet de modéliser les paires de modalités utilisées en tant qu'échantillons de comparaison, et une autre permet de modéliser la création d'une référence multimodale en associant deux références monomodales.

Le logiciel d'enregistrement des résultats a subi une importante refonte quant à son principe de fonctionnement. En effet, précédemment le *software* chargé de l'enregistrement des résultats était une application de bureau embarqué dans un ordinateur de bureau, ou dans une tablette PC. Néanmoins, motivé par un besoin d'être en mesure de réaliser plusieurs évaluations en parallèle, une architecture serveur-client a été adoptée afin de permettre à plusieurs clients d'enregistrer des résultats en parallèle sur une même base de données. Au niveau du client, un portage vers la plateforme *Android* a été fait, les enregistrements prennent donc place sur des dispositifs mobiles (*smartphone* ou tablette).

3.3.3.4 Résultats

Cette évaluation peut être scindé en deux phases : l'une visant à l'estimation d'une borne supérieure à la valeur de FAR, et l'autre visant à estimer le taux de FRR observé pour le système sous test. Cette section se propose de présenter dans un premier temps l'estimation du FAR et dans un second celle du FRR.

Estimation du FAR

L'estimation du FAR repose sur la *rule of 3* présentée dans la section 2.6.3.2. Des transactions biométriques ont ainsi été collectées par le biais des tentatives d'imposture réalisées par les participants à l'évaluation. Au cours de l'évaluation, aucun cas de FAR n'a pu être observé au cours des 3 105 tentatives d'imposture effectuées.

Ainsi, si l'on se réfère à la "*rule of 3*", le système biométrique sous test peut revendiquer un taux de FAR inférieur ou égal à 1.10^{-3}, avec un indice de confiance de 95%. Cette affirmation n'est pas en contradiction avec une évaluation en boîte blanche réalisée à plus grande échelle, qui permettait au système de revendiquer un taux de FAR inférieur ou égal à 0.0015% [2]. L'évaluation en boîte blanche permet de confirmer la revendication de performances réalisées par l'évaluation en boîte noire, et cette dernière n'invalide pas les performances déterminées par la première.

Estimation du FRR

Au cours de cette évaluation, des tentatives d'authentification légitimes ont été réalisées. Les résultats collectés sur ces dernières permettent d'estimer la propension du système à admettre des cas de faux rejets. Le protocole de cette évaluation définit qu'un participant présente cinq fois ses modalités contre une référence multimodale pour la collecte des résultats.

Les résultats issus de l'évaluation en boîte noire sont présentés dans le tableau 3.5. Le taux de FRR estimé est ainsi de 45.39%, avec une marge de ±15.46% pour son intervalle de confiance.

Système sous test	Taux de FRR estimé	Marge d'erreur de l'intervalle de confiance	Borne inférieure	Borne supérieure
MBAD	45.39%	±15.46%	29.93%	60.85%

TABLE 3.5 – Ce tableau présente l'intervalle de confiance obtenu à l'issu de l'évaluation.

L'évaluation boîte blanche reportée par [2], estime un taux de FRR de 8.28% pour le même paramétrage du système. Ces deux évaluations ne permettent pas de confirmer mutuellement les revendications de performances au regard des faux rejets. En effet, l'évaluation boîte noire aurait tendance à infirmer l'attestation de performances estimée au cours de l'évaluation en boîte blanche.

Néanmoins, des différences notables quant au protocole d'évaluation sont en mesure d'expliciter une telle différence. En effet, pour le protocole de capture de l'évaluation boîte blanche, la capture des échantillons est réalisés en une seule session. Il est ainsi probable que ce protocole de capture induise un biais, en fournissant une base d'échantillons multimodaux plus favorable au système biométrique sous test. Le protocole boite noire préconise un temps de latence, un délai entre la création des références et les tentatives d'authentification légitime a été observé, la durée de ce dernier pouvant aller de quelques heures à plusieurs jours (en effet, toutes les références sont créées au début de l'évaluation, et les tentatives d'authentification sont réalisées dans les jours qui suivent). Cette différence de protocole est susceptible d'induire un biais dans l'estimation des performances de faux rejet. En raison de la difficulté de disposer de participants disponibles, aucun des deux protocoles n'a été en mesure de suivre les recommandations fournies par Mansfield *et. al.* [66] sur la réalisation de plusieurs (au moins deux) séances de capture d'échantillons biométriques séparées par une période de temps similaire pour tous les participants (de l'ordre de deux semaines).

Néanmoins, les protocoles d'évaluation ne sont pas les seules sources de biais dans ces deux évaluations. En effet, chaque évaluation a été réalisée sur deux instances différentes du MBAD, et il s'avère que le MBAD utilisé au cours de l'évaluation en boîte noire possédait un défaut qui ajoutait du bruit aux échantillons vocaux. Cet accès aux échantillons vocaux est possible étant donné que le MBAD n'est pas une véritable boîte noire, mais qu'un accès aux échantillons est possible. Ce bruit peut s'avérer en mesure de perturber le bon fonctionnement de l'algorithme de reconnaissance vocale.

3.4 Conclusion

L'évaluation d'un système en boîte noire nécessite la mise en place d'un processus spécifique afin d'obtenir des résultats, ainsi ce chapitre a présenté une telle méthodologie. La mise en œuvre de ce type d'évaluation s'avère particulièrement couteux à mettre en place, tant en termes de délais, que d'immobilisation et de sollicitation de la population de test.

La mise au point de cette méthodologie a nécessité la mise en œuvre de deux expérimentations : 1. selon un protocole nomade et 2. selon un protocole assisté. Le retour d'expérience de ces deux expérimentations a permis de mettre au point une méthodologie permettant de réaliser l'évaluation d'un système en boîte noire avec des ressources (délai, taille de la population de test ...).

Une évaluation sur un ensemble de systèmes biométriques opérationnels a pu être réalisé en suivant cette méthodologie. En raison de taux de FAR particulièrement bas, il n'était pas raisonnable (et très difficilement possible) de mettre en place une estimation de ce taux d'erreur. Cette méthodologie boîte noire propose alors comme alternative de garantir une borne supérieure à ce taux d'erreur, par le biais de l'utilisation de la *rule of 3*. Il a été ainsi possible de déterminer que les bornes supérieures au FAR pour cinq systèmes biométriques sous test étaient légèrement inférieures à un pour dix mille. Les faux rejets sont quant à eux un phénomène assez courant pour qu'une évaluation de dimension réduite soit en mesure de produire une estimation du taux d'erreur correspondant. Ainsi, la mise en pratique de cette évaluation a permis d'estimer le taux de FRR pour trois systèmes.

Une seconde évaluation en boîte noire a été réalisée sur un système biométrique multimodale, la mise en place de cette dernière a nécessité de modifier quelque peu la méthodologie d'évaluation en boîte noire.

Le protocole proposé permet d'estimer les performances d'un système biométrique fonctionnant en boîte noire. Au cours de ces expérimentations, la faisabilité de ce

protocole a été soulignée, cependant il est nécessaire de mettre en œuvre une validation de ce protocole. En effet, les résultats obtenus peuvent ne pas être conformes à des évaluations réalisées en boîte blanche, ces biais peuvent provenir des conditions de capture (présence d'un bruit, ou conditions de température et d'humidité haute), de la population de test (population non représentative, hétérogénéité/homogénéité des personnes...), et doivent être mesuré avant de mettre en œuvre ce protocole dans le cadre d'un schéma de certification. Pour ce faire, ce protocole doit être utilisé afin de réaliser l'évaluation d'un système en boîte blanche aux performances connues, afin de déterminer si le protocole présenté dans ce papier introduit un biais dans l'évaluation. Cette expérimentation permet d'étudier les éventuels biais introduits au cours de la capture des échantillons, et par la démographie de la population de test. Les éventuels biais détectés permettraient alors de définir des recommandations et des exigences quant à la mise en œuvre d'une telle évaluation et pour le recrutement de la population de test.

En plus de cette confirmation du protocole, la mise en place d'une boîte blanche parallèle permettra d'étudier la population de test, tout en déterminant et estimant l'introduction d'un éventuel biais par cette dernière. Une boîte blanche parallèle devrait se présenter sous la forme d'un ensemble de capteurs et de systèmes biométriques en boîte blanche aux performances connues, et qui permettraient de déterminer ou tout du moins d'estimer la difficulté d'une population de test. Les échantillons capturés à l'aide de cette boîte blanche pourront en effet servir à mesurer la difficulté d'une population, en mesurant "l'écart-type" pour chaque testeur dans la présentation de sa modalité, la difficulté intrinsèque de ces dernières. De même, il est possible de mesurer la difficulté que pose une population en terme d'imposture, en cherchant à produire des paire d'utilisateurs présentant des modalités suffisamment proche pour que certains systèmes constituant la boîte blanche de référence permettent d'observer des occurrences d'impostures. Les résultats obtenus par le biais de cette boîte blanche parallèle contextualisent les résultats obtenus par le biais du protocole en boîte noire.

Chapitre 4

Étude de la qualité de la voix

Ce chapitre se propose de présenter les travaux réalisés dans le domaine de l'évaluation de la qualité des données biométriques pour la reconnaissance du locuteur. Pour ce faire, des caractéristiques sont extraites d'un signal audio et ces caractéristiques sont ensuite fusionnées afin d'obtenir une valeur scalaire. Les paramètres utilisés lors de la fusion sont issus d'un processus d'apprentissage.

Sommaire

4.1 Introduction

L'être humain utilise naturellement la voix à la fois comme moyen de communication, mais aussi comme support de reconnaissance et de distinction des individus. Le modèle perceptuel humain permet ainsi de distinguer les individus dans une certaine mesure, la capacité de distinction dépendant des caractéristiques auditives et vocales des acteurs impliqués dans une reconnaissance. Néanmoins, la reconnaissance d'un individu n'est impactée que dans une certaine mesure. En effet la majorité des

affections modifiant le canal vocal ne se répercute que de manière marginale sur le caractère reconnaissable de la voix. Ainsi, une personne est, dans une certaine mesure, capable de distinguer deux individus dont la voix est proche que ce soit par proximité naturelle, ou par une modulation volontaire (imitateur, tentative d'imposture), mais aussi de reconnaître une personne en dépit d'affections ou de variations (voix enrouée, extinction de voix, bruit ambiant ...).

La voix présente les qualités nécessaires afin d'être utilisée en tant que modalité biométrique. Les systèmes biométriques s'appuyant sur cette modalité biométrique, sont désignés sous le terme de systèmes de reconnaissance du locuteur. La mise en œuvre de la reconnaissance du locuteur permet d'effectuer des reconnaissances que ce soit sous la forme d'authentification ou d'identification. À l'instar des autres modalités biométriques, le signal vocal est sensible aux conditions environnementales dans lesquelles il est émis et enregistré. En particulier le cas du bruit ambiant parasite particulièrement les signaux vocaux, compromettant ainsi la fiabilité des décisions d'un système de reconnaissance du locuteur. La détermination de la qualité des échantillons biométriques permet de déterminer dans quelle mesure un échantillon est susceptible d'être utilisable dans le cadre de la reconnaissance biométrique.

Ce chapitre s'applique à étudier une méthodologie permettant d'estimer une métrique de qualité spécifique à la voix, afin de fournir un a priori sur les performances d'un système étant donné un échantillon de voix et la mesure de qualité associée.

4.2 Reconnaissance du locuteur

La reconnaissance du locuteur est un domaine de biométrie visant à étudier et à développer des méthodes permettant de reconnaître un individu à partir de sa voix. Ainsi, les implémentations de schéma de reconnaissance du locuteur peuvent servir autant en tant que solution d'authentification que d'identification.

La voix est une modalité biométrique associant à la fois des caractéristiques physiologiques et comportementales. En effet, la voix est un phénomène sonore issu du mouvement de l'air dans le conduit vocal. Le conduit vocal est un ensemble complexe de différents organes qui permet de moduler les sons produits par ce mouvement d'air (réalisé par le biais d'une inspiration ou d'une expiration). Une grande partie du système respiratoire est impliquée dans la création du flux vocal, la partie haute de ce dernier est présentée par le biais d'une coupe sagittale visible en figure 4.1. Les principaux organes impliqués dans la formation et la modulation de la voix sont :

- la langue,

- la mâchoire et les dents,
- les lèvres,
- le voile du palais,
- le larynx et la trachée,
- les fosses nasales,
- les cordes vocales.

La combinaison des caractéristiques de ces différents organes permet de créer des signaux vocaux distinctifs pour chaque individu.

De ce signal de voix, il est possible d'extraire différents types d'informations qui peuvent être divisées en deux groupes, d'après Rosenberg *et. al.*[85] :

- Les **informations bas niveau** incluant des informations spectrales et acoustiques qui peuvent être aisément extraites d'un signal sonore par le biais de méthodes de traitement du signal.
- Les **informations haut niveau** représentant des caractéristiques qui sont difficilement extractibles par le biais de traitement du signal. Ainsi, le vocabulaire et la sémantique, la prosodie, la syntaxe ... constituent des informations haut niveau présentes dans le discours.

Les informations de bas niveau sont susceptibles d'être extraites par le biais de méthode du traitement du signal. La capture et la représentation de ces informations de bas niveau sont réalisées par des mesures de traits du spectre à court terme. Ces caractéristiques à court terme sont obtenues par l'application d'une décomposition en différentes bandes spectrales (*filter banks* ou banque de filtres), ou par l'utilisation du codage linéaire prédictif (*LPC* : *Linear Predictive Coding*).

Ces caractéristiques sont utilisées afin de constituer une représentation de la voix d'un individu. De nombreuses méthodes ont ainsi été définies, et peuvent selon Chowdhury *et. al.* [18] être divisées en deux approches différentes :

— une approche générative : cette approche crée une représentation des caractéristiques biométriques par le biais de modèles probabilistes. Ainsi, ces approches peuvent par exemple recourir aux modèles cachés de Markov (*HMM* : *Hidden Markov Models*), aux modèles de mélange de Gaussiennes (*GMM* : *Gaussian Mixture Model*) ou encore aux réseaux Bayésiens.

— une approche discriminative : cette approche inclut les techniques connexionnistes telles que la quantification vectorielle, les méthodes de partitionnement des données ou encore les machines à vecteurs de support (*SVM* : *Support Vector Machine*).

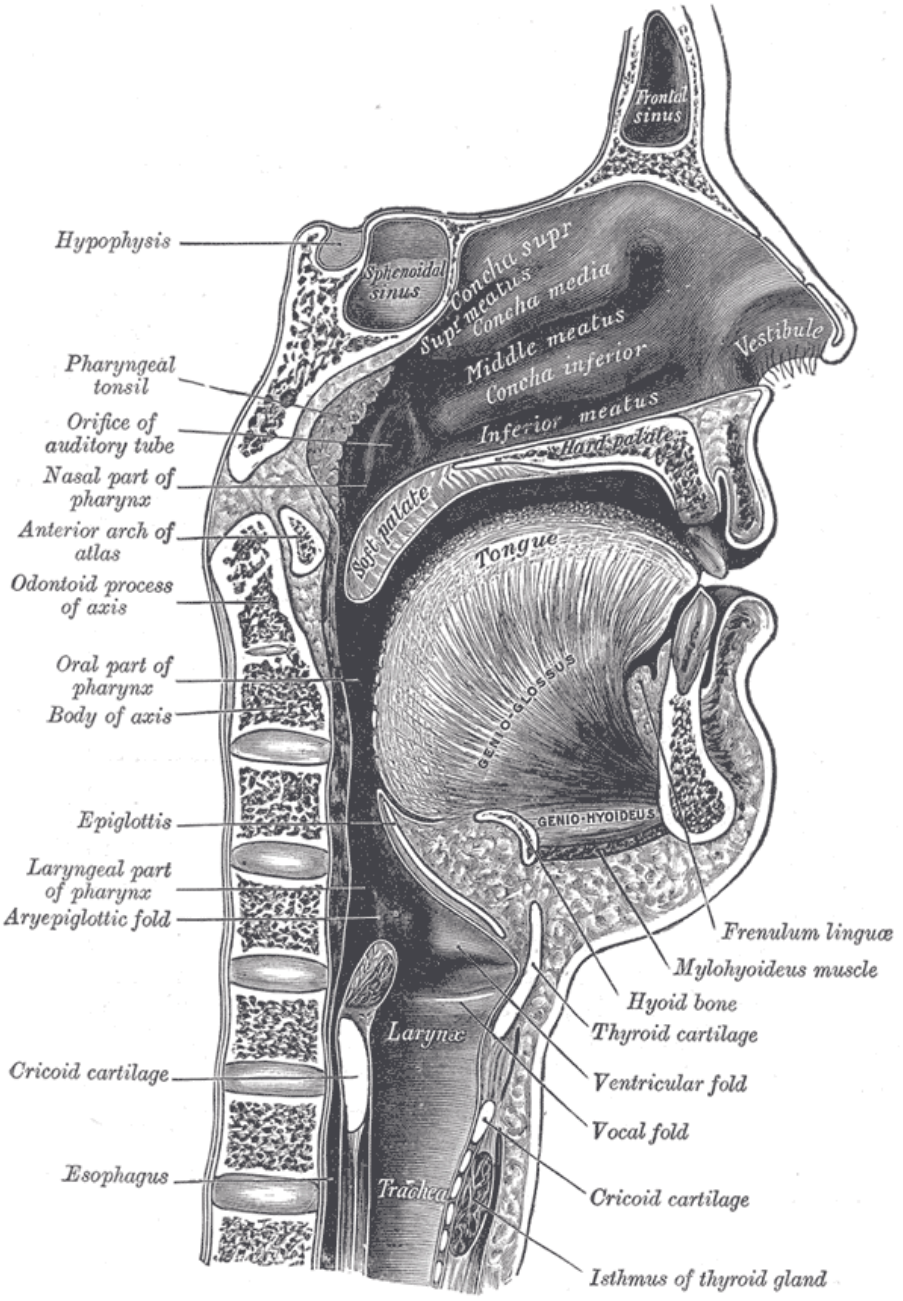

FIGURE 4.1 – Coupe sagittale mettant en évidence tous les organes et tissus autour du conduit vocal, extrait de [34].

Ces modèles utilisés pour représenter la modalité biométrique reposent généralement sur l'extraction de différentes caractéristiques biométriques.

À partir de ces caractéristiques, deux familles d'algorithmes se sont imposées :

1. les algorithmes dépendants du texte ou à texte fixe et

2. les algorithmes indépendants du texte ou à discours libre.

Les systèmes de reconnaissance du locuteur à texte fixe nécessitent qu'un individu énonce une phrase ou un mot prédéfini, afin que le processus de reconnaissance soit en mesure de s'exécuter correctement.

D'après Hébert [37], en dépit des différences fondamentales existant entre les implémentations à texte fixe et à discours libre, un recouvrement des techniques utilisées existe. La reconnaissance du locuteur dépendante du texte peut être vu comme un sous-ensemble de la reconnaissance du locuteur, en effet les techniques applicables pour la reconnaissance indépendante du texte peuvent être employées après quelques modifications. Particulièrement, les méthodes dépendantes du texte imposent que le *lexique* utilisé lors des comparaison soit un sous-ensemble du *lexique* utilisé lors de l'enregistrement.

4.3 Estimation de la qualité de la voix

L'estimation de la qualité d'un signal vocal est un sujet qui a fait l'objet de nombreux travaux de recherche. Historiquement, ces premiers travaux sur l'estimation de la qualité d'échantillons de voix ont porté sur la qualité auditive d'un échantillon de voix, en vue de sa transmission et de son enregistrement (numérisation du son, ou transmission sur le réseau téléphonique ou internet grâce à la *VOIP*). Ces mesures de qualité furent particulièrement employées afin de déterminer automatiquement la qualité d'un échantillon une fois transmis par le biais d'un réseau, ou après avoir subi une compression.

Dans le cadre de la reconnaissance du locuteur, des méthodes d'estimation de la qualité des échantillons biométriques ont aussi été développées, afin de permettre d'estimer la fiabilité d'une décision en fonction de la qualité estimée pour l'échantillon biométrique correspondant.

Cette section se propose de présenter les méthodes d'estimation de la qualité suivant respectivement ces deux objectifs.

4.3.1 Qualité auditive de la voix

L'étude de la qualité auditive d'un échantillon de voix s'est développée avec l'essor des télécommunications. Ainsi, différentes méthodologies pour l'estimation de la qualité auditives ont été développées et ce selon différentes approches :

- Approche objective : au cours de cette approche l'estimation se fait par le biais d'une méthode automatique.

- Approche subjective : dans cette approche, l'estimation de la qualité est réalisée par un opérateur humain.
- Méthode mono-extrémité ou sans référence : l'estimation de la qualité dans ces méthodes se fait sans requérir à l'échantillon de référence.
- Méthode avec référence : l'estimation de la qualité d'un échantillon se fait par le biais d'une comparaison entre ce dernier et la référence (soit le signal vocal d'origine).

Néanmoins, parmi ces approches certaines apparaissent comme plus adaptées à l'estimation de la qualité en vue de la création d'une métrique de qualité, que d'autres. Ainsi, l'approche objective de la qualité est particulièrement adaptée à la définition d'une métrique de qualité. Dans le cadre de la définition d'une métrique de qualité, des processus d'apprentissage sont susceptibles d'être utilisés afin d'établir et de maximiser une relation de corrélation entre la qualité mesurée par la méthode développée et la « qualité réelle ». Dans le cadre de la qualité auditive de la voix, les mesures de cette dernière sont issues de votes sur la qualité d'écoute d'un échantillon de voix. Ces valeurs de qualité servent ainsi de référence pour les processus d'apprentissage utilisés pour entraîner une métrique de qualité.

Les méthodes mono-extrémités se focalisent sur la détermination de la qualité d'un échantillon sans qu'aucune connaissance de la source ou de la référence ne soit nécessaire. Dans le cadre de l'estimation de la qualité, cette approche est particulièrement adaptée au cas d'une métrique de qualité biométrique. En effet, une métrique de qualité biométrique peut être utilisée afin d'indexer des échantillons issus d'une base de données biométriques, ou à la suite de leur capture.

Dans les méthodes d'estimation de la qualité auditive des échantillons de voix, la norme ITU P.563[81] correspond à ces deux critères. Cette norme fut définie par l'Union Internationale des Télécommunications afin de déterminer l'audibilité d'un signal vocal après sa transmission au travers d'un réseau de téléphonie mobile (GSM).

La métrique d'audibilité décrite dans cette norme repose sur l'extraction de plusieurs caractéristiques permettant de discriminer certaines déformations susceptibles de se produire lors de la transmission d'un signal au travers d'un réseau de téléphonie mobile. Les mesures relevant d'un même type de déformation sont regroupées au sein d'un bloc fonctionnel (voir figure 4.2). Les résultats issus de ces blocs fonctionnels sont ensuite fusionnés par le biais d'une combinaison linéaire, menant à l'obtention d'un scalaire représentant la qualité vocale globale pour l'audition.

Néanmoins, cette métrique d'audibilité est entraîné afin de prédire l'audibilité au regard du modèle perceptuel auditif humain. Ainsi, cette métrique de qualité ne

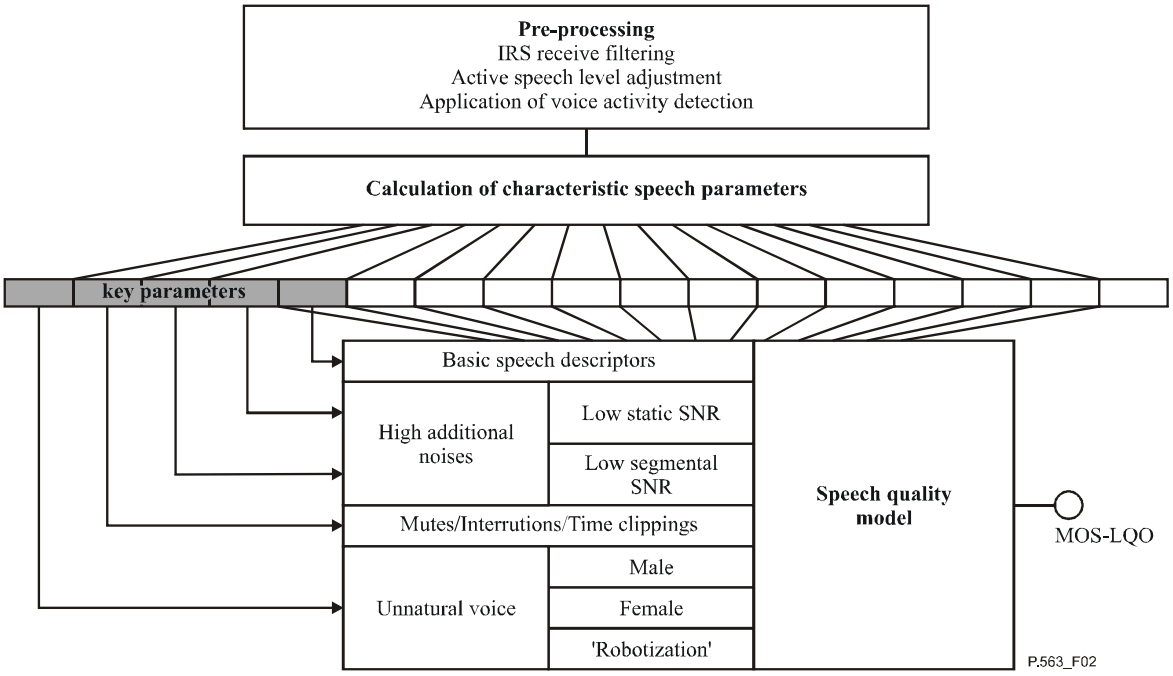

FIGURE 4.2 – Cette figure illustre le fonctionnement de la métrique de qualité proposée par la norme ITU P.563[81], telle que présentée dans cette dernière. Cette figure présente la structure de la méthode proposée par l'ITU, et particulièrement la structure en blocs. Cette méthode s'axe sur l'extraction de caractéristiques clés desquelles sont extraites des caractéristiques secondaires. Ces dernières sont ensuite comparées à un modèle afin de déterminer la qualité perceptuelle de l'échantillon de voix considéré.

peut être utilisée telle quelle afin de fournir un a priori sur les performances admises par un système biométrique pour un échantillon.

4.3.2 Qualité biométrique de la voix

La problématique de la qualité des échantillons de voix dans le cadre de la reconnaissance du locuteur est particulièrement prégnante dans la littérature. En effet, les échantillons sonores sont particulièrement sensibles aux conditions environnementales (présence de bruit additif ou convolutif) et à la qualité du signal d'origine (hauteur d'élocution, articulation ...). Ainsi, plusieurs travaux se concentrent sur l'estimation de la qualité d'un échantillon de voix afin d'estimer la fiabilité d'une décision

biométrique sur cet échantillon.

Villalba *et. al.*[103] proposent une méthode d'estimation de la fiabilité d'une décision biométrique. Cette méthode repose sur l'extraction de plusieurs caractéristiques, leur association est réalisée à l'aide d'un réseau bayésien préalablement entraîné. Parmi les caractéristiques extraites d'un signal vocal, les auteurs ont choisi d'utiliser

- des mesures de la gigue ou *jitter* et du scintillement (*shimmer*),
- une estimation du ratio signal sur bruit,
- le nombre de trames de discours détectées par un algorithme de *VAD* (*Voice Activity Detection* : Détection d'activité vocale),
- une mesure de l'entropie spectrale,
- l'index de modulation,
- la similarité d'un échantillon avec un modèle sous-jacent universel.

Le recours à un modèle sous-jacent universel (*UBM* : *Universal Background Model*) permet de représenter les caractéristiques générales d'une modalité biométrique [82], indépendamment des spécificités individuelles. Ainsi, la méthode proposée par Villalaba *et. al.*[103] requiert l'entraînement d'un *UBM* afin de représenter le domaine de la voix. Pour ce faire, un apprentissage d'un modèle de mélange de gaussiennes (*GMM*) sur des échantillons de discours a été réalisé. Les échantillons de discours utilisés pour l'entraînement sont des échantillons de voix non bruités, afin que l'*UBM* obtenu soit en mesure de représenter la distribution des caractéristiques de la parole pure. La différence des caractéristiques extraites d'un échantillon de parole bruitée avec le modèle permet d'obtenir une valeur de similarité, cette valeur est alors utilisée au travers du réseau bayésien afin de contribuer à la mesure de qualité de la parole.

La mesure de qualité obtenue est utilisée pour indiquer la fiabilité d'une décision et en conséquent de l'écarter en cas de qualité insuffisante. Le rejet de ces tentatives biométriques permet de faire baisser l'EER du système testé.

D'autre part, les travaux de Parthasarathy et Busso[76] proposent d'estimer la qualité d'un échantillon de parole par le biais du traitement des émotions. En effet, les performances d'un système biométrique sont susceptibles d'être influencées par l'état émotionnel du locuteur. Cette méthode utilise un ensemble de descripteurs de bas niveau, et particulièrement des caractéristiques dérivées de la fréquence fondamentale, des coefficients MFCC (*Mel-Frequency Cepstral Coefficient*), du taux de passage à zéro (*zero-crossing rate*) ... Cette méthode procède ensuite à une classification en trois classes à l'aide d'un réseau profond de neurones (*DNN* : *Deep Neural Network*). Les trois catégories définies classifient les différents signaux en tant que : 1. signaux fiables, 2. signaux incertains et 3. signaux non-fiables. Chacune de

ces classes permet de représenter un degré de fiabilité sur les décisions issues d'un système de reconnaissance biométrique, et par là même sur la qualité intrinsèque de l'échantillon.

4.4 Métrique de qualité pour l'estimation de la qualité d'échantillons de voix

Les systèmes de reconnaissance biométrique se basent sur les informations contenues dans les échantillons biométriques afin de réaliser l'authentification ou l'identification d'un individu. Ainsi, la qualité biométrique est particulièrement importante car représentant la quantité d'information contenue ou extractible d'un échantillon. La définition d'une métrique de qualité a ainsi nécessité de définir et d'effectuer certains choix, cette section se propose donc de présenter en détail la méthode d'estimation de la qualité proposée.

4.4.1 Objectifs de la méthode

La qualité des échantillons biométriques a pour objectif de déterminer a priori l'utilité d'un échantillon dans le processus de reconnaissance. Cette métrique a pour objectif de déterminer l'utilité d'un échantillon au regard d'un système de reconnaissance biométrique. Ainsi au cours de l'entraînement de cette dernière le retour d'information utilisé est un score de comparaison d'un échantillon biométrique avec une référence issue d'un même individu. L'apprentissage de cette métrique a pour objectif de corréler le score de comparaison, obtenu à l'issue de la confrontation des deux échantillons par un algorithme biométrique, avec la valeur calculée à partir de caractéristiques extraites du signal vocal.

L'évaluation des performances d'une métrique de qualité se fait généralement par le biais d'une base d'échantillons annotés. Cette méthode est particulièrement utilisée dans le cadre de la mise au point des métriques de qualité perceptuelle. Ainsi, le corpus d'échantillon est au préalable annoté, l'entraînement de la métrique ayant alors pour objectif de corréler cette dernière avec l'ensemble des valeurs de qualité annotées. La définition de la métrique de qualité perceptuelle PESQ[1], proposée par Rix *et. al.*[83], met en œuvre une corrélation entre la qualité mesurée et un ensemble de qualités préalablement annotées (valeur *MOS* : *Mean Opinion Score* ou Score d'Opinion Moyenne). Néanmoins de telles base annotées ne sont pas disponibles pour définir la qualité biométrique d'échantillons vocaux.

1. *Perceptual Evaluation of Speech Quality*, ou Èvaluation Perceptuelle de la Qualité Vocale)

Ainsi, une mesure alternative de la qualité intrinsèque a été utilisée avec l'utilisation des scores de comparaison biométrique, dans le cadre d'une tentative d'authentification légitime. Les scores de comparaison biométrique mesurent principalement la similarité entre deux échantillons différents, et que cette dernière est intrinsèquement liée à la quantité d'information utilisable dans le cadre d'une reconnaissance biométrique. Ainsi, la mise en œuvre de cet entraînement repose sur le postulat suivant : "un score de comparaison biométrique entre un échantillon et une référence de bonne qualité reflète la qualité de cet échantillon". La mise en œuvre de ce postulat a déjà été fait au cours de la définition d'une métrique de qualité, en particulier par Grothet et Tabassi[35].

4.4.2 Méthode proposée

L'estimation de la qualité des échantillons de voix fut étudiée, et mena à la définition d'une nouvelle méthode permettant d'estimer la qualité d'un signal de voix enregistré d'un point de vue biométrique (voir figure 4.3). Pour ce faire différentes caractéristiques sont extraites d'un signal de voix. Et afin de se conformer à la recommandation sur le facteur de forme de la métrique de qualité présenté par Grother et Tabassi [35], les caractéristiques extraites sont ensuite fusionnées par le biais d'une combinaison linéaire sous la forme d'une valeur scalaire. En effet, ces derniers définissent des recommandations sur les métriques de qualité, et sur le facteur de forme que ces dernières doivent adopter (en l'occurrence une valeur scalaire comprise entre 0 et 100). Grother et Tabbassi[35] fournissent aussi des recommandations sur la nature de la corrélation entre la métrique et la qualité des échantillons, ainsi une corrélation d'ordre étant selon eux une condition *sine qua non* à la mise en œuvre d'une métrique de qualité. En effet une relation de proportionnalité peut être difficile à obtenir. L'apprentissage des coefficients pour la métrique a pour objectif d'établir une corrélation de rang entre la métrique de qualité et d'autre part le(s) score(s) de comparaison issus de l'échantillon considéré. L'apprentissage de ces coefficients est effectué par le biais d'un algorithme génétique.

4.4.3 Extraction des caractéristiques

Afin d'estimer la qualité des échantillons de voix, la méthode proposée repose sur l'extraction de différentes caractéristiques d'un signal vocal. En fonction du facteur de forme de la caractéristique extraite sur le signal de voix, des valeurs dérivées ont été calculées. Dans le cas des caractéristiques locales, ces dernières ont été extraites sur un ensemble de fenêtres temporelles glissantes, dont la durée a été fixée à $256ms$,

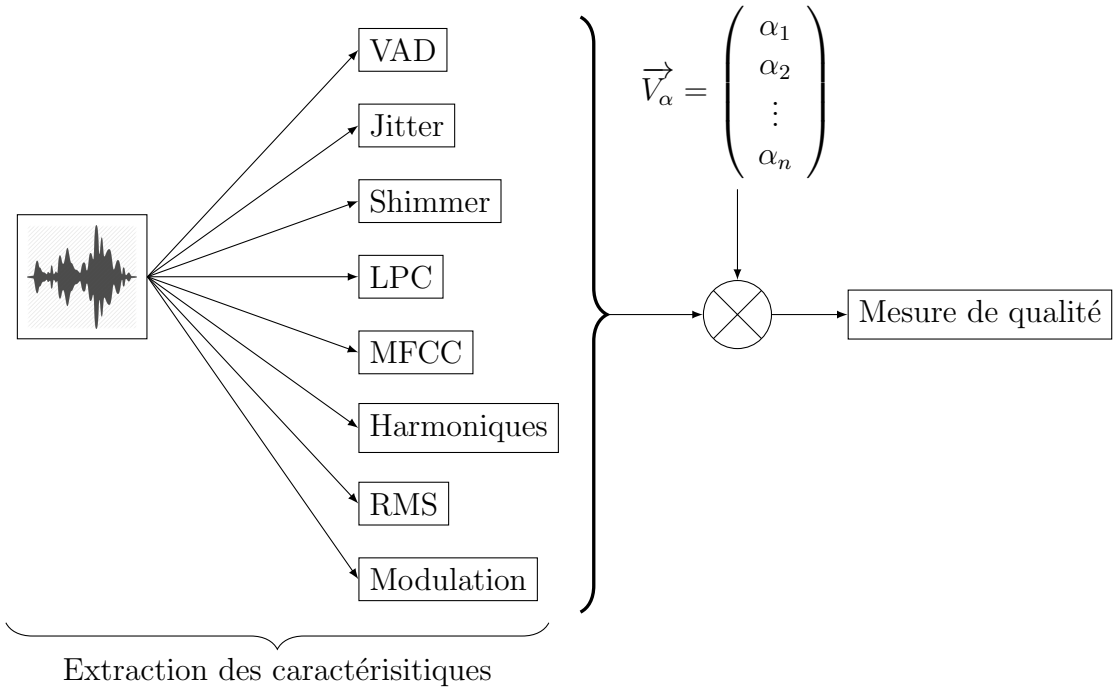

FIGURE 4.3 – Cette figure illustre le processus de mesure de la métrique de qualité. Tout d'abord des caractéristiques sont extraites du signal vocal, avant de subir une combinaison linéaire avec un vecteur de coefficients. Cette combinaison permet d'obtenir une mesure de la qualité de l'échantillon vocal considéré.

et à recouvrement bilatéral de $64ms$. L'ensemble de ces familles de caractéristiques est récapitulé au sein du tableau 4.1.

4.4.3.1 Jitter

La gigue ou *jitter* désigne dans le cadre de l'électronique la variation de la fréquence fondamentale d'un signal au cours du temps. Cette variation est généralement déterminée sur des périodes consécutives, et permet de quantifier l'effet du bruit de modulation fréquentielle. Teixeira *et. al.*[95] définissent la gigue relative comme :

$$Jitter = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}|T_i - T_{i-1}|}{\frac{1}{N}\sum_{i=1}^{N} T_i} \tag{4.1}$$

où N est le nombre de périodes dans la fenêtre considérée, T_i représente la durée de la i^{me} période.

4.4.3.2 Shimmer

Le scintillement ou *Shimmer* représente la variation de l'amplitude sur les périodes successives, et représente la modulation d'amplitude. Teixeira *et. al.*[95] définit le scintillement local comme :

$$Shimmer = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}|A_i - A_{i-1}|}{\frac{1}{N}\sum_{i=1}^{N} A_i} \tag{4.2}$$

où N est le nombre de période dans la fenêtre temporelle considérée, et A_i représente la i^{me} amplitude des pics de hauteur (*pitch*).

4.4.3.3 Asymétrie et kurtosis des coefficients LPC et MFCC

L'asymétrie et le kurtosis permettent d'étudier la répartition de coefficients, ainsi la norme ITU P.563 [81] indique que l'aplatissement (ou kurtosis) et l'asymétrie des coefficients LPC (*Linear Predictive Coding*) et MFCC (*Mel-Frequency Cepstral Coefficient*) sont particulièrement utiles pour l'estimation des propriétés des signaux vocaux.

Présentation des coefficients LPC et MFCC

Les coefficients LPC représentent des informations de l'enveloppe spectrale d'un signal. Ces informations sont susceptibles d'être utilisées comme des informations biométriques, ainsi une implémentation alliant les coefficients LPC et MFCC dans le cadre de la reconnaissance du locuteur est proposée par Subhashini et Pratap [93]. Les coefficients LPC sont particulièrement utilisés en raison de leur capacité à estimer les récurrences dans la vibration des cordes vocales. Le terme de « prédiction linéaire » des LPC (*Linear Predictive Coding*) se réfère au mécanisme de combinaison linéaire des p éléments précédents afin d'approximer, de « prédire » l'élément courant. Ainsi, un nombre restreint de coefficients LPC permet de représenter efficacement une plus longue séquence du signal d'après Deng et O'Shaughnessy [20].

Les coefficients MFCC sont une caractéristique employée dans le cadre de l'analyse vocale, étant porteurs d'informations distinctives sur la voix d'un individu. En conséquence, ces derniers sont régulièrement utilisés pour la représentation des caractéristiques biométriques de la voix. La détermination des MFCC requiert d'appliquer les étapes suivantes à un signal :

- fenêtrage du signal,
- application d'une transformée de Fourier rapide (*FFT* : *Fast Fourier Tansform*) sur chacune des fenêtres,

- une banque de filtres utilisant l'échelle des mels est appliquée sur la transformée de Fourier,
- une DCT est ensuite appliquée à la transformée filtrée.

Les coefficients MFCC sont alors l'amplitude de la transformée obtenue, néanmoins cette caractéristique est sujette à des variations si le signal est bruité par des perturbations de basse énergie.

Asymétrie et Kurtosis

La mesure de l'asymétrie et du Kurtosis permettent de représenter la distribution des coefficients précédemment décrits. D'une part, la valeur de *skewness* représente le degré d'asymétrie dans la distribution des coefficients. Ainsi, les valeurs de *skewness* peuvent être positives, négatives ou nulles. Une valeur nulle indique une distribution symétrique, tandis qu'une valeur positive (respectivement négative) indique une distribution décalée à gauche (respectivement à droite) de la médiane, et donc une queue de distribution étalée vers la droite (respectivement vers la gauche).

Le kurtosis étudie une distribution, afin de déterminer si celle-ci est semblable à une distribution normale (cas particulier d'une distribution *mesokurtique*), ou si elle relève d'une distribution *leptokurtique* ou *platikurtique*. Une distribution *leptokurtique* est marquée par des queues plus épaisses ; tandis que pour une distribution *platikurtique* celle-ci sera plus aplatie.

4.4.3.4 Détection de l'activité vocale

Les algorithmes de détection de l'activité vocale permettent de classifier une fenêtre temporelle comme comportant ou non de la parole, et/ou de déterminer la probabilité de présence de parole. L'algorithme de *VAD* choisi est une implémentation de l'algorithme proposé par Sohn *et. al.* [90]. Celui-ci définit deux classes possibles pour une fenêtre temporelle :

1. une fenêtre sans parole et
2. une fenêtre comportant de la parole.

Dans le cas d'une fenêtre sans parole, le signal qu'elle contient est considéré comme ne comportant que du bruit. Tandis que dans les fenêtres classées comme contenant de la parole, le signal est considéré comme un mélange entre du discours et du bruit ambiant. Le ratio de probabilité est extrait du signal par le biais des coefficients issus d'une transformée de Fourier discrète (*DFT* : *Discrete Fourier Transform*) et des fonctions de densité de probabilité de chacune des deux classes définies précédemment.

Sohn *et. al.* considèrent que deux fenêtres consécutives sont corrélées. Afin d'éviter une classification erronée d'une fenêtre ne comprenant que partiellement de la parole, un processus de *hang-over* est implémenté. Les auteurs ont opté pour une chaine de Markov de premier ordre, ainsi la probabilité qu'une fenêtre contienne de la parole, est intrinsèquement liée à celle de la fenêtre précédente.

4.4.3.5 Estimation de la fréquence fondamentale

Afin d'estimer la qualité d'un échantillon biométrique, la contribution de la fréquence fondamentale estimée a été étudiée. Dans l'estimation de la fréquence fondamentale sur une fenêtre temporelle, l'algorithme DYPSA (*Dynamic programming projected phase-slope algorithm*)[72] a été utilisé afin de détecter les instants de fermeture glottale à partir du signal vocal. Cet algorithme identifie les candidats pour l'instant de fermeture glottale, et sélectionne le plus probable par le biais de la programmation dynamique. L'estimation de la fréquence fondamentale est réalisée à l'aide des instants de fermeture glottale estimée sur chacune des fenêtres.

La fréquence fondamentale estimée subit ensuite un double seuillage afin de déterminer si celle-ci est cohérente avec la plage de fréquences fondamentales cohérente avec la voix humaine.

4.4.3.6 Mesure harmonique

Afin d'estimer la qualité d'un signal vocal, plusieurs mesures relatives aux harmoniques et à des caractéristiques spectrales ont été extraites.

La rugosité de la voix est extraite de son signal, cette mesure permet de caractériser la dissonance sensorielle du signal. La rugosité permet de caractériser le phénomène de battement apparaissant quand une paire de sinusoïdes sont proches en fréquence. L'estimation de la rugosité se fait par le biais de l'étude des pics présents dans le spectre obtenu par l'application d'une transformée de Fourier rapide (FFT), et en calculant la différence moyenne entre ces pics.

L'inharmonicité caractérise la non harmonicité d'un signal, c'est-à-dire dans quelle mesure les partiels de ce signal ne sont pas des harmoniques. Ainsi, tous les partiels d'un signal harmonique sont un multiple de la fréquence fondamentale de ce signal. L'inharmonicité représente donc le ratio d'énergie du signal en dehors de la série harmonique idéale. La sélection de cette caractéristique a été motivée par les travaux de Matteson et Lu[68] et de Milivojevix *et. al.*[69] qui ont établi une corrélation entre l'inharmonicité et les caractéristiques d'une voix humaine non-pathologique.

L'entropie telle que définie par Shannon [89] représente la désorganisation et l'incertitude dans une source d'information. L'entropie spectrale traite le spectre

d'un signal comme une source d'information, la sélection de cette caractéristique est motivée par son usage dans le cadre de la reconnaissance automatique du locuteur proposée par Misra *et. al.*[70]. L'implémentation choisie est l'entropie relative de Shannon, qui représente la quantité moyenne d'information dans le spectre d'un signal.

4.4.3.7 Valeur efficace

La valeur efficace (ou *RMS : Root Mean Square* soit la moyenne quadratique) permet d'estimer la puissance d'un signal sur une fenêtre temporelle. L'étude de la valeur efficace sur tout un signal permet d'estimer les variations de cette puissance au cours du temps. Cette méthode étudie en particulier l'influence du bruit ambiant sur la qualité d'un signal vocal, la valeur efficace semble donc particulièrement adaptée dans cette optique. Les *RMS* ont été utilisés comme une caractéristique biométrique dans un système de reconnaissance du locuteur par Chauhan[17]

4.4.3.8 Indice de modulation

L'indice de modulation caractérise la variation d'amplitude d'un signal au cours du temps selon Carlson[15]. Cette mesure permet de déterminer pour un signal dans quelles proportions l'amplitude varie autour du niveau non-modulé de ce dernier. L'indice de modulation est particulièrement sensible au interférences introduites par la présence d'un bruit additif. De plus, Steeneken et Houtgast[91] ont postulé un lien entre l'indice de modulation et le ratio signal sur bruit.

$$Indice = \frac{v_{max}(t) - v_{min}(t)}{v_{max}(t) + v_{min}(t)} \tag{4.3}$$

où $v_{max}(t)$ représente l'amplitude maximale admise par l'enveloppe du signal, et $v_{min}(t)$ l'amplitude minimale sur cette même enveloppe pour la fenêtre temporelle t.

4.4.4 Entrainement de la métrique de qualité

Les caractéristiques extraites (présentées dans la section 4.4.3) permettent de collecter de nombreuses informations sur le signal vocal considéré. Néanmoins, afin de créer une caractéristique biométrique se conformant aux recommandations émises par Grother et Tabassi[35] (à savoir se présentant sous la forme d'un scalaire), il est nécessaire de réaliser une fusion de ces différentes caractéristiques. Pour ce faire, la fusion de ces différentes sources d'information est réalisée par le biais d'une combinaison linéaire pondérée, prenant la forme suivante :

TABLE 4.1 – Ce tableau présente les différentes caractéristiques extraites du signal de voix afin de déterminer la métrique de qualité de la voix.

Famille de caractéristiques	Description
Jitter	Cette valeur caractérise la variation temporelle de la fréquence fondamentale
Shimmer	Cette valeur caractérise la variation périodique de l'amplitude
LPC	Coefficients de prédiction linéaire, représentant des informations de l'enveloppe spectrale d'un signal. En particulier, les valeurs d'asymétrie et de kurtosis de ces coefficients sont extraites d'une fenêtre glissante sur le signal audio.
MFCC	Les coefficients cepstraux sont extraits d'une transformée de Fourier rapide (FFT), ces coefficients sont utilisés comme caractéristiques biométriques. En particulier, l'asymétrie et le kurtosis de ces coefficients ont été extraits de fenêtres temporelles glissantes.
Détection de l'activité vocale	Cette caractéristique regroupe la décision de présence d'activité vocale sur des fenêtres temporelles, et d'autre part la probabilité qu'une fenêtre temporelle contienne un extrait de discours. Différentes caractéristiques sont dérivées de ces deux mesures.
Mesures Harmoniques	Ces mesures étudient les caractéristiques du spectre vocal, et particulièrement les fréquences harmoniques dérivées de la fréquence fondamentale. La voix étant partiellement un phénomène harmonique, l'étude de la proportion de spectre en dehors de l'inharmonicité permet de caractériser le niveau de bruit et de qualité d'un échantillon vocal.
Valeur efficace	La valeur efficace étudie la puissance d'un signal, et permet donc de caractériser sa variation au cours du temps. La valeur efficace est particulièrement affectée par la présence de bruit additif, et permet de caractériser efficacement un signal vocal bruité.
Indice de modulation	L'indice de modulation caractérise la variation d'amplitude d'un signal au cours du temps, ce dernier est particulièrement sensible à l'ajout d'un bruit additif au sein d'un signal.
Fréquence fondamentale	En tant que phénomène harmonique, la voix est caractérisée par l'existence d'une fréquence fondamentale.

$$QM = \sum_{i=1}^{N} \alpha_i . x_i \tag{4.4}$$

où N représente le nombre de caractéristiques extraites, α_i représente le $i^{ème}$ coefficient de pondération, et x_i la valeur admise par la $i^{ème}$ caractéristique extraite ; QM étant alors la métrique de qualité candidate.

Les coefficients intervenant dans cette combinaison linéaire doivent être précautionneusement choisis afin que celle-ci soit cohérente avec les propriétés nécessaires à une métrique de qualité biométrique. Une métrique de qualité devrait présenter une relation avec les performances biométriques observées, il est nécessaire d'entrainer ces coefficients afin qu'ils se conforment à une vérité terrain.

L'apprentissage de ces coefficients se présente comme un problème d'optimisation, ainsi l'objectif est de maximiser la corrélation entre la métrique de qualité candidate et la qualité biométrique perçue. La fonction déterminant la corrélation entre la métrique et la qualité de l'ensemble des échantillons d'entraînement et qui doit être minimisée, est une fonction multidimensionnelle non linéaire, et admettant plusieurs minima locaux. Ainsi, la stratégie de détermination d'une métrique de qualité ne doit pas se limiter à la détermination d'un minimum local, mais au contraire doit déterminer le minimum global permettant d'obtenir une métrique de qualité optimale pour les caractéristiques utilisées.

L'exploration par force brute de toutes les combinaisons possibles de coefficients n'est pas envisageable en pratique, principalement en raison de sa complexité directement dépendante de la taille du vecteur de coefficients. Il est donc nécessaire de recourir à une méthode d'optimisation permettant d'explorer efficacement l'espace des solutions, et robuste aux extrema locaux (et particulièrement aux minima dans ce cas précis).

Parmi les méthodes d'optimisation, l'Algorithme Génétique est une méthode de recherche stochastique basée sur la population permettant de déterminer ou tout du moins d'approximer une solution à un problème d'optimisation. La force d'une telle méthode est de travailler sur la population globale des solutions en effectuant un « échantillonnage » aléatoire, et en sélectionnant les meilleures solutions. Cette méthode permet de palier le problème des minima locaux, en ré-échantillonnant en permanence une partie de la population.

4.4.4.1 Algorithme génétique

Les algorithmes génétiques sont une méthode de résolution de problèmes d'optimisation issue de la famille des algorithmes évolutionnistes, qui ont été introduits

par John Holland[40] et qui sont inspirés de la théorie de l'évolution. En effet un algorithme génétique procède par itération ou génération sur une population de solutions candidates, ainsi d'une génération à une autre des opérateurs sont appliqués afin de s'approcher de l'optimum. Dans le cas de l'entrainement de cette métrique de qualité, chaque élément de la population peut être représenté comme un vecteur dont la dimension est égale au nombre de coefficients à déterminer. Par la suite, le terme de composante se réfère à la composante d'un vecteur candidat.

Ainsi, Mitchell[71] définit trois opérateurs qui sont appliqués à chaque génération :

- **Sélection** : cet opérateur se charge de déterminer quels sont les candidats les plus prometteurs, ces candidats sont ainsi conservés pour la génération suivante. Le critère de sélection est la minimisation d'une fonction d'évaluation qui détermine si le candidat est proche de l'optimum ou non.
- **Croisement** : cet opérateur réalise un croisement entre deux candidats sélectionnés afin d'obtenir deux nouveaux éléments pour la population de la prochaine génération. Le croisement opère une sélection aléatoire de composantes des deux candidats,
- **Mutation** : cet opérateur produit un nouveau candidat en changeant une des composantes d'un vecteur candidat sélectionné.

Ces opérateurs sont utilisé pour créer une portion (généralement spécifiée par l'utilisateur) de la prochaine génération. Le reste de la population est générée de manière aléatoire.

Les algorithmes génétiques disposent de différentes conditions d'arrêt, ainsi cet algorithme s'arrête si un des vecteurs sélectionné atteint la valeur optimale. Un candidat est déterminé comme optimal, lorsqu'il minimise la fonction d'évaluation (traditionnellement avec une valeur de zéro). Néanmoins, la génération aléatoire et les opérateurs peuvent ne pas être en mesure de déterminer une solution optimale ou encore celle-ci peut ne pas être atteinte, les algorithmes génétiques possèdent alors des conditions d'arrêt alternatives. Ainsi, lorsque les itérations successives de l'algorithme ne permettent pas d'obtenir de progression significative, l'algorithme s'arrête suite à une stagnation. Une autre méthode d'arrêt est de fournir un nombre limite d'itération que le système effectue avant de s'arrêter. À l'arrêt de l'algorithme, celui-ci renvoie le meilleur candidat sélectionné au cours de son exécution.

Fonction d'évaluation

La détermination des coefficients peut être vu comme un problème d'optimisation, dans lequel on essaie de maximiser la corrélation entre la métrique de qualité candidate QM (4.4), et une performance biométrique. La performance biométrique est dans

cette méthode le score de comparaison d'un système de reconnaissance du locuteur prenant en entrées : a) un échantillon de référence qui est un échantillon non bruité et b) un échantillon de comparaison auquel du bruit a pu être ajouté. Les transactions biométriques impliquées dans la mesure des performances biométriques sont des tentatives légitimes, c'est-à-dire que les échantillons impliqués dans chacune des transactions proviennent d'un même individu. L'ensemble des valeurs des coefficients de pondération est alors utilisé comme données d'entrée pour une fonction d'évaluation (ou fonction de *fitness*), qui doit être alors minimisée afin de déterminer si celle-ci est proche ou non de l'optimum.

Afin de mettre en œuvre un entrainement de ces coefficients, une fonction d'évaluation a été implémentée. Une fonction d'évaluation retourne une valeur que l'algorithme génétique tente de minimiser, et l'objectif de cet entrainement est au contraire de maximiser la corrélation entre les valeurs obtenues par la combinaison linéaire des caractéristiques extraites, et les performances observées sur un système biométrique de référence. À partir du vecteur de coefficients, cette fonction calcule une valeur pour la métrique candidate QM (4.4) et détermine ensuite la corrélation entre les deux suites de valeurs :

1. les différentes valeurs de la métrique candidate et
2. les scores de comparaison correspondant au même échantillon.

Ainsi, un produit scalaire est réalisé entre un vecteur contenant les coefficients de pondération de la métrique candidate courante, et le vecteur de caractéristiques de chaque échantillon. L'ensemble des scalaires ainsi obtenus et l'ensemble des performances biométriques correspondantes sont utilisés comme variables pour le coefficient de corrélation. Néanmoins, ce dernier n'est pas directement utilisé comme retour de la fonction d'évaluation, puisque cette dernière se doit d'être minimisée par l'algorithme génétique. Dans le cadre d'une métrique de qualité, l'on souhaite observer une corrélation positive entre la qualité et les performances escomptées. La valeur de retour de la fonction d'évaluation est donc $1 - \rho_s$ où ρ_s représente la valeur du coefficient de corrélation de Spearman.

Coefficient de corrélation de Spearman

L'entrainement des coefficients de pondération intervenant dans la combinaison linéaire s'est fait en cherchant à établir une relation entre les performances biométriques observées (scores de comparaison d'un algorithme de reconnaissance du locuteur), et les valeurs admises par une métrique de qualité. L'étude de la relation entre ces deux variables s'est faite par le biais d'un coefficient de corrélation.

La corrélation parfaite de deux variables aléatoires implique que les variations de ces deux variables soient toujours de même sens, ou au contraire pour une corrélation négative toujours de sens inverse. Les coefficients de corrélation sont compris entre -1 et 1, une valeur de 1 dénotant une corrélation positive parfaite et une valeur de -1 dénotant au contraire une corrélation négative parfaite. Une valeur de 0 dénotant ainsi une absence de corrélation, ne permet d'établir de relation entre les deux variables. Plusieurs coefficients de corrélation ont été définis dans le cadre des statistiques. Dans le cadre de la définition de cette métrique de qualité, les coefficients de Spearman, Pearson et Kendall ont été testés au cours de la définition de cette métrique de qualité.

Le coefficient de Spearman fut choisi pour déterminer la corrélation entre la métrique de qualité candidate et les scores de comparaison légitime. Le coefficient de Spearman est défini par Kendall[50], comme une corrélation de rang ; il ne s'agit ainsi pas d'étudier une relation de type affine entre les deux variables (à savoir la valeur de la métrique, et les performances biométriques), mais de déterminer dans quelle mesure leur relation peut se décrire comme une fonction monotone. Ce coefficient est défini par :

$$\rho_s = 1 - \frac{6}{n(n^2-1)} \sum_{i=1}^{N} \Big(rg(X_i) - rg(Y_i) \Big)^2 \tag{4.5}$$

où n est le nombre de valeurs pour les variables X et Y, et où $rg(X_i)$ et $rg(Y_i)$ représentent respectivement les rangs des valeur X_i et Y_i, $rg(X_i) - rg(Y_i)$ étant donc la différence de rang entre les deux valeurs observées X_i et Y_i. L'utilisation de ce coefficient de corrélation requiert néanmoins que les n rangs soient distincts, il n'est ainsi pas possible d'utiliser cette expression du coefficient de corrélation si il existe des doublons dans au moins une des deux séries associées aux variables X et Y.

Une autre expression de ce coefficient de corrélation, n'imposant pas de restriction sur la distinction des rangs, se présente sous la forme :

$$\rho_s = \frac{cov(rg(X_i), rg(Y_i))}{\sigma_{rg(X_i)} . \sigma_{rg(Y_i)}} \tag{4.6}$$

où $cov(rg(X_i), rg(Y_i))$ désigne la covariance des rangs des valeurs X_i et Y_i, et $\sigma_{rg(X_i)}$, $\sigma_{rg(Y_i)}$ sont les écarts types observés pour les rangs de ces mêmes valeurs.

4.4.5 Base d'apprentissage

L'entrainement des coefficients impliqués dans la combinaison linéaire nécessite des données d'entrainement. En conséquence, un corpus d'échantillons de voix a été

utilisé, ceux-ci sont des enregistrements de parole en anglais dans un environnement sonore dépourvu de bruit ambiant. Afin d'entrainer les coefficients de la métrique de qualité, il est nécessaire de disposer d'échantillons bruités, et de savoir dans quelle mesure ces derniers sont perturbés par ce bruit additif.

Ainsi, l'ajout de bruit s'est fait de manière à avoir pour chaque échantillon un ratio signal sur bruit moyen connu (SNR). Les valeurs de SNR choisies sont : 1. $6dB$, 2. $3dB$, 3. $0dB$ et 4. $-3dB$. Les échantillons sonores utilisés comme bruit additif sont issus de différentes sources afin de modéliser différents environnements. Ainsi, des échantillons enregistrés dans des rues afin de capturer le son de la circulation urbaine, dans des restaurants ou lieux très fréquentés afin d'utiliser des sons de discussions, ou encore un enregistrement d'un journal télévisé ont été utilisés comme base de bruit ambiant. Le choix de ces enregistrements a été motivé par la grande diversité d'environnement dans lesquels un système biométrique (en particulier ceux à usage personnel) est susceptible d'être utilisé.

4.4.6 Résultats préliminaires

La mise en œuvre de l'entrainement de la métrique a permis de déterminer des valeurs pour les coefficients impliqués dans la combinaison linéaire. Cet entrainement a pour objectif de maximiser la corrélation entre la valeur de la métrique et les performances biométriques observées.

Afin de déterminer quelles sont les familles de caractéristiques (présentées dans la section 4.4.3) les plus pertinentes, chacune de ces familles a été simplement corrélée avec les performances biométriques correspondantes. Les différentes familles de caractéristiques définies sont :

— les caractéristiques dérivant de la VAD,

— les mesures effectuées sur le scintillement et la gigue (respectivement le *shimmer* et le *jitter*),

— celles extraites des mesures des coefficients LPC,

— les valeurs associées aux coefficients cepstraux (MFCC),

— les valeurs dérivées des mesures harmoniques (inharmonicité et rugosité),

— la famille correspondant aux valeurs extraites sur la valeur efficace (ou RMS),

— l'ensemble de l'étude de la fréquence fondamentale sur un signal audio,

— les informations relevant de l'entropie spectrale,

— celles représentant les données issues de l'indice de modulation en amplitude.

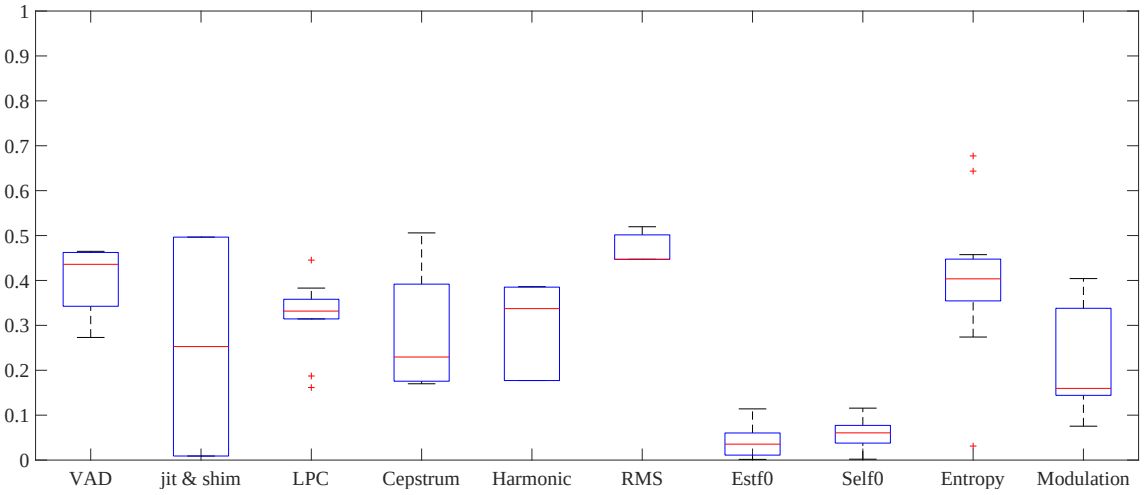

FIGURE 4.4 – Cette figure étudie la relation qu'entretient chacune des famille de caractéristique avec les performances biométriques. Ainsi, pour chacune des familles de caractéristiques, la distribution des coefficients de corrélation est présentée ; afin de faciliter la compréhension, la valeur absolue des coefficients de corrélation est utilisée.

La distribution des coefficients de corrélation associés à chaque famille de caractéristiques est présentée dans la figure 4.4. Ainsi selon cette figure, deux caractéristiques extraites de l'entropie spectrale sont en mesure d'atteindre un niveau de corrélation de 0.6, la meilleure atteignant un degré de corrélation de 0.67. Cette figure illustre que la valeur efficace (RMS) et la mesure de la détection de l'activité vocale (VAD) sont porteurs d'informations en corrélation avec les performances biométriques observées.

En revanche, les mesures portant sur l'estimation de la fréquence fondamentale ne montre qu'une faible corrélation avec les performances biométriques. En effet les différents coefficients de corrélation des familles correspondant à l'estimation de la fréquence fondamentale et aux fréquences sélectionnées ne sont que très peu corrélés avec les performances biométriques (inférieur à 10% de corrélation). Leur contribution apparait comme insuffisante pour les considérer comme pertinente.

À l'issue de cette étape d'entrainement, la meilleure métrique de qualité obtenue est en mesure d'atteindre un degré de corrélation entre la combinaison linéaire et les performances biométriques observées de 83%. Ce degré de corrélation dénote des relations intéressantes entre les caractéristiques sélectionnées et la qualité des échantillons biométriques (*i.e.* les scores de comparaison entre l'échantillon considéré et la référence d'un même utilisateur).

4.5 Obtention des résultats expérimentaux

Suite aux résultats préliminaires encourageants, une recherche plus exhaustive est menée afin de déterminer une métrique de qualité plus fortement corrélée avec les performances de comparaison. La détermination des coefficients de la combinaison linéaire, se faisant par le bais d'un algorithme génétique, est considérée comme un processus aléatoire. Des techniques statistiques ont ainsi été mises en œuvre afin de déterminer un candidat intéressant pour l'estimation de la qualité d'un échantillon biométrique.

La détermination d'une meilleure métrique de qualité candidate s'est faite en exploitant la théorie des grands nombres. La mise en œuvre de ces techniques nécessite de réaliser un grand nombre d'essai, ainsi un processus de *bootstrap* est utilisé.

4.5.1 Lois des grands nombres

Les lois des grands nombres sont issues du domaine des probabilités, en effet celles-ci découlent des observations empiriques sur les résultats d'un processus ou d'une action aléatoire. Les lois des grands nombres formalisent des intuitions quant aux résultats et tendances que l'on peut observer au cours d'expériences mettant à profit les probabilités.

Une exemple simple permettant d'illustrer la loi des grands nombres, est le lancer d'une pièce de monnaie. En effet, lorsqu'un grand nombre de lancers sont effectués avec une pièce de monnaie (non truquée), le nombre d'observations de cas "pile" et "face" devraient s'être stabilisé de manière à ce que ces deux événements apparaissent comme équiprobables.

Deux lois différentes ont été formulées afin de formaliser cette constatation empirique, ces dernières expriment des conditions sur la suite de variables aléatoires. Ces deux lois expriment que la moyenne d'une suite de variables aléatoires converge vers l'espérance de cette variable aléatoire. Ainsi, ces deux lois sont exprimés, dans *Probabilités pour les non-probabiliste* de W. Appel[4], comme :

- **la loi *faible* des grands nombres :** Soit $(\Omega, \mathfrak{T}, \mathbf{P})$, un espace probabilisé. Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires indépendantes, de même espérance m et de même variance σ^2. Alors la suite de terme général :

$$Y_n = \frac{X_1 + X_2 + \cdots + X_n}{n} \tag{4.7}$$

converge en probabilité vers (la variable aléatoire constante égale à) m :

$$\forall \epsilon > 0, \qquad \lim_{n\to\infty} P\big\{|Y_n - m| \geqslant \epsilon\big\} = 0 \tag{4.8}$$

- **la loi *forte* des grands nombres de Kolmogorov :** Soit X une variable aléatoire. Soit $(X_n)_{n\geqslant 1}$ une suite de variables aléatoires, indépendantes et de même loi que X. Enfin, soit m un réel. Alors la suite de terme général

$$Y_n = \frac{X_1 + X_2 + \cdots + X_n}{n} \tag{4.9}$$

  converge presque sûrement vers m si et seulement si X est intégrable et $E(X) = m$.

Ces deux lois formulent que la moyenne d'une variable aléatoire permet d'obtenir une estimation de l'espérance de cette variable aléatoire.

La détermination du vecteur de coefficients par le biais d'un algorithme génétique apparait comme un processus aléatoire. Au cours de différentes sessions d'entraînement, des convergences vers différents vecteurs finaux ont été observées au terme de l'exécution de l'algorithme biométrique, postulat a été fait que ces vecteurs sont susceptibles de tendre vers un vecteur de coefficients offrant une meilleure corrélation entre la qualité estimée et réelle des échantillons vocaux. Ainsi, si l'on considère que l'obtention d'un vecteur de coefficient est un processus aléatoire et qu'il admet une espérance finie, l'utilisation des lois des grands nombres permet d'estimer ce vecteur en répétant un grand nombre de fois cette expérimentation. Néanmoins, l'application des lois des grands nombres nécessite de disposer d'un grand nombre de résultats et donc d'expérimentations. L'application de ces lois nécessite donc la mise en œuvre de méthodes statistiques permettant de multiplier artificiellement le nombre d'exécutions de l'algorithme génétique.

4.5.2 Méthode de Bootstrap

La mise en œuvre de méthodes mettant à profit les lois des grands nombres nécessitent de disposer d'un grand nombre de résultats expérimentaux. Néanmoins devant la difficulté de disposer d'une base de données biométrique conséquente, le recours à des méthodes de rééchantillonage peut s'avérer nécessaire. De telles méthodes ont été introduites au cours des années cinquante avec la méthode du "*jackknife*" par Quenouille[79] et par Tukey[100]. Le "*jackknife*" consiste à créer plusieurs sous-ensembles de données à partir d'un seul jeu d'échantillons. Particulièrement, cette méthode crée autant de sous-ensembles qu'il y a d'échantillons dans l'ensemble d'origine, en excluant un échantillon différent pour chaque sous-ensemble.

Au cours des années dix-neuf cent soixante-dix, une évolution de ces méthodes a été proposée par Efron[24]. Le rééchantillonnage se fait en réalisant des tirages avec remise au sein de l'ensemble d'échantillons original, cette méthode présente

l'avantage de permettre la création d'un plus grand nombre de sous-ensembles. La méthode de "*bootstrap*" commence par définir une distribution de probabilité des échantillons $\hat{F}$, qui assigne un poids de $\frac{1}{n}$ à chacune des observations $(x_1, x_2, \ldots, x_n)$. La procédure définit un nouvel ensemble d'échantillons de taille n à partir de la distribution $\hat{F}$, comme :

$$X_i^* = x_i^*, \quad X_i^* \sim_{ind} \hat{F} \qquad i = 1, 2, \ldots, n \tag{4.10}$$

L'obtention de ces nouveaux ensembles d'échantillons permet de multiplier le nombre d'exécutions de l'algorithme génétique proposé, permettant ainsi d'augmenter le nombre de résultats obtenus.

4.5.3 Résultats expérimentaux

À l'issue du processus de *bootstrap* sur l'exécution de l'algorithme génétique, un nombre important de vecteurs a ainsi été obtenu. Pour chacun de ces vecteurs, un coefficient de corrélation a été calculé sur un ensemble d'échantillons de confirmation se composant de plusieurs échantillons de voix issus de plusieurs individus et ayant subi plusieurs altérations par ajout de bruit ambiant. La base de confirmation est ainsi composée de manière similaire à la base d'apprentissage, mais les échantillons vocaux et de bruit utilisés ne sont pas ceux utilisés pour cette première. La distribution des scores est ainsi visible sur la figure 4.5.

Le coefficient de corrélation maximale observé est de $83,83\%$ sur la base de confirmation, le degré de corrélation moyen est de $74,16\%$. La métrique candidate la plus faible n'est corrélée qu'à 65% avec la qualité observée. Le vecteur issu de la moyenne des vecteurs issus de ces différents apprentissages parvient à obtenir un score de corrélation de $78,42\%$.

L'entraînement de cette métrique s'est fait sur un ensemble non sélectionné de caractéristiques. De nombreuses caractéristiques sont extraites des différentes familles définies (voir la section 4.4.3 pour la présentation de ces dernières). Ainsi afin de déterminer lesquelles contribuent le plus à la corrélation de cette métrique avec la qualité biométrique observée, un processus de sélection par vote majoritaire a été mis en place afin de sélectionner les dix caractéristiques les plus prédominantes.

Ainsi, sur les données obtenues au cours du processus de bootstrap, un processus de vote majoritaire a été mis en place afin de déterminer quels sont les caractéristiques apportant le plus d'information dans le processus de corrélation. Ainsi, la sélection par ce biais des dix meilleurs pourcents des caractéristiques a permis de déterminer les performances de la métrique de qualité à partir des vecteurs précédemment entraînés. Ainsi, une corrélation maximale de $70,99\%$ a pu être observée, la corrélation maximale

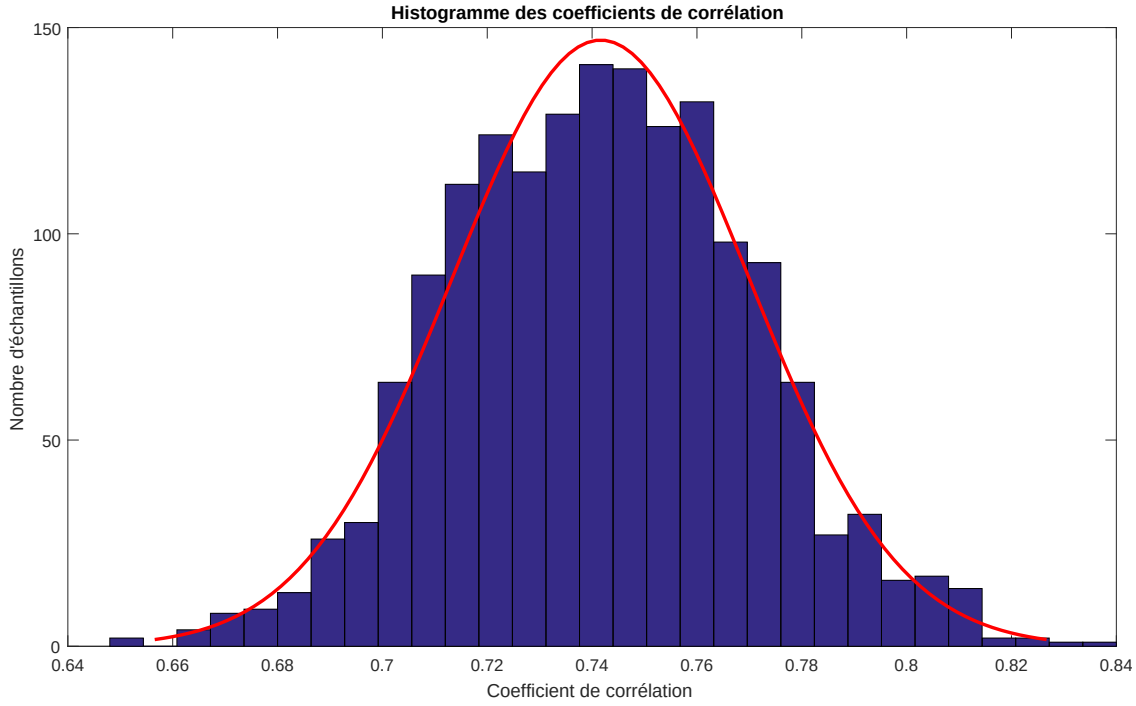

FIGURE 4.5 – Cette figure présente l'histogramme des coefficients de corrélation obtenus à l'issue du processus de *bootstrap*. La distribution de ces scores de corrélation s'approche d'une gaussienne, soit une distribution normale, visible en rouge sur la figure.

est de $76,81\%$. Le vecteur issu de la moyenne des vecteurs issus de ces différents apprentissages parvient à obtenir un score de corrélation de $73,09\%$. La métrique candidate la plus faiblement corrélée n'obtient un score de corrélation que de 55%.

4.6 Conclusion

Dans le cadre de l'évaluation des systèmes biométriques, la qualité des données biométriques permet de donner un a priori sur les performances observables. Ainsi, le contrôle de la qualité des échantillons est utilisé au cours de plusieurs processus biométriques, et particulièrement au cours de la création d'une référence biométrique. En effet, une référence ou un échantillon de qualité insuffisante sont susceptibles de ne pas contenir suffisamment d'informations biométriques, ou que ces dernières ne soient pas aisément extractibles ou sont masquées par diverses perturbations.

La voix est une modalité dont le mode de capture est particulièrement susceptible d'être sujet à des perturbations extérieures, en particulier le bruit ambiant s'additionne aisément au signal de voix capturé. La détermination de la qualité biométrique d'un

échantillon est ainsi particulièrement importante. Ce chapitre propose donc une étude de cette problématique, et a mené à l'obtention d'une métrique candidate montrant une corrélation significative avec les performances biométriques. En effet, à l'issue de l'entrainement une corrélation de 83% a pu être obtenue ; néanmoins, d'autres caractéristiques et méthodes de combinaison de ces dernières devraient être expérimentées afin d'augmenter encore le degré de corrélation entre la métrique de qualité candidate et les performances biométriques observées.

Néanmoins, l'indice de qualité proposé ne peut être utilisé en l'état, et constitué une étude préliminaire permettant d'étudier différentes familles de caractéristiques. Une sélection des caractéristiques les plus pertinentes, par le biais d'un vote majoritaire sur l'ensemble des indices candidats, doit permettre de réduire le nombre de caractéristiques extraites, tout en garantissant la sélection des caractéristiques les plus pertinentes.

La combinaison des différentes informations, par le biais de l'extraction de caractéristique, a été réalisée à l'aide d'une combinaison linéaire. Le choix de cette dernière a été motivé par la simplicité d'utilisation, et le nombre limité de coefficients de pondération à entraîner. En effet, les performances obtenues avec cette méthode ont permis d'attester d'une relation statistique entre les différentes caractéristiques extraites et la qualité observée. L'utilisation de cette méthode de combinaison a permis de déterminer la pertinence des différentes familles de caractéristiques. Néanmoins, d'autres méthodes de combinaison doivent être envisagées afin d'établir plus finement et moins naïvement une relation statistique entre les caractéristiques extraites et la qualité observée.

Conclusion

En raison des progrès technologiques de ces dernières décennies (tant sur la miniaturisation des différents composants que sur l'augmentation de la puissance de calcul des processeurs et microcontrolleurs), la reconnaissance biométrique a atteint une maturité suffisante afin d'être intégrée dans le marché grand public. Ainsi, la reconnaissance biométrique est devenue une méthode d'authentification à usage personnel sur des dispositifs portables (*e.g. smartphones*), s'intégrant aussi afin de réaliser des contrôles d'accès logiques et physiques au sein d'entreprises, ou encore dans le cadre régalien permettant d'automatiser le contrôle aux frontières. Certains *smartphones* et cartes de paiement proposent d'ores et déjà le remplacement du *PIN* par une authentification biométrique. D'autre part, des projets de recensements biométriques de la population ont vu le jour ou sont encore en gestation, le plus notable étant le projet Aadaar en Inde. Néanmoins, la biométrie n'est pas une méthode de reconnaissance des individus infaillible, celle-ci est en effet susceptible d'admettre des erreurs. En fonction du type d'erreur, de l'usage et de l'environnement dans lequel s'intègre un système biométrique, les conséquences d'une erreur peuvent être plus ou moins importantes. Ainsi, l'intégration d'un système biométrique pour un usage est conditionné par sa bonne conformité aux exigences de l'usage.

La détermination de l'adéquation d'un système avec un usage nécessite de connaître ses limites, principalement en termes de taux d'erreurs, de performances. Ces estimations de performances sont réalisées par le biais d'un processus d'évaluation, le domaine de l'évaluation biométrique a été présenté dans le deuxième chapitre de cette thèse. Cette thèse s'inscrit dans le domaine de l'évaluation des systèmes biométriques, et s'est structurée autour de deux axes principaux. Ainsi, face au manque de méthodologies pour l'évaluation de systèmes biométriques en boîte noire, le premier axe s'est défini autour de cette problématique.

Bilan

Les travaux réalisés au cours de cette thèse se sont structurés autour de deux axes, avec d'une part l'évaluation des systèmes biométriques en boîte noire, et d'autre part la détermination de la qualité des échantillons de voix.

Évaluation des boîtes noires biométriques

Au cours de cette thèse, le premier axe abordé fut l'évaluation des boîtes noires biométriques, en effet les systèmes sous-test ne permettent pas d'accéder aux données internes du système et imposent ainsi la mise en place d'une méthodologie spécifique. Ainsi, une méthodologie d'évaluation d'une boîte noire biométrique a été définie au cours de la réalisation de deux expérimentations et de deux évaluations.

Les deux premières expérimentations ont permis de définir et mettre en place une méthodologie d'évaluation en termes de déroulement, et de population. Ces deux expérimentations suivent deux méthodologies différentes : 1. un protocole nomade et 2. un protocole assisté. En particulier, cette dernière nécessitait l'implémentation et l'utilisation d'un outil de test dédié permettant d'enregistrer les résultats de chacune des transactions biométriques. L'aspect itinérant de l'expérimentation selon le protocole nomade fut conservé pour la définition d'une méthodologie d'évaluation des boîtes noires.

Ainsi ces travaux ont permis de réaliser une évaluation d'un système boîte noire sur cinq différents systèmes. Le dimensionnement de la population de testeurs a été réalisé afin d'obtenir un nombre suffisant de transactions biométriques, et d'en assurer la représentativité. Cette méthodologie a été généralisée afin de permettre d'estimer les performances biométriques d'un système multimodal en boîte noire.

La mise en œuvre de cette méthodologie d'évaluation boîte noire permet d'estimer les taux d'erreur soit sous forme d'un intervalle de confiance, soit sous la forme d'une borne supérieure. Néanmoins, cette mise en œuvre présente l'inconvénient de nécessiter une population d'une trentaine de personnes sur une période de temps étendue. Ainsi, cette méthodologie s'avère particulièrement couteuse en termes de temps et d'immobilisation d'une trentaine de personnes.

Évaluation de la qualité des échantillons de voix

Une autre problématique étudiée au cours de cette thèse est l'évaluation de la qualité d'un échantillon de voix. La qualité au sens biométrique permet d'estimer la pertinence d'un échantillon biométrique dans un processus de reconnaissance biométrique, ou en tant que référence. Dans le cadre de la reconnaissance du locuteur,

le signal audio est particulièrement susceptible d'être dégradé par la présence de bruit additif, en effet l'environnement sonore est susceptible de masquer des informations contenues dans le signal de voix, et de modifier le comportement du locuteur.

Les travaux centrés sur cet axe, se sont appliqués à déterminer quelles sont les caractéristiques susceptibles de dénoter la présence de bruit au sein d'un signal vocal. La définition de cette métrique nécessita la mise en place d'un entrainement par le biais d'un algorithme génétique, visant à maximiser la corrélation entre cette métrique et les performances biométriques. Une métrique candidate a été ainsi établie afin d'estimer la qualité des différents échantillons afin de procurer un a priori sur les performances attendues.

Perspectives

Perspectives pour l'évaluation des boîtes noires

Le protocole d'évaluation des systèmes biométriques en boîte noire proposé permet d'obtenir une estimation des performances des systèmes sous-test. Néanmoins, contrairement à une évaluation en boîte blanche, cette méthode ne permet pas d'accéder et d'étudier les échantillons a posteriori, afin d'estimer la difficulté de la base ou de les réutiliser sur un système biométrique de référence.

L'étude des populations de test peut se faire parallèlement à l'évaluation en boîte noire. En fonction des modalités utilisées par le système biométrique évalué, la capture peut se faire en parallèle de ce dernier pour des modalités comme le visage ou la voix, ou peut nécessiter des sessions de capture indépendantes comme pour les empreintes digitales. Ces captures ainsi réalisées peuvent ainsi être analysées afin de déterminer la difficulté d'une population, tant pour l'évaluation du FRR et FAR. Cette « boîte blanche parallèle » devrait permettre de réaliser l'intégralité de la comparaison croisée afin d'estimer la difficulté posée par la population de test utilisée en déterminant la répartition selon les classes définies par Doddington[21], et d'utiliser des méthodes de mesure de la difficultés similaire à celles définies dans l'ISO 29198[47], ou définies par Li *et. al.*[59, 60].

Perspectives pour la qualité de la voix

Dans le cadre de la définition d'une métrique de qualité pour la voix, le degré de corrélation entre la combinaison linéaire des caractéristiques extraites et les performances biométriques observées montrent une relation certaine et intéressante. Néanmoins celle-ci peut s'avérer insuffisante pour la définition d'une métrique de

qualité fiable. Ainsi, plusieurs possibles améliorations doivent être étudiées afin d'en augmenter le degré de corrélation. En premier lieu, il peut être intéressant d'étudier et de sélectionner de nouvelles caractéristiques à extraire d'un échantillon vocal, ainsi la décomposition du signal par le bais d'ondelettes peut s'avérer pertinent, en effet Hu et Loizou ont proposé une méthode d'amélioration de la parole utilisant des ondelettes et des filtres *multitaper*[42].

Une autre piste à étudier dans le cadre de l'estimation de la qualité de la voix est la définition d'un modèle sous-jacent universel (UBM). En effet, la définition d'*UBM* afin de modéliser la parole humaine est un sujet ayant fait l'objet de nombreuses recherches, ainsi plusieurs stratégies sont envisageables. La mise en place d'un modèle biométrique sous-jacent universel nécessite de procéder à un entrainement de ce dernier sur une base conséquente d'échantillons vocaux collectés auprès d'une importante population. Ainsi, le modèle de mélange de gaussienne (*GMM*) est particulièrement utilisé afin de définir des *UBM*. Cette stratégie est mise en œuvre par Villalba *et. al.*[103], Omar et Pelecanos[74] , Hasan et Hansen[36] ou encore Reynolds[82]. Des réseaux de neurones profonds (*DNN Deep Neural Network*) sont aussi susceptibles d'être utilisés afin de constituer des modèles sous-jacents universels, d'après Lei *et. al.*[58].

En particulier, la combinaison linéaire utilisée peut ne pas s'avérer satisfaisante afin de réaliser la fusion des informations portées par les différentes caractéristiques. Il peut ainsi être intéressant d'étudier d'autres types de combinaisons, par exemple en élevant les caractéristiques extraites à un coefficient déterminé au cours de l'entrainement par l'algorithme génétique. Ainsi, la fonction de combinaison des caractéristiques s'exprimerait comme :

$$QM = \sum_{i=1}^{N} \alpha_i . x_i^{\beta_i} \tag{4.11}$$

où N représente le nombre de caractéristiques extraites, α_i représente le i^{me} coefficient de pondération, β_i représente le i^{me} exposant, et x_i la valeur admise par la i^{me} caractéristique extraite ; QM étant alors la métrique de qualité candidate.

D'autres méthodes sont susceptibles d'apporter de nouveaux résultats intéressants, et particulièrement les machines à vecteurs de support (ou *SVM* : *Support Machine Vector*), et particulièrement les méthodes de régression qui y sont associées. L'utilisation de *SVM* est particulièrement intéressante afin de déterminer la relation liant les caractéristiques extraites du signal vocal à la qualité biométrique observée.

Bibliographie

[1] Chaos Computer Club breaks Apple TouchID. https://www.ccc.de/en/updates/2013/ccc-breaks-apple-touchid, 2013. [Online ; accédé le 09-Septembre-2018].

[2] Mba device project : Multimodal biometric authentication device. 2017.

[3] A. Anjos, L. El-Shafey, and S. Marcel. Beat : An open-source web-based open-science platform. *arXiv preprint arXiv :1704.02319*, 2017.

[4] W. Appel. *Probabilités pour les non-probabilistes*. H&K éditions, 2013.

[5] E. Bailly-Bailliére, S. Bengio, F. Bimbot, M. Hamouz, J. Kittler, J. Mariéthoz, J. Matas, K. Messer, V. Popovici, F. Porée, et al. The banca database and evaluation protocol. In *International conference on Audio-and video-based biometric person authentication*, pages 625–638. Springer, 2003.

[6] M. Ballantyne, R. S. Boyer, and L. Hines. Woody bledsoe : His life and legacy. *AI magazine*, 17(1) :7, 1996.

[7] A. Bertillon. *La photographie judiciaire : avec un appendice sur la classification et l'identification anthropométriques*. Gauthier-Villars, 1890.

[8] A. Bertillon. *Identification anthropométrique : instructions signalétiques*, volume 1. Impr. administrative, 1893.

[9] B. Bhanu and V. Govindaraju. *Multibiometrics for Human Identification*. Cambridge University Press, 2011.

[10] I. J. S. Biometrics. Iso 29794-1 biometric sample quality. *Committee Draft*, 1, 2007.

[11] V. Blanz and T. Vetter. A morphable model for the synthesis of 3d faces. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pages 187–194. ACM Press/Addison-Wesley Publishing Co., 1999.

[12] V. Blanz and T. Vetter. Face recognition based on fitting a 3d morphable model. *IEEE Transactions on pattern analysis and machine intelligence*, 25(9) :1063–1074, 2003.

[13] R. Cappelli, M. Ferrara, A. Franco, and D. Maltoni. Fingerprint verification competition 2006. *Biometric Technology Today*, 15(7) :7–9, 2007.

[14] R. Cappelli, D. Maio, D. Maltoni, J. L. Wayman, and A. K. Jain. Performance evaluation of fingerprint verification systems. *IEEE transactions on pattern analysis and machine intelligence*, 28(1) :3–18, 2006.

[15] A. B. Carlson. *Communication Systems : An Introducton to Signals and Noise in Electrical Communication*. McGraw Hill, 1986.

[16] E. Carson. Could your kid's face unlock your iPhone X ? https://www.cnet.com/news/kid-unlock-iphone-x-face-id/, 2017. [Online ; accédé le 09-Septembre-2018].

[17] N. Chauhan. Speaker recognition using pattern recognition neural network and feedforward neural network. *International Journal of Scientific & Engineering Research*, 8(3) :1444–1446, March 2017.

[18] F. Chowdhury, S.-A. Selouani, and D. O'Shaughnessy. Voice biometrics : Speaker verification and identification. *Signal and Image Processing for Biometrics*, pages 131–148.

[19] H. Cummins. Ancient finger prints in clay. *Journal of Criminal Law and Criminology (1931-1951)*, 32(4) :468–481, 1941.

[20] L. Deng and D. O'Shaughnessy. *Speech processing : a dynamic and optimization-oriented approach*. CRC Press, 2003.

[21] G. Doddington, W. Liggett, A. Martin, M. Przybocki, and D. Reynolds. Sheep, goats, lambs and wolves : A statistical analysis of speaker performance in the nist 1998 speaker recognition evaluation. Technical report, DTIC Document, 1998.

[22] G. R. Doddington, M. A. Przybocki, A. F. Martin, and D. A. Reynolds. The nist speaker recognition evaluation–overview, methodology, systems, results, perspective. *Speech Communication*, 31(2) :225–254, 2000.

[23] B. Dumas, C. Pugin, J. Hennebert, D. Petrovska-Delacrétaz, A. Humm, F. Evéquoz, R. Ingold, and D. Von Rotz. Myidea-multimodal biometrics database, description of acquisition protocols. *Proc. Third COST*, 275 :59–62, 2005.

[24] B. Efron. Bootstrap methods : another look at the jackknife. In *Breakthroughs in statistics*, pages 569–593. Springer, 1992.

[25] M. El-Abed, R. Giot, B. Hemery, J.-J. Schwartzmann, and C. Rosenberger. Towards the security evaluation of biometric authentication systems. *IACSIT International Journal of Engineering and Technology*, pages 315–320, 2012.

[26] G. Fant. *Acoustic theory of speech production : with calculations based on X-ray studies of Russian articulations*, volume 2. Walter de Gruyter, 2012.

[27] F. A. Fernandez. *Biometric sample quality and its application to multimodal authentication systems.* PhD thesis, 2008.

[28] B. Fernandez-Saavedra, R. Alonso-Moreno, J. Uriarte-Antonio, and R. Sanchez-Reillo. Evaluation methodology for analyzing usability factors in biometrics. In *Security Technology, 2009. 43rd Annual 2009 International Carnahan Conference on*, pages 347–354. IEEE, 2009.

[29] B. Fernandez-Saavedra, R. Sanchez-Reillo, R. Alonso-Moreno, and R. Mueller. Evaluation methodology for analyzing environment influence in biometrics. In *2008 10th International Conference on Control, Automation, Robotics and Vision*, pages 1342–1346, Dec 2008.

[30] M. B. Fernández Saavedra. *Evaluation methodologies for security testing biometric systems beyond technological evaluation.* PhD thesis, 2013.

[31] M. G. Ferrari. Dissemination of the argentine dactyloscopy system in the early twentieth century : Local, regional and international dimensions. In *Identification and Registration Practices in Transnational Perspective*, pages 44–59. Springer, 2013.

[32] F. Galton. *Fingerprint directories.* Macmillan and Company, 1895.

[33] S. Garcia-Salicetti, C. Beumier, G. Chollet, B. Dorizzi, J. L. Les Jardins, J. Lunter, Y. Ni, and D. Petrovska-Delacrétaz. Biomet : A multimodal person authentication database including face, voice, fingerprint, hand and signature modalities. In *International Conference on Audio-and Video-based Biometric Person Authentication*, pages 845–853. Springer, 2003.

[34] H. Gray and C. M. Goss. Anatomy of the human body. *American Journal of Physical Medicine & Rehabilitation*, 53(6) :293, 1974.

[35] P. Grother and E. Tabassi. Performance of biometric quality measures. *IEEE transactions on pattern analysis and machine intelligence*, 29(4) :531–543, 2007.

[36] T. Hasan and J. H. Hansen. A study on universal background model training in speaker verification. *IEEE Transactions on Audio, Speech, and Language Processing*, 19(7) :1890–1899, 2011.

[37] M. Hébert. Text-dependent speaker recognition. In *Springer handbook of speech processing*, pages 743–762. Springer, 2008.

[38] E. HENRY. Classification and uses of finger prints.[sl] : George routledge and sons, 1900.

[39] E. H. Holder, L. O. Robinson, and J. H. Laub. *The fingerprint sourcebook*. US Department of Justice, Office of Justice Programs, National Institute of Justice, 2011.

[40] J. H. Holland. *Adaptation in natural and artificial systems : an introductory analysis with applications to biology, control, and artificial intelligence*. MIT press, 1992.

[41] L. Hong, A. K. Jain, and S. Pankanti. Can multibiometrics improve performance ? In *Proceedings AutoID*, volume 99, pages 59–64. Citeseer, 1999.

[42] Y. Hu and P. C. Loizou. Speech enhancement based on wavelet thresholding the multitaper spectrum. *IEEE transactions on Speech and Audio processing*, 12(1) :59–67, 2004.

[43] I. ISO. Iec 19795-1 : Information technology–biometric performance testing and reporting-part 1 : Principles and framework. *ISO/IEC, Editor*, 2006.

[44] I. ISO. Iec 19794-2 : Information technology-biometric data interchange formats-part 2 : Finger minutiae data. *ISO/IEC, Editor*, 2011.

[45] I. ISO. Iec 19794-5 : Information technology-biometric data interchange formats-part 5 : Face image data. *ISO/IEC, Editor*, 2011.

[46] I. ISO. Iec 19794-2/a1 : Information technology-biometric data interchange formats-part 2 : Finger minutiae data-amendment 1 : Conformance testing methodology and clarification of defects. *ISO/IEC, Editor*, 2013.

[47] I. ISO. Iec iso/iec tr 29198 – information technology - characterization and measurement of difficulty for fingerprint databases for technology evaluation. *ISO/IEC, Editor*, 2013.

[48] A. K. Jain, A. Ross, and S. Prabhakar. An introduction to biometric recognition. *IEEE Transactions on circuits and systems for video technology*, 14(1) :4–20, 2004.

[49] M. Kaluszynski. Alphonse bertillon et l'anthropométrie judiciaire. l'identification au cœur de l'ordre républicain. *Criminocorpus. Revue d'Histoire de la justice, des crimes et des peines*, 2014.

[50] M. G. Kendall et al. The advanced theory of statistics. *The advanced theory of statistics.*, (2nd Ed), 1946.

[51] B. Klare, A. A. Paulino, and A. K. Jain. Analysis of facial features in identical twins. In *Biometrics (IJCB), 2011 International Joint Conference on*, pages 1–8. IEEE, 2011.

[52] K. Krishan, T. Kanchan, and G. S. Bumbrah. The fingerprint sourcebook. *Journal of Forensic and Legal Medicine*, 19(3) :182–183, 2012.

[53] W. Kruskal and F. Mosteller. Representative sampling, iii : The current statistical literature. *International Statistical Review/Revue Internationale de Statistique*, pages 245–265, 1979.

[54] E. Kukula, S. Elliott, H. Kim, and C. S. Martin. The impact of fingerprint force on image quality and the detection of minutiae. In *2007 IEEE International Conference on Electro/Information Technology*, pages 432–437, May 2007.

[55] E. P. Kukula, C. R. Blomeke, S. K. Modi, and S. J. Elliott. Effect of human-biometric sensor interaction on fingerprint matching performance, image quality and minutiae count. *International Journal of Computer Applications in Technology*, 34(4) :270–277, 2009.

[56] E. P. Kukula, S. J. Elliott, and V. G. Duffy. The effects of human interaction on biometric system performance. In *International Conference on Digital Human Modeling*, pages 904–914. Springer, 2007.

[57] E. P. Kukula, M. J. Sutton, and S. J. Elliott. The human–biometric-sensor interaction evaluation method : Biometric performance and usability measurements. *IEEE Transactions on Instrumentation and Measurement*, 59(4) :784–791, 2010.

[58] Y. Lei, N. Scheffer, L. Ferrer, and M. McLaren. A novel scheme for speaker recognition using a phonetically-aware deep neural network. In *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, pages 1695–1699. IEEE, 2014.

[59] S. Li, C. Jin, H. Kim, and S. Elliott. Assessing the difficulty level of fingerprint datasets based on relative quality measures. In *Hand-Based Biometrics (ICHB), 2011 International Conference on*, pages 1–5. IEEE, 2011.

[60] S. Li, H. Kim, C. Jin, S. Elliott, and M. Ma. Assessing the level of difficulty of fingerprint datasets based on relative quality measures. *Information Sciences*, 268 :122–132, 2014.

[61] J. Mahier, B. Hemery, M. El-Abed, M. El-Allam, M. Bouhaddaoui, and C. Rosenberger. Computation evabio : A tool for performance evaluation in biometrics. *International Journal of Automated Identification Technology (IJAIT)*, page 24, 2011.

[62] D. Maio, D. Maltoni, R. Cappelli, J. L. Wayman, and A. K. Jain. Fvc2000 : Fingerprint verification competition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(3) :402–412, 2002.

[63] D. Maio, D. Maltoni, R. Cappelli, J. L. Wayman, and A. K. Jain. Fvc2002 : Second fingerprint verification competition. In *Pattern recognition, 2002. Proceedings. 16th international conference on*, volume 3, pages 811–814. IEEE, 2002.

[64] D. Maio, D. Maltoni, R. Cappelli, J. L. Wayman, and A. K. Jain. Fvc2004 : Third fingerprint verification competition. In *Biometric Authentication*, pages 1–7. Springer, 2004.

[65] D. Maltoni, D. Maio, A. K. Jain, and S. Prabhakar. *Handbook of fingerprint recognition*. Springer Science & Business Media, 2009.

[66] A. J. Mansfield and J. L. Wayman. Best practices in testing and reporting performance of biometric devices. 2002.

[67] T. Mansfield, G. Kelly, D. Chandler, and J. Kane. Biometric product testing final report. *Contract*, 92(4009) :309, 2001.

[68] S. Matteson and F.-L. Lu. Vocal inharmonicity analysis : A promising approach for acoustic screening for dysphonia. *The Journal of the Acoustical Society of America*, 125(4) :2638–2638, 2009.

[69] Z. Milivojevic, D. Brodic, and D. Blagojevic. The impact of the acute hypoxia to speech inharmonicity. *Elektronika ir Elektrotechnika*, 20(5) :136–143, 2014.

[70] H. Misra, S. Ikbal, H. Bourlard, and H. Hermansky. Spectral entropy based feature for robust asr. In *Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on*, volume 1, pages I–193. IEEE, 2004.

[71] M. Mitchell. *An introduction to genetic algorithms*. MIT press, 1998.

[72] P. A. Naylor, A. Kounoudes, J. Gudnason, and M. Brookes. Estimation of glottal closure instants in voiced speech using the dypsa algorithm. *IEEE Transactions on Audio, Speech, and Language Processing*, 15(1) :34–43, 2007.

[73] L. O'Gorman. Comparing passwords, tokens, and biometrics for user authentication. *Proceedings of the IEEE*, 91(12) :2021–2040, 2003.

[74] M. K. Omar and J. W. Pelecanos. Training universal background models for speaker recognition. In *Odyssey*, page 10, 2010.

[75] J. M. Pandya, D. Rathod, and J. J. Jadav. A survey of face recognition approach. *International Journal of Engineering Research and Applications (IJERA)*, 3(1) :632–635, 2013.

[76] S. Parthasarathy and C. Busso. Predicting speaker recognition reliability by considering emotional content. In *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 434–439, Oct 2017.

[77] A. Perala. OnePlus 6 Face Unlock Fooled by Black-and-White Photo. https://mobileidworld.com/oneplus-6-face-unlock-fooled-photo-905307/, 2018. [Online ; accédé le 09-Septembre-2018].

[78] P. J. Phillips, H. Wechsler, J. Huang, and P. J. Rauss. The feret database and evaluation procedure for face-recognition algorithms. *Image and vision computing*, 16(5) :295–306, 1998.

[79] M. H. Quenouille. Notes on bias in estimation. *Biometrika*, 43(3/4) :353–360, 1956.

[80] N. K. Ratha, J. H. Connell, and R. M. Bolle. Enhancing security and privacy in biometrics-based authentication systems. *IBM systems Journal*, 40(3) :614–634, 2001.

[81] I. Rec. P. 563 : Single-ended method for objective speech quality assessment in narrow-band telephony applications. *International Telecommunication Union, Geneva*, 2004.

[82] D. Reynolds. Universal background models. In *Encyclopedia of biometrics*, pages 1349–1352. Springer, 2009.

[83] A. W. Rix, J. G. Beerends, M. P. Hollier, and A. P. Hekstra. Perceptual evaluation of speech quality (pesq)-a new method for speech quality assessment

of telephone networks and codecs. In *Acoustics, Speech, and Signal Processing, 2001. Proceedings.(ICASSP'01). 2001 IEEE International Conference on*, volume 2, pages 749–752. IEEE, 2001.

[84] C. Roberts. Biometric attack vectors and defences. *Computers & Security*, 26(1) :14–25, 2007.

[85] A. E. Rosenberg, F. Bimbot, and S. Parthasarathy. Overview of speaker recognition. In *Springer Handbook of Speech Processing*, pages 725–742. Springer, 2008.

[86] A. Ross and A. Jain. Information fusion in biometrics. *Pattern recognition letters*, 24(13) :2115–2125, 2003.

[87] R. Sanchez-Reillo, D. Sierra-Ramos, R. Estrada-Casarrubios, and J. A. Amores-Duran. Strengths, weaknesses and recommendations in implementing biometrics in mobile devices. In *Security Technology (ICCST), 2014 International Carnahan Conference on*, pages 1–6. IEEE, 2014.

[88] G. Saporta. *Probabilités, analyse des données et statistique*. Editions Technip, 2006.

[89] C. E. Shannon. A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, 5(1) :3–55, 2001.

[90] J. Sohn, N. S. Kim, and W. Sung. A statistical model-based voice activity detection. *IEEE signal processing letters*, 6(1) :1–3, 1999.

[91] H. J. M. Steeneken and T. Houtgast. A physical method for measuring speech-transmission quality. *The Journal of the Acoustical Society of America*, 67(1) :318–326, 1980.

[92] S. M. Stigler. Galton and identification by fingerprints. *Genetics*, 140(3) :857, 1995.

[93] P. Subhashini and T. Pratap. Text-independent speaker recognition using combined lpc and mfc coefficients.

[94] Z. Sun, A. A. Paulino, J. Feng, Z. Chai, T. Tan, and A. K. Jain. A study of multibiometric traits of identical twins. In *Biometric Technology for Human Identification Vii*, volume 7667, page 76670T. International Society for Optics and Photonics, 2010.

[95] J. P. Teixeira, C. Oliveira, and C. Lopes. Vocal acoustic analysis–jitter, shimmer and hnr parameters. *Procedia Technology*, 9 :1112–1122, 2013.

[96] M. Theofanos, B. Stanton, C. Sheppard, R. Micheals, N. Zhang, J. Wydler, L. Nadel, and W. Rubin. Usability testing of height and angles of ten-print fingerprint capture. *NISTIR, June*, 20(42) :210, 2008.

[97] M. F. Theofanos, B. C. Stanton, and C. Wolfson. Usability and biometrics : Ensuring successful biometric systems. In *International Workshop on Usability and Biometrics*, number International Workshop on Usability and Biometrics, 2008.

[98] M. Trauring. Automatic comparison of finger-ridge patterns. *Nature*, 197(4871) :938, 1963.

[99] M. Triplett and S. Everist. *Fingerprint Dictionary : An Examiner's Guide to the Who, What, and Where of Fingerprint Identification.* CreateSpace Independent Publishing Platform, 2015.

[100] J. Tukey. Bias and confidence in not quite large samples. *Ann. Math. Statist.*, 29 :614, 1958.

[101] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of cognitive neuroscience*, 3(1) :71–86, 1991.

[102] M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. In *Computer Vision and Pattern Recognition, 1991. Proceedings CVPR'91., IEEE Computer Society Conference on*, pages 586–591. IEEE, 1991.

[103] J. Villalba, A. Ortega, A. Miguel, and E. Lleida. Analysis of speech quality measures for the task of estimating the reliability of speaker verification decisions. *Speech Communication*, 78 :42–61, 2016.

[104] J. L. Wayman. Confidence interval and test size estimation for biometric data. In *Proceedings of the IEEE AutoID Conference*, 1999.

[105] L. Wiskott, J.-M. Fellous, N. Krüger, and C. Von Der Malsburg. Face recognition by elastic bunch graph matching. In *International Conference on Computer Analysis of Images and Patterns*, pages 456–463. Springer, 1997.

Contributions à l'évaluation opérationnelle des systèmes biométriques multimodaux.

Le développement et la multiplication de dispositifs connectés, en particulier avec les *smartphones*, nécessitent la mise en place de moyens d'authentification. Dans un soucis d'ergonomie, les industriels intègrent massivement des systèmes biométrique afin de garantir l'identité du porteur, et ce afin d'autoriser l'accès à certaines applications et fonctionnalités sensibles (paiements, *e-banking*, accès à des données personnelles : correspondance électronique...). Dans un soucis de garantir, une adéquation entre ces systèmes d'authentification et leur usages, la mise en œuvre d'un processus d'évaluation est nécessaire.

L'amélioration des performances biométriques est un enjeux important afin de permettre l'intégration de telles solutions d'authentification dans certains environnement ayant d'importantes exigences sur les performances, particulièrement sécuritaires. Afin d'améliorer les performances et la fiabilité des authentifications, différentes sources biométriques sont susceptibles d'être utilisées dans un processus de fusion. La biométrie multimodale réalise, en particulier, la fusion des informations extraites de différentes modalités biométriques.

Cette thèse se propose d'aborder cette problématique, et de décrire une méthodologie d'évaluation permettant de mettre en place un procédé d'évaluation des performances sur des systèmes biométriques opérationnels. Dans le cadre d'un projet de création d'un dispositif d'authentification biométrique, des travaux concernant l'évaluation de ce dispositif ont permis de mettre en œuvre cette méthodologie sur un système biométrique multimodale.

L'évaluation du système multimodal a permis de mettre en exergue l'importance de la qualité des échantillons biométriques. La constatation pratique de cette influence a, ainsi, motivé des travaux sur l'estimation de la qualité des échantillons de voix dans le cadre de la reconnaissance du locuteur. Cette thèse propose, dans un second axe, une méthodologie permettant d'estimer la présence de bruit additif dans des échantillons de voix, afin de prédire leur utilité dans le cadre de la reconnaissance biométrique.

Contributions to the operationnal evaluation of multimodal biometric systems

Development and spread of connected devices, in particular smartphones, requires the implementation of authentication methods. In an ergonomic concern, manufacturers integrates biometric systems in order to deal with logical control access issues. These biometric systems grant access to critical data and application (payment, e-banking, privcy concerns : emails ...). Thus, evaluation processes allows to estimate the systems' suitabilty with these uses. In order to improve recognition performances, manufacturer are susceptible to perform multimodal fusion.

In this thesis, the evaluation of operationnal biometric systems has been studied, and an implementation is presented. A second contribution studies the quality estimation of speech samples, in order to predict recognition performances.

Mots Clefs : Biométrie / Évaluation / Empreintes digitales / Boîte noire / Reconnaissance du locuteur / Métrique de qualité / Logiciels - évaluation

Spécialité Informatique

Laboratoire GREYC - UMR CNRS 6072 - Université de Caen Basse-Normandie - Ensicaen
6 Boulevard du Maréchal Juin - 14050 CAEN CEDEX