



## Rapport : Base de données

Anthony Larcher<sub>1</sub>

1 : Laboratoire d'Informatique d'Avignon - Université d'Avignon  
Tél : +33 (0) 4 90 84 35 55 - Fax : + 33 (0) 4 90 84 35 01  
[anthony.larcher@univ-avignon.fr](mailto:anthony.larcher@univ-avignon.fr)

14 février 2008

## Introduction

En reconnaissance du locuteur comme dans de nombreuses applications biométriques, la variabilité intra et inter-locuteur impose une validation rigoureuse des systèmes et des résultats. Or, pour être validées, ces approches biométriques statistiques doivent être testées sur un nombre suffisant d'individus et dans des conditions les plus proches possibles des conditions d'utilisation réelles des systèmes. De ce fait, le développement d'un système biométrique multimodal se heurte aux insuffisances des bases de données existantes.

La définition et la réalisation d'une base de données originale, utilisable dans ce cadre représente un investissement financier et humain conséquent qui apparaît comme trop coûteux et trop long dans le cadre du projet BIOBIMO. Jusqu'à présent, la base de données audio/vidéo ValidDB a été utilisée dans le cadre du projet. Néanmoins elle a démontré d'importantes limites, obligeant les partenaires à rechercher une base de données mieux adaptée aux critères et exigences scientifiques de ce projet. Il a donc fallu remplacer la base de données ValidDB [1] jugée insuffisante par une base répondant mieux aux critères de ce projet. Ce document présente les choix effectués par les partenaires du projet BIOBIMO afin de remplacer ValidDB. On y présente également l'investissement qu'ont impliqué ces choix et le travail réalisé au cours du deuxième semestre 2007 au Laboratoire Informatique d'Avignon.

## La base de données ValidDB et ses lacunes

Choisie dans un premier temps pour son caractère "libre de droits" et ses caractéristiques audio/vidéo intéressantes, la base de données ValidDB s'est avérée présenter des inconvénients et des insuffisances qui justifiaient son remplacement.

Dans le but de développer une application de reconnaissance du locuteur audio-vidéo utilisant des mots de passe, la base de données doit répondre à certaines attentes.

- Pour la variabilité inter-locuteur : regrouper un nombre important de locuteurs
- Pour la variabilité intra-locuteur : contenir plusieurs enregistrements d'une personne prononçant les mêmes mots de passe à des intervalles de temps importants.
- Pour la variabilité phonétique, les mots de passe prononcés par les locuteurs doivent être nombreux et variés.
- Afin de tester la robustesse du système développé aux attaques d'imposteurs, les locuteurs doivent prononcer les mots de passe des autres utilisateurs pour simuler une véritable tentative d'usurpation.
- Enfin pour tester les conditions réelles ainsi que le véritable impact de l'utilisation de mots de passe pour un tel système, les locuteurs doivent également prononcer des phrases différentes des mots de passe des clients qu'ils impostent.
- Pour des raisons techniques, les différentes phrases doivent être segmentées et indexées de façon précise afin de respecter le protocole établi.

La base de données ValidDB si elle contient un nombre relativement important de locuteurs (77 hommes) pour une base de données audio/vidéo (certaines

bases de données audio uniquement contiennent plus de 600 locuteurs) ne répond pas aux autres critères. Ainsi les locuteurs de cette base ne prononcent que 2 séquences qui sont : une phrase d'environ trois secondes et une série de digits. Ces deux séquences rendent difficiles de véritables tests d'imposture puisque leur contenu lexical est trop différent pour être considéré comme représentatif des mots de passe choisis par les utilisateurs du système. De plus on ne dispose que de deux mots de passe différents. La quantité de données disponible ne permet pas d'effectuer un nombre de tests suffisant pour obtenir des résultats fiables statistiquement.

## Présentation de MyIdea et réalité

La base de données MyIdea réalisée à l'université de Fribourg (Suisse) est une base de données multimodale qui contient entre autres des données audio-vidéo. Les caractéristiques présentées dans un rapport scientifique et avancées par nos interlocuteurs de Fribourg faisaient état d'une centaine de locuteurs, enregistrés lors de 3 ou 4 sessions dans des conditions réelles et adverses pour la tâche visée. Ces locuteurs devaient prononcer un nombre important de phrases différentes, choisies selon leur contenu linguistique. De plus, chaque locuteur possédait ses propres phrases et prononçait également les phrases d'autres locuteurs, de façon à simuler les différents types d'impostures recherchés.

Une fois le choix de MyIdea validé par l'ensemble des partenaires, la demande a été exprimée à l'université de Fribourg. La base de données n'a cependant été reçue que 3 mois après cette demande. Cette base s'est avérée ne pas correspondre exactement à sa description.

La plupart des données n'étaient pas validées et les différentes sessions n'étaient pas segmentées (chaque enregistrement contenait une trentaine de phrases prononcées par un locuteur). Il a donc été nécessaire de valider les données disponibles en retirant dans un premier temps les sessions incomplètes et les enregistrements présentant un quelconque problème. Le tableau 1 oppose la description de la base MyIdea à sa réalité après validation. Après sélection, une grande partie de la base de données s'est avérée être inutilisable, notamment les données enregistrées dans des environnements réels et dans des conditions adverses.

En conséquence, un important travail a dû être réalisé au LIA afin de rendre les données disponibles utilisables dans le cadre de BIOBIMO.

## Travail effectué au LIA

Les données réellement disponibles dans la base de données MyIdea s'avérant tout de même suffisantes pour l'évaluation d'un système bimodal, et considérant que la réalisation complète d'une base de données s'avérerait trop coûteuse, les trois partenaires ont pris la décision de finaliser MyIdea. Le fait qu'une base de données réalisée entièrement au sein du projet aurait certainement souffert des mêmes aléas que MyIdea a conforté les partenaires dans ce choix.

Afin de valider, segmenter et indexer la base de données MyIdea, le travail suivant a dû être réalisé au LIA :

- Les données ont été triées afin de retirer les locuteurs et les sessions non utilisables

DESCRIPTIF ANNONCÉ	BASE RÉELLE
<ul style="list-style-type: none"> <li>• Une centaine de locuteurs</li> <li>• 2 parties aux contenus distincts (BIO-MET et BANCA)</li> <li>• Enregistrements webcam et camera SONYDV</li> <li>• Enregistrements en conditions réelles et adverses</li> <li>• Base de données segmentée et finalisée</li>   <li>• Base de données déjà utilisée dans une évaluation internationale</li> </ul>	<ul style="list-style-type: none"> <li>• 32 francophones et 22 anglophones</li> <li>• Partie BANCA inutilisable</li>   <li>• Enregistrements Webcam inutilisables</li> <li>• Enregistrements en condition studio</li>   <li>• Données non validées et non segmentées</li> <li>• Base jamais utilisée</li> </ul>

TAB. 1 – Différence entre la base MyIdea annoncée et la version reçue au LIA

- Les fichiers ont été segmentés et indexés de façon semi-automatique à l'aide d'une application développée par l'université de Fribourg.
- Le son des séquences vidéo a été extrait et ré-échantillonné
- Des échanges ont eu lieu avec l'université de Fribourg afin d'améliorer la base et de récupérer des données inutilisables actuellement.
- La base de données segmentée et validée a été transmise à EURECOM
- Un protocole d'expérimentation a été développé afin de tester et d'évaluer le système.

## Protocole expérimental

Sont utilisés :

- 30 locuteurs hommes francophones
- 2 locuteurs hommes francophones incomplets pour le modèle du monde
- 3 sessions
- 2 phrases longues (environ 6 secondes de parole) par session
- 10 phrases courtes fixes "mots de passe" (environ 3 secondes de parole) par session
- 13 phrases courtes utilisées pour apprendre le modèle du monde par session
- 1 série de digits par session

Afin de compenser le faible nombre de locuteurs, les 30 hommes sont séparés en deux groupes (A et B) de 15. Les deux groupes sont tour à tour utilisés comme clients-imposteurs et modèle du monde. Si les locuteurs du groupe A sont utilisés comme clients-imposteurs, le protocole d'entraînement est le suivant :

- Un modèle du monde est créé à partir de l'ensemble des données des 15 locuteurs du groupe B, auxquelles sont ajoutées les données des deux locuteurs incomplets.
- Chacun des 15 locuteurs du groupe A est utilisé tour à tour comme client. Trois modèles de locuteurs sont appris pour chaque homme, chacun avec les deux phrases longues d'une des trois sessions d'enregistrement.
- Pour chaque modèle de locuteur, 10 mots de passe sont appris en adaptant ce modèle avec l'une des 10 phrases courtes fixes

Les tests sont effectués selon le protocole décrit ci-dessous :

- chacun des 30 mots de passe d’un locuteur est testé avec les 3 occurrences du mot de passe utilisé, prononcé par chacun des 14 autres locuteurs du groupe A soit : 42 tests par mot de passe, 1260 tests imposteurs avec le bon mot de passe par locuteur.
- ces mêmes mots de passe sont testés avec 3 phrases différentes du mot de passe choisies au hasard pour chacun des 14 autres locuteurs du groupe A. Soit 1260 tests imposteurs avec un mauvais mot de passe par locuteur.
- les test clients sont effectués en testant les 30 mots de passe de chaque locuteur avec les 2 occurrence des autres sessions de chaque mot de passe. On obtient un total de 60 tests par locuteur.

En intervertissant les rôles des groupes A et B on double le nombre de tests et l’on peut résumer le nombre de tests réalisés dans 2.

	Clients	Imposteurs bon mot de passe	Imposteurs mauvais mot de passe
Nombre de tests	1800	37 800	37 800

TAB. 2 – Nombre de tests réalisés pour une évaluation

## Conclusion

La partie de la base de données MyIdea validée et segmentée nous permet maintenant d’évaluer notre système bimodal audio-vidéo. La préparation de cette base a cependant demandé un investissement important au LIA. 6 mois ont été nécessaires depuis la demande de la base de données à l’université de Fribourg pour finaliser la base, définir un protocole et paramétrer notre système et revenir au niveau des évaluations faites sur ValidDB.

Le changement de base de données s’est cependant accompagné de résultats prometteurs puisque l’on retrouve actuellement des résultats équivalents à ceux obtenus sur ValidDB, ce qui montre que notre système n’est que peu sensible au changement de base et confirme les résultats obtenus précédemment en les légitimant du fait du nombre accru de tests effectués et du contenu linguistique plus riche et plus proche de l’application visée.

# Bibliographie

- [1] Niall A. Fox, Brian A. O'Mullane, and Richard B. Reilly, "The realistic multi-modal valid database and visual speaker identification comparison experiments," in *International Conference of Audio and Video-Based Person Authentication, AVBPA*, New York (US), July 2005.