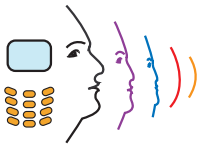


MOBIO

Mobile Biometry

http://www.mobioproject.org/

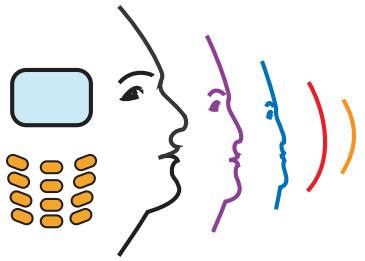
Funded under the 7th FP (Seventh Framework Programme)
Theme ICT-2007.1.4
[Secure, dependable and trusted Infrastructure]

D2.2: Report on the specifications of the database

Due date: 30/04/2008 **Submission date:** 30/04/2008
Project start date: 01/01/2008 **Duration:** 36 months
WP Manager: Christopher McCool **Revision:** 3

Author(s): A. Hadid (OULU)

Project funded by the European Commission in the 7th Framework Programme (2008-2010)		
Dissemination Level		
PU	Public	Yes
RE	Restricted to a group specified by the consortium (includes Commission Services)	No
CO	Confidential, only for members of the consortium (includes Commission Services)	No



D2.2: Report on the specifications of the database

Abstract:

Within the MOBIO project, we seek a database that has the following characteristics: synchronized face and speech data collected with mobile devices over an extended time period (2 years) and publicly available for research purposes. While there are a growing number of bi-modal and multi-modal databases, which have been surveyed here, none satisfies all the above requirements. This interim report clarifies the specific choices that we have adopted in designing the MOBIO database. In particular, it deals with the following issues: the number of persons, sessions and shots per person; the acquisition conditions; the recording formats and file naming convention; text and language choices for speech recording; the recording devices; the impostor attacks; and legal issues related to database distribution and privacy concerns.

Contents

1 Introduction

Portable electronic devices such as mobile phones and PDAs are becoming important means to provide wireless access to the Internet and other telecommunication networks. Very often, such access requires the verification of the user's identity in order to ensure that the person is really whom he or she claims to be. Traditional authentication methods such as PINs or passwords have several problems, for instance they can be forgotten, easily compromised when shared, copied or stolen. In comparison, biometric authentication is a more effective alternative because it is a natural, reliable and friendly means of authentication.

In this context, the FP7 project MOBIO (Mobile Biometry) aims to develop and evaluate new mobile services that are secured by bi-modal speech and face biometrics. Due to the device mobility, the problem of biometric authentication is much more challenging for two reasons. First, it has to deal with changing and often uncontrolled environments. Second, mobile devices have limited memory and CPU resources. This provides a constraint to the size of biometric template, or model, and the type of processing algorithms.

The MOBIO project aims to address these two challenges by considering five key areas. These key areas are robust-to-illumination face authentication, robust-to-noise speaker authentication, joint bi-modal authentication, model adaptation and scalability. The innovative techniques developed for these key areas have to be assessed with an appropriate audio-video biometric database.

The audio-video database must fulfil at least four requirements. First, the face and speech data must be synchronized and collected with mobile devices. Second, it should contain a reasonably long life span (2 years) in order to study possible performance degradation over time. Third, it must be publicly available for research purposes since it is well known in biometrics that two algorithms cannot be compared if they are not assessed on the same database. Fourth, the database should contain realistic and common environmental variations associated with the usage of mobile devices.

While there are a growing number of databases, none satisfies all the above requirements. This interim report explains and justifies the specific choices that we have adopted in designing the new database. In particular, it addresses such issues as when, where and who will record the database; the number of persons along with the number of sessions and shots per person; age and gender balance, ethnic diversity and acquisition environmental conditions; recording devices (mobile phones, PDAs, laptops, cameras, microphones); recording format and file naming convention; text and language choices for audio recording; and the legal and privacy issues related to public distribution of the database. In addition to this, it also provides a comprehensive review of the currently available bi-modal and multi-modal databases.

2 Review of Audio-Video and Multi-Modal Databases

There exist several audio-video (face and speech) and also multi-modal (face, speech, signature, hand, fingerprint and iris) databases [7, 1, 6, 8, 9, 4, 2, 3, 5]. Some of these databases

are publicly available for research purposes while others are not. Below is a description of the major databases.

M2VTS (Audio-Video)
The M2VTS [7] database consists of audio recordings and video sequences of 37 subjects uttering digits 0 through 9 in five sessions spaced apart by at least one week. The subjects were also asked to rotate their head to the left and then to the right in each session in order to obtain a head rotation sequence that can provide 3-D face features to be used for face recognition purposes. The main drawbacks of this database are its small size and limited vocabulary.

XM2VTS (Audio-Video)
The XM2VTS (extended M2VTS) database [6] consists of audio recordings and video sequences of 295 subjects uttering three fixed phrases, two ten-digit sequences and one seven-word sentence, with two utterances of each phrase, in four sessions taken at one month interval. Fig. 1 shows examples of face images from XM2VTS database. The main drawbacks of this database are its limitations to uniform background, controlled illuminations and text-dependent systems. Both M2VTS and XM2VTS databases have been frequently used in the literature for comparison of different biometric systems.

Figure 1: Example of images from XM2VTS database

BANCA (Audio-Video)
The BANCA database [8] consists of audio recordings and video sequences of 208 subjects (half men and half women) recorded in three different scenarios, controlled, degraded and

adverse, over a period of three months. The subjects were asked to say a random 12-digit number, their name, their address and date of birth, during each recording. The BANCA database was captured in four European languages (English, French, Italian and Spanish) but only the English part was made publicly available. Both high- and low-quality microphones and cameras were used for recording. The BANCA database provides realistic and challenging conditions and allows for comparison of different systems with respect to their robustness. Fig. 2 shows examples of face images from BANCA database.

Figure 2: Example of images from BANCA database in 3 scenarios: controlled (1st row), degraded (2nd row) and adverse (3rd row)

VidTIMIT (Audio-Video)

The VidTIMIT database [9] consists of audio and video recordings of 43 subjects (19 female and 24 male), reciting short sentences. It is used for research on topics such as automatic lip reading, multi-view face recognition, multi-modal speech recognition and person identification. The dataset was recorded in 3 sessions with an average delay of a week between sessions. Each person utters ten sentences. The first two sentences are the same for all subjects while the remaining eight are generally different for each person. All sessions contain phonetically balanced sentences. The recording was done in an office environment using a broadcast quality digital video camera. The video of each person is stored as a numbered sequence of JPEG images with a resolution of 512 x 384 pixels. The corresponding audio is stored as a mono, 16 bit, 32 kHz WAV file. Fig. 3 shows some frames from one sequence of VidTIMIT database. This database is publicly available for research purposes.

Figure 3: Example of frames from a sequence of VidTIMIT database

AVICAR (Audio-Video)
AVICAR [4] is a large, publicly available speech corpus database recorded in a car environment and contains audio and video recordings of 100 speakers (50 male and 50 female). The utternaces are all in English and consist of isolated digits, isolated letters, phone numbers, and short sentences. Data are collected through a multi-sensory array consisting of eight microphones and four video cameras under five different car noise conditions: idling, driving at 35 mph with windows open and closed, and driving at 55 mph with windows open and closed. The AVICAR database provides different challenges for tracking and extraction of visual features and can be utilized for analysis of the effect of nonideal acoustic and visual conditions on audio-video speaker recognition performance.

CRIM (Audio-Video)
CRIM [1] is large set of 591 face sequences showing 20 French Canadian individuals (10 males and 10 females) reading broadcast news for a total of approximately 5 hours. The database was originally collected for audio-visual recognition and there are between 23 and 47 video sequences for each individual. This database is publicly available for research purposes.

BIOMET (Audio, Video, Signature, Hand, Fingerprint)
BIOMET [2] is a multi-modal database including five different modalities: face, voice, fingerprint, hand and signature data. It was recorded to study the combination of several modalities for person identification. This database consists of three recording sessions (of 130 subjects in the first session, 106 in the second and 91 in the last one) spaced by three and five months. The face sequences were recorded with three different sensors: conventional camera, infrared camera and 3D acquisition system. The subjects were uttering in French their identification number, digits from 0 to 9, digits from 9 to 0, “oui”,“non” and 12 phonetically balanced sentences. Unfortunately, this database is not publicly available for research purposes.

BIOSECURE (Audio, Video, Signature, Hand, Fingerprint, Iris)
BioSecure[1] is a European project whose aim is to integrate multi-disciplinary research efforts in biometric-based identity authentication. The BioSecure database is collected by 11

[1]http://www.biosecure.info

university institutes across Europe and consists of three parts, simulating the use of biometrics in remote-access authentication via the Internet (termed the "Internet" scenario), physical access control (the "desktop" scenario), and authentication via mobile devices (the "mobile" scenario). The Internet dataset contains talking faces of over 1000 volunteers recorded through the Internet in 2-3 sessions (3rd session optional), where the period between the 2 sessions is minimum 2 weeks. The desktop dataset contains data (face, iris, speech, signature, fingerprint and hand modalities) of over 400 volunteers recorded with a PC in a laboratory in 2 sessions, where the period between each session is minimum 2 weeks. The Mobile dataset contains data (voice, face, signature and fingerprint modalities) of over 500 volunteers recorded with PDA and Laptop in both an indoor and outdoor environment. A weakness about the BioSecure database is that the right to use or distribute the data remains with the collecting European sites. This prevents even a modest distribution of the database for the research community.

SECUREPHONE PDA (Audio, Video, Signature)
The Secure phone PDA database [3] consists of data (face, speech and signature modalities) of 60 subjects (30 female and 30 male) recorded with a PDA in 3 sessions separated by at least one week. Each session comprised 2 indoor recordings and 2 outdoor. The 2 indoor recording (face and speech) conditions were "light-clean" and "dark-noisy". The 2 outside recordings were "light-noisy" and "dark-noisy". Handwritten signature conditions were always good. In order to test the effect of prompt length and prompt type, video recordings were made for 3 types of prompt (5-digit, 10 digit and short phrase), with 6 examples from each prompt type. This database has not been made publicly available for research purposes.

M3 CORPUS (Audio, Video, Fingerprint)
M3 (multi-biometric, multi-device and multilingual) Corpus [5] aims to support research in multi-biometric technologies for pervasive computing using mobile devices. The corpus includes three biometrics (facial images, speech and fingerprints); three devices (the desktop PC with plug-in microphone and webcam, Pocket PC and 3G phone) as well as three languages of geographical relevance in HongKong (Cantonese, Putonghua and English). The data was recorded in both an indoor and outdoor environment and includes 39 subjects in three sessions with at least one month interval between sessions. Unfortunately, this database is not yet publicly available for research purposes. Another drawback of the database lies in the limited number of subjects it contains.

3 Motivation for Recording the MOBIO Database

The MOBIO project aims to develop new mobile services which are secured by biometric authentication. To assess the effectiveness of any developed technique a multi-modal database is needed which fulfils the following criteria: it should be publicly available for

research purposes, recorded with mobile devices (and desktops or laptops for enrollment) in natural environments during a long time period (e.g., 2 years) and contain a large number of subjects with several recordings (shots) per subject. In addition, the audio and video data should be synchronized and instead of uttering simple fixed phrases the recordings should consist of rich text phrases or dialogues so that the system is less vulnerable to impostor attacks. Unfortunately no such database currently exists and so there is a great need to record a new database (the MOBIO database). During the recording phase, MOBIO partners can use BANCA dataset for training and testing the developed algorithms. Each partner may also consider other databases. Naturally, later developments and final demonstrations will be assessed on the MOBIO database. After recording, the MOBIO database will be publicly available to the research community. This will provide the research community with a realistic multi-modal database and enable a fair comparison of future multi-modal biometric authentication systems.

4 Recording Methodology

4.1 Recording Devices

The MOBIO database will be recorded on mobile devices which have the same characteristics to ensure data consistency. The chosen mobile device needs to be able to: record synchronised audio and video, allow for capture from the frontal camera and record speech from both the built-in microphone and from a headset. Once the mobile device has been chosen, IDIAP will then purchase these mobile devices and pre-install the recording software before distributing them to the partners. In some cases data will be captured on an alternate device, for instance in the bank order scenario (see Deliverable 2.1) the enrollment data is captured from a desktop, in these cases the recording devices will be kept relatively consistent. More details on the recording devices will be given in Deliverable 6.1.

4.2 Number of Subjects, Sessions and Shots

The MOBIO database will be recorded in two phases. During the first phase, only clients will be recorded while impostor will be simulated. This first part of the database will be recorded from August 2008 to March 2009. It has been agreed to collect MOBIO database from 160 subjects with a significant number of shots and sessions per subject. All partners (except EPM and IDEA) will participate in recording data and each partner is responsible for finding volunteers and ensuring their availability. Table.1. shows the minimum number of subjects, the local contact person at each recording site and the minimum number of male and female participants.

Four recording stages spaced by several weeks are planned during the first recording phase (i.e. from August 2008 to March 2009):

Recording Stage #1: This initial recording stage will consist of collecting data for

Site	Min. number of subjects	Min. number of female participants	Contact person
IDIAP	30	10	C. Mc Cool
UMAN	30	10	T. Cootes
UNIS	30	10	N. Poh
BUT	30	10	H. Cernocky
UOULU	20	7	T. Ahonen
LIA	20	7	D. Matrouf

Table 1: Number of subjects per site and contact persons

enrolment (1 session of 2 shots from both a laptop and a mobile), free speech data (1 session from a mobile), as well as the first test session (2 sessions of 2 shots from a mobile).

Recording Stage #2: The second recording stage will consist of collecting a second test session (1 session of two shots).

Recording Stage #3: The third recording stage will consist of collecting a third test session (1 session of two shots).

Recording Stage #4: The final recording stage will consist of collecting a forth test session (2 sessions of two shots).

The recording schedule and milestones are described in Section 4.7.

4.3 Ethnic Diversity, Age and Gender Balance

Ethnic diversity, age balance and gender balance are important considerations for capturing a database. The issue of ethnic diversity is addressed by having institutes from several countries capturing the data, this means that there will be at least some cross European diversity in terms of faces and accents. To ensure a balanced collection of data in terms of gender (male/female) and age of the subjects each MOBIO partner is requested to have at least 30% of males and 30% of females and to consider different categories of ages (18-65).

4.4 Environmental Conditions of the Recording

The database will be collected in conditions which reflect the environment that the system will be operating in. The environment condition will depend on the particular scenario, however, in all cases the illumination and facial expressions will not be controlled. It is noted that since the system will be interacting with cooperative users then it can be assumed that the faces will be frontal (or near frontal). For speech recording, both the built-in microphone of the mobile and headset will be used in order to experiment with both noisy and clean conditions.

4.5 Dialogue Manager for Speech Recording

There are several ways that speech, for speaker recognition, can be recorded. Most of the existing systems use fixed phrases for speaker recognition. Fixed phrase based systems are

generally trained on the phrases that are also used for testing which makes them vulnerable to impostor attacks because an impostor can easily play a recorded phrase uttered by the user. Alternatively, a system can ask the user to utter a sentence which is not seen in the training phase. These systems are less vulnerable to impostor attacks but need interfaces (or dialogue managers) for prompting sentences. A third category of systems use free text for testing which makes them difficult to train and therefore may be vulnerable to impostor attacks.

In MOBIO, free speech and dialogue oriented scheme will be considered. The language of all audio recording will be English.

Dialogue Manager

To collect the data a method for prompting the user is needed. To achieve this in the most natural setting there will be a dialogue manager (DM) which will prompt the participant (P), either by asking specific questions about his or her fake name and fake age or by asking open ended questions. For instance, every subject may have to answer 10 questions randomly taken from a predefined list of 100 questions. The goal is to collect about two minutes of quasi-spontaneous speech for each speaker. The subjects will be elicited to give rather longer answers (of about ten seconds) but with no special guidelines. The answers to the questions may vary across speakers. A preliminary list of questions are given in Table.2. The DM is responsible for presenting the questions to the participant. These questions will be decided upon prior to the data collection and will depend on whether free speech or a prompted response is required.

For specific scenarios such as Kindergarten and bank order, the Dialogue Manager will provide with simulated conversations between the participant and the representative of the bank or the Kindergarten. For instance, in case of the Kindergarten, the Dialogue Manager will provide with the following simulated conversation between the teacher, replaced by the DM, and the parent, replaced by the participant (P), who needs to verify his or her identity.

DM: Hello, this is Ms Watson, KinderGarten "Behind the Corner"
P: ... introduces him/herself and formulates the request about the kid (for example another person will pick him/her this afternoon).
DM: I am sorry I am new in the classroom, would you repeat the name of your kid and describe him to me?
P: ... gives the details.
DM: what was actually your request?
P: ... repeats the request.
DM: This is a serious thing and I have to verify your identity, would you tell me if you enrolled in the system, with whom and under which conditions?
P: ... answers.
DM: I still need to hear more from you, would you give me details on who brought

the kid in the morning and when?
P: ... answers.
DM: Is there any other way to verify that I am really speaking with the right person?
P: ... tries to find out
DM: And can you tell me why you didn't give my colleague a signed paper in the morning, as it is requested by the manager?
P: ... desperately tries to find out ...
DM: OK, one last question, give me some secret that only you and he can know. I will ask him.
P: ... desperately tries to find out ...
DM: Thank you, your identity was checked, the request was accepted, please next time let us a written note.
P: ... apologizes and says good bye.

The descriptions of other scenarios are given in Deliverable 2.1.

4.6 Impostor Attacks

During the first phase, only clients will be recorded while the impostor attacks will only be simulated. Realistic and various scenarios for possible impostor attacks will be considered during the second phase of data recording.

4.7 Outline of Recording Plan and Schedule

The MOBIO database will be recorded in two phases. The recording is planned to start on the 25th of August 2008 and all partners (except EPM and IDEA) will participate. An IDIAP representative will be responsible for coordinating data collection and validation with the local support of partners. The IDIAP representative will travel to each recording site twice. On the first visit they will assist in initiating the recordings and on the second visit they will assist with finalising the data collection. The initial recording phase will consist of collecting data for enrollment (1 session of 2 shots from both a laptop and a mobile), free speech data (1 session from a mobile), as well as the 2 first test sessions of 2 shots from a mobile. The final 2 test sessions (of 2 shots from a mobile) will be partially captured before the second visit of the IDIAP representative. Within this interval every partner will need to collect 2 sessions of 2 shots. Note that the speech will be recorded both from the built-in microphone of the mobile and from a headset.

The IDIAP representative will be going to each site to assist the people who will be collecting the data. This means that the representative will help run the first few enrollment captures at each site and help to run the first few test captures at each site. However, finding volunteers and the bulk of the data capture from mobiles will be the responsibility of each partner. The role of the IDIAP representative for the first week is as follows, they will

- collect the original signed consent form for each participant,
- assist with setting up and initiating data recording on mobile, and
- perform data capture for laptop enrollment to ensure consistent data capture.

The partners collecting the data (IDIAP, BUT, UNIS, OULU, UMAN and LIA) will provide their list of participants, with a copy of the signed consent form, by the beginning of June 2008. The original of the signed consent form will be collected by an IDIAP representative during the first week of data collection.

Each partner has been assigned one week where all the signed participants will need to be available for the beginning of data capture. On this week, IDIAP will send a representative to the partner's site to assist with the beginning of the data collection. Prior to the start of data collection each partner needs to:

- have at least one person who is responsible for data collection and submission,
- provide a full list of participants, and
- provide copies of the signed consent forms for each participant.

Figure 4 highlights the schedule of the IDIAP representative along with the deadlines for when each partner will need to submit the recorded data for a specific set of collections.

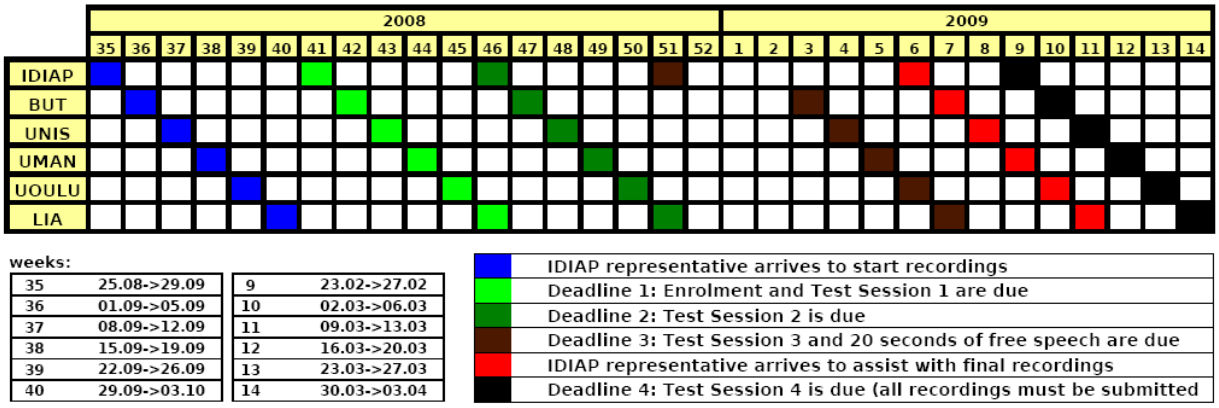


Figure 4: Recording schedule and deadlines for submitting recorded data at each site

5 Technical Specifications

5.1 Recording Format

The audio and video recording options and formats depend on the chosen mobile device and recording tools. Once the choice of the mobile device is done, EPM will provide input

about the recording formats, however, it is anticipated that the video will eventually be encoded as an AVI file and that the audio will be encoded as a "wav" file. The details such as the number of frames per second for video and the audio sampling frequency for speech will depend on the device but it is anticipated that the video will consist of approximately 15 frames per second and the audio sampling frequency will be 16 kHz. These details will depend on the performance of the chosen device and more details will be provided in deliverable D6.1.

5.2 File Naming Convention

To facilitate the processing and interpretation of the data, one should adopt meaningful and self descriptive file names. For MOBIO database, we will adopt the following convention for naming the files:

PersonID_Recording_ShotNum_Conditions-Channel.(avi/wav)

where,
PersonID = Gender + Institute + ID
Recording = Session
ShotNum = Shot
Conditions = Environment + Device
Channel = ChannelID
(.avi/.wav) = .avi for video .wav for audio

and
Institute: 0=Idiap, 1=Manchester, 2=Surrey, 3=Oulu, 4=Brno, 5=Avignon
Gender: m=Male, f=Female
ID: from 01 to 99 for each site
Session: x= test session, e= enrollment, f= free speech + ID from 01 to 99
Shot: ID from 1 to 9
Recording device: 0=Mobile, 1=Laptop
Environment: i=Inside, o=Outside
ChannelID: ID 0 to 9 (0 - first video/audio channel, 1 - second video/audio channel)

For example, the file name of *the second shot of the fourth test session of the subject (male) number 13 from UMAN taken on a mobile phone in an inside environment* is: **m113_x04_2_i0-0.avi**

Note that other useful information such as mother tongue, age and dialogue number will be stored in meta-data files.

5.3 Data Validation

Data validation is an important aspect which ensures that all modalities have been correctly recorded. The process of data validation is made quick and easy by having a set of automated data validation tools. These tools will be provided to each site who will then perform data validation on the data they collect. The data validation tools will include speech detection, face detection and sound quality verification (to ensure the audio recording has not been clipped). By performing data validation at each site any corrupted data can be quickly and easily fixed (usually by recording the data again). Once validated, the data will be uploaded to IDIAP where final data validation will occur. In the case of validation failures a new recording will be requested by IDIAP.

6 Legal and Privacy Issues Related to Database Distribution

The MOBIO database will be publicly available to research community. This will provide the research community with a realistic multi-modal database and will enable a fair comparison of future multi-modal biometric authentication systems. The database will be distributed with the several utilities. These utilities include the acquisition tool (if EPM agree) and baseline systems provided by the partners. Consent form (like the one shown in Fig. 5) will be signed by all participants prior to the start of the recording. Only people who sign the Consent form will be included in the MOBIO database. For legal and privacy issues the distributed MOBIO database will not contain any other personal information (e.g. names, addresses, phone numbers etc.) of the participants.

7 Summary

This document described specifications of the phase 1 of MOBIO database recording and explained the specific choices that we have adopted in designing the database. A summary is given in Table.3.

Acknowledgments

I would like to thank, in alphabetical order, T. Ahonen, H. Cernocky, B. Crettol, S. Marcel, C. Mc Cool, M. Pietikäinen, and N. Poh for their contributions.

References

[1] CRIM. http://www.crim.ca/.

[2] J. Kittler and M. S. Nixon, editors. *Audio-and Video-Based Biometrie Person Authentication*, volume 2688 of *Lecture Notes in Computer Science*. Springer, 2003.

[3] J. Koreman, A. Morris, S. Jassim, H. Sellahewa, G. Chollet, G. Aversano, S. Salicetti, and L. Allano, editors. *Multi-modal biometric authentication on the SecurePhone PDA*, 2006.

[4] B. Lee, M. Hasegawa-Johnson, C. Goudeseune, S. Kamdar, S. Borys, M. Liu, and T. Huang. Avicar: Audio-visual speech corpus in a car environment. In *Proc. Conf. Spoken Language*, 2004.

[5] H. Meng, P. C. Ching, T. Lee, M. W. Mak, B. Mak, Y. S. Moon, M. H. Siu, X. Tang, H. Hui, A. Lee, W. K. Lo, and B. Ma, editors. *The Multi-Biometric, Multi-Device and MultiLingual (M3) Corpus*, 2006.

[6] K. Messer, J. Matas, J. Kittler, J. Lüttin, and G. Maitre. XM2VTSDB: The extended M2VTS database. In *Audio- and Video-based Biometric Person Authentication, AVBPA'99*, pages 72–77, 1999. Washington, D.C., March 1999. 16 IDIAP–RR 99-02.

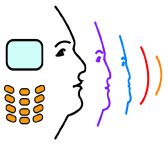
[7] S. Pigeon and L. Vandendorpe. The m2vts multimodal face database (release 1.00). In *AVBPA '97: Proceedings of the First International Conference on Audio- and Video-Based Biometric Person Authentication*, pages 403–409, London, UK, 1997. Springer-Verlag.

[8] V. Popovici, J. Thiran, E. Bailly-Bailliere, S. Bengio, F. Bimbot, M. Hamouz, J. Kittler, J. Mariethoz, J. Matas, K. Messer, B. Ruiz, and F. Poiree. The BANCA Database and Evaluation Protocol. In *4th International Conference on Audio- and Video-Based Biometric Person Authentication, Guildford, UK*, volume 2688 of *Lecture Notes in Computer Science*, pages 625–638. SPIE, 2003.

[9] C. Sanderson and K. K. Paliwal. Noise compensation in a person verification system using face and multiple speech feature. *Pattern Recognition*, 36(2):293–302, 2003.

Culture	What was the last movie you saw?
	What was the last book you read?
	Are you listening to classical music?
	What is your favorite CD?
Education	What was your favorite subject at grammar school?
	Who was your favorite teacher?
	Would you like to study abroad?
	Do you think the state is doing enough for the schools?
Politics	What do you think about the last electoral campaign?
	Who is you favorite politician?
	Do you think EU is a good thing?
Transport	Which public transport do you prefer?
	How do you think are the drivers in this country?
	How could be your ideal car?
Weather	Describe the weather yesterday.
	Which weather do you like the most?
	Do you personally experience global warming?
Education	Imagine a restaurant you would like to go to.
	If you like cooking, what is the favorite dish? if not, why don't you like it?
	Describe the food you would never take into your mouth
Family	Which member(s) of your family do you like the most?
	Why do you think the population in Europe declines?
	In what you're the most different from your parents?
Sports	What is your favorite sport?
	Do you think the soccer player merrit the money the earn?
	Are there enough sport facilities in your city?
Society	What do you think about reality shows running on the TV?
	Do you like the singers contests?
	Would you like to be a celebrity?
Environment	What do you see around you?

Table 2: A preliminary list of questions which will be used for prompting the user to elicit non-scenario speech



MOBIO Consent Form

MOBIO project partners are conducting advanced multi-modal research projects. In this context, they collect and exchange data and research results.

Multi-modal research requires large amounts of acoustic recordings of spoken language, along with video, and other multi-modal data recordings. MOBIO partners intend to compile such a corpus. This corpus will include a large number of non-native English speakers and will therefore be unique from those compiled by other institutions.

We are asking for your assistance in collecting a multi-modal database. This database will be collected using a mobile device (primarily a mobile phone). By signing this form you authorise the inclusion of your data within this corpus.

The recorded data will initially be used by the MOBIO project partners for research purposes. It is possible that some or all of the data will be made available to the wider research community, in both transcribed and digitised formats. Packaging and distribution of the data may be entrusted to specialised providers. Data may be put online for access in connection with research activities.

Please note that you remain solely responsible for the content of your statements and behaviour. Please avoid defamatory and otherwise derogative statements. If you are concerned about the content of your personal input, please advise us immediately.

By signing this form, you agree to allow the recorded data to be used without limitation in accordance with the above statements.

I, (please print name) .....

have read and understood this form and agree to authorise use of the recorded data on the terms indicated.

Signature: ..... Date: .....

Age (optional): ..... Sex: .....

Are you a native English Speaker? .....

Yes, please indicate your country and region: .....

.....

No, please indicate your native language: .....

.....

How long have you spent living in an English speaking country? .....

.....

Please list any other language influences (other languages spoken, dialects, etc)

.....

Please Provide your email address (or other contact information) so that we can contact you if necessary

.....

MOBIO project partners see http://www.mobioproject.org/partners

Figure 5: MOBIO Consent Form

Name:	MOBIO Database
Modalities:	Face and Speech
Recording Devices	Mobile device (see Deliverable D6.1) and laptop
Recording Period	Phase 1: August 2008-March 2009 Phase 2: March 2009 -
Responsible of Recording	IDIAP with active local help of partners
Sites of Recording	IDIAP, UMAN, UNIS, BUT, UOULU, and LIA
Minimum Number of Subjects per Site	IDIAP (30), UMAN(30), UNIS(30), BUT(30), UOULU(20), and LIA(20)
Total Number of Subjects:	160
Number of Sessions	Phase 1: 12; Phase 2: to be determined later
Speaker Recognition	Dialogue based
Audio-Video Recording	Synchronised
Audio Recording	from both built-in microphone and headset
Language(Audio)	English
Ethnics, Gender and Age	Balanced
Recording Environment:	Natural, indoor and outdoor
Text (Dialogue)	See Deliverable D2.1
Recording Format	See Deliverable D6.1
File Naming Convention:	PersonID_Recording_ShotNum_Conditions-Channel.(avi/wav)
Impostor Attacks	Simulated in phase 1 and studied in phase 2
Data Validation	Yes (using basic tools)
Public Distribution of the Database	Yes (and consent forms should be signed by all subjects)

Table 3: Summary of MOBIO Database Specifications