The M2VTS project: towards Multi Modal Verification for Teleservices and Security Applications

G. Richard[♣1], Y. Menguy, J. Boudy, F. Rigoulet, P. Lockwood,
Matra Communication, Rue JP Timbaud, 78392 Bois d'Arcy, France
D. García-Plaza, Fernando Fernández,
Ibermática, Avenida del Partenon, 16-18, Campo de las Naciones, 28042 Madrid, Spain
C. Kotropoulos, I. Pitas,
Aristotle University of Thessaloniki, PO BOX 451, Thessaloniki 54006, Greece
P. Ryser,
Cerberus, CH 8708 Mannedorf,Switzerland
C. Beumier,
Renaissance, Electrical Engineering Dept.
Av. de la renaissance, 30, B 1040 Bruxelles, Belgium

1. Introduction

The objectives of the M2VTS project is to address the issue of secured access to local and centralised services in a multimedia environment. The major problem in user authentication is to achieve on the one hand toll performance : false acceptance rate as low as possible (minimise access to impostors), and false rejection rate as low as possible (a registered user should access to his system in any case), and on the other hand stand the wide range of conditions of use of such systems as well as provide ergonomically viable solutions. The use of multiple modalities will help to overpass potential technical limitations of individual modalities as well as take benefit of the emerging multimedia environment : workstations, network computers, smart phones are more and more equipped with audio and video capabilities. The research is driven by the application needs and user requirements. Therefore, work has essentially been driven by three main goals :

1. Develop platforms for evaluation, implementation and fast prototyping of technology. Submit these platforms to user tests in real situations, in order to measure the adequation between user requirements and current maturity of the technology.
2. Develop algorithmic solutions for user authentication in a multimodal context. Implement these solutions on the software platforms for fast prototyping. Refinements of the algorithms based on the results of the user tests in real situations
3. Develop prototypical applications for end users.

This paper is organised as follows: the next section will include a description of all prototyping platforms and field test results for two of them. In section 3, a general scope and some anchor points in the litterature are given concerning the innovative Multimodal Verification techniques developed within the project. In section 4, a short comment is given concerning the current elaboration of a large multimodal databases. Section 5 describes three of the main multimodal applications that the project should deliver.

2. Prototyping platforms

The aim of such platforms is to allow a fast prototyping of new algorithms and to test them in real conditions. As a matter of fact, there is often a dramatic decrease of the authentication performances between a system placed in ideal conditions (clean speech, white background, uniform lightning, image centered and focused, correct zooming,….) and the same system placed in adverse conditions. Often, even if only one of the ideal

♣ corresponding author: gael.richard@matra-com.fr

condition is not fulfilled the performances are significantly affected. It is thus very important on the course of algorithm refinement and optimisation to integrate this variability. Furthermore, these prototyping platforms allow to integrate in the process of algorithm optimisation important information related to the acceptance of the systems such as unacceptable constraints (high verification time for instance). Finally, it is an essential tool to build a "real condition" multimodal database that will allow to work on the algorithm robustness against changing and sometimes difficult environmental conditions.

2.1 Demonstrators

A flexible software and hardware platform has been realised. From this platform, three main demonstrators have been developed, evaluated and installed at several end-user sites. These demonstrators are ordered in Levels. Level 1 represents a system with a single modality[1] for authentication, where at the opposite a level 3 system contains several authentication experts in each modality (speech and image).

2.1.1 The Level 1 demonstrator: A network-based voice mail system

This demonstrator performs secured access to a Voice Mail system. The User claims his identity by saying his full name. Verification is performed by using a single modality : the so-called Password modality (speaker-dependent speech recognition, Matra Communication technology).

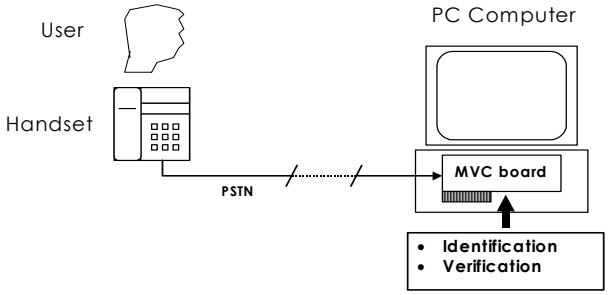

Figure 1: Physical Implementation of the Level 1 Demonstrator: The Demonstrator runs on a Windows-95 based PC. The PC includes a Matra MVC voice acquisition and recognition board which is connected to the phone network.

The Application scenario is the following in Access mode :

1. The user calls the Voice Mail system.
2. He can then claim his identity by saying his full name. He has 6 chances to say it correctly. If none of the utterances are correct, he must exit. But, if the full name is recognised, a message is played to indicate that the name has been recognised.
3. Password verification is performed. The user has 6 chances to say his password. The system computes a score which reflects the matching between the recognised password and the user's password. If none of the utterances is correct, he must exit. But, if the password is recognised, a "successful access" message is played.
4. After the Password is accepted, the user has access to his voice account.

2.1.2 The Level 2 network-based voice mail system

This demonstrator targets the same application, i.e. secured access to a Voice Mail system. However, two other verification algorithms for the speech modality have been included:

[1] A modality is associated to a mode of communication . In our case, i.e. for verification it could be either speech or image. In each of this mode however several techniques could be used in parallel: speech recognition, text dependent/ text independent speaker verification for the speech modality and front face verification, profile verification, lip movement,…. For the image modality.

- the IDIAP text-dependent (TD) modality ([1]), for which the user pronounces the 5-digit code of his mailbox, spelt digit by digit.
- the IDIAP text-independent (TI) modality), for which the user pronounces his professional address.

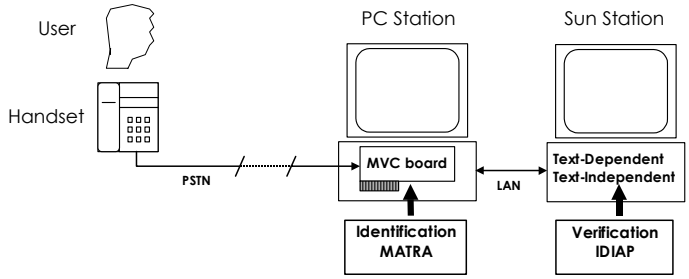

Figure 2: Physical Implementation of the Level 2 Demonstrator: The application runs on a PC under Windows-95. The PC includes a Matra MVC voice acquisition and recognition board, which is connected to the phone network. The IDIAP algorithms run on a remote Sun Sparc, connected to the PC through a LAN network.

The application scenario for the access mode is the following:

1. The **user dials the phone number** giving access to the system.
2. The **user claims his identity** by saying his full name. If the system does not recognise a known name, the user may repeat his name up to a maximal number of trials (6 trials). Whenever a known name is recognised, the system starts the text-dependent voice verification (step 3). If the maximal number of trials is reached, the system hangs up.
3. In the text-dependent verification step, the **user says his code, digit by digit**. The system checks the code by speech recognition. If the recognised code is the user's one, the system checks the recorded speech against the user's voice model. The verification results (code OK, similarity measure) are stored and text-independent verification is started (step 4).
4. For the text-independent verification, the **user says his address**. The system checks the recorded speech against the user's voice model with the text-independent method. The similarity measure is stored.
5. The **system** computes the decision based on the verification results of individual modalities and **notifies the user whether his access is accepted or rejected**.

2.1.3 The Level 3 demonstrator: A complete multi-modal verification system

This application implements access control to a door. Verification is performed by use of two modalities : speech (by means of a password and speaker-dependent speech recognition from Matra Communication technology) and image (front and/or profile images); the scores obtained by both modalities are fused in a weighted sum. The user claims identity by presenting a badge.

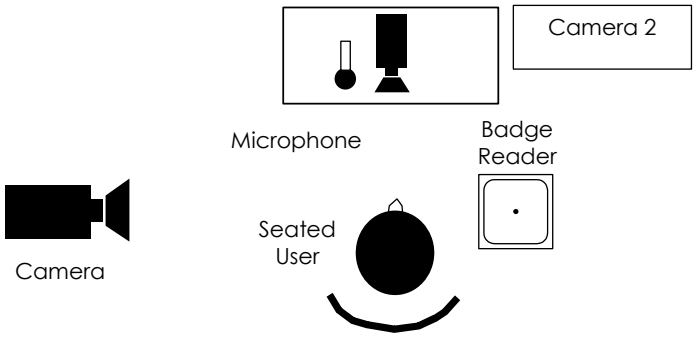

Figure 3: Physical Implementation of the Second Level 3 Demonstrator: The Demonstrator runs entirely on a Windows-95 based PC, with a Matra MVC voice acquisition and recognition board (connected to a microphone and speaker), a Matrox

Meteor video grabbing board (connected to an autofocus B&W camera and the computer screen). The PC is also connected to a Matra Security Badge reader and pilot system.

The typical application scenario is the following (in Access mode) :

1. The user sits on a chair and adjusts its height in order to be properly positioned in the display window on the system screen.
2. The user presents his badge which is recognised by the badge reader. If the person is an impostor, he has first notified the Application through the GUI (for statistical purposes !). The user can also manually type the claimed ID number through the GUI if he is not currently carrying a badge.
3. Password verification is performed. The user has 3 chances to say something recognisable by the system. The system then computes a score which reflects the matching quality between the recognised password and the user's password. Then the user is informed by a double beep that Password verification is completed (but not of whether the correct password was recognised).
4. Front face verification is performed and eventually profile verification.
5. The speech and image scores are fused for access authorisation. A red or green light indicates authorisation status.

2.2 Field tests results

Field tests were carried out on all three demonstrators. We provide below the results obtained with the Level 1 system and initial results on the Level 3 system obtained with the latest technology in Image and speech recognition developed within the project.

2.2.1 Results on the network based voice mail (level 1)

2.2.1.1 Normal access

The results given below indicate the performances of the Level 1 system for the users who tried to enter their voice account (successful full name and successful password), in normal conditions. A successful call means that both the full name identification and password verification stages were successful.
The results are given for numbers of trials between 1 and 5. It means that the user can say his full name and password at most 5 times if necessary. For example, if the number of trials is 2, we will count the number of calls that necessitated one or two repetitions for the full name and/or the password. So, with those data, computing False Rejection and True Acceptance rates for each case is possible and very interesting : it is a way of finding the optimum number of trials that should be authorized for a product using this modality.

Number of trials	**Secondary results (without *EE**) ⇨ 107 calls**		
	Successful calls	**FR rate**	**TA rate**
1	**98**	**8,5%**	**91,5%**
2	**106**	**0,94%**	**99,06%**
3	**106**	**0,94%**	**99,06%**
4	**107**	**0**	**100%**
5	**107**	**0**	**100%**

Tableau 1 : Normal conditions : FR and TA rates (without Electrical Echoe (EE[1]))

[1] By analyzing the primary results we noticed that they were spoiled by Electrical Echoes (EE), a known phenomenon in telecommunication. This electrical echo has a negative effect on voice recognition : a password that has been correctly pronounced by a user can be rejected because of EE while it would have been accepted without that problem of EE. These secondary results don't take into account the calls spoiled by EE: (20 out of 127 calls).

These results clearly show that at least 2 trials are necessary to achieve very good performances.

2.2.1.2 Impostor access

It must be said that, during the test sessions, the impostors knew the name and password of the user's voice account they were trying to have access to. In that sense, users had more information than most of the impostors would have had in real life.

The results given below indicate the performances of the system against impostors. In the table shown below, two cases are distinguished.

- the stage of the full name is successfully passed
- the stages of the full name and of the password are successfully passed.

This distinction is interesting in the case of imposture accesses since the quality of True Rejection for each stage can be measured. The results are given according to the number of trials permitted. As a matter of fact, it is particularly interesting to evaluate the robustness of the algorithm by letting the impostors a high number of imposture trials.

In impostor conditions, we count:

- Total number of calls : 194
- Number of impostors : 12

Number of trials	Number of calls (Name OK)	Number of calls (Name & Pwd OK)	FA rate	TR rate
1	**15**	**5**	**2,57%**	**97,43%**
2	**16**	**6**	**3,0%**	**97,0%**
3	**17**	**8**	**4,1%**	**95,9%**
4	**18**	**9**	**4,6%**	**95,4%**
5	**21**	**9**	**4,6%**	**95,4%**
6	**26**	**11**	**5,6%**	**94,4%**

Tableau 2 : Impostor tests

In this table, we notice that 11 out of 194 calls have let impostors to have access to a voice account. It represents only 5,6% in the worst case (six trials).

The fact that the user may try six times in a row to access another user's account doesn't 'help' very much: the FA rate is only twice the figure for one trial (5,6% against 2,57%). We can extrapolate and say that if it happens that an impostor breaks the system, it will be done very quickly or not. But in all cases, the False Acceptance rate is kept low. The system is globally very good against impostors : only 2,57% of False Acceptance for one authorized trial and 5,6% for six authorized trials.

This results table shows also that two stages are necessary for a well secured access to a Voice Mail system. But the only full name stage protects the system very strongly: 15 calls out of 194 for 1 trial and 26 calls out of 194 for 6 trials can access to the password stage. The password stage is a good complement to the full name stage because less than half of them were able to pass the password stage.

2.2.2 Results on Multimodal verification system (level 3 with project developed technology)

First Field tests result were obtained on two different implementation and configuration of level 3 systems and were based on background multimodal verification techniques. General trends showed the Level 3 system was not well accepted by users (which was not the case for level 1).

Users felt that the constraints were too high (response time, performances) compared to the service provided (Secured access). However, it is very importance to say that this demonstrator was based on background technologies. The initial field tests results thus clearly showed the need for innovative multi-modal verification techniques.

We give below, initial results on a Level 3 demonstrator concerning the image modality for three pre-selected algorithms among the techniques developed within the project (the algorithms have been called A, B and C since further testing are needed in order to give fully relevant results). These results are obtained on a rather small "real condition" database (17 people). For each of these persons, **one** training image has been acquired and, a few days later, **one** verification image (all images have 256 grey levels, and their size is 144x192). In summary, 17 authentic tests, and 272 imposture tests have been performed. The results are depicted in Figures 4, 5 and 6. It is clear that the results are only preliminary in nature but they a first idea about algorithms performances. For these field tests, optimal conditions have been chosen (uniform white background, uniform lighting, face centered in the image, fixed zoom). The unsmooth shape of the FFR curves is due to the low number of authentic tests. Therefore, the evaluation of the Equal Error Rates is not very precise yet :

- 12% for the algorithm A
- 7% for the algorithm B
- 3% for the algorithm C

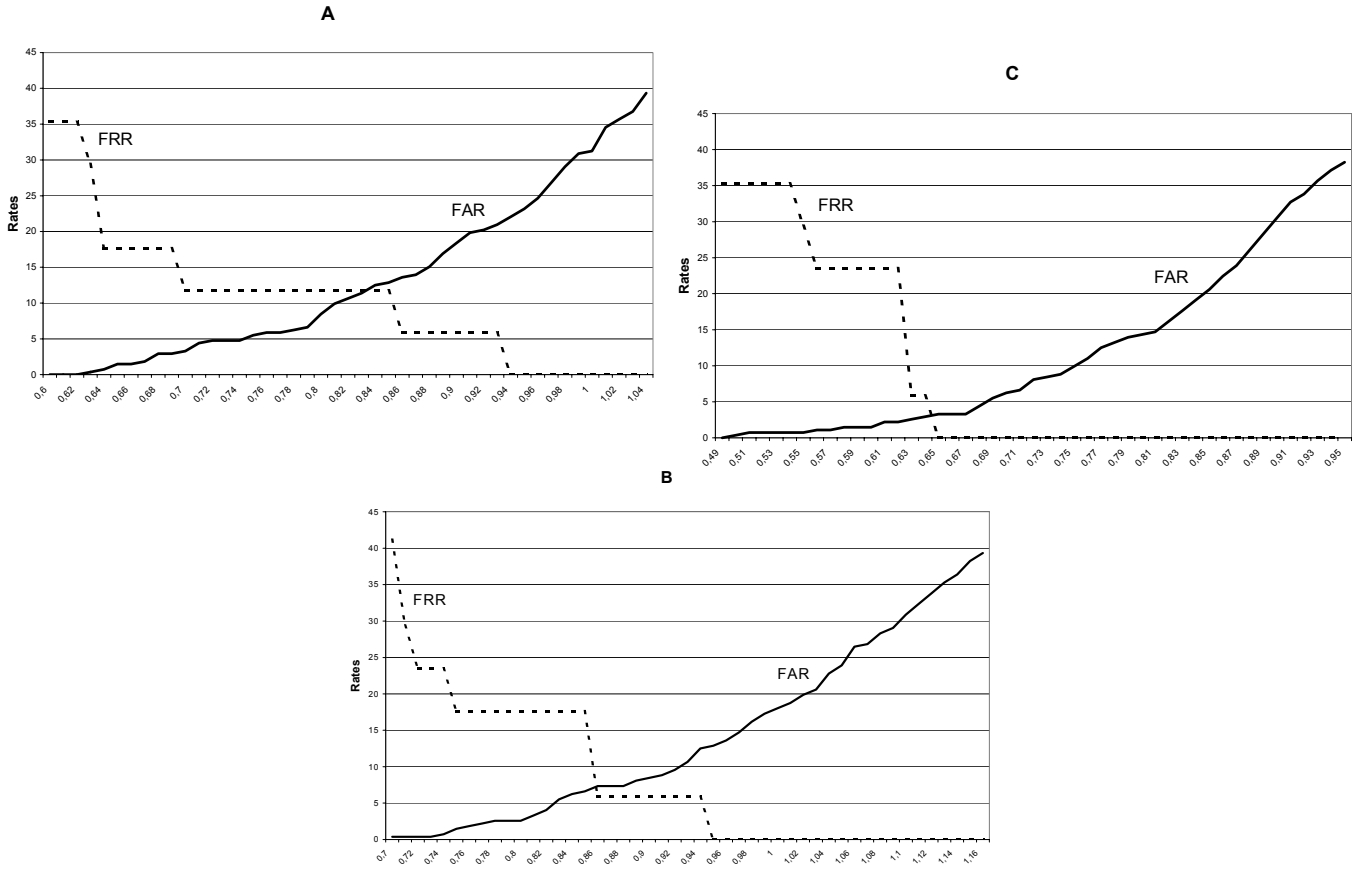

Figure 4 : Preliminary field test results from three M2VTS algorithms for Image verification

Since these algorithms already achieved far better performance than the previous technology used, there is no doubt that the acceptance by users will be significantly higher.

3. Innovative Multimodal Verification Techniques

Concurrently to the development of the different platforms, work was carried out to develop innovative multi-modal verification techniques. Speech and face recognition have exhibited a tremendous growth for more than

two decades. A critical survey of the literature related to human and machine face recognition is found in [4], and in [2,5,6,7,8] for speech recognition/speaker recognition. The purpose of this paper is not to give a detailed overview of the techniques developed in the project but rather to give the general scope and some important anchor points in the litterature. The key techniques introduced include:
Frontal face recognition algorithms with very low error rate (4.9% to 7% Equal Error rate (ERR) on a database of 37 persons for the best algorithms). Most of these techniques run very efficiently (less than a few seconds on modern processors) (for further information one may consult: [9,10,11,12,13,14]

- Profile recognition with very low error rate on "ideal conditions" images (7% ERR) [15].
- Lip tracking techniques [16,17]
- Speech verification techniques leading to very low error rates (less than 5% ERR). ([2],[3],[20][21][22][23]).
- Fusion techniques that permit to achieve multi-modal verification with Equal Error Rate as low as 0.7% (obtained on a database of 37 persons) [15,18].
- Facial surface analysis by 3D capture and analysis [19].

All Equal Error Rate given above were obtained on a database of 37 persons where a full description could be found in [10].
Some of these techniques have now been integrated in the fast prototyping platforms and are currently tested in real conditions (see section 2.2.2 for preliminary results). These tests already permit to iterate a back and forth collaboration between industrial and academic experts in order to optimise and to enhance the robustness of the algorithms in real conditions.

4. Multimodal Databases

The project already deliver a short multi modal database providing frontal images, profiles images along with speech for 37 persons [see [10] for a detailed description). To better assess the performances of the verification algorithms, an extended multimodal database is currently being recorded (300 persons, 20-30s of video sequence (head rotation from left to right), 10 s of speech) in clean condition (blue background, controlled lighting, etc…). A more detailed description will be included in the final paper.

Concurrently, "real conditions" databases are also built during field tests. In particular, situations such as non uniform lighting, smiling faces, or non centered images are represented.

5. Applications.

Nowadays there exist several applications of person recognition and/or verification. These applications are mainly based on the use of biometrical information (recognition of hand geometry, recognition of fingerprint patterns, recognition of eye iris,...), and on the use of identification codes (PIN) and/or passwords, but only using one of these techniques.
In the framework of M2VTS, a set of verification applications that use multi-modal information are being developed (voice, image, PIN codes,..). They cover the following functionalities:

- Control of accesses to information systems and teleservices.
- Access control to secured areas or buildings.
- Alarm verification.

These applications are being developed and will be tested by end-users of the project : U.T.A.P. (Auxiliary Technical Unit of Policeman), the B.B.V. (Bank Bilbao Vizcaya) and C.E.T. (Compagnie Européenne de Télésécurité).
Each of these applications is briefly described below.

5.1 Control of Accesses to Information Systems and Teleservices.

The aim of such an application will be to control and regulate the access of different users to secured information systems (Access to servers, access to computer nets, private sites,... [Fig. 4]), and to control the access to teleservices (automatic cashiers, electronic trade,... [Fig. 5]). The security level of the application will have to be the possible maximum, with an objective of 0% of false acceptances..
Nowadays, the most developed way to secure a system is by using a password or by identifying the machine. We propose to compare information of the face and the voice of the user, with the identity information provided by the password.
The system will include a camera (like a videoconference camera), that will take the frontal image of the face of the user, and a microphone to record the voice of the user. The application is based on verification which means that the user will claim a certain identity and the application will diagnose if the user is an impostor or not.

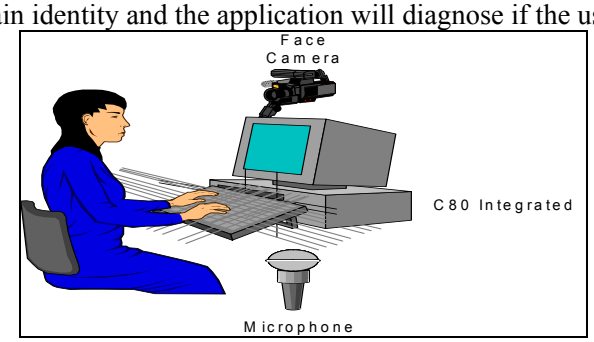

Figure 5 : Person verification for access to secured Information systems

The procedure to complete an access trial is the following:

- The user introduces in the system his identification code through keyboard or magnetic card
- One or several photos of the users are taken. Those images are either transmitted to the Information site where facial features are extracted and compared to the one obtained from the training of the claimed user. The other possibility is to extract the facial features on the local server or computer before sending them to the Information site.
- The user is also asked to pronounce a short sentence. As above, the speech signal is either transmitted to the Information site where voice features are extracted and compared to the one obtained from the training of the claimed user. The other possibility is to extract the voice features on the computer or (local station) before sending them to the Information site.
- Finally, the system gives an access authorization based on the fusion on both verification algorithm (face recognition and speech recognition)

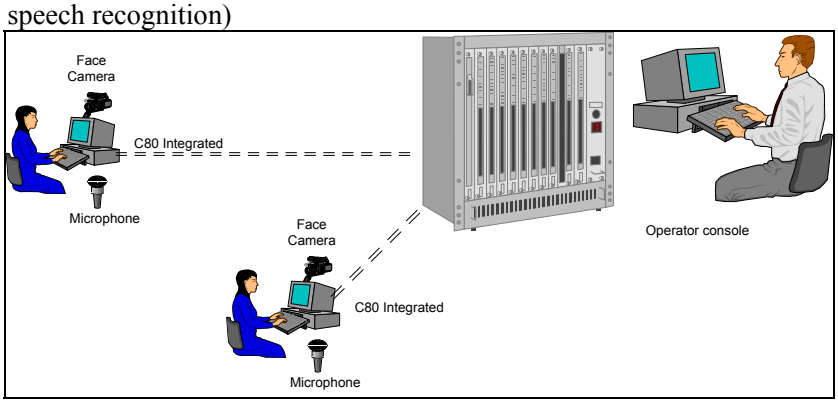

Figure 6: Teleservice application (such as automatic cashiers, banking through Internet, electronic trade…)

This application will be installed in the headquarters of the U.T.A.P. (to access to the private computer net), and in some services of the B.B.V. (automatic cashier, banking through Internet).

5.2 Access Control to secured Building.

The objective of such an application is to control the entrance of secured buildings (central headquarters of banks, Companies,…) or restricted areas (CPD,...). In this application it will be very important to keep the security level, near to 0% of false acceptances.
The system includes a camera, that takes the frontal image of the face of the user, a microphone to record the voice of the user, and a magnetic badge (or card) reader to claim an identity [Fig. 6]).

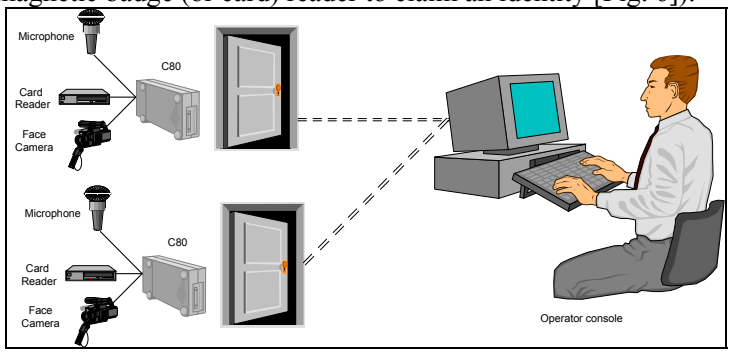

Figure 7 : Figure 3.- Access Control secured Building

The complete procedure to complete a access trial is the following:

- The user introduces in the system his identification code through a magnetic card or badge
- One or several photos of the users are taken. Those images are either transmitted to the Central station where facial features are extracted and compared to the one obtained from the training of the claimed user. The other possibility is to extract the facial features on the local server or computer before sending them to the Information site.
- The user is also asked to pronounce a short sentence. As above, the speech signal is either transmitted to the Information site where voice features are extracted and compared to the one obtained from the training of the claimed user. The other possibility is to extract the voice features on the computer or (local station) before sending them to the Information site.
- Finally, the system gives an access authorization based on the fusion on both verification algorithm (face recognition and speech recognition)

This application will be used to control and regulate the access to the central headquarter of the BBV and to control the entrance to the U.T.A.P. department.

5.3 Application for Alarm Verification

This application is designed for Alarm Verification in situations such as home intrusions by thieves or any intruders in secured building. In order to lower the number of false alarms, this application is built around a broad range of application softwares, including sustained signal, speech, and image processing and communication standards.
The main functionalities of this platform are:

Image and Sound Acquisition / Recording: Face and voice recognition as well as the alarm verification task needs to acquire and store audio and video files on the transmitter station as well as at the receiver station. Each

board composing the remote system is connected to a number of Multimedia devices (up to six cameras with microphones, one display and one loudspeaker which both play an important role for interactive interventions, i.e. providing communication means between the operator of the centralised system and the remote site). The analog signals are converted into a raw digital format with the assistance of specialised integrated circuits. These raw data are stored temporary to be converted into compressed files by compression algorithms

Image, Sound and Data Transfer: The transfer should be done (at most) over one 64kBps ISDN channel which obviously requires very good compression schemes. The current communication standard is H320.

Receiver station The receiver station can be a standard video telephone but the software around its card supports several functions like video recording, storage and playback. Today's third party video conferencing cards are not delivered with an application programming interface (API), i.e. it is not possible to implement additional features like accessing the H.320 binary data channel or storing the received video data.

Multiple receiver stations: The basic configuration is one simple receiver station, this is in fact not secure enough for a security system, but nevertheless possible. For that reason, we link several receiver stations checking each other if they are still running.

One may see below a prototype that shows an example of graphical interface for such a receiver station [Fig. 8].

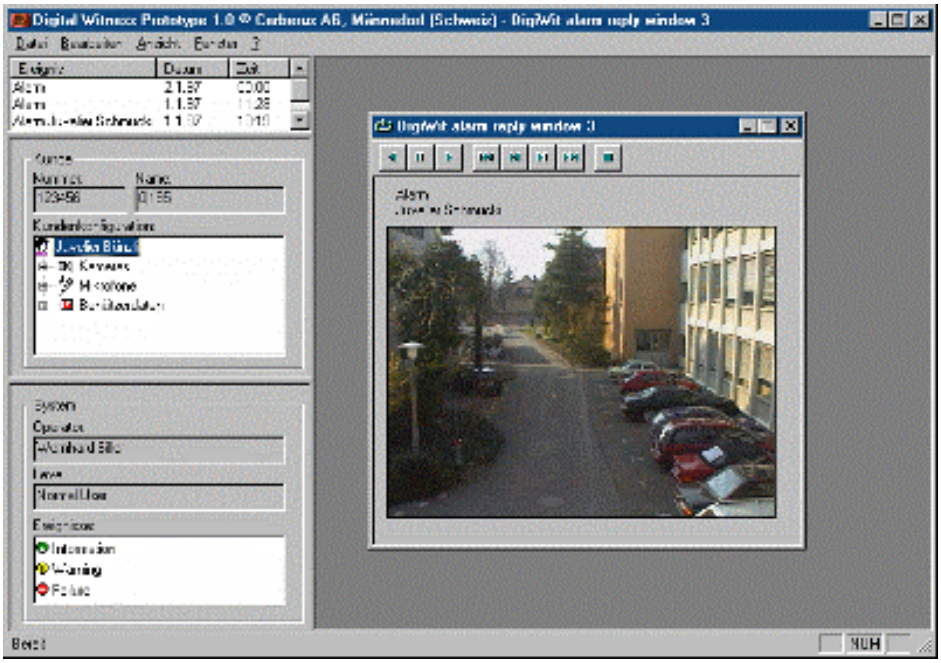

Figure 8 : Receiver Station Software Prototype

This prototype above just shows alarm verification functionality. In fact we expect that the user interface would not change much if any recognition functionality is added.

6. Conclusion

M2VTS is a project where a tight collaboration can be seen between industrial and academic world which can explain the successful results so far obtained. On the one hand, innovative multimodal techniques have been developed, implemented and evaluated according to a common scheme on a pre-recorded database. The best technique for each modality have been integrated in fast prototyping platforms for further testing in real conditions (different illumination, different face expressions (smiles,…),….). The preliminary results already

permit to point out the most sensitive part of each algorithm (different lighting robustness, ….). Finally, several applications have been specified and prototypes should be ready by the end of the project.

7. Acknowledgements

This work is partly supported by the European Community through the ACTS program (Project AC 102 – M2VTS). The authors of this paper also wish to thank all persons who actively participate to the project but who are not mentioned in the author list and especially Nelly suaudeau, Isabelle Guis,Remo Heimgartner, Andreas Meuly, Dieter Wieser, Eric Meurville, Cesar Fernandez, Elena Rodriguez, Josef Bigun, Juergen Luettin, Gilbert Maitre, Gerard Chollet, George Matas, Kenneth Jonsson.

8. References

[1] D. Genoud, F. Bimbot, G. Gravier, and G. Chollet. “Combining methods to improve the phone based speaker verification decision”, in ICSLP’96, vol 3, pp 1756-1759, 1996.
[2] D. Reynolds, “ Speaker Identification and Verification using Gaussian mixture speaker model”, Elsevier. Speech Communication. Vol 17, pp 91-108 (1995).
[3] E. Rodriguez, B. Ruiz, A. Garcia-Crespo and F. Garcai, “Speech/Speaker Recognition using a HMM/GMM Hybrid model, AVBPA’97, Crans-Montana,.
[4] R. Chellapa, C.L. Wilson, and S. Sirohey, “Human and machine recognition of faces: A survey," *Proceedings of the IEEE*, vol. 83, no. 5, pp. 705-740, May 1995.
[5] J Picone, “Signal Modeling Techniques in Speech Recognition”, Proceedings of the IEEE, vol 81, No9 Sept. 1993.
[6] L. Rabiner, “A tutorial on Hidden Markov Models and Selected Applications in Speech Recognition”, Porceedings of the IEEE, Vol 77, No 2 , Feb. 1989
[7] M.N. Jayant, “Speaker Verification: A Tutorial”, IEEE Communications Magazine, Jan. 1990.
[8] J. Campbell, “Speaker Recognition: A Tutorial”, proc. Of the IEEE, Vol 85, No 9, Sept. 1997.
[9] B. Duc, S. Fischer, and J. Bigun, "Face authentication with Gabor information on deformable graphs*," IEEE Trans. on Image Processing*, submitted 1997.
[10] C. Kotropoulos, and I. Pitas, "Face authentication based on morphological grid matching," *in Proc. of the IEEE Int. Conf. on Image Processing (ICIP 97)*, pp. I-105--I-108, Santa Barbara, California, U.S.A., 1997.
[11] C. Kotropoulos, A. Tefas and I. Pitas, "Face authentication based on morphological shape decomposition," in *Proc. of the IEE Int. Conf. on Image Processing and Its Applications*, pp. 794--798, Dublin, Ireland, 1997.
[12] C. Kotropoulos and I. Pitas, ``Linear projection algorithms and morphological dynamic link architecture for frontal face authentication" submitted in European Signal Processing Conference 1998.
[13] C. Kotropoulos, A. Tefas and I. Pitas, ``Face authentication using variants of elastic graph matching based on mathematical morphology that incorporate local discriminant coefficients," submitted in 1998 IEEE Int. Conf. on Acoustics, Speech and Signal Processing.
[14] J. Matas, K. Jonsson and J. Kittler, ``Fast face localisation and verification" in Proc. of British Machine Vision Conference, 1997.
[15] S Pigeon and L Vandendorpe, "Multiple Experts for Robust Face Authentication", To appear in the Proceedings of Conference on Optical Security and Counterfeit Deterrence Techniques (part of SPIE's Photonics West'98 Electronic Imaging Symposium), San Jose, CA, January 29-30, 1998.
[16] M U Ramos Sanchez, J Matas and J Kittler , “ Lip shape modelling and tracking for security and video coding applications”, Proc. of 7th Nation. Symposium on Pattern Recognition and Image Analysis, A Sanfeliu, J J Villanueva and J Vitria Eds., pp 73-78, 1997
[17] M. U. Ramos Sanchez and J. Matas and J. Kittler , “ Statistical chromaticity-based lip tracking with B-splines”, proc. of IEEE-ICASSP97, pp 2973-2976, Vol 4., 1997 (see also http://www.ee.surrey.ac.uk/Projects/M2VTS/experiments/lip_tracking/index.html).
[18] J. Kittler and M. Hatef and R.P.W Duin and J. Matas, “On combining classifiers”, accepted in IEEE Trans. Pattern Analysis and Machine Intelligence and to be published in 1998.
[19] C. Beumier and Marc Acheroy, "Automatic Face Recognition by 3-D Analysis", ORBEL 11 (Belgian Operations Research Society), Namur, Belgium, Jan. 1997
[20] Jourlin, P, Luettin, J., Genoud, D., and Wassner, H., “Acoustic-Labial Speaker Verification”, in Pattern Recognition Letters, to appear.

[21] Benoît Duc, Gilbert Maître, Stefan Fischer, and Josef Bigün,"Person Authentication by Fusing Face and Speech Information" in Proceedings of the First International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA'97), 1997.
[22] D. Genoud, F. Bimbot, G. Gravier, and G. Chollet, "Combining methods to improve speaker verification decision", in "Proceedings of The Fourth International Conference on Spoken Language Processing", 1996.
[23] J. Luettin, Neil A. Thacker, and S. W. Beet, "Learning to recognise talking faces" in Proceedings of the International Conference on Pattern Recognition (ICPR'96), 1996.

[10] S. Pigeon, and L. Vandendorpe, "The M2VTS multimodal face database," *in Lecture Notes in Computer Science: Audio- and Video- based Biometric Person Authentication (J. Bigun, G. Chollet and G.Borgefors, Eds.)*, vol. 1206, pp. 403--409, 1997.