Multi-modal biometric authentication on the SecurePhone PDA

J. Koreman, A.C. Morris, D. Wu
Saarland University
{jkoreman, amorris, daleiwu}@coli.uni-saarland.de

S. Jassim, H. Sellahewa, J. Ehlers
Buckingham University
{sabah.jassim, harin.sellahewa, johan-hendrik.ehlers}@buckingham.ac.uk

G. Chollet, G. Aversano, H. Bredin
École Nationale Supérieure des Télécomm.
{chollet, aversano, bredin}@tsi.enst.fr

S. Garcia-Salicetti, L. Allano, B. Ly Van, B. Dorizzi
Institut National des Télécommunications
{sonia.salicetti, lorene.allano, bao.ly_van, bernadette.dorizzi}@int-evry.fr

Abstract

We present an overview of the development of the SecurePhone mobile communication system in which multimodal biometric authentication gives access to the system's built-in e-signing facilities, enabling users to deal m-contracts using a mobile call in an easy yet secure and dependable way. Authentication uses an original combination of non-intrusive, psychologically neutral biometrics: the user reads a prompt into a camera and microphone, and signs on a touch screen. The state of the art techniques used for each biometric modality were initially developed using the benchmark databases BANCA (audio-visual) and BIOMET (signature). A suitable PDA was then selected and a multimodal database was recorded on the device itself. Several fusion techniques were tested for biometric evidence combination. Best performance achieved for voice, face, signature and fused modalities was 2.3, 17.3, 4.3 and 0.6% EER for BANCA/BIOMET and 3.2, 27.6, 8.0 and 0.8% EER for the PDA database.

1. Introduction

This article describes a multi-modal user authentication system which has been implemented on a PDA as part of the SecurePhone project. The aim of the project is twofold. The first aim is to enable the secure exchange of written and spoken documents. By using private and public keys, a PDA user can send a document securely to another PDA user, who can then edit the document and send it back for further editing, until a final form of the document has been agreed. The second aim is to use biometric authorisation (rather than PIN) to confirm that the user is the registered owner before electronically signing the document. This relies on three modalities: voice, face and signature. These modalities were chosen because they are easy to acquire on a standard PDA and are all characterised by a high user acceptance. All preprocessing of the signals is performed on the PDA, while storage and processing of the client's biometric profile will all be done on the SIM-card in the PDA. Data on the SIM-card is accessible only to the service provider, so in this way the security of the biometric authentication is maximised. However, given the storage and processing limitations of presently available SIM-cards, strong restrictions are placed on the biometric authentication methods which can be used.

In Section 2 we describe the databases which were used to develop suitable authentication techniques. Section 3 presents the best techniques which were found to date for each separate modality. Section 4 presents the best method used for score fusion. Section 5 shows test results for these methods when applied to two multimodal databases. This is followed by a discussion of the overall authentication process and a conclusion.

2. Data

Development of the SecurePhone user authentication system required a database with voice, face and signature data. The first database used was made up from the English section of the BANCA audio-visual database [16] together with the BIOMET on-line signature database [8]. It is possible to combine signature data with the video data of unrelated subjects into virtual subjects in this way because signatures can be assumed to be independent of voice and face data. Results from previous work on signature and voice data from BIOMET support this independence assumption [9].

The second multimodal database [15] was recorded directly on the PDA which was adopted for the SecurePhone project (Qtek2020). This database, and the test protocol and automatic test procedure which accompany it, were specifically designed to test the fixed prompt authentication protocol which was developed using BANCA/BIOMET for the adopted PDA.

2.1 Voice data

The voice is of course a natural modality to use on a PDA or mobile phone. As authentication on the SecurePhone system is PDA rather than server based, the voice data for biometric verification can be taken directly from the microphone. This signal is of high quality, as it has not been transmitted.

2.1.1 BANCA voice data

BANCA voice data is recorded under three noise conditions (termed controlled, degraded and adverse) using both a high- and a low-quality microphone. Data from 82 speakers (52 for development and evaluation plus 30 for UBM (Universal Background Model) training) [17], are recorded at 32 kHz with 16-bit amplitude resolution. Speaker verification tests were carried out for every test in the BANCA protocol (see Table 1) and for both the high- and low-quality microphone data.

Table 1. Test names as used in the BANCA protocol (M = matched, U = unmatched, P = pooled, G = grand). Test Q was added for our own interest.

Con = controlled Deg = degraded Adv = adverse		Training			
		Con	**Deg**	**Adv**	**All**
Test	**Con**	MC			Q
	Deg	UD	MD		
	Adv	UA		MA	
	All	P			G

Of these tests, the Pooled (P) and Grand (G) tests are the more realistic as the PDA will be used for biometric authentication in varying environmental conditions.

In BANCA each subject reads out a random 12-digit number followed by a name and address, once for their own name and address and once for that of another subject. Each recording lasts about 15 seconds. The prompt is repeated in each of the three noise conditions, the 12-digit number changing each time. When training uses data from all 3 noise conditions the words spoken are therefore not exactly repeated. While text independent voice authentication would be desirable for security reasons, this would require up to 30 minutes of voice training data and the model size required to capture all possible speech sounds would not fit into the PDA secure memory. However, for text dependent modelling with Gaussian mixture models (GMMs), best results are obtained when exactly the same words are repeated, and verification performance does not generally improve for utterances longer than around 6 seconds. BANCA is therefore not ideally suited to the envisaged SecurePhone PDA fixed-prompt authentication protocol.

2.1.2. PDA voice data

To enable testing with data as close as possible to that for the PDA in real use, a new database was recorded on the PDA itself. Data was collected for 30 male and 30 female subjects from three age groups (under 30, 30-45, over 45) and consists of 5-digit prompts, 10-digit prompts and short phrases (six examples of each), recorded in quiet and noisy environments, both inside and outside. 5-digit and phrase prompts were taken from [3]. Data was recorded in two sessions separated by at least one week. Like BANCA, subjects were divided into three groups: one for UBM training, and two other groups, g1 and g2. For any given FA/FR (false acceptance to false rejection) cost ratio, thresholds can then be optimised on g1 and evaluated on g2, and vice versa.

Voice data can be taken directly from the microphone at up to 44 kHz, rather than having to make use of the 8 kHz signal which is transmitted over the network, as in a server based system. However, while voice data was recorded at 44 kHz, as preprocessing time increases with sampling frequency and BANCA tests had shown that, with the techniques which we had so far developed, verification accuracy was not affected by using data downsampled to 22 kHz (which the Qtek2020 can also provide directly), the voice data used in PDA tests was at 22 kHz. It is possible that techniques which we have not yet tested could make use of speech quality information from above 22 kHz.

2.2. Face data

As most PDAs have a video camera, face data can be obtained non-intrusively together with the speech data as the client reads a prompt from the PDA, while positioning their face image inside a box on the screen. The ease with which face data can be acquired is one of the reasons why it has also been adopted for European passports.

2.2.1. BANCA face data

The videos of the BANCA database were recorded with two different cameras: a cheap analogue web cam and a high quality digital camera. The image was lossy compressed and, though full video was recorded on tape at 25 frames per second, only 5 image frames per recording were made available in the published database. In order to enable both speaking face verification and liveness tests, as well as to avoid the problem of information loss through compression, the English part of the BANCA database, which we used here, was redigitised to use full uncompressed video data [14].

We report BANCA results for test protocols P and G for all modalities (for a more detailed presentation, including results for BANCA protocols MC, MD, etc., please contact the authors).

2.2.2. PDA face data

PDA face data were recorded simultaneously with voice data (Section 2.1.2) and use variable backgrounds and lighting conditions (inside office and outside on the street), in order to provide a level of variation compatible with the expected use scenario.

2.3. Signature data

Signature data can be captured from the PDA touch screen. Clients are already accustomed to signing for authentication purposes, so user acceptance is high.

2.3.1. BIOMET signature data

Signature verification was developed with the BIOMET database, for which signatures were acquired on a WACOM Intuos2 A6 digitiser, with an ink pen on paper, at a sampling rate of 100 Hz. At each sampled point of the signature, the digitizer captures the (x, y) coordinates, the pressure p and the two angles necessary to encode the position of the pen in space (see Fig. 1).

Signatures in BIOMET were acquired from 84 subjects in two sessions, with five months between them. In the first session, 5 genuine signatures and 6 forgeries per person were captured. In the second session, there were 10 genuine signatures and 6 forgeries per individual.

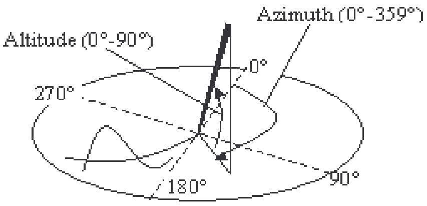

Figure 1. Azimuth and altitude angles

The 12 forgeries of each person's signature were made by 4 different impostors, 3 per impostor from 2 impostors in each session. To the 5 raw signature features for each point, curvature, line angle and a number of other derived statistics, together with first and second time differences for some of these, were added to obtain a total of 25 features per point [12]. When coupling with BANCA subjects to create virtual subjects for the fusion experiments, coupling was gender-dependent.

2.3.2. PDA signature data

BIOMET signature data does not ideally match that of the PDA. This is because the use of a touch screen in place of a writing pad can affect the quality of signatures due to differences in smoothness of the surfaces. A signature database was therefore also captured directly from the PDA. No angle measurements are available on the PDA. Twenty signatures were obtained from each of sixty subjects, with 4 signature experts each providing twenty impostorisations for 4 different sets of fifteen subjects. As with BIOMET signature data, before data modelling a number of derived features were added to the two raw signature features to obtain a total of 19 features per point.

With video data recorded in the UK and signatures in France, each subject was assigned to a video subject of the same gender and age group to create virtual subjects for the fusion experiments.

3. Authentication techniques

In this section, the user authentication methods applied for each individual modality are presented.

3.1. Voice authentication

As in [17], Mel-frequency cepstral coefficients (MFCCs) were computed every 10ms from a 20ms window, using a pre-emphasis factor of 0.97, a Hamming window and 20 Mel scaled filter-banks. All

20 MFCC coefficients were used except c0. Features were obtained using HTK [21].

The SecurePhone voice verification system uses a state of the art GMM [6], with a client model trained on speech features from the client and a universal background model (UBM) trained on features from a number of other speakers. A gender-independent UBM was used both for client model initialisation and score normalisation [17]. GMMs were trained and tested using the Torch machine learning API [4]. GMM training used k-means clustering followed by EM iteration. The client model was trained by MAP adaptation from the UBM, updating Gaussian means only [13][18]. For BANCA the optimum UBM prior weight was 0.0 and the optimum number of Gaussians was 300. For tests with the PDA database best performance used an optimum UBM prior weight of 0.3 and 128 Gaussians.

3.2. Face authentication

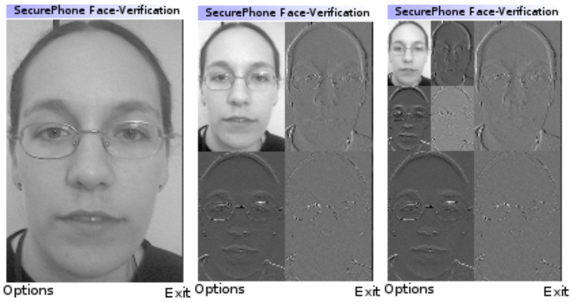
SecurePhone Face-Verification
Options
Exit
SecurePhone Face-Verification
Options
Exit
SecurePhone Face-Verification
Options
Exit

Figure 2. Original image and its stage 1 and stage 2 wavelet decompositions

The Wavelet Transform (WT) is a technique for multi-resolution signal analysis. The discrete WT (DWT) is a special case of the WT which compactly represents a signal in time and frequency, decomposing it into frequency subbands at different spatial scales. The most commonly used wavelet image decomposition is known as the pyramid scheme. At a resolution depth of k, an image I is decomposed into $3k + 1$ subbands, $\{LL_k, LH_k, HL_k, HH_k, \ldots ,HH_1\}$, where LL_k is the k-level resolution approximation of I while LH_1, HL_1, and HH_1 contain the finest scale coefficients (see Figure 2). Based on earlier tests for ORL and BANCA databases [10][19][20], we use the Haar wavelet filter. As variation in lighting conditions has a negative impact on verification accuracy, histogram equalization is used to normalise image luminosity. The face region was automatically localised.

The face feature vectors used here are the LL_4 subband of the histogram-equalised images. A client face template consists of 24 120-coefficient feature vectors representing randomly selected frames as prescribed by the relevant protocol. For the BANCA P protocol these are selected from a single client video, but for the G protocol we selected 8 frames from each of the 3 client videos. For the PDA database we select 6 frames from each of the four client videos. For testing 10 frames are randomly selected from each test video. For each test frame F_i, its minimum City-Block distance d_i is calculated from the template frames. The score of the test video is the minimum d_i value.

3.3. Signature authentication

A continuous left-to-right HMM [6], with diagonal covariances and 4 Gaussians per state to model the emission probability densities, was chosen to model each writer's signing process (cf. Fig. 3).

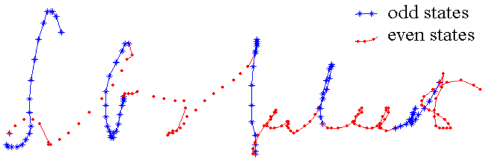

Fig. 3. The segmentation of a training signature modelled by a 10-state HMM

The number of states in the client HMM is determined according to the total number of sampled points available in the signatures used for HMM training. A detailed description of the HMM structure can be found in [12]. A normalization similar to Z-norm, but performed separately for each writer, was applied to each signature feature.

In a departure from the normal use of HMMs for signature verification, the usual normalised log-likelihood information (NLLRs, obtained by dividing the raw log-likelihood by the length of the signature) is combined with state occupancy vectors (SOV, the number of frames assigned to each state by the Viterbi segmentation). Both are normalised to the [0, 1] interval before combination by arithmetic mean. The total score for the writer’s claimed identity is taken as the average of the normalised NLLR and SOV scores. This improves system performance because the two scores reflect complementary information [12].

For BIOMET data, signature scores are obtained using the protocol described in [12]: training uses 5 signatures from session 2 and testing uses signatures from both sessions 1 and 2. For the PDA database, 5 signatures are used to train the model and the remaining are used for test purposes.

4. Fusion method

In this section we describe the method used to combine the biometric authentication evidence across modalities. Tests showed that fusion by the concatenation of voice and face features led to substantially lower performance than voice verification alone. In any case, as signatures cannot usefully be time aligned with video recordings, combination of all three modalities must use some form of late fusion. In late fusion the biometric data from each modality is processed to produce a measure of the fit of the data to the client profile which we refer to here as the "score" for each modality. These scores are then combined into a single score and the claimed identity is accepted if this combined score is above some preset threshold.

In the best scoring method for scores fusion which we present here, a GMM (1) is used to model the joint distribution of the 3 client scores *s=(s1 s2 s3)* from each modality [1]. The GMM has 3 Gaussians with diagonal covariance,

$$p(s|C) \cong \sum_{i=1}^{N} \alpha_i N(s, \mu_i, \Sigma_i) \qquad (1)$$

where N(.) is the multivariate normal (Gaussian) distribution. Before training the GMM, a "Min-Max" normalisation is applied to the client scores [7]. If μ_{cl} and σ_{cl} are the sample mean and standard deviation of the client scores for one modality, then for score s, the normalised score is n = (s-m)/(M-m), where $M = \mu_{cl}+2\sigma_{cl}$ and $m = \mu_{imp}-2\sigma_{imp}$. Normalised values outside [0, 1] are truncated to this interval. The same type of normalisation and GMM training is also applied to the impostor scores. From the trained client and impostor score distributions the joint posterior client probability can then be obtained using Bayes' rule (2),

$$P(C|s_1, s_2, s_3) = \frac{P(C)p(s_1, s_2, s_3|C)}{P(C)p(s_1, s_2, s_3|C) + P(I)p(s_1, s_2, s_3|I)} \qquad (2)$$

The client prior probability P(C) was set to 0.5; P(I) = 1-P(C). A detailed description and results for a number of other fusion techniques which were also tested are given in [1].

5. Test results

Test results are presented only for the best performing authentication technique used with each modality, and for the best performing fusion technique.

As the coupling between signers and video subjects is arbitrary for both databases tested, fusion tests were repeated with 100 gender-dependent random couplings. For the PDA database the random couplings were also age-group dependent.

Fusion results across the 100 couplings report EER (Equal Error Rate) means and standard deviations. For the PDA database, the WER (Weighted Error Rate) value (3) is also given for three values of the cost ratio R = CFA/CFR (cost of false acceptance over cost of false rejection).

$$WER = \frac{CFA.FAR + CFR.FRR}{CFA + CFR} \qquad (3)$$

Unlike EER, which is threshold independent, WER is computed on the evaluation set g2 for the threshold which minimises WER on the development set g1 (a-priori thresholds), and vice versa. WER reported is averaged over these 2 values.

5.1. BANCA/BIOMET results

For BANCA/BIOMET we report results for both the P and G test protocols. In both cases test data is from all three conditions (controlled, degraded and adverse), but for the P test training uses only controlled data, while for the G test training uses data from all 3 conditions (also 3 times as much data). Table 2 only shows results for the lower quality microphone 2 which may be expected to have the most similar characteristics to the PDA microphone. The corresponding DET curves are shown in Fig. 4.

Table 2 and Figure 4 show that, in both P and G tests, score fusion provides a strong improvement in verification performance over any single modality.

Table 2. EER% for BANCA/BIOMET tests P and G, microphone 2, for individual and fused modalities

	P	G
Voice	10.26	2.30
Face	26.36	17.31
Signature	4.29	
Fusion (mean)	1.92	0.57
Fusion (sd)	0.87	0.58

5.2. PDA database results

For tests with the PDA database we report results for all three prompt types (5 digits, 10 digits, phrases).

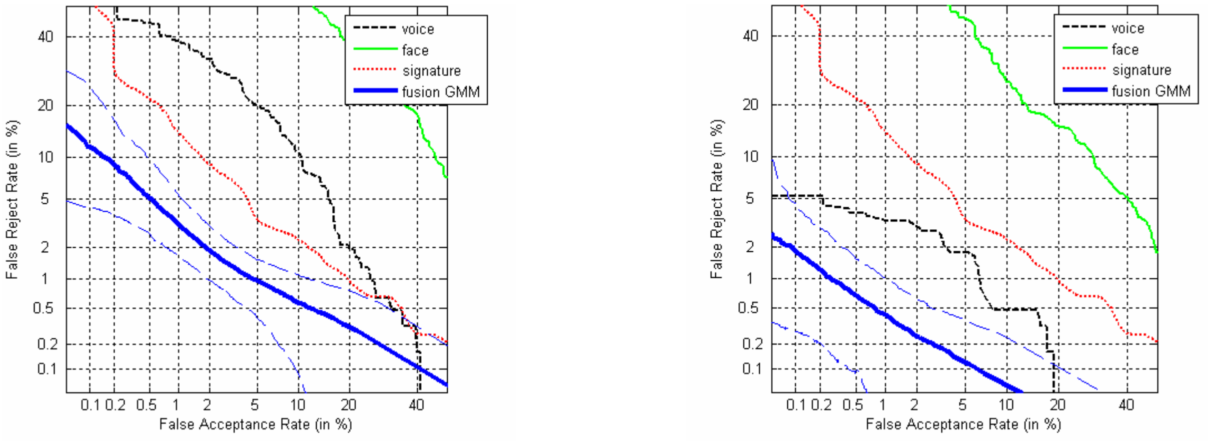

Fig. 4. DET curves for voice, face, signature and score fusion with the BANCA/BIOMET database, microphone 2. Test P (left) and G (right). Fusion curves shows mean +/- one standard deviation (dashed lines).

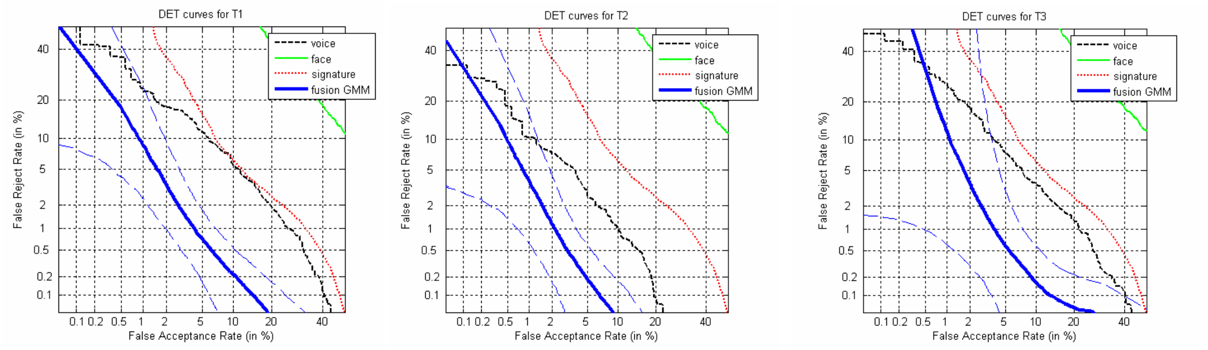

Fig. 5. DET curves for voice, face, signature and score fusion with the PDA database, for 5 digits (left), 10 digits (centre) and phrases (right). Fusion curves show mean +/- one standard deviation (dashed lines).

The results for each test are averaged over results for 6 different prompt examples. All tests were text dependent, with data used for model training coming only from the prompt being tested and only from a single session. Table 3 confirms the findings obtained for BANCA, showing a strong improvement of biometric verification when the scores for the three modalities are fused.

Table 3. EER % for PDA database, all 3 prompt types for individual and fused modalities. Training data from 1 session only

	5-digit	10-digit	Phrase
Voice	7.21	3.24	5.54
Face	28.40	27.55	28.33
Signature	8.01		
Fusion (mean)	2.39	1.54	2.30
Fusion (sd)	0.96	0.83	1.85

Table 4 reports weighted error rates (3) for 3 different R values: 1, 0.1 and 10, together with their corresponding false acceptance and false rejection rates. Results show that low error rates can be obtained for a wide range of cost ratios. Results for 10-digit prompts are better than for either 5-digit prompts or phrases, which show similar performance.

Table 4. Fusion FAR, FRR and WER % for PDA database, all 3 prompt types, for different false acceptance to false rejection cost ratios R

R	5-digit			10-digit			Phrase		
	FAR	FRR	WER	FAR	FRR	WER	FAR	FRR	WER
0.1	4.97	1.56	1.87	3.05	1.20	1.37	4.54	1.78	2.03
1.0	1.57	3.24	2.40	0.89	3.32	1.60	1.61	3.14	2.37
10.0	0.43	6.95	1.02	0.25	4.37	0.63	0.38	6.34	0.92

6. Discussion

From the test results reported above, and also from further tests concerning the complete system which we have not reported, a number of implications arise concerning the design of the user enrolment procedure and the level of performance which can be obtained when hardware constraints are taken in account.

6.1 Application scenario

BANCA P and G test results show that, for tests with multi-condition data, training should also use multi-condition data. This should be reflected in the procedure used for client enrolment. Varied environments could be simulated by playing a range of suitable noises within a single studio-based enrolment session.

The BANCA Q voice test (not reported above), unlike the MC test, gave zero error (EER and WER for all R values). This means that, for models trained with multi-condition data, authorisation in a quiet environment should obtain 100% accuracy.

PDA voice model training with data from 2 well-separated sessions (results not reported here) gave better performance than with the same amount of data from one session (results reported). Enrolment should therefore aim to simulate time separation.

Digits are semantically neutral and therefore more easily accepted than phrases. 10 digits would be preferable to 5, but only if processing time permits.

Signature verification on the PDA database is greatly reduced compared to BIOMET, probably due to the difficulty of signing on a glass surface.

6.2 Implementation issues

6.2.1 Common use of GMM

The verification procedure would be simplified if all three modalities use GMM based models. For speech mode the GMM is already the model of choice. Tests not reported here have indicated that for face mode GMMs give a comparable performance to that reported above. For signature verification GMM performance is similar to NLLR based HMM, but up to 50% lower than for the NLLR+SOV HMM technique. The best fusion approach already uses a GMM. Common use of the GMM would therefore be open to question if memory constraints would permit HMM scoring to be implemented on the SIM card.

6.2.2 Online processing

To maximise the speed of the authentication response, parts of the verification algorithm will be implemented in an “online” fashion. For speech preprocessing online versions for silence detection and CMS (Cepstral Mean Subtraction for MFCC generation) have both been implemented, with silence detection using a running estimate of the additive noise level and CMS using a running estimate of the convolutive noise level. Online CMS has been tested on both BANCA and PDA data and has shown improved performance over offline CMS. It may also be possible to start the verification process before audio-visual acquisition is completed.

6.2.3 Integerisation

The PXA263 PDA processor is fixed-point (integer) with only simulated floating-point operations. While tests reported here have used floating point arithmetic, a 16-bit fixed-point preprocessing implementation has shown a factor of 3.5 speed-up against only a 3% relative EER degradation on the Pooled BANCA protocol. 32-bit fixed-point tests showed a factor of 1.5 speed up and an accuracy comparable to floating-point.

Both scores calculation for each modality and scores fusion should run on the SIM card, which provides only 16-bit fixed-point arithmetic operations, no other operations (log, exp, etc.) and no floating-point simulator. Suitable verification procedures, both for GMM based models as well as the simpler weighted distance models tested above for face verification, have been implemented on the SIM. These verification procedures are not computationally intensive, but further tests need to be made to check the effect which 16-bit fixed-point calculations will have on authentication accuracy.

6.2.4 Imposture scenarios

The possibility for impostorisation depends on to what extent the impostor is prepared to go. For higher security applications it would be easy to add further modalities, such as iris and fingerprint, but this would risk alienating the casual user. It is important to recall that, for many types of transaction, the traditional level of security given by a single PIN or signature is acceptable. For such applications the security provided by the SecurePhone should be sufficient. If a photograph of the owner’s face and signature plus an audio recording of their reading the fixed prompt was obtained, then successful impostorisation would normally be possible, but a liveness test we have proposed, measuring the correlation between mouth opening and speech energy [2], should counter this possibility. This would not prevent impostorisation using a video recording. This could be countered by use of text-independent modelling with random prompts, but secure memory limitations mean that this would only be possible on a server-based system.

7. Conclusion

We have presented the current status of the SecurePhone project which will permit documents to be interactively modified and agreed in a mobile environment, after which multi-modal biometric authorization will give access to an e-signature facility which will enable legally binding contracts to be signed. It has been shown that, using state-of-the-art verification techniques, the combination of the non-intrusive biometrics of voice, face and signature can achieve a level of authorisation accuracy which should be acceptable for the wide range of applications which is normally secured by a PIN or signature.

Acknowledgments

This work was supported by the EC SecurePhone project IST-2002-506883.

References

[1] Allano, L., Garcia-Salicetti, S., Ly-Van, B., Morris, A.C., Koreman, J., Sellahewa, H., Jassim, S. & Dorizzi, B., "Non intrusive multi-biometrics on a mobile device: a comparison of fusion techniques", Proc. SPIE conference on Biometric Techniques for Human Identification III, Orlando (in press).

[2] Bredin, H., Miguel, A., Witten, I.H. & Chollet, G., "Detecting replay attacks in audiovisual identity verification", Proc. ICASSP 2006 (in print).

[3] Cole, R., Noel, M. & Noel, V., "The CSLU Speaker Recognition Corpus", Proc. ICSLP, Sydney, pp.3167-3170, 1998.

[4] Collobert, R., Bengio, S. & Mariéthoz, J., "Torch: a modular machine learning software library", Technical Report IDIAP-RR 02-46, 2002.

[5] Dolfing, J.G.A., "Handwriting recognition and verification, a Hidden Markov approach", Ph.D. thesis, Philips Electronics N.V., 1998.

[6] Duda, O., Hart, P.E. & Stork, D.G., Pattern classification, Wiley, 2001.

[7] Fierrez-Aguilar, J.M Ortega-Garcia, J. & Gonzalez-Rodriguez, J., "Target dependent score normalization techniques and their application to signature verification". IEEE Trans. on Systems, Man and Cybernetics, part C 35, 2005.

[8] Garcia-Salicetti, S., Beumier, C., Chollet, G., Dorizzi, B., Leroux-Les Jardins, J., Lunter, J., Ni, Y. & Petrovska-Delacretaz, D., "BIOMET: a Multimodal Person Authentication Database Including Face, Voice, Fingerprint, Hand and Signature Modalities", Proc. 4th Conf. on AVBPA, pp. 845-853, Guildford, UK, July 2003.

[9] Garcia-Salicetti, S., Mellakh, M.A., Allano, L., Dorizzi, B., "Multimodal biometric score fusion: the Mean rule vs. Support Vector Classifiers", Proc. EUSIPCO, Antalya, Turkey, Sept. 2005.

[10] ICBA competition – CSU results, 2004. http://www.ee.surrey.ac.uk/banca/icba2004/csuresults.html

[11] Indovina, M., Uludag, U., Snelick, R., Mink A. & Jain, A., "Multimodal biometric authentication methods : a COTS approach", Proc. MMUA 2003, pp. 99-106, Santa Barbara, California, USA, Dec. 2003.

[12] Ly Van, B., Garcia-Salicetti, S. & Dorizzi, B., "Fusion of HMM's Likelihood and Viterbi Path for On-line Signature Verification", Biometric Authentication Workshop (BioAW), Lecture Notes in Computer Science (LNCS) 3087, pp. 318-331, Prague, Czech Republic, May 2004.

[13] Mariéthoz, J., Lindberg, J. & Bimbot, F., "A MAP approach, with synchronous decoding and unit-based normalisation for test-dependent speaker verification", Proc. ICASSP 2000.

[14] McTait, K., Bredin, H., Colon, S., Fillon, T. & Chollet, G., "Adapting a high quality audiovisual database to PDA quality", Proc. ISPA, 2005.

[15] Morris, A.C., Koreman, J., Sellahewa, H., Ehlers, J, Jassim, S., Allano, L. & Garcia-Salicetti, S., "The SecurePhone PDA database, experimental protocol and automatic test procedure for multimodal user authentication", Tech Report, Jan. 2006. http://www.coli.uni-saarland.de/SecurePhone/documents/PDA_database_and_test_protocol.pdf

[16] Porée, F., Mariéthoz, J., Bengio, S. & Bimbot, F., "The BANCA Database and experimental protocol for speaker verification", IDIAP-RR 02-13, 2002.

[17] Reynolds, D.A., "Speaker identification and verification using Gaussian mixture speaker models", Speech Communication, Vol.17, pp.91-108, 1995

[18] Reynolds, D.A., Quatieri, T.F. & Dunn, R.B., "Speaker verification using adapted gaussian mixture models", Digital Signal Processing, Vol. 10, Issues 1-3, pp. 19-41, 2000.

[19] Sadeghi, M., Kittler, J., Kostin, A. & Messer, K., "A Comparative Study of Automatic Face Verification Algorithms on the BANCA Database," Proc. AVBPA Int'l Conf. Audio-and Video-Based Biometric Person Authentication, pp. 35-43, June, 2003.

[20] Sellahewa, H. & Jassim, S.,"Wavelet-based Face Verification for constrained platforms", Proc. SPIE on Biometric Technology for Human Identification II, Florida, Vol. 5779, pp 173-183, 2005.

[21] Young, S., Evermann, G., Gales, M., Hain, T., Kershaw, D., Moore, G., Odell, J., Ollason, D., Povey, D. & Valtchev, V. HTKbook (V3.3), Cambridge University Engineering Dept., 2005.