Scenario and Survey of Combined Handwriting and Speech Modalities for User Authentication

A. Humm, J. Hennebert and R. Ingold
Department of Informatics
University of Fribourg, Switzerland
e-mail: {andreas.humm, jean.hennebert, rolf.ingold}@unifr.ch

Abstracts: *We report on our developments towards building a novel user authentication system using combined handwriting and speech modalities. In our project, these modalities are simultaneously recorded by asking the user to utter what he is writing. We introduce two potential scenarios that we have identified as candidates for applications and we describe the database recorded according to these scenarios. We then report on a usability survey that we have conducted while recording the database. Finally, we present preliminary performance results obtained on the database using one of the scenario.*

Keywords: multimodal biometrics, speech, signature, handwriting, usability survey

1 Introduction and Motivations

We are interested in building multimodal authentication systems using speech and handwriting as modalities. Speech and handwriting are indeed two major modalities used by humans in their daily transactions and interactions. We propose here to investigate scenarios in which these modalities are recorded simultaneously, i.e. the user is asked to read what he is writing. We have named our approach Combined Handwriting And Speech Modalities (CHASM) to stress the fact that both modalities are recorded at the same time, allowing for combined modelisation techniques[1]. In this work, we have been defining two scenarios. In the first one, a bimodal signature with voice is acquired. In this case, the user is actually asked to say the content of his signature. In the second scenario, the user is asked to write and read synchronously the content of a given text.

Prior to investigating technical issues, we have tried to answer several non-technical questions such as : Is it acceptable for the user to read and write at the same time in the context of a biometric system? How many words would a user accept to read and write to be recognized? Is it possible to ask the user to read his signature? We have proceeded to an acquisition campaign of CHASM data to allow us to get a better feeling of how users react. We also performed a survey to ask users their own feeling about CHASM recordings. The acquisition campaign and the survey allowed us to partially answer to the previously stated questions.

As for other multimodal biometrics approaches, the main motivation of CHASM is to increase performances, i.e. the accuracy of CHASM is going to be better than monomodal systems based on speech [2, 11] or handwriting [7, 10] alone. The motivation of performing synchronized acquisition is multiple. Firstly, it avoids doubling the acquisition time. For the user, it is important to reduce as much as possible the enrolment time and the test time. Secondly, the synchronized acquisition will give better robustness against intentional imposture. We can

[1] We also point out that CHASM recordings can also be used for content recognition. This is however out of the scope of this paper where we focus on potential biometric authentication applications.

6th International Conference on Recent Advances in Soft Computing (RASC 2006), Canterburry, Kent, UK. 2006. pp. 496-501

reasonably expect that imitating simultaneously the voice and the writing of somebody else is more difficult considering the extra cognitive load. Finally, the synchronization patterns (i.e. where do users synchronize) or the intrinsic deformation of the inputs (mainly the slowdown of the speech) may be dependent to the user, therefore bringing useful biometrics information.

Several related works have already shown that using speech and signature acquired independently and modelled together permits to improve significantly the authentication performances in comparison to systems based on speech or signature alone. In [3], an on-line signature verification system and a speaker verification system are joint. These two systems are first tested separately, then the scores of each systems are successfully fused together to reach better accuracy. For this test, fictitious users are built by randomly associating signature and speech samples from two independent databases. In [8], similar conclusions are reached for a system modelling speech and signature together where the data are taken from the same user in the BIOMET database [4]. Speech and signature streams are however not recorded simultaneously, as we propose in our CHASM methodology here.

The remainder of this paper is organized as follows. In section 2 we introduce two potential scenarios for CHASM acquisition and we report on the acquisition campaign that was performed according to these scenarios. In section 3, we present the results from the usability survey that was conducted during the acquisition campaign. Section 4 summarizes preliminary performance results obtained using one of the scenario of use. Finally, conclusions, discussions and future work are presented in section 5.

2 Scenario definition and data acquisition

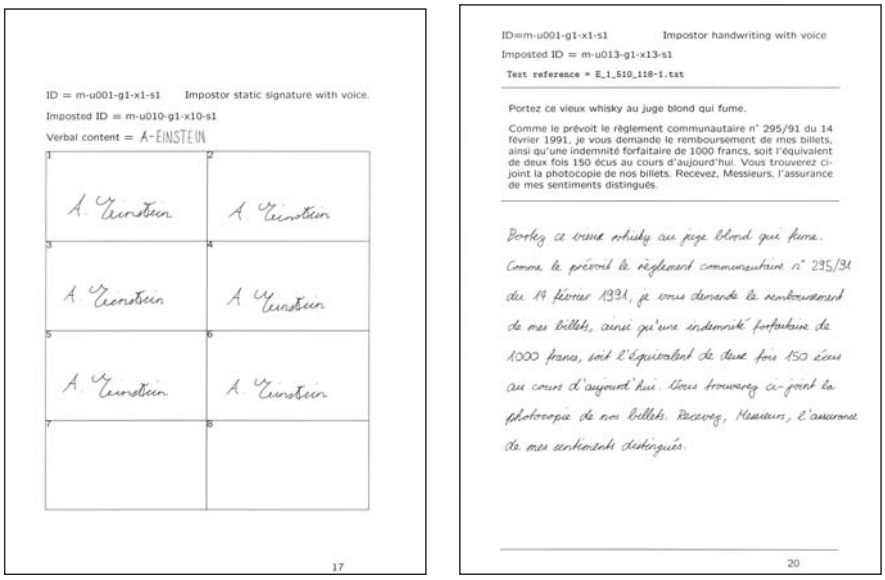
Two scenarios have been defined: **CHASM signature** and **CHASM handwriting**. In CHASM signature, a bimodal signature with voice is acquired. In this case, the user is actually asked to say synchronously the content of his own signature. Strictly invariant synchronization will of course not be possible, however, the subject is asked to sign in such a way that the written symbols correspond roughly in time with the uttered phonemes. If the signature contains flourish or non-readable signs, the subject is simply asked to utter his name while signing. In CHASM handwriting, we assume that the system prompts the subject to write and say a random piece of text each time an access is performed. This kind of scenario allows to make the system more secure against spoofing attacks where the forger could try to play back a pre-recorded version of the genuine data. This scenario has also the advantage to be very convenient for the subject who does not need to remember any password.

CHASM data have been acquired according to these two scenarios in the framework of the MyIDea multimodal data collection [1] lead in Fribourg, Switzerland. The MyIDea database also contains other modalities such as fingerprint, talking face, palmprint, etc. About 70 users have been recorded over three sessions with an interval of time between sessions ranging from one week to several month. This procedure eased the planning of recordings and actually corresponds to a real life situations where users get authenticated at random frequencies. The CHASM data set used in this article has been given reference MYIDEA-CHASM-SET1. Signatures and handwritings are acquired with a WACOM Intuos2 graphical tablet. A WACOM InkPen is used to write on standard sheets of paper positioned on the tablet. The writing feeling is therefore close to the one of writing on a standard sheet of paper using a regular pen.

2.1 CHASM signature

As illustrated on Figure 1 part (a), template papers are used for recording signatures. Subjects sign six times per session for a total of 18 genuine signatures, using the cells on the template.

The two remaining cells are used in the case of missed signatures that need to be redone. Prior to the recording, the subject is allowed to train for a few bimodal signatures to get used with the sensors and procedure. During each session, the subject is also asked to imitate the bimodal signature of another subject. In order to do this, the subject has access, during a limited time of two minutes, to the *static* image and to the *verbal content* of the signature to imitate. In other words, access to the voice recording is not given to the impostor and there is no intention to imitate the voice. Six imitations of the signature of another subject are performed per session for a total of 18 impostor signatures after the three sessions.



(a) (b)

Figure 1: Example of a signature (a) and handwriting (b) acquisition page.

2.2 CHASM handwriting

During each session, the subject is asked to synchronously write and utter the content of a randomly chosen piece of text. The content of the text is composed of 50 to 100 words chosen from a corpus of articles in different domains. The layout of the forms used for guiding the acquisitions is inspired from the procedure used in the IAM database [9], see example in Figure 1, part (b). The sheets were automatically generated to guarantee that all forms are processed and generated in the same way. A guiding paper with 15 mm spaced rulers was placed under the forms to help the writer perform the writing. The subject is allowed to train for a few lines on a separated sheet in order to accustom with the procedure of talking and writing in the same time. After the three sessions, a total of three true handwriting acquisitions are available per subject. During each session, the subject is also asked to imitate the handwriting of another subject and to synchronously utter the content of the text. The subject has access during two minutes to the *static* handwriting data of the subject to imitate. Note that the access to the voice recording of the imitated subject is not given. The subject has a limited time of two minutes to train to imitate the handwriting while uttering the content of the text. This procedure leads to a total of three impostor attempts on different subjects after the three sessions.

2.3 Comments on the acquisition

We noticed that **all** 70 users could perform the signature and handwriting acquisition. The fact that they had to read and sign or read and write at the same time did not prevent any

Table 1: Questions and results of the usability survey.

	question	result
1	Did you find it simple/difficult to write on a tablet?	s ●○○○○○ d
2	Do you think that you wrote faster, at the same speed or more slowly than usual (without simultaneous speaking)?	s ○●○○○ f
3	Did you find it simple/difficult to speak and sign at the same time?	s ○○○●○○ d
4	Did you find it simple/difficult to speak and write at the same time?	s ○○●●○○ d
5	How many lines of text would you accept to say and write in order to perform your own identification in a banking environment?	[0.5 – 10] **2.0**
6	Do you think that the act of speaking and writing at the same time affected your capacities to imitate the writing?	y ○●●○○○ n

acquisition to happen. As the speech production is generally faster than the handwriting, the speech signal is deformed due to its slowdown and re-synchronization occurs at specific times. For the signature, most of the users synchronized the written symbols with syllables. For the handwriting, re-synchronization is happening most of the time at syllables boundaries and sometimes at end of words. While the deformation of the speech signal was clearly identified, we were not able to visually state if the handwriting signal is also deformed. Many signatures contained some pre or post flourishes that were not said by the user, without giving specific instruction about this. Very few users were having signatures with no readable signs. These users were then asked to simply utter their name while signing.

3 Usability survey

We asked each subject taking part to the recording of the database to answer some questions about CHASM acquisition. The questionnaire included the questions listed in table 1. For each question, subjects were asked to answer according to a pre-defined scale. An extra field for free comments was also left to the users on the questionnaire. The answers given by all subjects are also schematically summarized in table 1. The main conclusions of the survey are the following.

- A large majority of users found it easy to write on a tablet. The use of regular papers stuck on the tablet may have help to reach this result.
- Users ranked as average the difficulty of writing and speaking at the same time. This is most probably due to the extra level of concentration needed to perform such acquisitions.
- All users were able to sign and utter the content of their signature. They however feel that signing and speaking is more difficult than writing and speaking. An interpretation can be given to this result considering the fact that a signature usually contains pre or post flourishes on which users cannot utter anything. Also, performing a signature is very much an automatic gesture for most people while speaking is a more coordinated act.
- Users feel they are writing at a slower speed when they are speaking in the same time. This feeling is potentially due to the fact that people need to slow down their speech production.
- Users would accept to write up to two lines of text to perform their authentication.

Table 2: Performance of a baseline GMM system on CHASM signature, 16 gaussian mixtures are used for the client model, 64 mixtures for the world model.

forgeries	random (%EER)	skilled (%EER)
signature	5.3	9.4
speech	14.0	19.5
sum fusion	**3.5**	**6.9**

- Interestingly, users feel that the act of speaking and writing at the same time affected their capacities to imitate signatures and handwriting. While this feeling is of course not related to the real capacity of the system to reject forgers, the perceived security of the CHASM procedure is potentially higher than for monomodal systems.

As all users were able to perform CHASM acquisitions and considering the above answers given to the survey, our current conclusion is that such CHASM acquisitions are acceptable from a usability point of view.

4 Evaluations

We have built a baseline system to model CHASM data. We have chosen to use standard Gaussian Mixture Models to model independently both streams of speech and handwriting data, followed by a simple fusion at the score level. For more details on this system, we refer to [6]. While this system uses straightforward feature extraction and modelling, it has allowed us to validate our approach and to draw our first conclusions. Realistic assessment protocols have been defined on the recorded database according to the two scenarios defined above. For more details on these protocols, we refer to [5]. We report here on results obtained with the CHASM signature scenario. For each subject, the six signatures from the first session are used to build the models. Genuine tests are performed on the six signatures of session two and session three, giving a total of 70 users * 12 accesses = 840 genuine tests. We consider two kinds of impostor attempts. In the first case, impostor attempts are randomly performed using one signature for each of the remaining subjects in the database, giving a total of 70 users * 69 accesses = 4830 random forgeries. In the second case, the 18 available skilled forgeries are used against each user, giving a total of 70 users * 6 accesses * 3 sessions = 1260 skilled forgeries[2].

Table 2 summarizes the results in terms of Equal Error Rates (EER) for the signature stream, for the speech stream and for both modalities. The following conclusions can be drawn. Skilled forgeries decreases significantly the performance in comparison to random forgeries. A drop of almost 100% relative performance is observed for the signature modality and about 40% for the speech modality. We have to note again here that the forger do not try to imitate the voice of the user but actually say the genuine verbal content, which seems enough to degrade the performances. The fusion brings systematically a clear improvement of the results which is in favor of the CHASM methodology.

5 Conclusions and Future Work

We have reported on scenarios definition of a novel CHASM-based authentication system. The two defined scenarios are CHASM signature and CHASM handwriting. To allow us to answer acceptability questions that arise when defining the scenario, we performed an initial data acquisition campaign together with a user survey. We experienced that all users were able to perform

[2]The figures given here are approximate numbers as some users did not perform all sessions.

the simultaneous acquisition of handwriting and speech. The survey shows that such simultaneous acquisitions seem to be acceptable for the user from a usability point of view. We also presented initial results that were obtained using a baseline GMM system run on CHASM signatures. Although the baseline system is rudimentary in terms of modelling and fusion scheme, the measured performances show a significant improvement for the CHASM approach in comparison to the modalities tested alone. In our future work, we plan to build more advanced CHASM modelling systems. We will also investigate if the biometrics performances are impaired due to the signal deformations induced by the simultaneous recordings. Another part of our work will be dedicated to pursue the acquisition and survey campaign with a second set of CHASM data, similar to the one presented here.

Acknowledgments. This work was supported by the Swiss National Science Foundation program "Interactive Multimodal Information Management (IM2)", as part of NCCR and by the EU BioSecure NoE.

References

[1] B. Dumas et al. Myidea - multimodal biometrics database, description of acquisition protocols. In *In proc. of Third COST 275 Workshop (COST 275)*, pages 59–62, October 27 - 28 2005. Hatfield (UK).

[2] F. Bimbot et al. A tutorial on text-independent speaker verification. *EURASIP Journal on Applied Signal Processing*, 4:430–451, 2004.

[3] M. Fuentes, D. Mostefa, J. Kharroubi, S. Garcia-Salicetti, B. Dorizzi, and G. Chollet. Identity verification by fusion of biometric data: On-line signature and speech. In *Proc. COST 275 Workshop: The Advent of Biometrics on the Internet*, pages 83–86, 2002. Rome.

[4] S. Garcia-Salicetti et al. Biomet: a multimodal person authentication database including face, voice, fingerprint, hand and signature modalities. In *4th International Conference on Audio and Video-Based Biometric Person Authentication (AVBPA)*. Springer-Verlag, 2003.

[5] A. Humm, J. Hennebert, and R. Ingold. Combined handwriting and speech modalities for user authentication. Tech. Report 06-05, University of Fribourg, Dept. of Informatics, 2006.

[6] A. Humm, J. Hennebert, and R. Ingold. Gaussian mixture models for chasm signature verification. *Accepted for publication in 3rd Joint Workshop on Multimodal Interaction and Related Machine Learning Algorithms*, Washington, 2006.

[7] F. Leclerc and R. Plamondon. Automatic signature verification: the state of the art–1989-1993. *Int'l J. Pattern Recognition and Artificial Intelligence*, 8(3):643–660, 1994.

[8] B. Ly-Van, R. Blouet, S. Renouard, S. Garcia-Salicetti, B. Dorizzi, and G. Chollet. Signature with text-dependent and text-independent speech for robust identity verification. In *Proc. Workshop on Multimodal User Authentication (MMUA)*, pages 13–18, 2003.

[9] U.-V. Marti and H. Bunke. A full english sentence database for off-line handwriting recognition. In *Proc. of the 5th ICDAR*, pages 705–708, 1999.

[10] R. Plamondon. The design of an on-line signature verification system: From theory to practice. *Int'l J. Pattern Recognition and Artificial Intelligence*, 8(3):795–811, 1994.

[11] Douglas Reynolds. An overview of automatic speaker recognition technology. In *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, volume 4, pages 4072–4075, 2002.