UNIVERSITÉ DE FRIBOURG SUISSE
UNIVERSITÄT FREIBURG SCHWEIZ

# Combined Handwriting and Speech Modalities for User Authentication

DIUF-RR 270 06-05

Andreas Humm[1] Jean Hennebert[2] Rolf Ingold[3]

March 30, 2006

Department of Informatics Research Report

Département d'Informatique - Departement für Informatik • Université de Fribourg - Universität Freiburg • Boulevard de Pérolles 90 • 1700 Fribourg • Switzerland

phone +41 (26) 300 84 65 fax +41 (26) 300 97 31 Diuf-secr-pe@unifr.ch http://diuf.unifr.ch

[1] DIVA-DIUF, University of Fribourg, Switzerland, andreas.humm@unifr.ch
[2] DIVA-DIUF, University of Fribourg, Switzerland, jean.hennebert@unifr.ch
[3] DIVA-DIUF, University of Fribourg, Switzerland, rolf.ingold@unifr.ch

**Abstract**

We report on our first 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 first report on a database that we have recorded according to this scenario. Then, we report on the results of a usability survey that we have conducted while recording the database. Finally, we present the assessment protocols for authentication systems defined on the database.

**Keywords:** biometrics, multimodal biometrics, signature verification, speaker verification

# 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 use a scenario in which these modalities are recorded simultaneously, i.e. by asking the user to read what he is writing. We have named our project Combined Handwriting And Speech Modalities (CHASM) to stress the fact that both modalities are recorded at the same time, allowing for combined modelisation techniques. 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.

The motivation of CHASM is to increase performances of biometric systems while keeping the level of usability acceptable by simultaneously recording both modalities. We expect that the accuracy of CHASM is going to be better than monomodal systems based on speech [1, 11] or handwriting [10, 5, 9] alone. We can also reasonably expect better performances than a multimodal system based on independent acquisition of speech and handwriting. Indeed, in our case, the correlation between both streams can potentially be exploited to perform combined modelisation upstream in the system, for example at the feature level or at the statistical level. From a usability point of view, we can reasonably forecast that subject enrolment and access procedures will be shorter than for each modality taken alone, therefore enhancing the acceptability of the system. We can also expect better robustness against imposture as imitating simultaneously the voice and the writing of somebody else is certainly not an easy task considering the extra cognitive load. Finally, from an industrial point of view, the requested sensors already exists and are nowadays low-cost: simple microphone and any handwriting acquisition device such as a graphic tablet. 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.

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 [6], 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. Therefore, to our knowledge, the use of CHASM data for user authentication is a novel approach.

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 feeling about CHASM recordings. The acquisition campaign and the survey allowed us to answer to the previously stated questions. As our goal is to build CHASM modelling techniques and to evaluate them, we have defined user authentication assessment protocols on the recorded database. These protocols have been crafted to correspond to realistic use of potential CHASM biometric applications. The protocols also try to put in evidence some modelling difficulties tied to time-variability.

The remainder of this paper is organized as follows. In section 2 we give an overview of the CHASM data acquisition campaign. In section 3, we present the results from the usability survey that was performed during the acquisition campaign. Section 4 describes the evaluation protocols that are defined on the database. Finally, conclusions, discussions and future work are presented in section 5.

# 2 Data acquisition

CHASM data have been acquired in the framework of the MyIDea multimodal data collection [2]. MyIDea data base also contains other modalities such as fingerprint, talking face, palmprint, etc. At the time of writing this article, about 70 users have been recorded over three sessions. The interval of time between sessions ranges between one week to several month. This procedure eased the planning of recordings and actually corresponds to a real life situation where users get authenticated at random frequencies. For CHASM data, two scenarios have been used. 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 text. The CHASM data set used in this article has been given reference MYIDEA-CHASM-SET1. In the MyIDea collection, two closely related data sets are also recorded on the same users. The first one has reference MYIDEA-SIGNATURE-SET1 and includes signature acquired alone. The second one has reference MYIDEA-HANDWRITING-SET1 and contains handwriting acquired alone.

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. This hardware is similar to the one used for other databases such as BIOMET [4], MCYT [8] and IAM [7] databases. The voice is recorded at 16kHz using a standard computer microphone mounted on a headset.

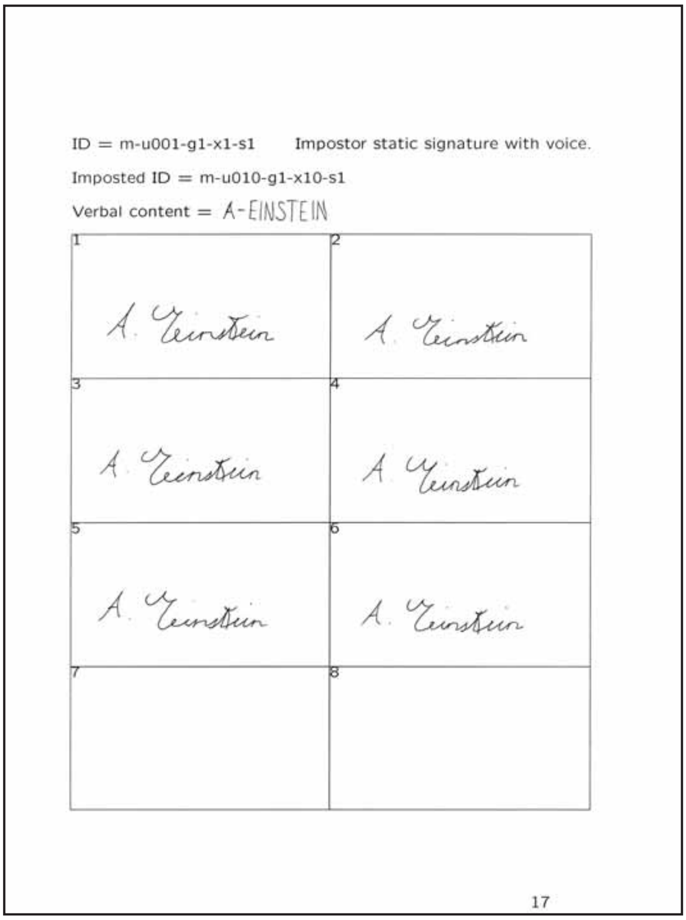
An acquisition software has been developed to perform the synchronized acquisition of handwriting and speech data. For each sampled text point, the tablet records five parameters at a frequency of 100 Hz: x,y-coordinates, pressure, and the two angles azimuth and altitude. We further record timestamps for each data packet sent by the tablet. Timestamps are also recorded for the beginning and end of speech acquisition. This procedure allows us to precisely synchronize speech and handwriting streams, even when pens-up are occurring while signing or writing.

## 2.1 Signature with voice

As shown on figure 1, template papers are used for recording signatures. Subjects sign six times per session, using the cells on the template. The two remaining cells are used in the case of missed signatures.

**True signatures.** The subject is asked to synchronously sign and utter the content of his own signature. Invariant synchronization is of course not 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 during the signature. During each session, the subject performs six bimodal signatures leading to a total of 18 true acquisitions after three sessions. Prior to the recording, the subject is allowed to train for a few bimodal signatures in order to get used with the procedure.

**Impostor signatures.** For each session, the subject is asked to imitate the bimodal signature of another subject. In order to do this, the subject has access to the *static* image and to the *verbal content* of the imitated signature. In other words, access to the voice recording is not given to the impostor. The subject has a limited time of two minutes to train to imitate the signatures. During each session, the subject performs six imitations of the signature of another subject. This procedure leads to a total of 18 impostor signatures after the three sessions, i.e. six signatures on three different subjects.



ID = m-u001-g1-x1-s1 Impostor static signature with voice.

Imposted ID = m-u010-g1-x10-s1

Verbal content = A-EINSTEIN

| | |
|---|---|
| 1<br>A. Einstein | 2<br>A. Einstein |
| 3<br>A. Einstein | 4<br>A. Einstein |
| 5<br>A. Einstein | 6<br>A. Einstein |
| 7 | 8 |

17

Figure 1: Example of a signature page.

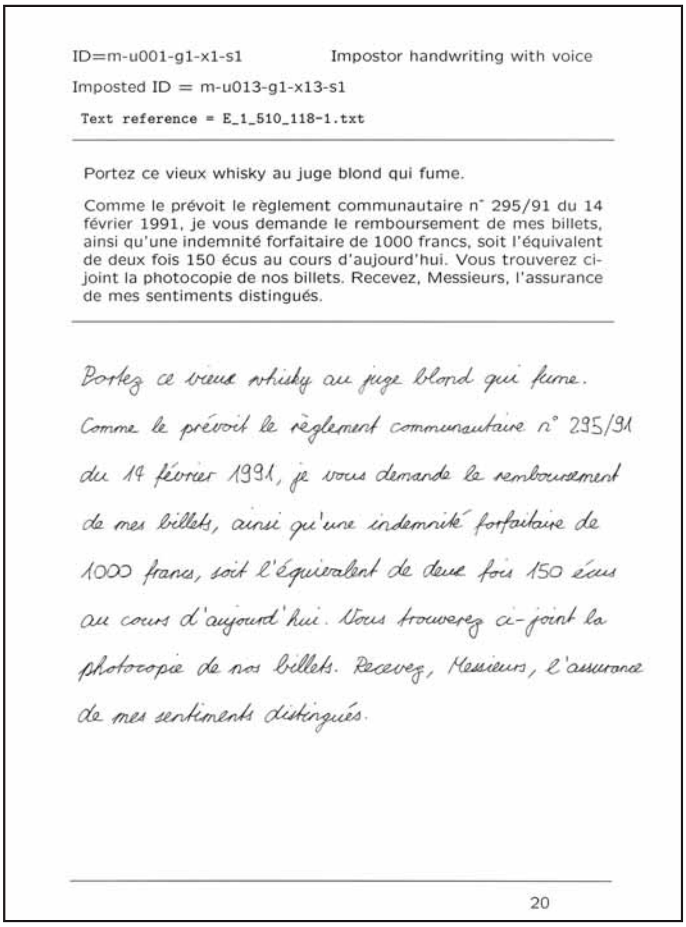
## 2.2 Handwriting with voice

The subject is asked to synchronously write and utter the content of a text. The content is composed of a fixed generic phrase containing all the letters of the alphabet, followed by a text of 50 to 100 words, randomly chosen from a corpus. The fixed phrase can be used to evaluate text-dependent systems and the random text can be used to evaluate text-independent systems.

The layout of the forms used for guiding the acquisitions is shown on figure 2. The sheets were automatically generated to guarantee that all forms are processed and generated in the same way. The layout of the form is inspired from the IAM database [7]. As our main focus is the use of handwriting data to perform identity verification, we wanted to make image preprocessing as easy as possible. Therefore, we decided that the writers have to use rulers. These guiding lines are 15 mm spaced and are printed on a separate sheet of paper which is put under the form.

**True handwriting.** For each session, the subject is asked to read and write the fixed phrase and the random text fragment. This leads to a total of three true handwriting acquisitions per subject. 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.

**Impostor handwriting.** For each session, the subject is asked to imitate the handwriting of another subject and to synchronously utter the content of the text. The subject has access to the *static* handwriting data of the subject to imitate but does not have access to the voice recording of that subject. 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.



ID=m-u001-g1-x1-s1 Impostor handwriting with voice

Imposted ID = m-u013-g1-x13-s1

Text reference = E_1_510_118-1.txt

Portez ce vieux whisky au juge blond qui fume.

Comme le prévoit le règlement communautaire n° 295/91 du 14 février 1991, je vous demande le remboursement de mes billets, ainsi qu'une indemnité forfaitaire de 1000 francs, soit l'équivalent de deux fois 150 écus au cours d'aujourd'hui. Vous trouverez ci-joint la photocopie de nos billets. Recevez, Messieurs, l'assurance de mes sentiments distingués.

Portez ce vieux whisky au juge blond qui fume.
Comme le prévoit le règlement communautaire n° 295/91
du 14 février 1991, je vous demande le remboursement
de mes billets, ainsi qu'une indemnité forfaitaire de
1000 francs, soit l'équivalent de deux fois 150 écus
au cours d'aujourd'hui. Vous trouverez ci-joint la
photocopie de nos billets. Recevez, Messieurs, l'assurance
de mes sentiments distingués.

20

Figure 2: Example of a handwriting acquisition page used for the impostor handwriting scenario. The subject id and imitated id are shown on the top of the page, as well as the original text reference.

### 2.3 Comments on the acquisition

We noticed that all recorded 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 acquisition to happen.

For the signature, most of the users could actually read the content of their signature, synchronizing the written symbols with syllables. Most of the signatures contained some pre or post flourishes that were not said by the user. Very few users were having signatures composed mostly of flourishes that were not readable. These users were then asked to simply utter their name while signing. The average length of the spoken part of the signatures have been measured to be around 2 seconds.

For the handwriting, we have noticed that the voice is slightly desynchronized with the writing. Re-synchronization is happening and corresponds roughly to syllables or end of words. Some users are pronouncing the punctuation while some users are not pronouncing it. The reading of numbers is variable from user to user.

## 3 Usability survey

We asked each subject of the database to answer some questions about CHASM acquisition. The questionnaire included the following questions:

1. Did you find it simple/difficult to write on a tablet?
2. Do you think that you wrote faster, at the same speed or slowlyer than usual (without simultaneous speaking)?
3. Did you find it simple/difficult to speak and sign at the same time?
4. Did you find it simple/difficult to speak and write at the same time?
5. How many lines of text would you accept to say and write in order to perform your own identification in the scenario of accessing to your banking account?
6. Do you think that the act of speaking and writing at the same time affected your capacities to imitate the writing?

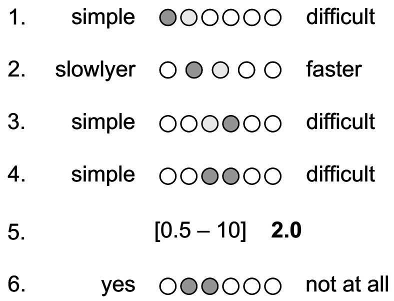


Figure 3: Results of the usability survey

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. Figure 3 presents a schematic view of the average answers given by all subjects. The main conclusions of the survey are the following. A large majority of users found it easy to write on a tablet. Users ranked as average the difficulty of writing and speaking at the same time. This observation is certainly due to the extra level of concentration that is requested to perform such acquisitions. All users were able to sign and utter the content of their signature, however they feel that signing and speaking is more difficult than writing and speaking. An interpretation can be given to this result considering the fact that signatures usually contain pre or post flourishes on which users cannot utter anything. Users feel they are writing at a slightly slower speed when they are speaking in the same time. Users would accept to write up to two lines of text to perform their authentication. Interestingly, they feel that the act of speaking and writing at the same time affected their capacities to imitate signatures and handwriting.

As all users were able to perform CHASM acquisitions and considering the answers given through the survey, we can reasonably conclude that such simultaneous acquisitions are acceptable from a usability point of view in most scenarios.

# 4 Evaluation protocols

User authentication assessment protocols have been defined on the recorded database. These protocols corresponds, as close as possible, to realistic use of potential CHASM biometric applications. Protocols have also been crafted to put in evidence difficulties tied to time-variability and to text-dependency.

## 4.1 Signatures with voice

**Without time variability.** For each subject in the database, three spoken signatures are sampled out randomly of the six genuine accesses from the first session to build the models. For testing, the three remaining signatures of the first session are used. The same procedure is repeated for session two and three, leading to a total of 70 users * 3 accesses * 3 sessions = 630 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 * 3 sessions = 14490 random forgeries. In the second case, the 18 available skilled forgeries (six forgeries performed by three subjects) are used against each user, giving a total of 70 users * 18 accesses * 3 sessions = 3780 skilled forgeries[1].

**With time variability.** 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. Random and skilled impostor attempts are performed in the same manner as above.

## 4.2 Handwriting with voice

**Text-dependent scenario.** In this scenario, we assume that the subject writes and says a fixed piece of text identical from access to access. The phrase which is used contains all the letters of the alphabet to allow modelling all the letters for each subject. The scenario corresponds to a text dependent (or password-based) system with the particularity that all users share the same text. For each subject in the database, the fixed phrase from the first session is used to train the system. For testing, the genuine fixed phrase from session two and three are used, leading to a total of 2 accesses * 70 users = 140 genuine tests. We also consider two kinds of impostor attempts. In the first case, we use random impostor attempts using one access for each of the remaining subjects in the database, giving a total of 70 users * 69 accesses = 4830 random forgeries. In the second case, 3 available skilled forgeries are used against each user, giving a total of 70 users * 3 accesses = 210 skilled forgeries.

**Text-prompted scenario.** In this scenario, 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 plays 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 phrase. For each subject in the database, the text from the first session is used to train the system. The available text for training is composed of about 5 lines including 50 to 100 words. Each genuine test uses one line of text taken from session two and session three. Therefore, on average, about 10 genuine tests can be performed per user, giving a total of 70 users * 10 accesses = 700 genuine tests. As for above, we consider two kinds of forgeries. Random forgeries are performed using one line of text of the remaining subjects, giving 70 users * 69 accesses = 4830 random forgeries. Skilled forgeries are performed using one line of text extracted from the 3 available imitations. As each imitation contains about 5 lines of text, we have a total of 70 users * 15 accesses = 1050 skilled forgeries.

# 5 Conclusions and Future Work

In this paper, we have reported on our first developments towards building a novel CHASM-based authentication system. An acquisition campaign of CHASM data has been conducted during which we have experimented that all users were able to perform the simultaneous acquisition of handwriting and speech. We have also experimented that CHASM data can be recorded following two scenarios: signature with voice and handwriting with voice. A usability survey which has been conducted during the acquisition shows that such simultaneous acquisitions are acceptable for the user from a usability point of view. We also presented the assessment protocols for user authentication that we have defined on the recorded database.

In our future work, we plan to build baseline and advanced CHASM modelling systems and to perform their evaluation using the protocols such as defined in this paper. We will also investigate the impact of simultaneous recordings. In this direction, we would like to determine if modalities recorded simultaneously bring different authentication information than modalities recorded independently. Another part of our work will be dedicated to pursue the acquisition campaign with a second set MYIDEA-CHASM-SET2 of CHASM data, similar to the one presented here. This second

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

set will be used to perform open-set evaluations of CHASM systems, allowing to develop and tune parameters of the model on the first set and to evaluate model configuration on the second set. We also plan to record a separate brute-force forgeries set to evaluate the impact of highly skilled attacks, making the assumption that forgers have access to the full dynamics of a given piece of CHASM data.

## Acknowledgments

This work was supported by the Swiss National Science Foundation program ”Interactive Multimodal Information Management (IM2)” in the Individual Project ”Access and Content Protection”, as part of NCCR.

## References

[1] F. Bimbot, J.-F. Bonastre, C. Fredouille, G. Gravier, I. Magrin-Chagnolleau, S. Meignier, T. Merlin, J. Ortega-Garcia, D. Petrovska, and D. A. Reynolds. A tutorial on text-independent speaker verification. *EURASIP Journal on Applied Signal Processing*, 4:430–451, 2004.

[2] B. Dumas, C. Pugin, J. Hennebert, D. Petrovska-Delacrétaz, A. Humm, F. Evéquoz, R. Ingold, and D. Von Rotz. 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).

[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 on The Advent of Biometrics on the Internet*, pages 83–86, November 2002. Rome, Italy.

[4] S. Garcia-Salicetti, C. Beumier, G. Chollet, B. Dorizzi, J. Leroux les Jardins, J. Lunter, Y. Ni, and D. Petrovska-Delacrétaz. Biomet: a multimodal person authentication database including face, voice, fingerprint, hand and signature modalities. In UK University of Surrey, Guildford, editor, *4th International Conference on Audio and Video-Based Biometric Person Authentication (AVBPA)*. Springer-Verlag, 2003.

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

[6] 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, December 2003. Santa Barbara, California, USA.

[7] U.-V. Marti and H. Bunke. A full english sentence database for off-line handwriting recognition. In *Proc. of the 5th Int. Conf. on Document Analysis and Recognition (ICDAR'99)*, pages 705–708, 1999.

[8] J. Ortega-Garcia, J. Fierrez-Aguilar, D. Simon, J. Gonzalez, M. Faundez-Zanuy, V. Espinosa, A. Satue, I. Hernaez, J.-J. Igarza, C. Vivaracho, D. Escudero, and Q.-I. Moro. Mcyt baseline corpus: a bimodal biometric database. *IEE Proc.-Vis. Image Signal Process.*, 150(6):395–401, December 2003.

[9] 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.

[10] R. Plamondon and G. Lorette. Automatic signature verification and writer identification - the state of the art. *Pattern Recognition*, 22(2):107–131, 1989.

[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.