IDIAP RESEARCH REPORT

IDIAP

Martigny - Valais - Suisse

LIKELIHOOD RATIO ADJUSTMENT FOR THE COMPENSATION OF MODEL MISMATCH IN SPEAKER VERIFICATION

Frédéric BIMBOT [+] Dominique GENOUD [*]

IDIAP–RR 97-05

August 1997

to appear in
Eurospeech, Rhodos 1997

Dalle Molle Institute for Perceptive Artificial Intelligence • P.O.Box 592 • Martigny • Valais • Switzerland

phone +41 – 27 – 721 77 11
fax +41 – 27 – 721 77 12
e-mail secretariat@idiap.ch
internet http://www.idiap.ch

+ ENST - Dépt Signal, CNRS - URA 820, 46 Rue Barrault, 75634 Paris cedex 13, FRANCE, European Union
* IDIAP, CP 592,CH-1920 Martigny

IDIAP Research Report 97-05

LIKELIHOOD RATIO ADJUSTMENT FOR THE COMPENSATION OF MODEL MISMATCH IN SPEAKER VERIFICATION

Frédéric BIMBOT   Dominique GENOUD

August 1997

TO APPEAR IN
Eurospeech, Rhodos 1997

Abstract. Cet article présente une méthode d'ajustement des seuils de vérification du locuteur basée sur un modèle Gaussien des distributions du logarithme du rapport de vraisemblance. L'article expose les hypothèses sous lesquelles ce modèle est valide, indique plusieurs méthodes d'ajustement des seuils, et en illustre les apports et les limites par des expériences de vérification sur une base de données de 20 locuteurs.

1 Introduction

Speaker verification systems rely on two main modules : a speaker modeling module and an acceptance/rejection decision module. When probabilistic models are used as speaker models (and non-speaker models), a classical decision rule is based on the likelihood ratio (LR) test, namely the comparison of the ratio between the speaker model and non-speaker model likelihoods to a pre-determined threshold. Usually, this threshold is set so as to optimise the overall system performance according to a particular cost function. In theory, it should not depend on the speaker.

In practice however, a mismatch between the model and the data is often observed, which invalidates the use of a pre-determined, speaker-independent threshold. Among the reasons for such a mismatch are the choice of an improper class of speaker models, the inappropriate dimensioning of the model with respect to the amount of training data, the non-representativity of the training material, the possible presence of outliers within the training utterances, etc...

In this paper, we present a way to adjust the LR test in order to correct for (some of) the model mismatch, under a few hypothesis concerning the statistical model. We show how an adjusted threshold can be estimated from the mean and standard deviation of the distribution of the frame-by-frame likelihood values at the output of the speaker model and of the non-speaker model. We compare several ways for estimating these means and standard deviations. We finally illustrate the benefits and the limits of this adjustment by a series of experiments in speaker verification on telephone data.

2 Theoretical aspects

2.1 Formalism

Let X denote a speaker, and $\mathcal{X}$ a probabilistic model of that speaker. Let $\bar{\mathcal{X}}$ denote the *non-speaker* model for speaker X, i.e the model of the rest of the population. Let Y be a speech utterance claimed as belonging to speaker X.

If we denote as $\hat{X}$ (resp. $\hat{\bar{X}}$) the acceptance (resp. rejection) decision of the system, and p_X (resp. $p_{\bar{X}}$) the a priori probability of the claimed speaker to be (resp. not to be) speaker X, the total cost function of the system is [1] :

$$C \;=\; C_{(\hat{X}|\bar{X})} \cdot p_{\bar{X}} \,.\, P\,(\hat{X}|\bar{X}) \;+\; C_{(\hat{\bar{X}}|X)} \cdot p_X \,.\, P\,(\hat{\bar{X}}|X) \tag{1}$$

where $P\,(\hat{X}|\bar{X})$ and $P\,(\hat{\bar{X}}|X)$ denote respectively the probability of a false acceptance and of a false rejection, while $C_{(\hat{X}|\bar{X})}$ and $C_{(\hat{\bar{X}}|X)}$ represent the corresponding costs[1].

If we now denote by $P_{\mathcal{X}}$ and $P_{\bar{\mathcal{X}}}$ the likelihood functions of the speaker and of the non-speaker models, the minimisation of C in equation (1) is obtained by implementing the likelihood ratio (LR) test [2] :

$$LR\,(Y) \;=\; \frac{P_{\mathcal{X}}\,(Y)}{P_{\bar{\mathcal{X}}}\,(Y)} \;\underset{reject}{\overset{accept}{\gtrless}}\; \Theta\,(R) \;=\; R \tag{2}$$

where $P_{\mathcal{X}}$ and $P_{\bar{\mathcal{X}}}$ denote the likelihood functions for the speaker and the non-speaker, and R is the risk ratio :

$$R \;=\; \frac{C_{(\hat{X}|\bar{X})}}{C_{(\hat{\bar{X}}|X)}}\,\frac{p_{\bar{X}}}{p_X} \tag{3}$$

[1] We assume a null cost for a true acceptance and a true rejection.

As can be seen from equation (3), the optimal threshold does not depend, in theory, of anything else than the false acceptance / rejection cost ratio and the impostor / client a priori probability ratio. In the particular case when the costs $C_{(\hat{X}|\bar{X})}$ and $C_{(\hat{\bar{X}}|X)}$ are equal, and when genuine speakers and impostors are assumed a priori equiprobable, the system is set to the *equirisk* condition, and the choice of $\Theta = 1$ as a decision threshold should then lead to a minimum *Total Error Rate* for the system : $TER = P(\hat{X}|\bar{X}) + P(\hat{\bar{X}}|X)$.

2.2 Adjustment of the LR test

In practice, it is often observed that the LR test with $\Theta = R$ as a decision threshold may not yield the minimum of the cost function C. In fact, the LR in equation (2) is calculated from estimations of the likelihood functions, which do not match the exact speaker and non-speaker distributions. As a consequence, it is usually beneficial to adjust the threshold of the LR test, in order to correct for the improper fit between the model and the data.

By denoting as $\hat{P}_{\mathcal{X}}$ and $\hat{P}_{\bar{\mathcal{X}}}$ the respective *model* likelihood functions for the speaker and the non-speaker, the LR test can be rewritten in a more general framework as :

$$\widehat{LR}(Y) \;=\; \frac{\hat{P}_{\mathcal{X}}(Y)}{\hat{P}_{\bar{\mathcal{X}}}(Y)} \;\underset{reject}{\overset{accept}{\gtrless}}\; \Theta_X(R, n) \tag{4}$$

with n corresponding to the number of frames in the test utterance Y. The function Θ_X can be viewed as a speaker-dependent adjusted threshold that accounts for the speaker and non-speaker model mismatch causing a difference between LR and $\widehat{LR}$ and for the influence of the utterance length on the distribution of $\widehat{LR}$.

In the general case, there is no straightforward way of modeling or estimating $\Theta_X(R, n)$. However, if we assume that the model log-likelihood function is obtained as the average of a large number of independent frame-based log-likelihood values, the adjusted threshold Θ_X relates directly to the mean and the variance of the distributions followed by the client and impostor frame-by-frame likelihoods at the output of the speaker and non-speaker models.

2.3 Distribution of the $\widehat{LR}$

In fact, for most conventional probabilistic models, the logarithm of the numerator in $\widehat{LR}$ can be rewritten as the average of a set of frame-based log-likelihoods[2] :

$$\log\, \hat{P}_{\mathcal{X}}(Y) \;=\; \frac{1}{n} \sum_{i\,=\,1}^{i\,=\,n} \log\, \hat{P}_{\mathcal{X}}(y_i) \tag{5}$$

where y_i denotes the i^{th} frame in utterance Y, of total length n. If n is large enough and if the frame-based log-likelihood values $\log \hat{P}_{\mathcal{X}}(y_i)$ are assumed independent, $\log \hat{P}_{\mathcal{X}}(Y)$ follows a Gaussian distribution $\mathcal{G}(\mu_{\mathcal{X}};\, \sigma_{\mathcal{X}}/\sqrt{n})$, where $\mu_{\mathcal{X}}$ and $\sigma_{\mathcal{X}}$ are the mean and variance of the distribution of the frame-by-frame likelihood, whatever type of distribution is actually followed by this function (Central Limit Theorem). The same property also stands for the logarithm of the denominator, i.e $\log \hat{P}_{\bar{\mathcal{X}}}(Y)$.

Therefore :

$$\begin{aligned} \log\, \hat{P}_{\mathcal{X}}(Y) &\longrightarrow \mathcal{G}(\mu_{\mathcal{X}};\, \sigma_{\mathcal{X}}/\sqrt{n}) \\ \log\, \hat{P}_{\bar{\mathcal{X}}}(Y) &\longrightarrow \mathcal{G}(\mu_{\bar{\mathcal{X}}};\, \sigma_{\bar{\mathcal{X}}}/\sqrt{n}) \end{aligned}$$

[2] A very similat approach can be readily adapted to classifiers using the *sum* instead of the average log-likelihood.

If we now distinguish between utterances which were actually uttered by speaker X and those that were not $(\bar{X})$, the numerator and the denominator follow distinct conditional distributions :

For client utterances :

$$\log \hat{P}_{\mathcal{X}}(Y|X) \longrightarrow \mathcal{G}\left(\mu_{\mathcal{X}}(X);\ \sigma_{\mathcal{X}}(X)/\sqrt{n}\right)$$
$$\log \hat{P}_{\bar{\mathcal{X}}}(Y|X) \longrightarrow \mathcal{G}\left(\mu_{\bar{\mathcal{X}}}(X);\ \sigma_{\bar{\mathcal{X}}}(X)/\sqrt{n}\right)$$

For impostor utterances :

$$\log \hat{P}_{\mathcal{X}}(Y|\bar{X}) \longrightarrow \mathcal{G}\left(\mu_{\mathcal{X}}(\bar{X});\ \sigma_{\mathcal{X}}(\bar{X})/\sqrt{n}\right)$$
$$\log \hat{P}_{\bar{\mathcal{X}}}(Y|\bar{X}) \longrightarrow \mathcal{G}\left(\mu_{\bar{\mathcal{X}}}(\bar{X});\ \sigma_{\bar{\mathcal{X}}}(\bar{X})/\sqrt{n}\right)$$

For instance, the notation $\mu_{\mathcal{X}}(\bar{X})$ represents the expected value of the log-likelihood of impostor utterances when scored by the speaker model.

Ultimately, under the assumption that $\log \hat{P}_{\mathcal{X}}(y_i)$ and $\log \hat{P}_{\bar{\mathcal{X}}}(y_i)$ are independent random variables, the log-$\widehat{LR}$ follows the two conditional distributions :

$$\log \widehat{LR}(Y|X) \longrightarrow \mathcal{G}\left(m_X;\ s_X/\sqrt{n}\right) = \mathcal{G}_X^{(n)}$$
$$\log \widehat{LR}(Y|\bar{X}) \longrightarrow \mathcal{G}\left(m_{\bar{X}};\ s_{\bar{X}}/\sqrt{n}\right) = \mathcal{G}_{\bar{X}}^{(n)}$$

with :

$$\begin{array}{l|l} m_X = \mu_{\mathcal{X}}(X) - \mu_{\bar{\mathcal{X}}}(X) & s_X = \sqrt{\sigma_{\mathcal{X}}(X)^2 + \sigma_{\bar{\mathcal{X}}}(X)^2} \\ m_{\bar{X}} = \mu_{\mathcal{X}}(\bar{X}) - \mu_{\bar{\mathcal{X}}}(\bar{X}) & s_{\bar{X}} = \sqrt{\sigma_{\mathcal{X}}(X)^2 + \sigma_{\bar{\mathcal{X}}}(X)^2} \end{array}$$

2.4 Expression of the adjusted threshold

If we now denote :

$$\mathcal{F}_X^{(n)}(\tau) = \int_{-\infty}^{\tau} \mathcal{G}_X^{(n)}(v)\, dv$$
$$\mathcal{F}_{\bar{X}}^{(n)}(\tau) = \int_{-\infty}^{\tau} \mathcal{G}_{\bar{X}}^{(n)}(v)\, dv$$

the functions $(1 - \mathcal{F}_{\bar{X}}^{(n)}(\tau))$ and $\mathcal{F}_X^{(n)}(\tau)$ can then be understood as models of the false acceptance and false rejection probabilities as a function of the threshold τ, and can be used for the minimisation of the overall cost of equation (1). In this case, $\Theta_X(R,n)$ is expressed as :

$$\log \Theta_X(R,n) = \\ Argmin_{\,\tau} \left\{ R\left(1 - \mathcal{F}_{\bar{X}}^{(n)}(\tau)\right) + \mathcal{F}_X^{(n)}(\tau) \right\} \qquad (6)$$

The adjusted threshold $\Theta_X(R,n)$ is thus estimated from the Gaussian model of the log-likelihood distributions yielded by the speaker and the non-speaker models, for speech data uttered by the true speaker and by other speakers.

3 Estimation of the threshold

3.1 Data sets

For each client speaker X, the speaker model $\mathcal{X}$ is trained from an enrollment set $\mathcal{E}$ and the non-speaker model $\bar{\mathcal{X}}$ is estimated from an other set of data $\bar{\mathcal{E}}$. This last set is composed of speech data produced by a given population of non-speakers. We also have an other set of data, denoted $\bar{\mathcal{I}}$ and produced by a third population of *pseudo*-impostors, which has no intersection with the two previous populations. It is then possible to estimate the means $\mu_{\mathcal{X}}(\bar{X})$ and $\mu_{\bar{\mathcal{X}}}(\bar{X})$ and the standard deviations

$\sigma_{\mathcal{X}}(\bar{X})$ and $\sigma_{\bar{\mathcal{X}}}(\bar{X})$ by scoring the pseudo-impostor data $\bar{\mathcal{I}}$ both with the speaker and the non-speaker models.

If we have an additional set of data $\mathcal{T}$, corresponding to speech material uttered by speaker X, which have not yet been used for the enrollment of model $\mathcal{X}$, these *tuning* data can be used in order to estimate the two remaining means and standard deviations $\mu_{\mathcal{X}}(X)$, $\mu_{\bar{\mathcal{X}}}(X)$ and $\sigma_{\mathcal{X}}(X)$, $\sigma_{\bar{\mathcal{X}}}(\bar{X})$, by simply scoring them with models $\mathcal{X}$ and $\bar{\mathcal{X}}$. In a first series of experiments reported below, we adopted this approach based on separate tuning data.

3.2 Estimation of the means and variances

In our experiments, we have used 3 distinct estimators of the mean and variance of the likelihood distributions :

1) *100 % data :* the maximum likelihood estimators are used, i.e the classical mean and standard deviation of a data population.

2) *95 % most typical :* the mean and standard deviation computed from the 95 % most typical frames (at the utterance level), i.e. after having removed the 2.5 % minimum and 2.5 % maximum frame likelihood values.

3) *95 % best :* the mean and standard deviation computed from the 95 % frames with the highest likelihood (at the utterance level), i.e. after having removed the 5 % minimum frame likelihood values.

When a 95 % frame selection approach was used for the mean and standard deviation estimations, the same approach was used during the test, before computing the average client likelihood of equation (5), and the average non-speaker likelihood.

3.3 Enrollment data as tuning data

In the context of practical applications, the need for collecting separate tuning data for each client can be a quite heavy constraint. It is indeed desirable to estimate the decision threshold from the training data themselves. In this context, we have considered a particular case, consisting in calculating $\mu_{\mathcal{X}}(X)$, $\mu_{\bar{\mathcal{X}}}(X)$ and $\sigma_{\mathcal{X}}(X)$, $\sigma_{\bar{\mathcal{X}}}(X)$ using the enrollment data $\mathcal{E}$.

For what concerns the means and standard deviations estimated for non-speaker data, we continue resorting to the separate set of pseudo-impostor data $\bar{\mathcal{I}}$, as this is not a severe constraint for a practical application.

3.4 Approximation of the adjusted threshold

An approximation of the solution of equation (6) can be obtained numerically using the approximation $\hat{\mathcal{F}}(\tau)$ of $\int_{-\infty}^{\tau} \mathcal{G}(v)\, dv$, calculated as follows [3] :

$$u = \frac{\tau - \mu}{\sigma} \;\Big|\; s = sgn\,(u) \;\Big|\; g = \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}u} \;\Big|\; t = \frac{1}{1 + a\,s\,u}$$

$$f = 1 - g\,(b_1 t + b_2 t^2 + b_3 t^3 + b_4 t^4 + b_5 t^5)$$

$$\text{and ultimately : } \hat{\mathcal{F}} = sf + \frac{1-s}{2}$$

with the following numerical constants :

$$\begin{array}{llll} a & = & 0.231641900 & b_1 = 0.319381530 \\ b_2 & = & -0.356563782 & b_3 = 1.781477937 \\ b_4 & = & -1.821255978 & b_5 = 1.330274429 \end{array}$$

4 Experiments

4.1 Database

The approach described above was tested in text-dependent speaker verification. The database is composed of 35 speakers who recorded up to 10 telephone sessions containing 5 times their 7-digit card number and 5 times the 10 (connected) digits in a random order. The language is French.

A sub-population $\mathcal{C}$ of 20 speakers (15 male and 5 female) made 10 calls, and are considered as clients. Each call is composed of a maximum of 4 utterances. The client data are split into 3 subsets : the enrollment set $\mathcal{E}$, corresponding to 1 or 2 calls times 4 utterances (resp. $\mathcal{E}_1$ and $\mathcal{E}_2$), a tuning set $\mathcal{T}$ consisting of the first utterance of 1, 2 or 3 *other* calls ($\mathcal{T}_1$, $\mathcal{T}_2$, $\mathcal{T}_3$) and a test set $\mathcal{A}$ composed of all valid utterances in the 5 *remaining* calls (374 tests altogether). Clients are also used randomly as impostors against other clients (set $\bar{\mathcal{A}}$: 3496 impostor attempts).

Part of the utterances from the 15 other speakers (10 male, 5 female) are used as the set of pseudo-impostors data. The non-speaker model is a world-model trained on a set of 500 speakers (distinct from the 35). Thus $\bar{\mathcal{I}}$ and $\bar{\mathcal{E}}$ are distinct and both independent of the client.

Pseudo-impostor and impostor (test) utterances against a given client are generated by rearranging digit segments from the 10-digit sequence in the same order as those in the client's 7-digit number. The segmentation is yielded by the world-model.

4.2 Speaker verification algorithm

The speech signal is represented by 12 LPC-Cepstrum coefficients plus the log-energy, together with their delta and delta-delta, totalling 39 acoustic features per frame[3]. Each speaker model and the world-model are word-based Left-Right Hidden Markov Models, with 3 states per phoneme and 1 Gaussian mixture per state.

Viterbi decoding is used for verification, and frame-based likelihoods are taken along the Viterbi path.

4.3 Results

Tables 1, 2 and 3 summarize results obtained on the test set, with various threshold tuning procedures. They correspond respectively to the 3 estimations of the means and variances exposed in section 3.2. The False Acceptance and False Rejection Rates (FAR and FRR) are computed as the average of each speaker's FAR and FRR. The threshold adjustment procedure is used to tune the system to a minimum TER, in the equirisk condition ($R = 1$).

In each table, the top part corresponds to enrollment set $\mathcal{E}_1$, whereas the bottom part corresponds to $\mathcal{E}_2$. Scores 1 and 2 are obtained with a threshold setting procedure that uses only the enrollment data. Scores 3, 4 and 5 resort to 1, 2 or 3 additional client (tuning) utterances. Scores 6 and 7 are obtained a posteriori, on the test data : they are reported for comparison with the others, but they are not relevant from an application point of view.

The difference between scores 1 and 2 illustrates the benefit of using a speaker-dependent adjusted threshold. Scores 3, 4 and 5 show that the use of 2 utterances of tuning data yields a lower TER than the one obtained with the enrollment data. Frame selection for computing the LR and adjusting the threshold seems an efficient strategy, with a slight advantage for the *95 % best*. The difference between scores 5 and 6 are owed to a better estimation of the test log-likelihood distribution parameters using the test data, whereas the difference between scores 6 and 7 translate the fact that the Gaussian model for the log-likelihood distribution does not match exactly the test data log-likelihood distribution.

[3]The frame size is 25.6 ms, with a frame shift of 10 ms. Pre-emphasis coefficient is 0.94. A Hamming window is used.

	Estim. of Θ	FAR (%)	FRR (%)	TER (%)
1)	$\Theta = 1$	0.40	28.11	28.51
2)	$\mathcal{E}_1$ and $\bar{\mathcal{I}}$	0.47	8.94	9.41
3)	$\mathcal{T}_1$ and $\bar{\mathcal{I}}$	5.37	3.18	8.55
4)	$\mathcal{T}_2$ and $\bar{\mathcal{I}}$	5.11	2.59	7.70
5)	$\mathcal{T}_3$ and $\bar{\mathcal{I}}$	4.81	1.58	6.39
6)	$\mathcal{A}$ and $\bar{\mathcal{A}}$	2.83	1.00	3.83
7)	min TER	0.11	1.29	1.40
1)	$\Theta = 1$	1.21	6.47	7.68
2)	$\mathcal{E}_2$ and $\bar{\mathcal{I}}$	0.37	6.68	7.05
3)	$\mathcal{T}_1$ and $\bar{\mathcal{I}}$	4.68	2.25	6.93
4)	$\mathcal{T}_2$ and $\bar{\mathcal{I}}$	3.84	2.25	6.09
5)	$\mathcal{T}_3$ and $\bar{\mathcal{I}}$	3.84	1.25	5.09
6)	$\mathcal{A}$ and $\bar{\mathcal{A}}$	2.44	1.00	3.44
7)	min TER	0.03	1.29	1.32

Table 1: False Acceptance Rate, False Rejection Rate and Total Error Rate for various adjustments of the threshold Θ. Top : 1 enrollment session - Bottom : 2 enrollment sessions. Estimation of the means and variances using the *100 % data* method.

5 Conclusions

The use of Gaussian models of the speaker and impostor log-likelihood ratio distributions provides a simple yet reasonably accurate procedure for a priori threshold setting in speaker verification. The means and variances of these models can be estimated from the enrollment data and/or from very few tuning data. Future work will aim at increasing the robustness of the method, will consolidate it with wider scale experiments, and will study its behaviour in other risk conditions.

References

[1] R.O. DUDA, P.E. HART : *Pattern Classification and Scene Analysis.* John Wiley & Sons, 1973.

[2] L.L. SCHARF : *Statistical Signal Processing. Detection, Estimation and Time Series Analysis.* Addison-Wesley Publishing Company, 1991.

[3] G. SAPORTA : *Probabilités, analyse des données et statistique.* Editions Technip, 1990.

	Estim. of Θ	FAR (%)	FRR (%)	TER (%)
1)	$\Theta = 1$	0.58	20.89	21.47
2)	$\mathcal{E}_1$ and $\bar{\mathcal{I}}$	0.76	5.49	6.25
3)	$\mathcal{T}_1$ and $\bar{\mathcal{I}}$	4.29	2.77	7.06
4)	$\mathcal{T}_2$ and $\bar{\mathcal{I}}$	3.21	2.47	5.68
5)	$\mathcal{T}_3$ and $\bar{\mathcal{I}}$	3.00	2.47	5.47
6)	$\mathcal{A}$ and $\bar{\mathcal{A}}$	2.90	1.00	3.90
7)	min TER	0.18	2.00	2.18
1)	$\Theta = 1$	1.74	5.88	7.62
2)	$\mathcal{E}_2$ and $\bar{\mathcal{I}}$	0.31	3.43	3.74
3)	$\mathcal{T}_1$ and $\bar{\mathcal{I}}$	3.13	1.84	4.97
4)	$\mathcal{T}_2$ and $\bar{\mathcal{I}}$	2.68	1.84	4.52
5)	$\mathcal{T}_3$ and $\bar{\mathcal{I}}$	2.45	1.84	4.29
6)	$\mathcal{A}$ and $\bar{\mathcal{A}}$	2.57	1.00	3.57
7)	min TER	0.03	1.00	1.03

Table 2: False Acceptance Rate, False Rejection Rate and Total Error Rate for various adjustments of the threshold Θ. Top : 1 enrollment session - Bottom : 2 enrollment sessions. Estimation of the means and variances using the *95 % most typical* method.

	Estim. of Θ	FAR (%)	FRR (%)	TER (%)
1)	$\Theta = 1$	1.08	16.54	17.62
2)	$\mathcal{E}_1$ and $\bar{\mathcal{I}}$	0.95	4.94	5.89
3)	$\mathcal{T}_1$ and $\bar{\mathcal{I}}$	4.26	2.00	6.26
4)	$\mathcal{T}_2$ and $\bar{\mathcal{I}}$	3.55	1.50	5.05
5)	$\mathcal{T}_3$ and $\bar{\mathcal{I}}$	2.89	1.00	3.89
6)	$\mathcal{A}$ and $\bar{\mathcal{A}}$	2.96	1.00	3.96
7)	min TER	0.05	1.54	1.59
1)	$\Theta = 1$	2.21	5.34	7.55
2)	$\mathcal{E}_2$ and $\bar{\mathcal{I}}$	0.71	4.09	4.80
3)	$\mathcal{T}_1$ and $\bar{\mathcal{I}}$	3.66	1.50	5.16
4)	$\mathcal{T}_2$ and $\bar{\mathcal{I}}$	2.95	1.25	4.20
5)	$\mathcal{T}_3$ and $\bar{\mathcal{I}}$	2.87	1.00	3.87
6)	$\mathcal{A}$ and $\bar{\mathcal{A}}$	2.65	1.00	3.65
7)	min TER	0.00	1.00	1.00

Table 3: False Acceptance Rate, False Rejection Rate and Total Error Rate for various adjustments of the threshold Θ. Top : 1 enrollment session - Bottom : 2 enrollment sessions. Estimation of the means and variances using the *95 % best* method.