Open access • Proceedings Article • DOI:10.1109/ODYSSEY.2006.248116

How to Deal with Multiple-Targets in Speaker Identification Systems?

— Source link

Yaniv Zigel, Moshe Wasserblat

Institutions: NICE Systems

Published on: 28 Jun 2006

Topics: Speaker diarisation, Speaker recognition, False alarm and Word error rate

How to Deal with Multiple-Targets in Speaker Identification Systems?

Yaniv Zigel and Moshe Wasserblat

NICE Systems Ltd., Audio Analysis Group, P.O.B. 690 Ra'anana 43107, Israel
yanivz@nice.com

Abstract

In open-set speaker identification systems a known phenomenon is that the false alarm (accept) error rate increases dramatically when increasing the number of registered speakers (models). In this paper, we demonstrate this phenomenon and suggest a solution using a new model-dependent score-normalization technique, called Top-norm. The Top-norm method was specifically developed to improve results of open-set speaker identification systems. Also, we suggest a score-normalization parameter adaptation technique. Experiments performed using speaker recognition corpora are described and demonstrate that the new method outperforms other normalization methods.

1. Introduction

Open-set speaker identification or multi-target detection [1] is the most challenging type of speaker recognition. It aims to determine if an input utterance (test) is from a specified speaker (target) which is modeled in a target-set (models, stack, or watch-list). The task is open-set since the test can be from non-target speakers (unknown speakers). Open-set speaker identification has applications to the audio database search task for recorded telephone interactions, recorded meetings, or other historical audio documents, where a set of speakers (targets) are searched.

Open-set identification systems suffer from the "false alarm (FA) error rate problem". The FA increases dramatically with increasing stack size. This problem does not exist in verification systems. In order to deal with this FA problem, a new speaker-dependent (model-dependent) score-normalization called Top-norm is presented. In this normalization method, only the top-impostor-scores which produce the FA errors in a low-FA tuned verification system are to be considered in the normalization process.

In speaker recognition systems based on stochastic models (such as GMM) likelihood scores are very sensitive to variations in text, speaking behavior, and recording conditions [2]. Score normalization has been introduced explicitly to cope with score variability and to make speaker-independent decision threshold tuning easier [3]. It has been shown that score normalization is very effective in improving open-set speaker identification system performances [4] as well as in speaker verification systems [2], [5].

In this paper, the open-set speaker identification problem is discussed and emphasized using some simulated results. We show that significantly better identification results can be achieved, theoretically, when decreasing the standard deviation of the non-target score distribution in the prototype verification system, even when the equal error rate (EER) is kept. Open-set speaker identification experiments using speaker recognition databases are presented and the proposed Top-norm score normalization method is compared with other methods such as world model normalization (WMN), Z-norm, unconstrained cohort normalization (UCN), and T-norm.

2. The False Accept Error Problem of Identification Systems

Given a set of known speakers (target-set) and a sample utterance of an unknown speaker (test), open-set speaker identification (OSI) is defined as a twofold problem [4]: 1) Identifying the speaker model in the set, which best matches the test utterance. 2) Decision whether the test utterance has actually been produced by the speaker associated with the best-matched model (target), or by unknown speaker outside the target-set (non-target). This decision is made using a threshold operated on the maximum score s^* produced by the target-set models. This open-set identification can be seen also as multi-target detector [1], where each target-model in the test phase is operated as a single target detector.

Suppose that M speakers (targets) are enrolled in the system and their statistical models are $\lambda_1, \lambda_2, \ldots, \lambda_M$. If $\mathbf{O}$ denotes the feature vectors sequence extracted from the test utterance then the open-set identification can be stated as follows:

$$s^* = \max_{1 \le m \le M} \left\{ s\left(\mathbf{O} \mid \lambda_m\right) \right\} \begin{cases} \ge \tau \\ < \tau \end{cases} \rightarrow \mathbf{O} \in \begin{cases} \arg\max_{1 \le m \le M} \left\{ s\left(\mathbf{O} \mid \lambda_m\right) \right\} \\ \text{unknown speaker} \end{cases}$$

where τ is a pre-determined threshold and $s\left(\mathbf{O} \mid \lambda_m\right)$ is a probabilistic score of $\mathbf{O}$ given a target model λ_m .

Three types of errors exist in an open-set identification system: 1) FA error - occurs when the maximum score s^* is above threshold given that the test is of a non-target speaker. 2) False reject error (FR or miss) - occurs for a target test when the maximum score s^* is below threshold. 3) Confusion error - occurs for a target test when the maximum score s^* is above threshold; however the model that yields the maximum score is not belong to the input tested speaker. We may consider the case of *group detector* (or top-*S* stack detector [1]), which is the case where we only want to know if the tested input speaker belongs to the target-set or not; the exact identity of the speaker is not necessary. Thus, the confusion error does not exist.

The overlapping between the distributions of the non-target scores and of the target scores in open-set identification is greater than the overlapping of impostor scores and target scores in speaker verification. This is because of the maximum score selection (between the models' scores) in the identification process; the bigger the target-set size (*M*), the greater is this overlapping.

In order to emphasize the problem of open-set identification versus verification, a simulated test was performed using equations of predicted miss and false alarm probabilities (1), (2). These equations were derived for the group detector (of size *M*) probabilities of miss $\hat{P}_{miss}(\tau)$ and false alarm (false accept) $\hat{P}_{fa}(\tau)$ using the miss $P_{miss}(\tau)$ and false alarm $P_{fa}(\tau)$ probabilities (τ - threshold) of the prototype verification system (single detector). For these equations, we assume that the (*M*) detectors operate independently of each other, that they all have the same miss and false alarm probabilities, and that the prior probabilities of the target classes are equal [1]

$$\hat{P}_{fa}(\tau) = 1 - (1 - P_{fa}(\tau))^M \quad (1)$$

$$\hat{P}_{miss}(\tau) = P_{miss}(\tau) \cdot (1 - P_{fa}(\tau))^{M-1} \quad (2)$$

In this test, the target scores and the impostor scores of the prototype single target detector (verification system) were simulated using Gaussian distributions. Fig. 1 shows the simulated score histograms of the first simulated experiment. The mean of the target scores (μ_T) is 1, the mean of the non-target (impostors) scores (μ_{nT}) is -3.9, and the standard deviation of both scores (σ_T, σ_{nT}) is 1.5. Fig. 2 shows the detection error trade-off (DET) curves of the predicted identification system using equations (1) and (2) calculated from the scores distribution of Fig. 1 for different stack (target-set) sizes ($M = 1, 2, ..., 10, 20, ..., 100, 200, ..., 1000$). From Fig. 2 one can see that the stack size has a major influence on the group detector performances. For an equal error rate (EER) of 5% we get EER of 12.8% for stack size of 10 and EER of 23.9% for a stack size of 100. In many applications, these error rates are too large to work with; especially if we have very large stack sizes (over 100).

Fig. 3 shows the group detector miss and false alarm probabilities for stack sizes $M = 1, 2, ..., 1000$ for a multi-target detector composed of prototypes (single-detectors) operating at $P_{miss}(\tau) = 0.2,\ P_{fa}(\tau) = 0.0075$. From this figure one can see that the false alarm rate increases dramatically with increasing stack size. These $(P_{miss}(\tau), P_{fa}(\tau))$ points can be seen also by the circles in Fig. 2. As has been observed also by others [1] [6], we need single detectors (verification systems) that produce very low FA rates in order to have a reasonable FA rate when combining these single-detectors into multi-target (identification) systems.

Fig. 4 shows another example of simulated score histograms. This time, the standard deviation of the non-target score distribution ($\sigma_{nT} = 0.8$) is less than the standard deviation of the target score distribution ($\sigma_T = 1.5$). The mean of the non-target score distribution was chosen to be $\mu_{nT} = -2.76$ in order to have the same EER for the verification results as in the previous simulated example (EER = 5%). Fig. 5 shows the DET curves of the predicted identification system using equations (1) and (2) calculated from the scores distribution of Fig. 4 for different stack (target-set) sizes ($M = 1, 2, ..., 10, 20, ..., 100, 200, ..., 1000$). From this figure one can see that significantly better identification results are achieved as compared to the previous example (Fig. 2), especially in the low FA areas. For example, for *M* = 100, the EER of this system is 14.8% as compared to 23.9% in the previous one, although both prototype verification systems have the same EER value (5%). This phenomenon is caused by the fact that when decreasing the standard deviation of the non-target score distribution, the DET curve is tilted counterclockwise. The "width" of the non-target score distribution right "tail" is narrower, which causes lower FA rates for a given FR rate (in the FR > EER area). Hence identification systems, whose DET curve points can be seen as transformed from the lower FA area points of the prototype verification system (see the circles in Fig. 5), perform better.

Score normalization techniques may cause this kind of "tilt" in the prototype verification DET curve [7]. In order to achieve this effect, we choose to use a score normalization method which is presented next.

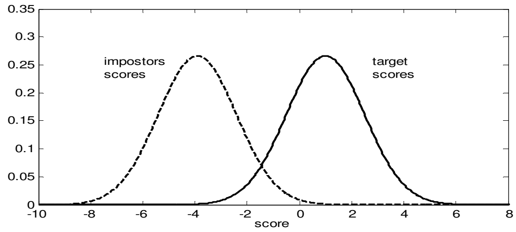

Figure 1: Histograms of the first simulated test scores.

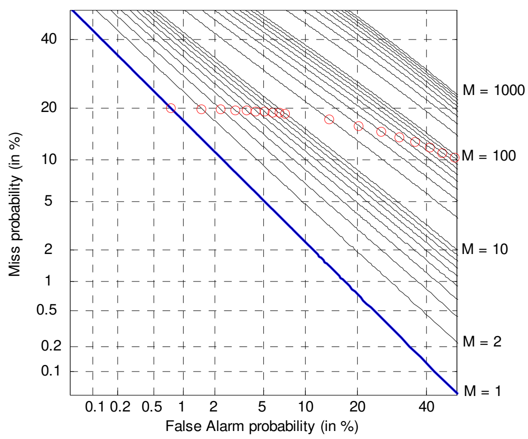

Figure 2: DET curves of the first simulated test.

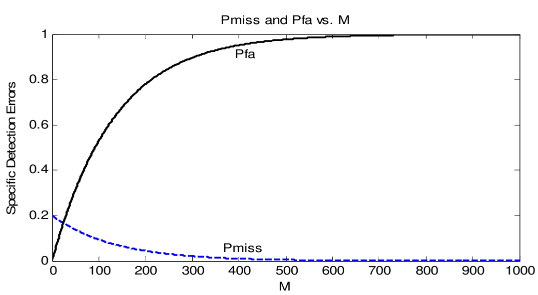

Figure 3: The group detector miss and false alarm probabilities vs. stack size for a group detector composed of prototypes operating at $P_{miss}(\tau) = 0.2,\ P_{fa}(\tau) = 0.0075$.

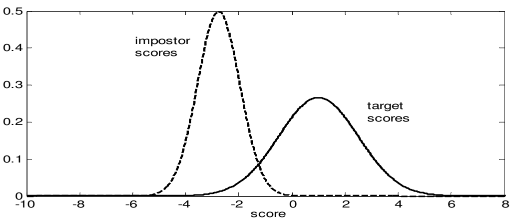

Figure 4: Histograms of the second simulated test scores.

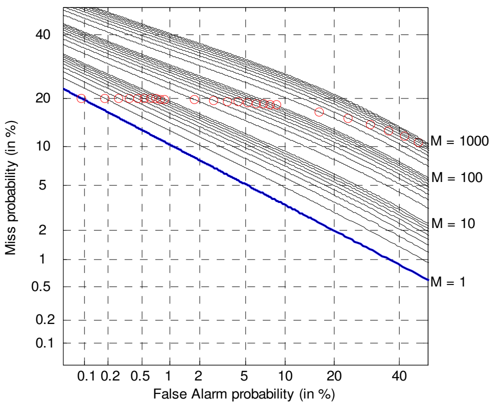

Figure 5: DET curves of the second simulated test.

3. The Top-Norm Method

Even after world model normalization (WMN) the score distribution is different for each speaker model (target). This causes relatively high FA errors for some targets and relatively high FR (miss) errors for other targets. Fig. 7 (left column) demonstrates this phenomenon; it shows an example of impostor score distributions of 20 different speakers (models), where the scores have WMN only. To compensate for this phenomenon, Z-norm score normalization method [8] has been proposed and used. In this method speaker-dependent mean and variance (normalization parameters) are estimated from the speaker- (model-) dependent impostor (non-target)-score distribution (see impostor Z-norm score distribution on Fig. 7 – middle column). However, this method assumes that this distribution for each model is Gaussian. Practically, it is not accurate. Most often the right tail may be considered Gaussian, but not the overall distribution. This is mainly (but not only) because of the existence of different channels and/or handset devices and, in some cases, both male and female trials. To deal with this phenomenon and still to perform speaker-dependent score-normalization, a new score-normalization technique called Top-norm is presented.

In the Top-norm method, only the top-impostor-scores which produce the FA errors in a low-FA tuned verification system are to be considered. Top-norm is similar to Z-norm, in means of calculating two parameters for each target-speaker: mean and standard-deviation, however, these parameters are not to be calculated from all the impostor scores, only from the top-scores (see Fig. 6). Hence, the normalized score is

$$s_n = \frac{s - \mu^{\%}}{\sigma^{\%}} \quad (3)$$

where $\mu^{\%}$ and $\sigma^{\%}$ are the top-scores mean and standard-deviation respectively, calculated from a pre-defined percent value (*p*, *topPercent*) of the top-scores.

Fig. 7 (right column) shows impostor score distributions of the different speakers as in the left columns, however the scores are normalized using the top-norm method (*topPercent* = 10%). From this figure one can see that all distributions are "right-aligned", causing equally dispersed top-scores for all the target-speakers (models). Fig. 8 shows DET curves of these three normalization methods (WMN, Z-norm, and Top-norm) combined in a verification system (database description in the next chapter). From this figure one can see that although all these three methods have almost the same EER, Top-norm is superior to the others in the low FA area ($P_{fa}(\tau) < 0.5\%$). As investigated in chapter 2, this implies better identification results using large stack sizes.

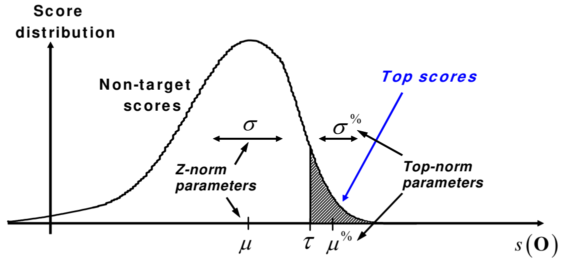

Figure 6: Illustration of the parameters of the Top-norm.

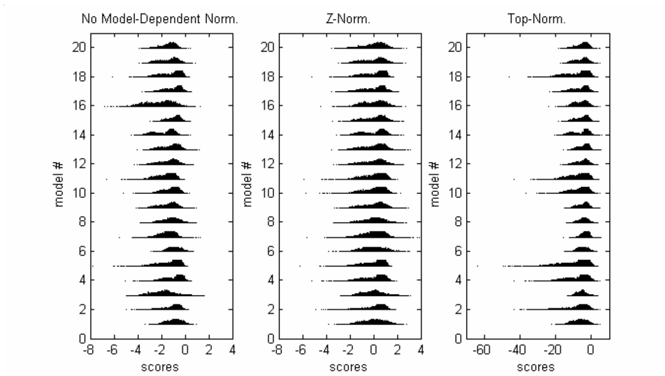

Figure 7: Speaker-dependent non-target score distributions for three normalization methods: 1) no speaker-dependent normalization (WMN only), 2) Z-norm, and 3) Top-norm.

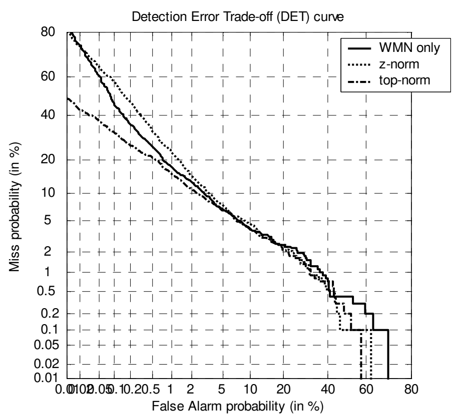

Figure 8: The DET curves of the verification system using: 1) WMN, 2) Z-norm, and 3) top-norm.

4. Adaptation of Normalization Parameters

In some speaker identification applications, such as fraudster-detection, most of the tested speakers are non-targets (non-fraudsters; not belong to the model-list), and hence produce non-target scores. We can exploit these (collected during test) non-target scores in order to better estimate the normalization parameters that are initially estimated in the training stage. This can be done using some adaptation techniques.

The idea behind the normalization-parameters adaptation is that we can estimate mean and variance (μ_N , and σ_N^2) of any score PDF using the previous values of the mean and variance (μ_{N-1} and σ_{N-1}^2) and the current *N*th score value, s_N , using the equations:

$$\mu_N = \frac{1}{N}\left[(N-1)\mu_{N-1} + s_N\right] \quad (4)$$

$$\sigma_N^2 = \frac{N-1}{N}\left[\sigma_{N-1}^2 + \frac{1}{N}\left(\mu_{N-1} - s_N\right)^2\right] \quad (5)$$

Next we propose normalization-parameters adaptation technique for Z-norm and for Top-norm methods.

4.1. Z-norm Parameter Adaptation Algorithm

For this method, for each model, we should store the parameters: *N*, μ_N , and σ_N^2 .

The basic algorithm for adapting Z-norm parameters:

1. Initialization Step (using initial N_i scores)

$$N = N_i$$

$$\mu_N = \frac{1}{N}\sum_{n=1}^{N} s_n$$

$$\sigma_N^2 = E\left[\left(s-\mu_N\right)^2\right] = E\left[s^2\right] - \mu_N^2 = \frac{1}{N}\sum_{n=1}^{N} s_n^{\ 2} - \mu_N^2$$

2. Adaptation iteration using new score (s_N)

$$N = N + 1$$

$$\sigma_N^2 = \frac{N-1}{N}\left[\sigma_{N-1}^2 + \frac{1}{N}\left(\mu_{N-1} - s_N\right)^2\right]$$

$$\mu_N = \frac{1}{N}\left[(N-1)\mu_{N-1} + s_N\right]$$

4.2. Top-norm Parameter Adaptation Algorithm

Adapting Top-norm parameters is more difficult task than adapting Z-norm parameters. In the suggested Top-norm adaptation algorithm, we need to store, for each model, two sets of distribution parameters (*N*, μ_N , and σ_N): 1) the general score distribution parameters (similar to Z-norm), 2) the top *p*% (*topPercent*) score distribution parameters ($N^{\%}$, $\mu_N^{\%}$, and $\sigma_N^{\%}$).

The basic algorithm for adapting Top-norm parameters:

1. Initialization Step (using initial N_i scores)

$$N = N_i$$

$$\mu_N = \frac{1}{N}\sum_{n=1}^{N} s_n$$

$$\sigma_N^2 = E\left[\left(s-\mu_N\right)^2\right] = E\left[s^2\right] - \mu_N^2 = \frac{1}{N}\sum_{n=1}^{N} s_n^{\ 2} - \mu_N^2$$

$$N^{\%} = p/100 \times N$$

$$\mu_N^{\%} = \frac{1}{N^{\%}}\sum_{n=1}^{N^{\%}} s_n \ ; \ s_n \in \left\{s^{\%}\right\} \text{ (top } p\% \text{ of scores)}$$

$$\sigma_N^{\%} = \sqrt{\frac{1}{N^{\%}}\sum_{n=1}^{N^{\%}} s_n^2 - \left(\mu_N^{\%}\right)^2} \ ; \ s_n \in \left\{s^{\%}\right\}$$

2. Adaptation iteration using new score (s_N)

$$N = N + 1$$

$$\sigma_N^2 = \frac{N-1}{N}\left[\sigma_{N-1}^2 + \frac{1}{N}\left(\mu_{N-1} - s_N\right)^2\right]$$

$$\mu_N = \frac{1}{N}\left[(N-1)\mu_{N-1} + s_N\right]$$

Calculate Top-norm score threshold τ_N

$$\left.\begin{array}{l} \text{if } N^{\%}/N > p/100 \ \& \ \tau_N < \tau_{N-1} \\ \quad \tau_N = \tau_{N-1} + \alpha \\ \text{if } N^{\%}/N < p/100 \ \& \ \tau_N > \tau_{N-1} \\ \quad \tau_N = \tau_{N-1} - \alpha \end{array}\right\} \text{Threshold correction} \quad (6)$$

if $s_N \geq \tau_N$

$$N^{\%} = N^{\%} + 1$$

$$\sigma_N^{\%} = \sqrt{\frac{N^{\%}-1}{N^{\%}}\left[\left(\sigma_{N-1}^{\%}\right)^2 + \frac{1}{N^{\%}}\left(\mu_{N-1}^{\%} - s_N\right)^2\right]}$$

$$\mu_N^{\%} = \frac{1}{N^{\%}}\left[\left(N^{\%}-1\right)\mu_{N-1}^{\%} + s_N\right]$$

else

$$\mu_N^{\%} = \mu_{N-1}^{\%}$$

$$\sigma_N^{\%} = \sigma_{N-1}^{\%}$$

Where the Top-norm threshold, τ_N , indicates the (*p*) top-percentage score-threshold (see figure 6). Since we do not store each one of the (test) scores, we can not accurately calculate this threshold, we can only estimate it. We used the threshold estimation equation:

$$\tau_N = \frac{\mu_N \sigma_{N-1}^{\%} + \mu_{N-1}^{\%}\sigma_N}{\sigma_N + \sigma_{N-1}^{\%}} \quad (7)$$

And in each iteration step, a threshold correction procedure was made in order to adjust this threshold to the required $N^{\%}/N$ ratio (6). The parameter α is an adaptation coefficient used for this threshold correction (e.g. $\alpha = 0.0001$).

5. Experiments and Results

Experiments for open-set speaker identification were performed using GMM-based system. 24-features were extracted from each 20msec frame (50% overlapping). The features were 12 Mel-frequency-cepstral-coefficients (MFCC) and 12 ΔMFCC. Cepstral mean subtraction (CMS) was added. The speech data used was part of NIST99 speaker evaluation (1SPK) and SwitchBoard (SB; release 2, phase 1) databases, which consist of one-sided telephone conversations. In order to make a speaker identification test, 183 speakers from NIST99 were chosen to be the target-set, these speakers (males and females) have more than six files. 183 (50-mixture) GMMs were trained; two minutes from two files. The identification test included a total of 5233 files (trials) from over 700 speakers (females and males) which included 984 target files and 4249 non-target files. For estimating Z-norm and Top-norm parameters, 1000 additional 1SPK files from SB database were used. The baseline system was normalized using gender-dependent WMN; these models (Two 256-mixture GMMs) were trained using the NIST98 database.

In many speaker identification applications, such as fraud detection, very low FA error rate is required and the EER point is less important; therefore, in the next reported performance results we will relate also to the FA error value, P_{fa}^{50}, in which the false reject error in this working point is equal to 50%. Fig. 9 shows curves of the EER and the false accept (P_{fa}^{50}) versus *topPercent* parameter (*p*) of the Top-norm method, combined in OSI system (M = 183). From this figure one can see that a good compromise between optimal EER and optimal P_{fa}^{50} is the choice of 10% for the value of *p*.

Figure 10 shows the Detection Error Trade-off (DET) curve of this open-set speaker identification system, using the mentioned databases and different target-set sizes: 1, 10, 20, 50, 100, and 183.

Other normalization methods were implemented and tested for comparison with the Top-norm method; among them: baseline (WMN only), Z-norm, T-norm [5], and unconstraint cohort normalization (UCN) [3]. Fig. 11 shows DET curves of the open-set speaker identification system (group detector; M = 183) using different normalization methods: 1) baseline (WMN), 2) Z-norm, 3) UCN (C = 2), and 4) Top-norm (p = 10%). Cohort size (C) of 2 was chosen for the UCN because it was the optimal value. From this figure one can see that Top-norm is superior to other methods at all working points. We may see also the counterclockwise "tilt" of the UCN DET curve relatively to the other DET curves. This may suggest a useful combination of the Top-norm and UCN methods. Note that Z-norm performs worse than the baseline (WMN) system. This is because the Z-norm parameters were estimated using gender-independent utterances; however, when we used gender-dependent utterances, Z-norm performances were better than the baseline system, but not better than the Top-norm system. T-norm (not shown here) was tested as well and yielded poorer results than UCN, probably because its sensitivity to mixed-gender model population.

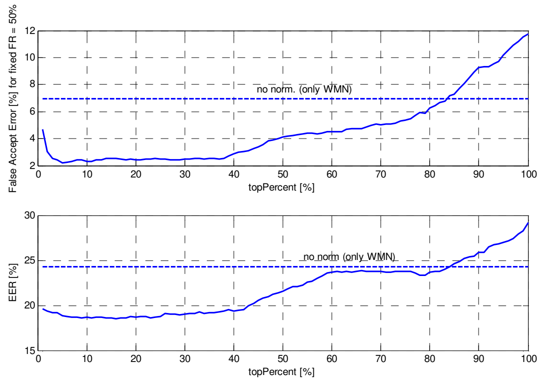

Figure 9: P_{fa}^{50} and EER versus *topPercent* parameter of the Top-norm method.

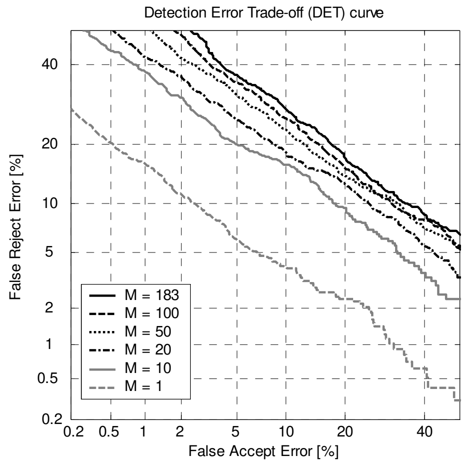

Figure 10: The DET curve of the open-set speaker identification system using top-norm method, and 1/10/20/50/100/183 models (target-set size).

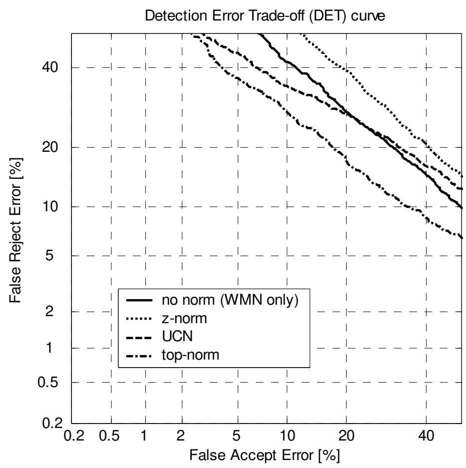

Figure 11: The DET curve of the OSI system (M = 183) using different normalization methods: 1) WMN, 2) Z-norm, 3) UCN, and 4) Top-norm.

Table 1 shows the EER and P_{fa}^{50} for the different normalization methods and for different target-set sizes. The identification test for M=183 included 984 target trials and 4249 non-target trials. For smaller M, cross-validation tests were performed; hence, target trials and non-target trials were increased. From this table we can see that UCN is useful only above 20 models (stack size) and only in the low FA working points (FR = 50%). Moreover, we can see from this table that Top-norm method is superior to the other methods in each of the tested target-set sizes.

Table 1: P_{fa}^{50} and EER of different normalization methods.

		Target-set size (M)					
Method		**1**	**10**	**20**	**50**	**100**	**183**
WMN (baseline system)	P_{fa}^{50} (%)	0.09	1.3	1.63	3.4	4.71	6.91
	EER (%)	6.09	13.5	16.1	19.8	21.9	24.2
UCN (*C*=2)	P_{fa}^{50} (%)		2.8	1.26	2.2	2.36	2.8
	EER (%)		19	17.8	21.1	22.2	24.1
Top-norm (*p*=10%)	P_{fa}^{50} (%)	***0.023***	***0.28***	***0.51***	***1.11***	***1.7***	***2.33***
	EER (%)	***5.58***	***13.4***	***14.8***	***16.2***	***17.4***	***18.7***

Fig. 12 shows DET curves of the open-set speaker identification system (M = 183) using Top-norm method (p = 10%): 1) without parameter adaptation (dashed line), 2) with parameter adaptation (solid line). The parameter adaptation was performed as in section 4.2 using $N_i = 400$ and $\alpha = 0.0001$. From this figure one can see that Top-norm parameter adaptation causes better performances in the desired working points (low FA rates).

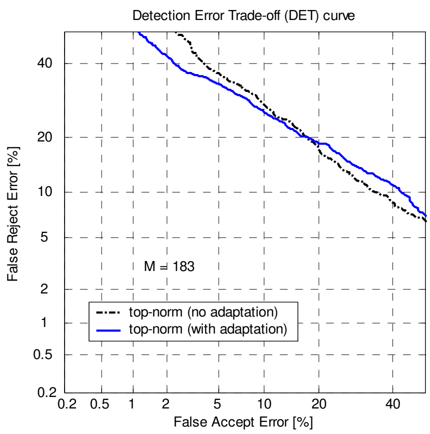

Figure 12: The DET curve of the OSI system (M = 183) using Top-norm method: 1) without parameter adaptation, 2) with parameter adaptation.

6. Conclusions

In this paper we presented and demonstrated the "false accept error problem" of identification systems, which is the phenomenon in which the false alarm rate increases dramatically when increasing the number of registered speakers (target-set size). We showed that significantly better identification results can be achieved, when decreasing the standard deviation of the non-target score distribution in the prototype verification system, even when the equal error rate (EER) is kept. We suggested a new model-dependent score-normalization method, called Top-norm. The Top-norm was developed especially for open-set speaker identification systems. It was shown that Top-norm is superior to other normalization methods. It is also faster than test-dependent normalization methods (such as UCN and T-norm) and it is not sensitive to mixed-gender population (such as Z-norm and T-norm). One drawback is the need for relatively many non-target files to estimate the Top-norm parameters, since we deal with the "right-tail" of the non-target score-distribution (the top-scores). However, this is not really a problem, since we can adapt the Top-norm parameters during the test phase. We also suggested an algorithm for Z-norm and Top-norm parameter adaptation. Although Top-norm method was developed especially for speaker identification, it can be used also in speaker verification systems. Combination of Top-norm with test-dependent score normalization method (such as T-norm or UCN) can yield better performances.

7. References

[1] E. Singer and D. Reynolds, "Analysis of Multitarget Detection for Speaker and Language Recognition," *Proc. Odyssey 2004 The Speaker and Language Recognition Workshop*, Toledo, 2004.

[2] Y. Zigel and A. Cohen, "On Cohort Selection for Speaker Verification," *Proceedings of Eurospeech 2003*, pp. 2977-2980, Geneva, 2003.

[3] F. Bimbot, J-F Bonastre, C. Fredouille, G. Gravier, I. Magrin-Chagnolleau, S. Meignier, T. Merlin, J. Ortega-Garcia, D. Petrovska-Delacretaz, and D. A. Reynolds, "A Tutorial on Text-Independent Speaker Verification," *EURASIP Journal on Applied Signal Processing*, vol. 4, pp. 430-451, 2004.

[4] J. Fortuna, P. Sivakumaran, A. M. Ariyaeeinia, and A. Malegaonkar, "Relative Effectiveness of Score Normalization Methods in Open-Set Speaker Identification," *Proc. Odyssey 2004 The Speaker and Language Recognition Workshop*, Toledo, 2004.

[5] R. Auckenthaler, M. Carey, and H Lloyd-Thomas, "Score Normalization for Text-Independent Speaker Verification Systems, *Digital Signal Processing*, vol. 10, pp. 42-54, 2000.

[6] J. Daugman, "Biometric Decision Landscapes," *Technical Report No. TR482, University of Cambridge Computer Lab*. http://www.cl.cam.ac.uk/users/jgd1000.

[7] J. Navratil, G. N. Ramaswamy, "The awe and mystery of T-Norm," *Proc. of EUROSPEECH-03*, Geneva, 2003.

[8] D. A. Reynolds, "Comparison of Background Normalization Methods for Text-Independent Speaker Verification," *in Proceedings of the Eurospeech'97*, pp. 963-966, 1997.

[9] Y. Zigel and A. Cohen "Text-Dependent Speaker Verification Using Feature Selection with Recognition Related Criterion," *Proc. Odyssey 2004 The Speaker and Language Recognition Workshop*, pp. 329-336, Toledo, 2004.

[10] J. M. Colombi, J. S. Reider, and J. P. Campbell, "Allowing Good Impostors to Test," *Conference Record of the Thirty-First Asilomar Conference on Signals, Systems & Computers*, 1997, pp. 296-300, 1998.

[11] G. N. Ramaswamy, J. Navratil, U. V. Chaudhari, R. D. Zilca, ”The IBM System for the NIST 2002 Cellular Speaker Verification Evaluation," *ICASSP-2003*, Hong Kong, April, 2003.