

A REVIEW OF VARIOUS SCORE NORMALIZATION TECHNIQUES FOR SPEAKER IDENTIFICATION SYSTEM

Piyush Lotia¹, M. R. Khan²

¹H.O.D. of E&I deptt. & ²Principal G. E. C. Raipur,
Shri Shankaracharya Technical Campus, Faculty of Engineering & Technology,
Junwani Bhilai, India

ABSTRACT

This paper presents an overview of a state-of-the-art text-independent speaker verification system using score normalization. First, an introduction proposes a modular scheme of the training and test phases of a speaker verification system. Then, the most commonly speech parameterization used in speaker verification, namely, cepstral analysis, is detailed. Normalization of scores is then explained, as this is a very important step to deal with real-world data. When acoustic and prosodic based systems are established, it is advantageous to normalize the dynamic ranges of the score dimensions, that is, likelihood scores from different quality of acoustic- and prosodic based models. Score normalization methods, linear scaling to unit range and linear scaling to unit variance, are applied to transform the output scores using the background instances so as to obtain meaningful comparison between speaker models. In this fusion system based on linear score weighting approach, the performance of speaker identification is further improved when incorporating prosodic level of information. The evaluation of a speaker verification system is then detailed, and the detection error trade-off (DET) curve is explained. Then, some applications of speaker verification are proposed, including won-site applications, remote applications, applications relative to structuring audio information, and games.

KEYWORDS: score normalization, cohort model, speaker verification, speaker adaptive normalization, DET curves.

I. INTRODUCTION

Numerous measurements and signals have been proposed and investigated for use in biometric recognition systems. Among the most popular measurements are fingerprints, face, and voice [1]. While each has pros and cons relative to accuracy and deployment, there are two main factors that have made voice a compelling biometric. First, speech is a natural signal to produce that is not considered threatening by users to provide. In many applications, speech may be the main (or only, e.g., telephone transactions) modality, so users do not consider providing a speech sample for authentication as a separate or intrusive step. Second, the telephone system provides a ubiquitous, familiar network of sensors for obtaining and delivering the speech signal. For telephone based applications, there is no need for special signal transducers or networks to be installed at application access points since a cell phone gives one access almost anywhere. Even for non-telephone applications, sound cards and microphones are low-cost and readily available. Additionally, the speaker recognition [1] area has a long and rich scientific basis with over 30 years of research, development, and evaluations. Over the last decade, speaker recognition technology has made its debut in several commercial products. The specific recognition task addressed in commercial systems is that of verification or detection (determining whether an unknown voice is from a particular enrolled speaker) rather than identification (associating an unknown voice with one from a set of enrolled

speakers). These generally employ what is known as text-dependent or text-constrained systems [1]. An example of this is background verification where a speaker is verified behind the scene as he/she conducts some other speech interactions. For cases like this, a more flexible recognition system able to operate without explicit user cooperation and independent of the spoken utterance (called text-independent mode) is needed [1]. This paper focuses on the technologies behind these text-independent speaker verification systems using score normalization. A speaker verification system is composed of two distinct phases, a training phase and a test phase [1]. Each of them can be seen as a succession of independent modules. Figure 1 shows a modular representation of the training phase of a speaker verification system. The first step consists in extracting parameters from the speech signal to obtain a representation suitable for statistical modelling as such models are extensively used in most state-of-the-art speaker verification systems. The second step consists in obtaining a statistical model from the parameters. This training scheme is also applied to the training of a background model. Figure 2 shows a modular representation of the test phase of a speaker verification system. The entries of the system are a claimed identity and the speech samples pronounced by an unknown speaker. The purpose of a speaker verification system is to verify if the speech samples correspond to the claimed identity. First, speech parameters are extracted from the speech signal using exactly the same module as for the training phase. Then, the speaker model corresponding to the claimed identity and a background model are extracted from the set of statistical models calculated during the training phase [1]. Finally, using the speech parameters extracted and the two statistical models, the last module computes some scores, normalizes them, and makes an acceptance or a rejection decision. The normalization step requires some score distributions to be estimated during the training phase or/and the test phase. Finally, a speaker verification system can be text dependent or text-independent. In the former case, there is some constraint on the type of utterance that users of the system can pronounce (for instance, a fixed password or certain words in any order, etc.). In the latter case, users can say whatever they want. This paper describes state-of-the-art text-independent speaker verification systems.

This represents the steps preceding score normalization. The last step in speaker verification is the decision making [1]. This process consists in comparing the likelihood resulting from the comparison between the claimed speaker model and the incoming speech signal with a decision threshold. If the likelihood is higher than the threshold, the claimed speaker will be accepted, else rejected. The tuning of decision thresholds is very troublesome in speaker verification. This uncertainty is mainly due to the score variability between trials, a fact well known in the domain. This score variability comes from different sources. First, the nature of the enrolment material can vary between the speakers. The differences can also come from the phonetic content, the duration, the environment noise, as well as the quality of the speaker model training. Secondly, the possible mismatch between enrolment data (used for speaker modelling) and test data is the main remaining problem in speaker recognition. Two main factors may contribute to this mismatch: the speaker him-/ her self through the intra-speaker variability (variation in speaker voice due to emotion, health state, and age) and some environment condition changes in transmission channel, recording material, or acoustical environment. On the other hand, the inter-speaker variability (variation in voices between speakers), which is a particular issue in the case of speaker-independent [1] threshold-based system, has to be also considered as a potential factor affecting the reliability of decision boundaries. Indeed, as this inter-speaker variability is not directly measurable, it is not straightforward to protect the speaker verification system (through the decision making process) against all potential impostor attacks. Lastly, as for the training material, the nature and the quality of test segments influence the value of the scores for client and impostor trials. Score normalization has been introduced explicitly to cope with score variability and to make speaker-independent decision threshold tuning easier.

Organisation of paper is as follows. After introduction the score normalisation is defined, then steps of normalisation is explained. After that speaker recognition is explained and then various method of

score normalisation is discussed. Application based methods have been categorized and finally a comparison of the various methods has been discussed.

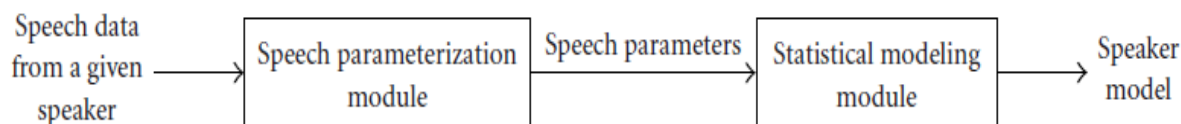


Fig1: Modulator representation of the training phase of a speaker verification system.[1]

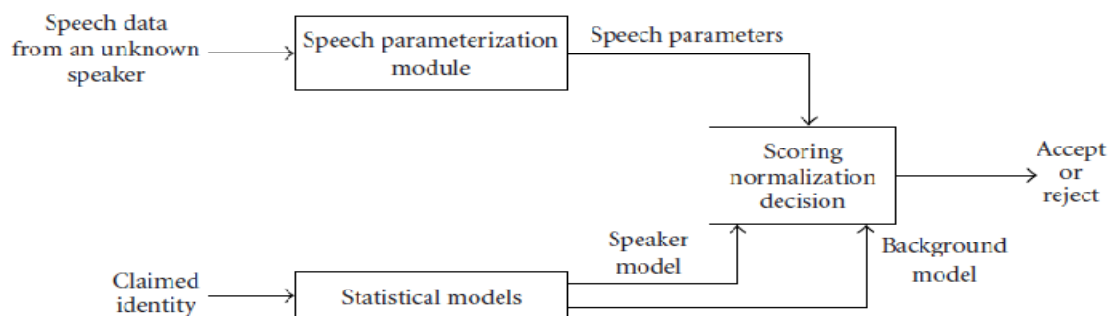


Fig2: Modulator representation of the test phase of a speaker verification system.[1]

II. WHAT IS SCORE NORMALIZATION?

Normalization at the score level is one of noise reduction methods, which normalizes log likelihood scores at the decision stage. A log-likelihood score, for short, score, is a logarithmic probability for a given input frame sequence generated based on a statistical model. Since the calculated log-likelihood scores depend on test environments, the purpose of normalization [2] aims at reducing this mismatch between a training and test set by adapting the distribution of scores to test environments, for instance, by shifting the means and changing the range of variance of the score distribution. The normalization techniques at the score level are mostly often used in speaker verification, though they can be also applied to speaker identification, because they are extremely powerful to reduce the mismatch between the claimant speaker and its impostors. Thus, in our introduction to normalization techniques at the score levels, we shall use some terminologies from speaker verification, such as claimant speaker/model, or impostor (world) speaker/model, without explicitly emphasizing these techniques being applied to speaker identification as well. The reader who is not familiar with these terminologies can refer to [3,4] for more details

III. STEPS BEFORE SCORE NORMALIZATION

Due to different quality of the speaker model training, possible mismatch and environment change among test utterances, the reliability of the likelihood scores of the reference speaker models cannot be ensured in testing. In order to normalize the score oscillation and obtain meaningful comparison, linear scaling to unit range and linear scaling to unit variance are applied using the total number of background score instances [5]. First, linear scaling rescales the output likelihood scores to the [0, 1] range when each test segment is scored against a set of speaker models. Then the likelihoods of the test segments given the target speaker is normalized according to the mean parameter and standard deviation of score distribution. Linear score weighting is employed to fuse the normalized acoustic and normalized prosodic scores. The best matching speaker is given as the identification result. . First, we

use linear scaling to unit range to normalize the range of likelihood scores. Linear scaling to unit range is described as in[6]

$$S_{ij} = \frac{S_{ij} - (S_i)_{\min}}{(S_i)_{\max} - (S_i)_{\min}} \quad (1)$$

where S_{ij} is the likelihood score of the i th speech utterance against j th speaker model and S_{ij}' is the linear-scaled value. Note that $(S_i)_{\max}$ and $(S_i)_{\min}$ are the maximal value and the minimal value for an array of likelihood scores of i th test segment against a set of target speaker models. The resulting normalized value lies in the closed interval from 0 to 1. So the acoustic-based and prosodic-based likelihood scores of a set of reference speaker models can be compared within the same dynamic range. Then, the mean and standard deviation of likelihood scores given j th speaker model are estimated to adjust the scores computed from all test segments against the speaker model. Linear scaling to unit variance is derived from the following equation:

$$S_{ij}'' = \frac{S_{ij}' - \mu_j}{\sigma_j} \quad (2)$$

Where μ_j' is the mean parameter and σ_j' is the standard deviation of the statistical distribution of linear-scaling transformed likelihood values at the first stage. Score normalization transforms each likelihood score by its value in the background distribution and performs a rescaling of the instances to obtain an approximately comparable distribution. Score normalization methods mentioned above are applied to speaker identification as follows. In testing, all of the likelihood scores of test utterances against the reference speaker models are saved as background instances. So we get a matrix of likelihood scores [2]. $[S_{ij}]$ is an I by J matrix of scores that each of the J speaker models calculated for each of I test segments. For each speech utterance, an array of likelihood scores is linear rescaled to unit range. Then, for each speaker model, the mean and standard deviation parameters are estimated to transform the likelihood value by the total number of background instances

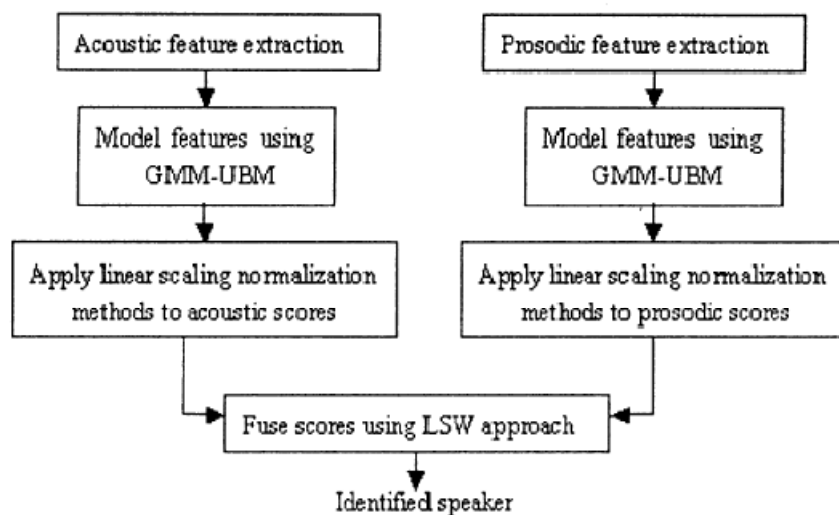


Fig 3. Block diagram of linear scaling normalisation based speaker identification fusion system [5]

IV. SPEAKER VERIFICATION VIA LIKELIHOOD RATIO DETECTION

Given a segment of speech Y and a hypothesized speaker S , the task of speaker verification, also referred to as detection, is to determine if Y was spoken by S . An implicit assumption often used is that

Y contains speech from only one speaker [1]. Thus, the task is better termed single speaker verification. If there is no prior information that Y contains speech from a single speaker, the task becomes multi-speaker detection. The single-speaker detection task can be stated as a basic hypothesis test between two hypotheses: H_0 : Y is from the hypothesized speaker S , H_1 : Y is *not* from the hypothesized speaker S . The optimum test to decide between these two hypotheses is a likelihood ratio [9]

$$\frac{p(Y|H_0)}{p(Y|H_1)} \begin{cases} > \theta, \text{accept } H_0, \\ < \theta, \text{accept } H_1, \end{cases} \quad (3)$$

(LR) test given by where $p(Y|H_0)$ is the probability density function for the hypothesis H_0 evaluated for the observed speech segment Y , also referred to as the “likelihood” of the hypothesis H_0 given the speech segment Y . The likelihood function for H_1 is likewise $p(Y|H_1)$. The decision threshold for accepting or rejecting H_0 is θ . One main goal in designing a speaker detection system is to determine techniques to compute values for the two likelihoods $p(Y|H_0)$ and $p(Y|H_1)$. Figure shows the basic components found in speaker detection systems based on LR.

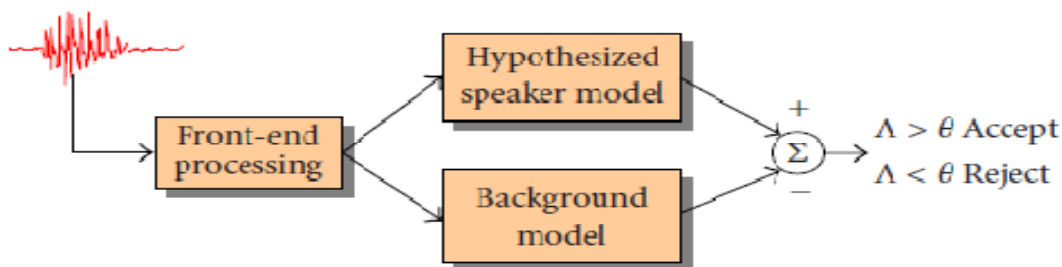


Fig.4 Likelihood ratio based speaker verification system [1]

Likelihood ratio-based-speaker verification system. Here, the role of the front-end processing is to extract from the speech signal features that convey speaker-dependent information [1]. In addition, techniques to minimize confounding effects From these features, such as linear filtering or noise, may be employed in the front-end processing. The output of this stage is typically a sequence of feature vectors representing the test segment $X = \{x_1, \dots, x_T\}$, where x_t is a feature vector indexed at discrete time $t \in [1, 2, \dots, T]$. There is no inherent constraint that features extracted at synchronous time instants be used; as an example, the overall speaking rate of an utterance could be used as a feature. These feature vectors are then used to compute the likelihoods of H_0 and H_1 . Mathematically, a model denoted by λ_{hyp} represents H_0 , which characterizes the hypothesized speaker S in the feature space of x . For example, one could assume that a Gaussian distribution best represents the distribution of feature vectors for H_0 so that λ_{hyp} would contain the mean vector and covariance matrix parameters of the Gaussian distribution. The model λ_{hyp} represents the alternative hypothesis, H_1 . The likelihood ratio statistic is then $p(X|\lambda_{hyp})/p(X|\lambda_{hyp})$. Often, the logarithm of this statistic is used

$$\Lambda(X) = \log p(x|\lambda_{hyp}) - \log p(X|\frac{\lambda}{hyp}). \quad (4)$$

giving the log LR While the model for H_0 is well defined and can be estimated using training speech from S , the model for λ_{hyp} is less well defined since it potentially must represent the entire space of possible alternatives to the hypothesized speaker. Two main approaches have been taken for this alternative hypothesis modelling. The first approach is to use a set of other speaker models to cover the space of the alternative hypothesis. In various contexts, this set of other speakers has been called

likelihood ratio sets, cohorts, and background speakers. Given a set of N background speaker models $\{\lambda_1, \dots, \lambda_N\}$, the alternative hypothesis model is represented by

$$P(X|\frac{\lambda}{hyp}) = f(p(X|\lambda_1), \dots, p(X|\lambda_N)), \quad (5)$$

Where $f(\cdot)$ is some function, such as average or maximum, of the likelihood values from the background speaker set. The selection, size, and combination of the background speakers have been the subject of much research [10,11,12,13]. In general, it has been found that to obtain the best performance with this approach requires the use of speaker-specific background speaker sets. This can be a drawback in applications using a large number of hypothesized speakers, each requiring their own background speaker set. The second major approach to the alternative hypothesis modeling is to pool speech from several speakers and train a single model. Various terms for this single model are a general model [14], a world model, and a universal background model (UBM) [15]. Given a collection of speech samples from a large number of speakers representative of the population of speakers expected during verification, a single model λ_{bkg} , is trained to represent the alternative hypothesis. Research on this approach has focused on selection and composition of the speakers and speech used to train the single model [16, 17]. The main advantage of this approach is that a single speaker-independent model can be trained once for a particular task and then used for all hypothesized speakers in that task. It is also possible to use multiple background models tailored to specific sets of speakers [17,18]. The use of a single background model has become the predominate approach used in speaker verification systems.

V. SCORE NORMALIZATION TECHNIQUES

Score normalization techniques have been mainly derived from the study of Li and Porter [8]. In score normalization, the “raw” match score is normalized relative to a set of other speaker models known as cohort. The main purpose of score normalization is to transform scores from different speakers into a similar range so that a common (speaker-independent) verification threshold can be used. Score normalization can correct some speaker-dependent score offsets not compensated by the feature and model domain methods. A score normalization of the form[7]-

$$S' = \frac{s - \mu_I}{\sigma_I} \quad (6)$$

is commonly used. Here s' is the normalized score, s is the original score, and μ_I and σ_I are the estimated mean and standard deviation of the impostor score distribution, respectively.

5.1 Z norm

The zero normalization (Z norm) technique is directly derived from the work done in [19]. The zero normalization (Z norm) technique [1] has been massively used in speaker verification in the middle of the nineties. In practice, a speaker model is tested against a set of speech signals produced by some impostor, resulting in an impostor similarity score distribution. Speaker-dependent mean and variance normalization parameters—are estimated from this distribution and applied [17] on similarity scores yielded by the speaker verification system when running. One of the advantages of Z-norm is that the estimation of the normalization parameters can be performed offline during speaker model training [1]. This is done by matching a batch of non-target utterances against the target model, and obtaining the mean and standard deviation of those scores. Concretely speaking, let $L(\mathbf{x}_i | S)$ be a log-likelihood score [2] for a given speaker model S and a given feature frame \mathbf{x}_i , where an overall utterance is denoted by $\mathbf{X} = \{\mathbf{x}_i\}$, $i \in [1, N]$. Here $L(\mathbf{x}_i | S)$ is also called raw score. We shall then refer to *L-norm*

$(\mathbf{x}_i | S)$ as the normalized log-likelihood score. Based on the notations, we have the following equation, [8]

$$L(\mathbf{x}|S) = \frac{1}{N} \sum_{i=1}^N L(\mathbf{x}_i|S), \quad (7)$$

and the normalized score,

$$L_{\text{norm}}(\mathbf{x}|S) = \frac{L(\mathbf{x}|S) - \mu_I}{\sigma_I}, \quad (8)$$

where μ_I and σ_I are the mean and standard deviation of the distribution of the impostor scores, which are calculated based on the impostor model SI .

5.2 H norm

H-norm or handset dependent score normalization technique is given in [20]. In examining the scores produced by the different recognition systems, it became clear that speaker models were producing different distributions of scores for the same test utterances, most significantly for the mismatched telephone number tests. Since a pooled (speaker-independent) threshold is being used, this caused significantly higher false alarm rates for a given miss rate. Based on earlier work [21, 22], we believed that handset differences associated with different telephone numbers was the root cause of the observed differences. Since handset information is not available, we created a handset detector to label the test utterances as being either from a carbon-button type handset (CARB) or an electrets type handset (ELEC). The handset detector is a simple maximum likelihood classifier in which handset dependent GMMs were trained using the HTIMIT corpus1 [23]. Using these labels, we did indeed observe that different claimant models responded differently to different handset types. This occurs because the claimant model not only represents the speaker but also the handset characteristics over which the training data was collected. Thus a claimant model trained on speech from a CARB handset would tend to score better to other utterances also collected over a CARB handset. There is a similar affinity for claimant models trained with ELEC speech to score well on ELEC test data. These observations and the utility of the handset labeler are supported by work reported in [24]. To normalize out these effects, we developed a handset score normalization technique called h-norm. In h-norm, we first determine the response of a claimant's model to speech with CARB and ELEC labels. The response to CARB speech is parameterized as the mean and variance of the likelihood ratios produced by the claimant model for development utterances labeled as CARB. Likewise for ELEC. Note that the speech used to determine the claimant's response is not from the claimant, but from non-claimant development speakers. Each claimant s then has two sets of parameters describing his/her model's response to CARB and ELEC type speech:

$$\{\mu_8(CARB), \sigma_8(CARB), \mu_8(ELEC), \sigma_8(ELEC)\} \quad (9)$$

During testing, an input utterance is first labeled as CARB or ELEC and unit standard deviation scores for non-claimant speech, independent of the handset characteristics of the test utterance or of those used in training the claimant model. In addition to helping normalize out handset-dependent biases for a particular claimant model, this normalization also makes a speaker independent threshold more effective for all claimant speakers. The h-norm [20] procedure was applied to the evaluation corpus. A comparison of the baseline UBM using claimant model adaptation with and without applying h-norm is shown in Figure 5. It is evident that h-norm produces a significant reduction in errors for the mismatched condition. At 10% miss rate, the false alarm rate decreases from 14.5% to 2.4% – an 83% reduction in error.

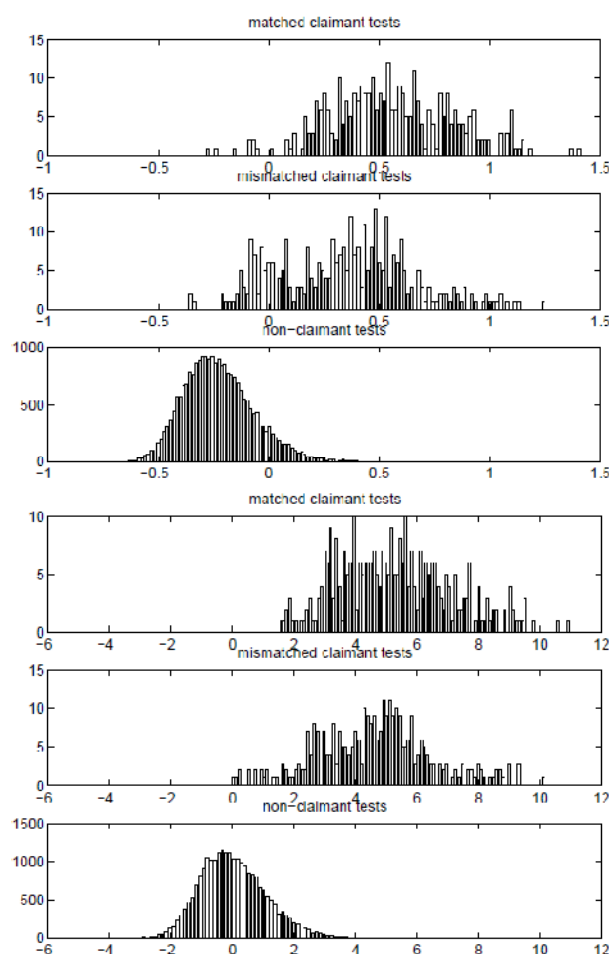


Fig.5 Distribution of log-likelihood ratio scores for matched claimant tests, mismatched claimant tests and non claimant tests. All scores are from the UBM with claimant adaptation. The upper three plots are baseline scores. The bottom three plots are for scores after h norm has been applied. [20]

5.3 T-norm

Still based on the estimate of mean and variance parameters to normalize impostor score distribution, test-normalization (T-norm) proposed in[25] differs from Z-norm by the use of impostor models instead of test speech signals. During testing, the incoming speech signal is classically compared with claimed speaker model as well as with a set of impostor models to estimate impostor score distribution and normalization parameters consecutively. If Z-norm is considered as a speaker-dependent normalization technique, T-norm is a test-dependent one. As the same test utterance is used during both testing and normalization parameter estimate, T-norm avoids a possible issue of Z-norm based on a possible mismatch between test and normalization utterances. Conversely, T-norm has to be performed online during testing. Normalized score is obtained by

$$L_{norm}(X|S) = \frac{L(x|S) - \mu_{I-test}}{\sigma_{I-test}} r \quad (10)$$

where μ_{I-test} and σ_{I-test} are the mean and standard deviation of the distribution of the impostor scores estimated on a test set. In contrast, for Z-norm, the corresponding μ_I and σ_I are estimated on the

training set. In T-norm, during the test stage the test utterance is scored against a pre-selected set of cohort models (pre-selection is based on the claimant model). The resulting score distribution is then used to estimate the normalization parameters in. The advantage of T-norm over Z-norm [26] is that any acoustic or session mismatch between test and impostor utterances is reduced. However, the disadvantage of T-norm is the additional test stage computation in scoring the cohort models. As shown in Fig.(a) for the NIST-2002 corpus, we observe considerable overlap among both the impostor and claimant score distributions thus resulting in verification errors and higher EER[27]. Using score normalization methods, the impostor score distribution can normalize to zero mean and unit variance. As shown in Fig.(b), we observe that the T-norm reduces the overlap among the distributions resulting in fewer verification errors and lower EER.

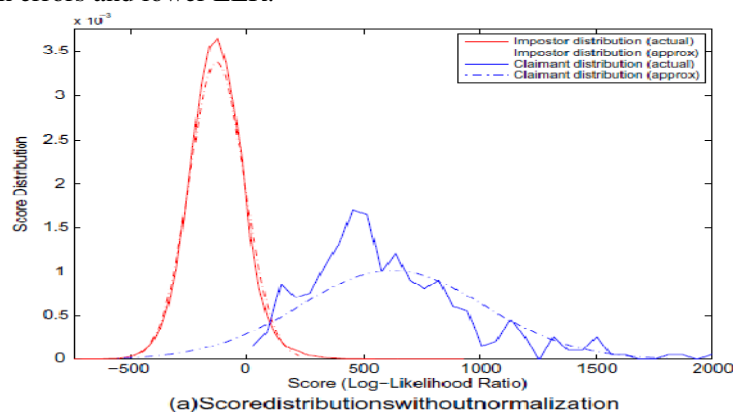


Fig.6 Score distribution without normalization [26]

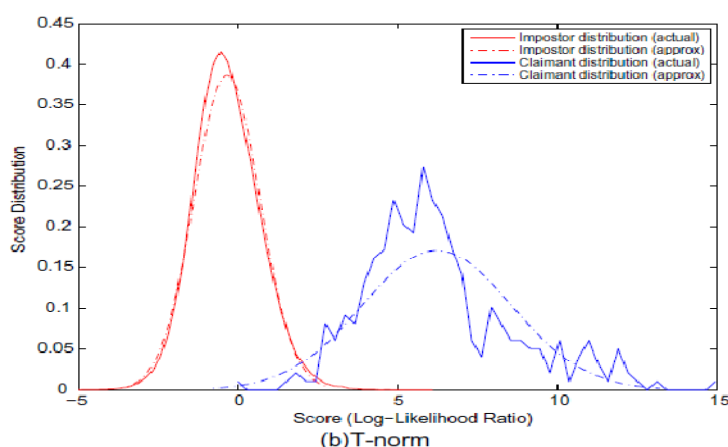


Fig.7 Score distribution without normalization [26]

5.4 C-norm

C-norm is referred to as cellular normalization which was proposed by [28] for compensation of channel effects of cellular phones. However, C-norm [2] is also called a method of feature mapping, because C-norm is based on a mapping function from a channel dependent feature space into a channel independent feature space. The final recognition procedure is done on the mapped, channel independent feature space. Following the symbols, which we used above, \mathbf{x}_t is denoted as a frame at time t in a channel dependent (CD) feature space, and a frame at time t in a channel independent (CI) feature space [2]. The GMM modeling for the channel dependent feature space is denoted G_{CD} as and

the GMM for the channel independent feature space is denoted as *GCI*. The Gaussian mixture to which a frame \mathbf{x}_t belongs is chosen according to the maximum likelihood criterion, i.e.

$$i = \underset{j}{\operatorname{argmax}} \{ \omega_j^{CD} \cdot p_j^{CD}(\mathbf{x}_t | \mu_j^{CD}, \sigma_j^{CD}) \} \quad (11)$$

Normalization and Transformation Techniques for Robust Speaker Recognition 319 where a Gaussian mixture is defined by its weight, mean and standard deviation $\{\omega_j^{CD}, \mu_j^{CD}, \sigma_j^{CD}\}$. Thus, by a transformation $f(\bullet)$, a CI frame feature \mathbf{y}_t is mapped from \mathbf{x}_t according to

$$\mathbf{y}_t = f(\mathbf{x}_t) = (\mathbf{x}_t - \mu_i^{CD}) \frac{\sigma_i^{CI}}{\sigma_i^{CD}} + \mu_i^{CI} \quad (12)$$

Where i is a Gaussian mixture to which \mathbf{x}_t belongs. After the transformation, the final recognition is conducted on the CI feature space, which is expected with the advantages of channel compensation.

5.5 D-norm

D-norm was proposed by Ben et al. in 2002. D-norm [3] deals with the problem of pseudo-impostor data availability by generating the data using the world model. A Monte Carlo-based method is applied to obtain a set of client and impostor data, using, respectively, client and world models. The normalized score is given by:-

$$\widetilde{L}_\lambda(X) = \frac{L_\lambda(X)}{KL2(\lambda, \bar{\lambda})}, \quad (13)$$

Where $KL2(\lambda, \bar{\lambda})$ is the estimate of the symmetrised Kullback-Leibler distance between the client and world models. The estimation of the distance is done using Monte-Carlo generated data. As for the previous normalizations, D-norm[3] is applied on likelihood ratio, computed using a world model. D-norm presents the advantage not to need any normalization data in addition to the world model. As D-norm is a recent proposition, future developments will show if the method could be applied in different applications like password-based systems.

5.6 WMAP

WMAP is designed for multi recognizer systems. The technique focuses on the meaning of the score and not only on normalization. WMAP [1], proposed by Fredouille et al. in 1999,[29] is based on the Bayesian decision framework. The originality is to consider the two classical speaker recognition hypotheses in the score space and not in the acoustic space. The final score is the a posteriori probability to obtain the score given the target hypothesis:

$$WMAP(L_\lambda(X)) = \frac{P_{Target} \cdot p(L_\lambda(X)|Target)}{P_{Target} \cdot p(L_\lambda(X)|Target) + P_{Imp} \cdot p(L_\lambda(X)|Imp)}, \quad (14)$$

Where P_{Target} (resp., P_{Imp}) is the a priori probability of a target test (resp., an impostor test) and $p(L_\lambda(X)|Target)$ (resp., $p(L_\lambda(X)|Imp)$) is the probability of score $L_\lambda(X)$ given the hypothesis of a target test (resp., an impostor test). The main advantage of the WMAP [1] normalization is to produce meaningful normalized score in the probability space. The scores take the quality of the recognizer directly into account, helping the system design in the case of multiple recognizer decision fusion. The implementation proposed by Fredouille in 1999[29] used an empirically approach and nonparametric models for estimating the target and impostor score probabilities.

VI. APPLICATION BASED NORMALIZATION APPROACHES

In this section, we present two normalization techniques which address the problem of constructing robust speaker scores when enrolment data for each speaker is unevenly distributed over the library of context-dependent phonetic events. The choice of normalization technique becomes especially important when the system is forced to synthesize an appropriate speaker score for a context-dependent phonetic event that has few or no training tokens in the enrolment data.

6.1 Speaker Adaptive (SA) Normalization

We originally described a speaker adaptive normalization approach in[31]. This technique relies on interpolating speaker dependent (SD) probabilities with speaker independent (SI) probabilities on a

per-unit basis. This approach learns the characteristics of a phone for a given speaker when sufficient enrolment data is available, but relies more on general speaker independent models in instances of sparse enrolment data. Mathematically, the speaker score can be written as:

$$Y(x, s) = \frac{1}{|X|} \sum_{x \in X} \log [\lambda_{s, \phi(x)} \frac{p(x|S\phi(X))}{p(x|\phi(X))} + (1 - \lambda_{s, \phi(x)}) \frac{p(X|\phi(X))}{p(X)}] \quad (15)$$

Here $\lambda_{s, \phi(x)}$ is the interpolation factor given by:

$$\lambda_{s, \phi(x)} = \frac{n_{s, \phi(x)}}{n_{s, \phi(x)} + \tau} \quad (16)$$

In this equation, $n_{s, \phi(x)}$ refers to the number of times the CD phonetic event $\hat{\phi}(x)$ was observed in the enrolment data for speaker S , and τ is an empirically determined tuning parameter that was the same across all speakers and phones. By using the SI models in the denominator of the terms in Equation second, the SI model set acts as the normalizing [4] *background model* typically used in speaker verification approaches. The interpolation between SD and SI models allows our technique to capture detailed phonetic-level characteristics when a sufficient number of training tokens are available from a speaker, while falling back onto the SI model when the number of training tokens is sparse. In other words, the system backs off towards a neutral score of zero when a particular CD phonetic model has little or no enrolment data from a speaker. If an enrolled speaker contributes more enrolment data, the variance of the normalized scores increases and the scores become more reflective of how well (or poorly) a test utterance matches the characteristics of that speaker's model.

6.2 Phone Adaptive (PA) Normalization

An alternative and equally valid technique for constructing speaker scores is to combine phone dependent and phone independent speaker model probabilities. In this scenario, the speaker-dependent phone-dependent [4] models can be interpolated with a speaker-dependent phone-independent model (i.e., a global GMM) for that speaker. Analytically, the speaker score can be described as:

$$Y(X, S) = \frac{1}{|X|} \sum_{x \in X} \log [\lambda_{s, \phi(x)} \frac{p(x|S\phi(X))}{p(x|\phi(X))} + (1 - \lambda_{s, \phi(x)}) \frac{p(X|S)}{p(X)}] \quad (17)$$

Here, $\lambda_{s, \phi(x)}$ has the same interpretation as before. The rationale behind this approach is to bias the speaker score towards the global speaker model when little phone-specific enrolment data is available. In the limiting case, this approach falls back to scoring with a global GMM model when the system encounters phonetic units that have not been observed in the speaker's enrolment data. This is intuitively more satisfying than the speaker adaptive approach, which backs off directly to the neutral score of zero when a phonetic event is unseen in the enrolment data.

6.3 User-specific score normalization and fusion for biometric person recognition

Every person is unique. This uniqueness is not only prevalent in his/her biometric traits, but also in the way he/she interacts with a biometric device. A recent trend in tailoring a biometric system to each user (client) is by normalizing the match score for each claimed identity[32]. This technique is called user- (or client-) specific score normalization. This concept can naturally be extended to the multimodal biometrics here several biometric devices and/or traits are involved. This application gives a survey on user-specific score normalization as well as compares several representative. It also shows how this technique can be used for designing an effective user-specific fusion classifier. The advantage of this approach, compared to the direct design of such a fusion classifier, is that much less genuine data is needed. Several potential research directions are also outlined.

6.3.1 Doddington's Menagerie

An automatic biometric authentication system operates by first building a reference model or template for each user (or enrollee). A template is a single enrollment data whereas a reference model, in a more general context, is a statistical model obtained from one or more enrollment samples. During the operational phase, the system compares a scanned biometric sample with the reference model of a

claimed identity in order to render a decision. Typically, the underlying probability distributions of genuine and impostor scores exhibit strong user model dependency. They also reflect the stochastic nature of the biometric matching process. Essentially, these user-dependent components of the distribution determine how easy or difficult it is to recognize an individual and how successfully he or she can be impersonated. The practical implication of this is that some reference models (and consequently the users they represent) are systematically better (or worse) in authentication performance than others. The essence of these different situations has been popularized by the so called Doddington's zoo, with individual users characterized by animal names such as [33]:

- Sheep: persons who can be easily recognized;
- Goats: persons who are particularly difficult to be recognized;
- Lambs: persons who are easy to imitate;
- Wolves: persons who are particularly successful at imitating others.

Goats contribute significantly to the False Reject Rate (FRR) of a system while wolves and lambs increase its False Acceptance Rate (FAR). A more recent work further [34] distinguishes four other semantic categories of users by considering both the genuine and impostor match scores for the same claimed identity simultaneously.

6.3.1.1 User-specific Class Conditional Score Distributions

To motivate the problem, it is instructive to show how the different animals are characterized by their match scores. In Figure below for the purpose of visualization, we fitted a Gaussian distribution to the match scores originated from a reference model, subjecting to genuine or impostor comparisons. The choice of Gaussian distribution is dictated by the small sample size of the data, especially the genuine match scores. In order to avoid cluttering the figure, we show only the distributions associated with 20 randomly selected enrolled identities (enrollees) out of 200. These scores are taken from the XM2VTS [35] benchmark database. Since there is one pair of distribution per enrollee (subjecting to being a genuine and an impostor comparison), there are a total of 40 distributions. The match scores used here (as well as throughout the discussion in this chapter) are likelihood ratio scores in the logarithmic domain. A high score implies a genuine user whereas a low score implies an impostor. Similarity scores can be interpreted in the same way. However, for dissimilarity scores, where a high (resp. low) value implies an impostor (resp. a genuine user), the interpretation is exactly the opposite. In this case, if y is a dissimilar score, one can use $-y$ in order to interpret it as a similarity score. Similarity or likelihood ratio match scores are thus assumed throughout this chapter. Referring to our discussion above, sheep (resp. goats) are characterized by high (low) genuine match scores. Hence, the genuine distributions with high mean values are likely to be sheep. On the other hand, the genuine distributions with low mean values are likely to be goats. Lambs are characterized by high impostor match scores. This implies that they have high impostor mean values. These characteristics are used to identify the animals. Wolves are not shown in Figure below. These are persons who "look" similar to all other enrollees in classification sense, i.e., similar in the feature representation. The presence of a large number of wolves will shift the impostor score distribution to the right, closer to the genuine score distributions. This will increase the amount of overlap between the two classes. Consequently, the classification error is increased. It should be noted that the so-called impostors here refer to zero-effort impostors, i.e., these persons do not have any knowledge about the claimed identity, e.g., possessing his/her biometric traits. While this is a common practice to assess biometric performance, in an authentication/verification application, a deliberate impostor attempt would be more appropriate. Examples of deliberate impostor attempts are gummy fingers [36], synthesized voice forgery via transformation [37], and animated talking faces [38]. This subject is an on-going research topic. For the rest of the discussion, we shall focus on zero-effort impostor attempts.

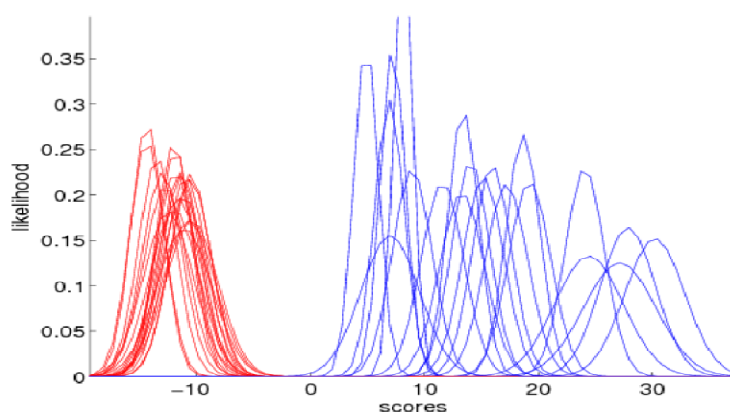


Fig.8 [32] User-specific class-conditional score distributions of a typical speech verification system. Shown here are the distributions of 20 enrollees. The right clusters (in blue) are for the genuine class whereas the left ones (in red) are for the impostor class.

VII. COMPARISONS BETWEEN SCORE NORMALIZATION TECHNIQUES

Table No. 1: EVALUATION

z-norm	t-norm	h-norm	d-norm	WMAP	c-norm
It uses test speech signal. It is a speaker dependent normalization. It was massively used in nineties.	It uses impostor models. It is a test dependent one. <u>Advantage:</u> Acoustic or session mismatch between test & impostor utterances is reduced. <u>Disadvantage:</u> 1]Additional test stage computation in scoring the cohort models. 2]Tnorm reduces the overlapping among the distributions resulting in fewer verification errors and lower EER.	<u>Advantage</u> Handset normalization improve performance during normalization parameter computation .Hnorm combined with Tnorm is better than other normalization by 2001&2002 NIST evaluation campaigns. <u>Disadvantage:</u> It is expensive in computational time.	It's a promising alternative to ht-norm since the computational time is reduced & no impostor population is required. D-norm is a advanced znorm technique.	Same level as znorm but without any knowledge about real target speaker normalization parameters are learned a priori using a separate set of speakers/tests. <u>Disadvantage</u> Difficult to apply in a target speaker mode since few speaker data are not sufficient to learn the normalization methods.	Cellular normalization is used to compensate channel effects of cellular phones.It uses feature mapping.

VIII. TYPES OF ERRORS FOR SCORE BASED DECISION OF TARGET SPEAKER

Two types of errors can occur in a speaker verification system, namely, false rejection and false acceptance. A false rejection (or non detection) error happens when a valid identity claim is rejected. A

false acceptance (or false alarm) error consists in accepting an identity claim from an impostor. Both types of error depend on the threshold θ used in the decision making process. With a low threshold, the system tends to accept every identity claim thus making few false rejections and lots of false acceptances. On the contrary, if the threshold is set to some high value, the system will reject every claim and make very few false acceptances but a lot of false rejections. The couple (false alarm error rate, false rejection error rate) is defined as the *operating point* of the system. Defining the operating point of a system, or, equivalently, setting the decision threshold, is a trade-off between the two types of errors. In practice, the false alarm and non detection error rates, denoted by P_{fa} and P_{fr} , respectively, are measured experimentally on a test corpus by counting the number of errors of each type. This means that large test sets are required to be able to measure accurately the error rates. For clear methodological reasons, it is crucial that none of the test speakers, whether true speakers or impostors [1], be in the training and development sets. This excludes, in particular, using the same speakers for the background model and for the tests. However, it may be possible to use speakers referenced in the test database as impostors. This should be avoided whenever discriminative training techniques are used or if across speaker normalization is done since, in this case, using referenced speakers as impostors would introduce a bias in the results.

8.1 DET curves and evaluation functions

As mentioned previously, the two error rates are functions of the decision threshold. It is therefore possible to represent the performance of a system by plotting P_{fa} as a function of P_{fr} . This curve, known as the system operating characteristic, is monotonous and decreasing. Furthermore, it has become a standard to plot the error curve on a normal deviate scale in [1] which case the curve is known as the detection error trade-offs (DETs) curve. With the normal deviate scale, a speaker recognition system whose true speaker and impostor scores are Gaussians with the same variance will result in a linear curve with a slope equal to -1 . The better the system is, the closer to the origin the curve will be. In practice, the score distributions are not exactly Gaussians but are quite close to it. The DET curve representation is therefore more easily readable and allows for a comparison of the system's performances on a large range of operating conditions.

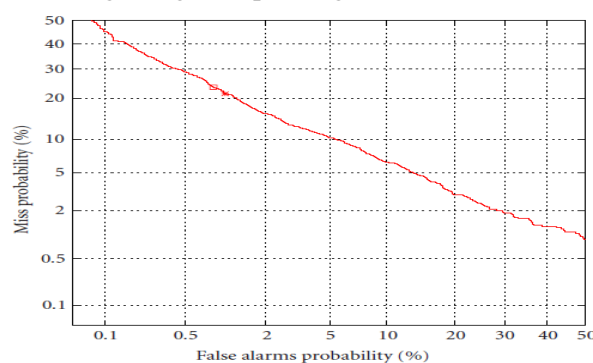


FIG 9: Example of a DET curve [1]

Figure 9 shows a typical example of a DET curves. Plotting the error rates as a function of the threshold is a good way to compare the potential of different methods in laboratory applications. However, this is not suited for the evaluation of operating systems for which the threshold has been set to operate at a given point. In such a case, systems are evaluated according to a cost function which takes into account the two error rates weighted by their respective costs, that is $C = C_{fa} P_{fa} + C_{fr} P_{fr}$. In this equation, C_{fa} and C_{fr} are the costs given to false acceptances and false rejections, respectively. The cost function is minimal if the threshold is correctly set to the desired operating point. Moreover, it is possible to directly compare the costs of two operating systems. If normalized by the sum of the error costs, the cost C can be interpreted as the mean of the error rates, weighted by the cost of each

error. Other measures are sometimes used to summarize the performance of a system in a single figure. A popular one is the equal error rate (EER) which corresponds to the operating point where $P_{fa} = P_{fr}$. Graphically, it corresponds to the intersection of the DET curve with the first bisector curve. The EER performance measure rarely corresponds to a realistic operating point. However, it is a quite popular measure of the ability of a system to separate impostors from true speakers. Another popular measure is the half total error rate (HTER) which is the average of the two error rates P_{fa} and P_{fr} [1]. It can also be seen as the normalized cost function assuming equal costs for both errors. Finally, we make the distinction between a cost obtained with a system whose operating point has been set up on development data and a cost obtained with a posterior minimization of the cost function. The latter is always to the advantage of the system but does not correspond to a realistic evaluation since it makes use of the test data. However, the difference between those two costs can be used to evaluate the quality of the decision making module (in particular, it evaluates how well the decision threshold has been set).

IX. APPLICATIONS OF SPEAKER VERIFICATION

There are many applications to speaker verification. Currently, most applications are in the banking since the speaker recognition technology is currently not absolutely reliable, such technology is often used in applications where it is interesting to diminish frauds but for which a certain level of fraud is acceptable. The main advantages of voice-based authentication are its low implementation cost and its acceptability by the end users, especially when associated with other vocal technologies.

9.1 On-site applications

On-site applications [1] regroup all the applications where the user needs to be in front of the system to be authenticated. Typical examples are access control to some facilities (car, home, warehouse), to some objects (locksmith), or to a computer terminal. Currently, ID verification in such context is done by means of a key, a badge or a password, or personal identification number (PIN). For such applications, the environmental conditions in which the system is used can be easily controlled and the sound recording system can be calibrated. The authentication can be done either locally or remotely but, in the last case, the transmission conditions can be controlled. The voice characteristics are supplied by the user (e.g., stored on a chip card). This type of application can be quite dissuasive since it is always possible to trigger another authentication mean in case of doubt.

9.2 Remote applications

Remote applications regroup all the applications where the access to the system is made through a remote terminal, typically a telephone or a computer. The aim is to secure the access to reserved services (telecom network, databases, web sites, etc.) or to authenticate the user making a particular transaction (e-trade, banking transaction, etc.). In this context, authentication currently relies on the use of a PIN, sometimes accompanied by the identification of the remote terminal (e.g., caller's phone number). For such applications, the signal quality is extremely variable due to the different types of terminals and transmission channels, and can sometimes be very poor. The vocal characteristics are usually stored on a server. Some commercial applications in the banking and telecommunication areas are now relying on speaker recognition technology to increase the level of security in a way transparent to the user.

9.3 Games

Finally, another application [1] area, rarely explored so far, is games: child toys, video games, and so forth. Indeed, games evolve toward a better interactivity and the use of player profiles to make the game more personal. With the evolution of computing power, the use of the vocal modality in games is probably only a matter of time. Among the vocal technologies available, speaker recognition certainly

has a part to play, for example, to recognize the owner of a toy, to identify the various speakers, or even to detect the characteristics or the variations of a voice (e.g., imitation contest). One interesting point with such applications is that the level of performance can be a secondary issue since an error has no real impact. However, the use of speaker recognition technology in games is still a prospective area.

X. CONCLUSIONS

The need of score normalization in speaker verification has been focussed including steps before normalization. Various score normalization techniques such as z-norm, t-norm, d-norm, h-norm, c-norm, WMAP have been presented with comparisons among them. The application based score normalization such as speaker adaptive, phone adaptive, user specific score normalization & fusion for biometric person recognition has been presented to have improved decision in terms of far & fir The DET(detection error tradeoffs) curve plot has been discussed lastly the various application has been discussed .

REFERENCES

- [1] A Tutorial on Text-Independent Speaker Verification". Received 2 December 2002; Revised 8 August 2003
- [2] "Normalization and Transformation Techniques for Robust Speaker Recognition" Dalei Wu, Baojie Li and Hui Jiang Department of Computer Science and Engineering, York University, Toronto, Ont., Canada
- [3] Wu, D. (2008a). Discriminative Preprocessing of Speech, VDM Verlag Press, ISBN: 978-3- 8364-3658-8.
- [4]Wu, D.; Li, J. & Wu, H. (2008b). Improving text-independent speaker recognition with locally nonlinear transformation. Technical report, Computer Science and Engineering Department, York University, Canada.
- [5] Relative Effectiveness of Score Normalization Methods in Speaker Identification Fusing Acoustic and Prosodic Information Rong ZHENG, Shuwu ZHANG, Bo XU Institute of Automation, Chinese Academy of Sciences, Beijing, 100080 Email: {rzheng, swzhang, xubo}{hitic.ia.ac.cn
- [6] Aksoy, S., Haralick, R.M.: Feature normalization and likelihood-based similarity measures for image retrieval. Pattern Recognition Letters, Vol.22 pp. 563-582, 2001.
- [7] An Overview of Text-Independent Speaker Recognition: from Features to Supervectors Tomi Kinnunen, a, Haizhou Lib aDepartment of Computer Science and Statistics, Speech and Image Processing Unit University of Joensuu, P.O.Box 111, 80101 Joensuu, FINLAND WWW homepage: <http://cs.joensuu.fi/sipu/> bDepartment of Human Language Technology, Institute for Infocomm Research (I2R) 1 Fusionopolis Way, #21-01 Connexis, South Tower, Singapore 138632 WWW homepage: <http://hlt.i2r.a-star.edu.sg/>
- [8] Li, K.-P., and Porter, J. Normalizations and selection of speech segments for speaker recognition scoring. In Proc. Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP 1988) (New York, USA, April 1988), pp. 595–598.
- [9] R. B. Dunn, D. A. Reynolds, and T. F. Quatieri, "Approaches to speaker detection and tracking in conversational speech," Digital Signal Processing, vol. 10, no. 1–3, pp. 93–112, 2000.
- [10] A.Higgins, L. Bahler, and J. Porter, "Speaker verification using randomized phrase prompting," Digital Signal Processing, vol. 1, no. 2, pp. 89–106, 1991.
- [11] A. E. Rosenberg, J. DeLong, C.-H. Lee, B.-H. Juang, and F. K. Soong, "The use of cohort normalized scores for speaker verification," in Proc. International Conf. on Spoken Language Processing (ICSLP '92), vol. 1, pp. 599–602, Banff, Canada, October 1992.
- [12] D. A. Reynolds, "Speaker identification and verification using Gaussian mixture speaker models," Speech Communication, vol. 17, no. 1-2, pp. 91–108, 1995.
- [13] T. Matsui and S. Furui, "Similarity normalization methods for speaker verification based on a posteriori probability," inProc. 1st ESCA Workshop on Automatic Speaker Recognition, Identification and Verification, pp. 59–62, Martigny, Switzerland, April 1994.
- [14] M. Carey, E. Parris, and J. Bridle, "A speaker verification system using alpha-nets," in Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '91), vol. 1, pp. 397–400, Toronto, Canada, May 1991.
- [15] D. A. Reynolds, "Comparison of background normalization methods for text-independent speaker verification," in Proc. 5th European Conference on Speech Communication and Technology (Eurospeech '97), vol. 2, pp. 963–966, Rhodes, Greece, September 1997.

- [16] T. Matsui and S. Furui, "Likelihood normalization for speaker verification using a phoneme- and speaker-independent model," *Speech Communication*, vol. 17, no. 1-2, pp. 109–116, 1995.
- [17] A. E. Rosenberg and S. Parthasarathy, "Speaker background models for connected digit password speaker verification," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '96)*, vol. 1, pp. 81–84, Atlanta, Ga, USA, May 1996.
- [18] L. P. Heck and M. Weintraub, "Handset-dependent background models for robust text-independent speaker recognition," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '97)*, vol. 2, pp. 1071–1074, Munich, Germany, April 1997.
- [19] K. P. Li and J. E. Porter, "Normalizations and selection of speech segments for speaker recognition scoring," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP '88)*, vol. 1, pp. 595–598, New York, NY, USA, April 1988.
- [20] Comparison of background normalization methods for text-independent speaker verification Douglas A. Reynolds Speech Systems Technology Group MIT Lincoln Laboratory E-mail: dar@sst.ll.mit.edu
- [21] D. Reynolds, M. Zissman, T. Quateri, G. O'Leary, and B. Carlson, "The effects of telephone transmission degradations on speaker recognition performance," *ICASSP*, pp. 329–332, May 1995.
- [22] D. A. Reynolds, "The effects of handset variability on speaker recognition performance: Experiments on the switchboard corpus," *ICASSP*, pp. 113–116, May 1996.
- [23] D. A. Reynolds, "HTIMIT and LLHBD: Speech corpora for the study of handset transducer effects," *ICASSP*, April 1997.
- [24] L. P. Heck and M. Weintraub, "Handset-dependent background models for robust text-independent speaker recognition," *ICASSP*, April 1997.
- [25] R. Auckenthaler, M. Carey, and H. Lloyd-Thomas, "Score normalization for text-independent speaker verification system," *Digital Signal Processing*, vol. 10, no. 1, 2000.
- [26] Speaker verification score normalization using speaker model clusters Vijendra Raj Apsingekar, Phillip L. De Leon * Klipsch School of Electrical and Computer Engineering, New Mexico State University, Las Cruces, NM 88003, USA Received 31 August 2009; received in revised form 6 July 2010; accepted 7 July 2010
- [27] Auckenthaler, R., Carey, M., Lloyd-Thomas, H., 2000. Score normalization for test-independent speaker verification system. *Digital Signal Process.* 10 (1), 42–54.
- [28] Reynolds, D. A. (2003). Channel robust speaker verification via feature mapping. *Proceedings of ICASSP'03*, Vol. 2, 53–56.
- [29] C. Fredouille, J.-F. Bonastre, and T. Merlin, "Similarity normalization method based on world model and a posteriori probability for speaker verification," in *Proc. European Conference on Speech Communication and Technology (Eurospeech '99)*, pp. 983–986, Budapest, Hungary, September 1999.
- [30] A Comparison of Normalization and Training Approaches for ASR-Dependent Speaker Identification Alex Park and Timothy J. Hazen MIT Computer Science and Artificial Intelligence Laboratory 32 Vassar Street, Cambridge, MA 02139, USA {malex, hazen}@sls.csail.mit.edu
- [31] A. Park and T. J. Hazen, "ASR dependent techniques for speaker identification," in *Proc. ICSLP*, Denver, Colorado, September 2002, pp. 2521–2524.
- [32] User-specific Score Normalization and Fusion for Biometric Person Recognition Norman Poh
- [33] G. Doddington, W. Liggett, A. Martin, M. Przybocki, and D. Reynolds. Sheep, Goats, Lambs and Wolves: A Statistical Analysis of Speaker Performance in the NIST 1998 Speaker Recognition Evaluation. In *Int'l Conf. Spoken Language Processing (ICSLP)*, Sydney, 1998.
- [34] N. Yager and T. Dunstone. Worms, chameleons, phantoms and doves: New additions to the biometric menagerie. *Automatic Identification Advanced Technologies*, 2007 IEEE Workshop on, pages 1–6, June 2007.
- [35] N. Poh and S. Bengio. Database, Protocol and Tools for Evaluating Score-Level Fusion Algorithms in Biometric Authentication. *Pattern Recognition*, 39(2):223–233, February 2005.
- [36] T. Matsumoto, H. Matsumoto, K. Yamada, and S. Hoshino. Impact of artificial gummy fingers on fingerprint systems. In *Proc. of SPIE 4677: Biometric Techniques for Human Identification*, pages 275–289, 2002.
- [37] P. Perrot, G. Aversano, R. Blouet, M. Charbit, and G. Chollet. Voice forgery using alisp: Indexation in a client memory. *Acoustics, Speech, and Signal Processing*, 2005. *Proceedings. (ICASSP '05)*. IEEE International Conference on, 1:17–20, 18–23, 2005.
- [38] B. Abboud and G. Chollet. Appearance based lip tracking and cloning on speaking faces. *Image and Signal Processing and Analysis*, 2005. *ISPA 2005. Proceedings of the 4th International Symposium on*, pages 301–305, Sept. 2005.

Authors Biography

Piyush Lotia, received the Master of Technology in Electronic and Telecommunication with specialization in control and instrumentation from BIT, Durg in 2006 and Bachelor of Engineering in Electronics engineering for NIT Raipur in 1997. He is working as Senior Associate Professor and Head Department of Electronics and Instrumentation in Shree Shankaracharya Technical Campus Bhilai. His Area of interest is Signal Processing and Wireless communication. He has published 24 papers in Journal and Conferences.



M.R. Khan, is a graduate in Electronics and Telecommunication from Govt. Engineering College, Jabalpur in 1985 and has done M.Tech from IIT Kharagpur in Telecommunication Systems Engineering in the year 1999. He completed his Ph.D. in Dec 2008. In the area of speech coding for telephone communication, from NIT, Raipur, Speech signal processing, communication and System simulation and modeling being his major areas of interest. He holds a teaching experience of more than 20 years in Govt. Engineering College and subsequently NIT, Raipur. He has got 16 research papers published in reputed National and International journals to his credit Currently he is working as Principal of Government Engineering college Raipur.,

