Odyssey 2016
21-24 Jun 2016, Bilbao, Spain

Multi-Bit Allocation: Preparing Voice Biometrics for Template Protection

M. Paulini[*], *C. Rathgeb*[†], *A. Nautsch*[†], *H. Reichau*[*], *H. Reininger*[*], *C. Busch*[†]

[*]atip – Advanced Technologies for Information Processing GmbH, Frankfurt, Germany
{marco.paulini,hermine.reichau,herbert.reininger}@atip.de

[†]da/sec – Biometrics and Internet Security Research Group, Hochschule Darmstadt, Germany
{christian.rathgeb,andreas.nautsch,christoph.busch}@h-da.de

Abstract

Technologies of biometric template protection grant a significant improvement in data privacy and increase the likelihood that the general public will effectively consent in the biometric system usage. Focusing on speaker recognition this area of research is still in its infancy. Previously proposed voice biometric template protection schemes fail in guaranteeing required properties of irreversibility and unlinkability without significantly degrading the recognition accuracy. A crucial step for accurate and secure template protection schemes is the feature type transformation which might be required to binarize extracted feature vectors.

In this paper we introduce a binarization technique for voice biometric features called *multi-bit allocation*. The proposed scheme, which builds upon a GMM-UBM-based speaker recogniton system, is designed to extract discriminative compact binary feature vectors to be applied in a voice biometric template protection scheme. In a preliminary experimental study we show that the resulting binary representation causes only a marginal decrease in biometric performance compared to the baseline system, confirming the soundness and aplicability of the proposed scheme.

Keywords: template protection, feature-type transformation, bit allocation, binarization

1. Introduction

Speaker recognition technologies [1] enable reliable recognition of individuals based on physical and/or behavior-related traits [2]. Given the increasing deployment of biometric technologies speaker recognition is developing into a rapidly growing field of research as voice represents the characteristic of choice in numerous application scenarios. In order to safeguard individuals privacy biometric system security as well as the protection of biometric reference data are of particular concern [3]. Providing a strong link between individuals and their biometric trait, (voice) biometric reference data is considered sensitive personal data [4]. Hence, biometric data protection is of utmost importance in order to prevent from serious privacy threats, e.g. identity theft or cross-matching. Technologies of biometric template protection [5], which are commonly categorized as biometric cryptosystems [6] and cancelable biometrics [7], offer solutions to privacy preserving biometric authentication. In accordance with the ISO/IEC IS 24745 [8] on biometric information protection, these technologies are designed to meet the major requirements: (1) *irreversibility*, i.e. knowledge of a protected template can not be exploited to reconstruct a biometric sample which is equal or close (within a small margin of error) to an original captured sample of the same source; (2) *unlinkability*, i.e. different versions of protected biometric templates can be generated based on the same biometric data (re-newability), while protected templates should not allow cross-matching.

Based on the standardized architecture [8] extracted biometric features serve as input of *pseudonymous identifier encoder* (PIE) of a biometric template protection scheme. PIE generates a pseudonymous identifier and corresponding auxiliary data which constitute the protected template. To complete the enrolment process the unprotected feature vector is deleted. At authentication the *pseudonymous identifier recorder* (PIR) takes a feature vector and a queried auxiliary data as input and calculates a pseudonymous identifier. Finally, a *pseudonymous identifier comparator* (PIC) is used to compare the generated pseudonymous identifier to the stored one. Depending on comparators, the comparison result s is either a binary decision (yes/no) or a similarity score which is then compared against a threshold t, in order to obtain a binary decision.

Commonly, biometric template protection schemes are categorized as *biometric cryptosystems* [6] and *cancelable biometrics* [7]. Biometric cryptosystems are designed to securely bind a digital key to a biometric characteristic or generate a digital key from a biometric characteristic [6]. The vast majority of biometric cryptosystems require the storage of biometric dependent public information applied to retrieve or generate keys, which is referred to as helper data [5]. Biometric comparisons are performed indirectly by verifying key validities, where the output of the authentication process is either a key or a failure message, i.e. unprotected templates are replaced through biometric-dependent public information which is used in order to release a key. Biometric cryptosystems operate on a certain form of input, e.g. point-sets or binary features. Hence, incompatibility issues arise when the type of intended biometric features does not match the acceptable input type of a biometric cryptosystem [9]. Ideally, biometric features can be transformed to the required feature-type in an appropriate manner. Moreover, feature-type transformations shall not cause a decrease in biometric performance compared to the original representation [5]. In particular, in mobile environments where devices are only equipped with restricted storage and processing capabilities, these transformations are essential for extracting compact discriminative biometric reference data [9]. Furthermore, biometric feature-type transformation comes into play when features of different biometric characteristics need to be fused at the feature level.

In this work we propose a feature-type transformation for voice biometric data which referred to as *multi-bit allocation* (MBA). This scheme is designed to extract a binary feature vector from a set of MFCC-based supervector coefficients of

doi: 10.21437/Odyssey.2016-42

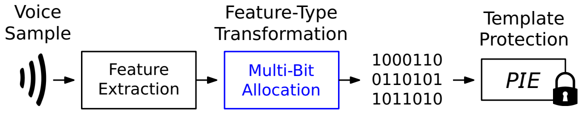

Figure 1: *Overview: the proposed feature-type transformation (multi-bit allocation) in the processing chain of a voice biometric template protection scheme.*

a GMM-UBM speaker recognition system. Extending previous work [10], feature warping is utilized in order to obtain uniform feature space for each supervector element. The feature space is generically divided into intervals which are encoded with multiple bits using a Gray code. In addition, so-called *relevant projections* are employed in order to detect most discriminative and stable supervector elements per enroled subject. Experiments, which are carried out on a text-independent digit corpus, demonstrate that the proposed MBA is capable of extracting compact discriminative binary voice biometric templates. In case of Hamming distance-based comparisons (deemed to most relevant features of a reference voice template) we observe a negligible decrease in biometric performance in terms of EER compared to the GMM-UBM speaker recognition system is almost maintained. Hence, extracted binary feature vectors can be applied effectively in a biometric template protection scheme operating on binary input, as illustrated in Fig. 1. Moreover, we provide an extensive review and discussion of the current state-of-the-art in research area of voice biometric template protection.

This paper is organized as follows: Sect. 2 summarizes related works with regard to voice biometric template protection and feature-type transformations in biometric systems. The employed baseline speaker recognition system is descried in Sect. 3. In detailed Sect. 4 we give a detailed description of the proposed MBA system. Experimental evaluations are presented in Sect. 5 and conclusions are drawn in Sect 6.

2. Related Work

In 1999, Monrose *et al.* [11] proposed a concept of enhancing the security of a password-based authentication system which they refer to as *password hardening*. In this concept, which has been applied to voice biometrics in [12], an existing password is "salted" with biometric data yielding a hardened password to be tested for login purposes or used as biometric key. By employing Shamir's secret sharing scheme [13] so-called shares, which are used to reconstruct the hardened password, are arranged in an instruction table. Most stable voice biometric features, which are referred to as *distinguishable features*, are detected and utilized to re-construct the shares at each authentication attempt. Further, by maintaining an encrypted constant-size history file as part of the biometric template, the system is designed to adapt to slight changes of voice patterns over time. The password hardening scheme represents the very first approach to voice biometric template protection implementing password salting for the purpose of biometric-dependent key generation [5].

Atah and Howells [14] use a combination of stable features extracted from voice data to directly generate biometric keys. The authors report high stability of keys, however, while the elimination of biometric templates improves privacy protection, re-newability of keys can not be achieved without storing any kind of auxiliary data. In [15] Teoh and Chong propose cancelable voice biometrics based on probabilistic random projections [16]. Features extracted from a GMM-UBM speaker recognition system are obscured by employing these projections, which can hide the actual speech features, where projections are parametrized based on subject-specific tokens. The incorporation of additional tokens yields two-factor authentication and enables re-newability/ unlinkability. It is shown that the presented scheme is capable of maintaining biometric performance even in the stolen-token scenario, where impostors are in possession of valid tokens.

Inthavisas and Lopresti [17, 18] present a voice biometric cryptosystem based on the fuzzy commitment scheme [19]. In the feature extraction stage a discrete Fourier transform is applied to the signals obtained from spoken pass-phrases. The system then performs dynamic time warping to minimize distance between the reference signal and training utterance. Next, the features of a subject's spoken pass-phrase are mapped to a binary string. For this purpose a fixed-length set of the most stable features, referred to as *distinguishing descriptor*, is detected and bound to a cryptographic key within a fuzzy commitment framework, together with a user-specific password. In order to harden their template, they perturb the system by successively running the mapping function, each time removing one of the stable features and using this new vector as the key in the dynamic time warping. In [10] we proposed a fuzzy commitment for Despite achieving the requirements of irreversibility and unlinkability, the system suffers from a significant drop in biometric performance caused by the binarization step in which in which each MFCC-based supervector element of a GMM-UBM system is allocated to a single bit.

Johnson *et al.* [20] proposed a voice verification system based on the concept of the fuzzy vault scheme [21]. Focusing on mobile devices they present a client-server protocol where no biometric information is stored on the server. In this scheme biometric data is split into several parts, which are referred to as blocklets. Then the same amount of chaff blocks are added and both sets which form the template are encrypted and sent to the server. The server decrypts and verifies the template and creates a locked challenge by choosing a bitstring based on which real and chaff blocklets are scrambled. Finally, the client reconstructs this bitstring by estimating comparison scores with respect to real and chaff parts of the template. However, vaulted voice verification, as originally proposed in [22], was not well suited for the exchange of larger keys. In [20] the authors extend the concept and present an index-based vaulted voice verification which significantly reduces communication overhead and allows the transmission of keys that are suitable for secure communication.

More recently, Vaquero and Rodríguez [23] discuss the need of biometric template protection in speaker recognition. As a general conclusion, the authors claim that text-dependent PLDA and HMM-based speaker recognition systems fulfil requirements of irreversibility and unlinkability (to a certain extent). Focusing on irreversibility, it is shown that success probabilities of presentation attacks based on voice samples reconstructed from compromised templates are close to random guessing. The fact that text-independent feature extraction techniques would allow cross-matching is omitted.

Focusing on biometric feature-type transformation, existing approaches can be classified based on the input/output representations and corresponding data-type [9], e.g. unordered-to-ordered transformation from real to binary feature vectors. In this paper we restrict to feature-type transformations which output a compact binary (ordered) feature vector. Binary biometric templates can be employed diverse template protection schemes which require a binary input, e.g. the fuzzy commitment scheme [19]. For a detailed survey on biometric feature-type transformation the reader is referred to [9]. With respect to voice biometrics the amount of introduced binarization methods turns out to be rather limited while numerous schemes have been presented for other biometric characteristics, e.g. face or fingerprint [24, 25, 26]. Focusing on speaker recognition early proposals suffered from a significant drop in biometric performance, e.g. [27, 28, 29]. More recent approaches have been shown to be capable of maintaining biometric performance of the baseline system providing compact template which enable rapid comparisons, e.g. [30, 31, 32]. It is important to note that existing binarization scheme have been designed for tasks other than biometric template protection such as rapid identification or speaker diarization.

3. Baseline System

A GMM-UBM-based speaker recognition system serves as baseline. The system comprises a feature extraction and comparison module which are described in this section.

3.1. Feature Extraction

Given a biometric observation $\mathbf{O}$ a total number of $\mathbf{M}$ feature vectors $\mathbf{o}_m\,[t]$ of length $\boldsymbol{T}$ are obtained in the feature extraction process, with $m = 1, \ldots, M$ and $t = 1, \ldots, T$. Feature vectors are modelled as a realization of a GMM by adapting the means of the UBM to the estimated means for the speaker using MAP adaptation. The GMM of a subject u is represented as the supervector $\boldsymbol{\xi}^{(u)}$ containing the mean vectors for each Gaussian distribution in the model.

Starting with an energy-based voice activity detection, MFCC features are extracted from the speech signal using window function of 20-30 ms. We obtain 12 coefficients and the Log-energy value for each frame and the first and second-order derivative, i.e. in total we obtain a feature vector with 39 components per frame. The obtained vectors are further processed by cepstral mean subtraction as well as feature warping [33] using a sliding window of 3 seconds. Feature warping maps Gaussian distributed features onto a standard Gaussian distribution in order to take between-sample mismatches into account. In the warping process the central feature within a sliding window is mapped with respect to its rank R within all N sliding window features, resulting in a warped value w which represents the corresponding expected value of a target distribution e.g., the standard Gaussian $\mathcal{N}(0, 1)$,

$$\frac{N - R + \frac{1}{2}}{N} = \int_{-\infty}^{w} \mathcal{N}(0, 1). \tag{1}$$

The process of feature warping involves a separate warping of each feature. Thereby, the feature space will result in R different feature values, which are similar-distributed to the original time-continuous biometric features. This mapping approach may be considered as recognizing the relative positions of each of the features as more important rather than their absolute feature values [33]. This full exploitation of a pre-defined feature space is essential to the proposed MBA scheme, see Sect. 4.

In order to derive an individual speaker model GMM from the general UBM, we only adapt the means of the UBM using the MAP adaptation. Assume that E biometric observations $\mathbf{O}^{(u)e}$, $e = 1, \ldots, E$ of a subject u are available during enrolment. We let the a posteriori probability for the component i, with $i = 1, \ldots, I$, of UBM $\boldsymbol{\Lambda}$ be,

$$P\left(i|\mathbf{x}_T, \boldsymbol{\Lambda}\right) = \frac{\mathbf{w}_i g(\mathbf{x}_T|\boldsymbol{\mu}_i, \sum_i)}{\sum_{j=1}^{J} \mathbf{w}_j g(\mathbf{x}_T|\boldsymbol{\mu}_j, \sum_j)}, \tag{2}$$

where $\mathbf{w}_i$ represents the mixture weights, $\boldsymbol{\mu}_i$ the means, $\mathbf{x}_T$ the training vectors for the desired model, and $g(\mathbf{x}_T|\boldsymbol{\mu}_i, \sum_i)$ the component's Gaussian density. Further, for the training vectors $\mathbf{x}_T$, we compute the Maximum-Likelihood for the mean parameters $E_i(\mathbf{x}_T)$. A model of a subject is then characterized by the MAP adapted means $\widehat{\boldsymbol{\mu}}_i(\mathbf{x}_T)$ using fixed relevance factors [34]. The adapted means are then combined to form the supervector $\boldsymbol{\xi}^u$ (biometric template) for subject u

$$\boldsymbol{\xi}^{(u)} = \left[\widehat{\boldsymbol{\mu}}_1, \widehat{\boldsymbol{\mu}}_2, \ldots, \widehat{\boldsymbol{\mu}}_I\right], \tag{3}$$

where I defines the number of components in the UBM.

3.2. Comparison

The comparison score between the probe of subject u and a reference template of subject $\tilde{u}$ is defined as the LLR of the according GMM $\boldsymbol{\Lambda}^{\tilde{u}}$ and the UBM $\boldsymbol{\Lambda}$, which is defined as,

$$\mathrm{LLR}(\mathbf{O}^{(u)}, \boldsymbol{\Lambda}^{\tilde{u}}, \boldsymbol{\Lambda}) = \sum_i \log\bigl(P(\mathbf{x}|\boldsymbol{\Lambda}_i^{\tilde{u}})\bigr) - \log\bigl(P(\mathbf{x}|\boldsymbol{\Lambda}_i)\bigr), \tag{4}$$

where $\boldsymbol{\Lambda}^{\tilde{u}}$ represents the GMM of subject $\tilde{u}$, $\boldsymbol{\Lambda}$ is the UBM, $\mathbf{w}_i$ are the mixture weights, $\boldsymbol{\mu}_i$ the means, and $\mathbf{x}$ the feature vectors extracted from $\mathbf{O}^{(u)}$.

4. Proposed System

The following subsections describe the proposed MBA in detail. In addition, the use of relevant projections and the comparison stage are summarized.

4.1. Multi-Bit Allocation

As described in the previous section, at the time of enrolment of subject u we record a number of E observations and process them by applying feature warping the MAP adaptation to derive the vectors $\boldsymbol{\xi}^{(u)e}$ with Z coefficients ($Z = 39\,I$). By assuming sufficient $\boldsymbol{\xi}^{(u)e}$ estimates for each observation, the mean vector $\mathbf{d}^{(u)}$ is then defined as,

$$\mathbf{d}^{(u)} = \frac{1}{E}\sum_{e=1}^{E} \boldsymbol{\xi}^{(u)e}. \tag{5}$$

Due to the application of feature warping the population mean of each vector element is expected to be zero. A binary vector $\mathbf{b}^{(u)}$ can be directly extracted by setting bits according to the sign of elements in $\mathbf{d}^{(u)}$,

$$\mathbf{b}^{(u)}[z] = \begin{cases} 1, \text{ if } \mathrm{sign}(\mathbf{d}^{(u)}[z]) = 1 \\ 0, \text{ otherwise} \end{cases} \qquad z = 1, \ldots, Z. \tag{6}$$

This *single-bit allocation* (SBA), which has been suggested in [30], has been found to cause a significant decrease in biometric performance. In other words, mapping each elements of $\mathbf{d}^{(u)}$ to a single bit results in a too coarse quantization and hence in binary feature vector exhibiting less discriminative power.

Motivated by this fact we suggest to divide the feature space into 2^k intervals and encode each interval with k bits according to a Gray code, where the chosen length of intervals is defined

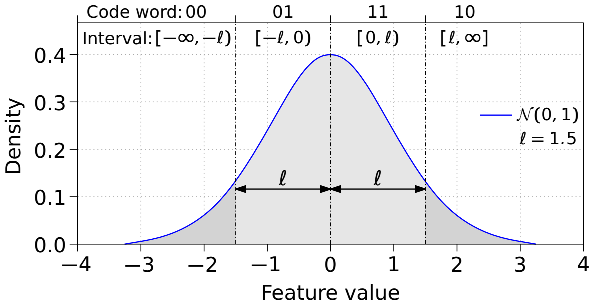

Figure 2: *MBA Example: the feature space is divided and encoded with 2 bits according to a target interval size of* $\ell = 1.5$.

by ℓ. Starting from zero 2^{k-1} intervals are set up in each direction where border-intervals take up the remaining feature space resulting in the following interval sequence,

$$\begin{aligned}&[-\infty, -2^{k-2}\ell), [-2^{k-2}\ell, -2^{k-3}\ell), \ldots,\\ &[-\ell, 0), [0, \ell], \ldots [2^{k-3}\ell, 2^{k-2}\ell), [2^{k-2}\ell, \infty].\end{aligned} \tag{7}$$

Subsequently, intervals are encoded by assigning one unique k-bit word to each interval such that the sequence of code words results in a Gray code. In a Gray code two successive code words differ in only one bit. Hence, in case feature values are 'close' but mapped into different intervals still cause only small increases in resulting binary strings. Fig. 2 illustrates an example for $k = 2$ and $\ell = 1.5$ where the feature space is divided into four intervals $[-\infty, -1.5)$, $[-1.5, 0)$,$[0, 1.5)$ and $[1.5, \infty]$. As will be shown in experimental evaluations, a total number of four intervals, which yields binary feature vectors of double length compared to a SBA scheme, suffices in order to obtain discriminative binarized templates.

It is important to note feature warping plays an important role within the proposed scheme. In case feature warping is not used, intervals have to be defined depending on standard deviations or variances of according feature elements. Obviously, this is not necessary if feature warping is applied, since feature values are mapped to a standard normal distribution.

4.2. Relevant Projection and Comparison

It might be required to generate binarized templates of a predefined size v, that comprise only bits exhibiting the highest possible discriminativity. According fixed-length binary vectors can be obtained by estimating the reliability measures $\boldsymbol{\varphi}^{(u)}[z]$,

$$\boldsymbol{\varphi}^{(u)}[z] = \frac{\left|\mathbf{d}^{(u)}[z]\right|}{\boldsymbol{\sigma}^{(u)}[z]}, \tag{8}$$

with the variance of the z-th feature estimated during the enrolment of subject u defined as,

$$(\boldsymbol{\sigma}^{(u)}[z])^2 = \frac{1}{E-1}\sum_{e=1}^{E}(\boldsymbol{\xi}^{(u)e}[z] - \mathbf{d}^{(u)}[z])^2. \tag{9}$$

This measure assigns greater relevance to those features which lie further away from the population mean ($\simeq$ zero) than others. The v most discriminative features (largest values) are then indexed by storing a bit mask pointing at these values which is referred to as relevant projection $\mathbf{RP}^{(u)}$, i.e. this bit mask contains 1s at positions of the v most discriminative features.

Table 1: *Performance evaluation: EERs of the proposed approach for* $k = 1, 2$ *and different settings of* ℓ *and* v.

System		v			
	ℓ	100	75	50	25
Single-bit	–	5.55	5.42	7.03	9.92
Multi-bit $k = 2$	0.50	13.29	13.05	14.16	14.19
	0.75	7.81	6.80	8.31	12.80
	1.00	6.13	4.57	5.62	8.72
	1.33	**4.17**	**3.56**	**3.87**	**6.21**
	1.50	5.13	4.70	5.36	8.42
	1.66	5.84	4.54	5.86	5.85
	2.00	5.22	5.53	6.04	9.60
	2.50	7.38	6.17	6.79	9.98
	3.00	6.20	7.06	7.74	10.29

In this preliminary study we measure the dicriminativity of extracted binary template by performing pair-wise Hamming distance-based comparisons. The comparator is implemented by the simple Boolean exclusive-OR operator (XOR) applied to a pair of binary vectors, masked (AND'ed) by the relevant projection of the reference template. The XOR operator $\oplus$ detects disagreements between any corresponding pair of bits while the AND operator $\cap$ ensures that only most discriminative bits (with respect to the reference template) are considered. For a binary template $\mathbf{b}^{(u)}$ of subject u, a reference template $\mathbf{b}^{(\tilde{u})}$ of subject $\tilde{u}$ and the corresponding relevant projection $\mathbf{RP}^{(\tilde{u})}$ we compute the fractional Hamming distance (HD) as a measure of the dissimilarity,

$$HD(\mathbf{b}^{(u)}, \mathbf{b}^{(\tilde{u})}) = \frac{||(\mathbf{b}^{(u)} \oplus \mathbf{b}^{(\tilde{u})}) \cap \mathbf{RP}^{(\tilde{u})}||}{v}, \tag{10}$$

where v is equal to the norm of $\mathbf{RP}^{(\tilde{u})}$, $v = ||\mathbf{RP}^{(\tilde{u})}||$.

5. Experiments

In the following subsection the experimental setup is described and obtained results are presented and discussed.

5.1. Experimental Setup

Experiments are carried out on a text-independent digit corpus database which comprises voice samples of a total number of 332 female (∼48%) and 369 male speakers (∼52%). For each subject 32 voice samples of length 3000ms-5000ms are available. The samples contain three to five spoken digits which vary for each sample. At enrolment, total number of M =30 samples are used to generate the according models for a subject and remaining samples are applied during authentication, where a total amount of 128 UBM components have been found to be an adequate choice. It is important to point out that single voice samples are rather short containing a reasonable amount of silence, i.e. the total amount of raw voice data (excluding silence) used to generate subject-specific models results in approximately 30-40 seconds. We use the same evaluation set of 339 subjects as described in [30] for the SBA scheme leading to a total number of 678 genuine and 229,164 impostor comparisons (all cross-comparisons). Thereby, obtained biometric performance results can be compared directly. It is important to note, that favourable acquisition conditions are considered a fundamental premise for biometric template protection [5].

In accordance with IS ISO/IEC 19795-1:2006 [35] biometric performance is estimated in terms of euqal error rate (EER).

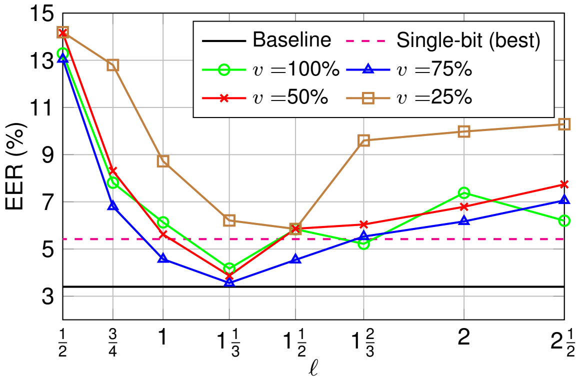

Figure 3: *Performance evaluation: EERs of the proposed approach for different values of v in relation to ℓ.*

5.2. Performance Evaluation and Discussion

Given the above described setup the (non-binary) baseline system achieves an EER of 3.40%. The SBA scheme extracts binary feature vectors consisting of $128 \times Z = 4992$ bits. In contrast, the proposed MBA generates binary feature vectors consisting of $128 \times Z \times k$ bits. For the considered numbers of intervals, $k = 2$, a total number of 9984 bits are extracted for each subject. For larger values of k we did not observe further performance gains. Obtained EERs of the SBA and MBA schemes are summarized in Table 1. Compared to the baseline system the SBA scheme suffers from a significant drop in biometric performance achieving EERs above 5%. Employing the proposed MBA scheme biometric performance is improved achieving EERs below 4%. Fig. 3 compares EERs of the baseline system and the best SBA scheme to different settings of the proposed MBA scheme. It can be observed that various configuration of the MBA outperform the SBA scheme. Interestingly, best performance is achieved for $\ell = 1.33$. Since feature values are distributed within $\mathcal{N}(0, 1)$, inner intervals are expected to cover distinctly more than half of the feature values for this setting. On the other hand, only features that exhibit a larger distance to the population mean are assigned to outer intervals which increases the discriminativity of resulting binary templates.

On the contrary, we also observe the for numerous configurations the proposed MBA scheme does not outperform the SBA scheme, e.g. ℓ<1 or ℓ>2. Both, rather small and large values of ℓ results in rather narrow inner and outer intervals, respectively, which decreases the decreases the stability of discriminative feature elements. As a consequence, biometric performance decreases significantly.

Moreover, we observe that the use of relevant projections can further improve biometric performance. For the majority of considered settings limiting comparisons to v =75% of the most relevant bits of the reference template turns out to be optimal. This means, employing less but more discriminative features improves biometric performance. However, for most settings using less than 50% of bits reveals a severe negative impact on biometric performance. The lowest EER of 3.56% is obtained for $\ell = 1.33$ and v =75%. That is, the proposed MBA is capable of recovering 92.1% of the performance gap between the baseline and the SBA system.

Extracted binary feature vectors are appropriate to be applied in biometric template protection schemes which require a binary input. For instance, generated binary templates could be used a fuzzy commitment scheme binding and retrieving keys of approximately 30-40 bits [30]. Focusing on unlinkability, we suggest to implement subject/application-specific random permutation of bits which are applied to the binary templates as well as corresponding relevant projection masks. Alternatively, random bit strings generated based on PRNGs could be XORed with extracted binary templates simulating one-time pad encryption. Moreover, it is important to note that compact binary feature representation obtained by MBA enables template protection in scenarios where storage capacities are limited.

6. Conclusion

In this work we gave an overview of the current state-of-the-art with respect to voice biometric template protection and provided an in-depth discussion on the related topic of biometric feature type transformation. Further, we introduce a ordered-to-ordered real-to-binary feature type transformation referred to as *multi-bit-allocation*. The proposed method extracts compact discriminative binary template suitable to be integrated to biometric template protection schemes. In experiments MBA is shown to almost maintain biometric performance compared to a GMM-UBM baseline system.

Future work will focus on the integration of obtained binary feature vector to generic template protection schemes. While favourable capture conditions is considered a fundamental premise in the area of biometric template protection [3], an evaluation of the proposed MBA system on more challenging datasets, which might require refined division of feature spaces, is also subject to future research. In addition, while we employed a GMM-UBM-based speaker recognition system, an integration of the proposed scheme to a more sophisticated feature extraction techniques might be considered, e.g. i-vectors [36]. Finally, further encoding strategies could be analysed.

7. Acknowledgments

This work has been partially funded by the Hesse government (project no. 467/15-09, BioMobile).

8. References

[1] T. Kinnunen and H. Li, "An overview of text-independent speaker recognition: From features to supervectors," *Speech Communication*, vol. 52, no. 1, 2010.

[2] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 14, 2004.

[3] K. Nandakumar and A. K. Jain, "Biometric template protection: Bridging the performance gap between theory and practice," *IEEE Signal Processing Magazine - Special Issue on Biometric Security and Privacy*, pp. 1–12, 2015.

[4] European Commission, "Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)," Dec 2015.

[5] C. Rathgeb and A. Uhl, "A survey on biometric cryptosystems and cancelable biometrics," *EURASIP Journal on Information Security*, vol. 2011, no. 3, 2011.

[6] U. Uludag, S. Pankanti, S. Prabhakar, and A. K. Jain, "Biometric cryptosystems: issues and challenges," *Proc. of the IEEE*, vol. 92, no. 6, 2004.

[7] N.K. Ratha, J.H. Connell, and R.M. Bolle, "Enhancing security and privacy in biometrics-based authentication systems," *IBM Systems Journal*, vol. 40, no. 3, 2001.

[8] ISO/IEC JTC1 SC27 Security Techniques, *ISO/IEC 24745:2011. Information Technology - Security Techniques - Biometric Information Protection*, International Organization for Standardization, 2011.

[9] M.-H. Lim, A. B. J. Teoh, and J. Kim, "Biometric feature-type transformation: Making templates compatible for template protection," *IEEE Signal Processing Magazine*, vol. 32, no. 5, 2015.

[10] S. Billeb, C. Rathgeb, H. Reininger, K. Kasper, and C. Busch, "Biometric template protection for speaker recognition based on universal background models," *IET Biometrics*, vol. 4, no. 2, 2014.

[11] F. Monrose, M. K. Reiter, and S. Wetzel, "Password hardening based on keystroke dynamics," in *Proc. 6th ACM Conf. on Computer and Communications Security*. 1999, pp. 73–82, ACM.

[12] F. Monrose, M. K. Reiter, Q. Li, and S. Wetzel, "Using Voice to Generate Cryptographic Keys," in *Proc. Speaker Odyssey 2001, The Speech Recognition Workshop*, 2001.

[13] A. Shamir, "How to share a secret," *Commun. ACM*, vol. 22, no. 11, 1979.

[14] J. A. Atah and G. Howells, "Key generation in a voice based template free biometric security system," in *Proc. of the Joint COST 2101 and 2102 Int'l Conf. on Biometric ID Management and Multimodal Communication*. 2009, Springer-Verlag.

[15] A. B. J. Teoh and L.-Y. Chong, "Secure speech template protection in speaker verification system," *Speech Communication*, vol. 52, no. 2, 2010.

[16] A. B. J. Teoh and Chong T. Y., "Cancelable biometrics realization with multispace random projections," *IEEE Trans. on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 37, no. 5, 2007.

[17] K. Inthavisas and D. Lopresti, "Speech cryptographic key regeneration based on password," in *Proc. of the Int'l Joint Conference on Biometrics (IJCB'11)*, 2011.

[18] K. Inthavisas and D. Lopresti, "Secure speech biometric templates for user authentication," *IET Biometrics*, vol. 1, no. 1, 2012.

[19] A. Juels and M. Wattenberg, "A fuzzy commitment scheme," in *Proc. 6th ACM Conf. on Computer and Communications Security*, 1999.

[20] R.C. Johnson and T.E. Boult, "With vaulted voice verification my voice is my key," in *Proc. Int'l Conf. on Technologies for Homeland Security (HST'13)*, 2013.

[21] A. Juels and M. Sudan, "A fuzzy vault scheme," in *Proc. IEEE Int'l Symp. on Information Theory*, 2002.

[22] R. C. Johnson, W. J. Scheirer, and T. E. Boult, "Secure voice based authentication for mobile devices: Vaulted voice verification," *CoRR*, vol. abs/1212.0042, 2012.

[23] C. Vaquero and P. Rodríguez, "On the need of template protection for voice authentication," in *Proc. of INTERSPEECH'15*. 2015, ISCA.

[24] C. Chen, R. Veldhuis, T. Kevenaar, and A. Akkermans, "Biometric quantization through detection rate optimized bit allocation," *EURASIP Journal on Advances in Signal Process*, vol. 2009, 2009.

[25] M.-H. Lim, A. B. J. Teoh, and K.-A. Toh, "An efficient dynamic reliability-dependent bit allocation for biometric discretization," *Pattern Recognitipon*, vol. 45, no. 5, 2015.

[26] A. Nagar, K. Nandakumar, and A.K. Jain, "Multibiometric cryptosystems based on feature-level fusion," *IEEE Trans. on Information Forensics and Security*, vol. 7, no. 1, 2012.

[27] X. Anguera and J.-F. Bonastre, "A novel speaker binary key derived from anchor models," in *Proc. of INTERSPEECH'10*. 2010, ISCA.

[28] J.-F. Bonastre, P. M. Bousquet, D. Matrouf, and X. Anguera, "Discriminant binary data representation for speaker recognition," in *Proc. Int'l Conf. on Acoustics, Speech and Signal Processing (ICASSP'11)*, 2011.

[29] G. Hernandez-Sierra, J.-F. Bonastre, and J. R. Calvo de Lara, "Speaker recognition using a binary representation and specificities models," in *Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications*, vol. 7441 of *LNCS*. 2012.

[30] S. Billeb, C. Rathgeb, M. Buschbeck, H. Reininger, and K. Kasper, "Efficient two-stage speaker identification based on universal background models," in *Proc. Int'l Conf. Biometrics Special Interest Group (BIOSIG'14)*, 2014.

[31] H. Delgado, C. Fredouille, and J. Serrano, "Towards a complete binary key system for the speaker diarization task," in *Proc. of INTERSPEECH'14*. 2014, ISCA.

[32] H. Delgado, X. Anguera, C. Fredouille, and J. Serrano, "Fast single- and cross-show speaker diarization using binary key speaker modeling," *IEEE/ACM Trans. on Audio, Speech, and Language Processing*, vol. 23, no. 12, 2015.

[33] J. Pelecanos and S. Sridharan, "Feature warping for robust speaker verification," in *Proc. of Speaker Odyssey - The Speaker Recognition Workshop*, 2001.

[34] C. You, H. Li, B. Ma, and K.-A. Lee, "Effect of relevance factor of maximum a posteriori adaptation for gmm-svm in speaker and language recognition.," in *Proc. of INTERSPEECH*. 2012, ISCA.

[35] ISO/IEC TC JTC1 SC37 Biometrics, *ISO/IEC 19795-1:2006. Information Technology – Biometric Performance Testing and Reporting – Part 1: Principles and Framework*, International Organization for Standardization and International Electrotechnical Committee, Mar. 2006.

[36] N. Dehak, P. J. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet, "Front-End Factor Analysis for Speaker Verification," *IEEE Trans. on Audio, Speech and Language Processing*, 2010.