Security Evaluation of i-Vector Based Speaker Verification Systems Against Hill-Climbing Attacks

Marta Gomez-Barrero, Javier Gonzalez-Dominguez, Javier Galbally, Joaquin Gonzalez-Rodriguez

ATVS Biometric Recognition Group, C/ Francisco Tomas y Valiente 11,
Universidad Autonoma de Madrid, 28049 Madrid, Spain

{marta.barrero,javier.gonzalez,javier.galbally,joaquin.gonzalez}@uam.es

Abstract

This work studies the vulnerabilities of i-vector based speaker verification systems against indirect attacks. Particularly, we exploit the one-to-one representation of speakers via their corresponding i-vectors to perform Hill-Climbing based attacks; under the hypothesis that the inherent low-dimensionality of i-vectors might represent a potential security breach to fraudulently access the system. The conducted attacks followed a standard experimental protocol already applied to other biometric systems based on face or signature; and they were tested against a state-of-art PLDA speaker verification system in the framework of the NIST SRE 2010 evaluation campaign. Specifically, up to 200 speakers, 100 female and 100 male, were attacked supplanting their corresponding i-vectors by those derived with the Hill-Climbing approach. Experiments show the success of the proposed attack compared with those based on brute force, achieving high Success Rates (up to 100%) and needing half as many comparisons as the brute force access attempts. These results evidence the vulnerability of i-vector based speaker verification systems in those scenarios where access to the matcher score is granted multiple times. As a countermeasure to minimize the effect of the attack score quantization is also evaluated.
Index Terms: speaker verification, i-vector, security, hill-climbing.

1. Introduction

The use of biometrics for automatic identification purposes is spreading day by day [1]. Being a relatively young technology, most efforts undertaken by the different parties involved in its development (researchers, industry, evaluators, etc.) have been mainly (but not exclusively) focused on the improvement of its performance [2]. This has left partially uncovered other important aspects involved in the complex biometric recognition problem.

In particular, as any other security-related application, biometric systems are vulnerable to external attacks, and it has not been until recently when biometric security assessment has emerged as a major field of research.

From a general point of view, attacks to biometric systems have traditionally been classified into *spoofing* attacks, also referred to as direct attacks (type 1 in Fig. 1), carried out at the sensor level, and *indirect* attacks, directed to the inner modules of the system (labelled 2-8 in Fig. 1).

Regarding speaker recognition, security related issues such as the impact of spoofing or tampering attacks are arising an increasing interest; the maturity reached by this technology has opened the door to develop voice driven applications. Focusing on spoofing attacks, the first attempts towards deceiving

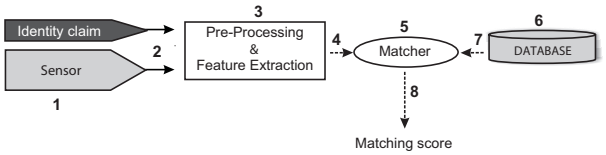

Figure 1: Architecture of an automated biometric verification system, with the possible points of attack identified in [20].

speaker verification systems used concatenated client speech [3] or synthetic speech [4, 5]. A similar approach was analysed in [6, 7] for state-of-the-art top-ranked systems. More recent articles focus on transforming the voice of an arbitrary person and thereby forging the identity of another person (tampering) [8, 9, 10, 11, 12]. The ability of professional impersonators to attack an automatic system has also been studied in [13, 14].

However, while those works analyse direct attacks, no attention has been paid up to date to indirect attacks against speaker verification systems. The goal of this paper is to evaluate the strength of state-of-art speaker verification systems to indirect attacks, specifically against those attacks where access to the matcher score is granted multiple times. For this purpose, a Probabilistic Linear Discriminant Analysis speaker verification system [15] where each speaker is represented by a low-dimensional vector, the so-called i-vector, has been attacked using a particular variant of the Hill-Climbing attack strategy.

Hill-Climbing strategies (HC) have been widely used in other traits to perform indirect attacks [16, 17, 18, 19]: taking advantage of the score given by the matcher, they iteratively change a synthetically generated template until the similarity score exceeds a fixed decision threshold, and thereby access to the system is granted. In this work, generated templates are used to replace i-vectors, and thus the speaker identities. To reach this objective, a variant of Hill-Climbing based on the Uphill-Simplex algorithm has been used. This approach has the main advantage of presenting a high biometric independence (i.e., it may be used to attack different modalities) as well as a great ability to adapt to different matchers that use fixed length feature vectors of real numbers.

Besides, this work also analyses possible countermeasures to prevent indirect attacks. In the field of voice biometrics, several studies have been conducted on anti-spoofing techniques against synthetic voice signals [21]. Most of those works base their decisions on information measured directly from the input signal [22, 23, 24], while other attempts combine visual and auditory information [25, 26] or model both synthetic and real signals in order to discriminate between them [27]. All those methods are very specific to the system at hand and the voice

trait, and do not guarantee protection against indirect attacks following a HC scheme. Therefore, we propose the use of score quantization, a standard countermeasure recommended by the BioAPI consortium [28], already applied to systems working on different biometric traits, such as fingerprints [19] or face [29].

The rest of this paper is structured as follows. The Hill-Climbing attack algorithm used in the experiments is outlined in Sect. 2, while the attacked system is presented in Sect. 3. The database and experimental protocol followed are described in Sect 4. The results of the attack and of the quantization scheme studied as countermeasure are detailed in Sect. 5. Conclusions are finally drawn in Sect. 6.

2. Hill Climbing Attack Based on the Uphill-Simplex Algorithm

In a generic HC attack, synthetic templates are generated and iteratively modified according to the similarity score given by a matcher, until the verification threshold δ is reached.

In the present contribution we use the attack based on the Uphill-Simplex algorithm, first presented in [30]. The core idea behind the algorithm is to iteratively change a randomly initialized simplex (a polygon with $n+1$ vertices in which each vertex is a synthetic n dimensional feature vector) so that it approaches the objective (the user account being attacked, defined as $\mathcal{C}$). In each iteration, the similarity score (s_j) from each simplex vertex (w_j) to the target ($\mathcal{C}$) is computed, according to a matching function ($\mathcal{J}$), $s_j = \mathcal{J}(\mathcal{C}, w_j)$, with $j = 1, \ldots, n+1$. The vertex furthest to the objective, w_l, is discarded and substituted by a new point, which can be computed in three different ways: i) *reflection*, according to a previously fixed α parameter; if reflection fails, either ii) *expansion* (dependent on the γ parameter) or iii) *contraction* (dependent on the β parameter) are used as a means to compute the new vertex. This process continues until the maximum score of the vertices exceeds the verification threshold or the maximum number of iterations allowed is reached.

Throughout the experiments, we will assume the same configuration found to be optimal for attacking a signature-based system in [30] and used as well to break a face-based application in [29], that is: $[\alpha, \beta, \gamma] = [1.1, 0.8, 1.1]$.

3. Speaker Verification System Attacked

The described HC attack based on the Uphill-Simplex algorithm is used to evaluate the security of a Probabilistic Linear Discriminant Analysis (PLDA) speaker verification system.

PLDA is a generative latent variable model that has been recently used to successfully model i-vectors [31]. It can be seen as a probabilistic version of classical LDA [15], where a specific i-vector i of a given speaker m is assumed to be decomposed as

$$w_{mi} = \mu + Fh_m + Gk_i + \epsilon_i$$

where F and G represent the new speaker and session variability subspaces respectively, h_m and k_i their respective latent variables associated and ϵ_i is a residual noisy term assumed to be normal distributed with zero mean and diagonal covariance matrix Σ.

Following the PLDA model the similarity measure or score S_{w_1,w_2} between two given i-vectors, w_1 and w_2, can be computed as the ratio of the two alternative hypothesis: H_0, both w_1 and w_2 belongs to a same identity (same h_m) and H_1, w_1 and w_2 belongs to different identities (different h_m). This ratio can be expressed as

$$\begin{aligned} s_{w_1,w_2} &= \frac{p(w_1, w_2|H_0)}{p(w_1|H_1)p(w_2|H_1)} \\ &= \frac{\int p(w_1, w_2|h)p(h)dh}{\int p(w_1|h_1)p(h_1)dh_1 \int p(w_2|h_2)p(h_2)dh_2}. \end{aligned}$$

Assuming Gaussian priors for the latent variables, it can be seen that the integrals involved in above equation turn out tractable and therefore the score, s_{w_1,w_2}, can be easily derived in a closed-form solution. Further details can be found in [15, 32].

In this work, the i-vector dimension was set to $n = 600$, and Length Normalization [32] was applied before scoring in order to avoid non-Gaussian behaviour of i-vectors.

4. Experimental Protocol

The success chances of most attacking approaches (either direct or indirect) are highly dependent on the operating point of the recognition system being evaluated. For this reason, in the present work we have followed a standard experimental protocol already applied in other biometric vulnerabilities evaluations [29, 30, 33], which comprises two successive steps, namely: i) performance evaluation, and ii) security evaluation. The first step allows us to establish the operating point at which the system works in terms of the False Acceptance Rate (FAR, percentage of impostors accepted as genuine users) and the False Rejection Rate (FRR, percentage of genuine users rejected by the system). Once the operating point is fixed, in the second step we are able to evaluate in a more objective way the robustness of the system against the attacking algorithm.

Furthermore, such a protocol also permits a more fair comparison with other security evaluations carried out on *different* systems working on the *same* operating points.

4.1. Database

Experiments were conducted on a subset of the core condition evaluated in NIST SRE10 [34]. In this task, the available train and testing data consists of a conversational telephone speech excerpt with a nominal duration of ∼150s. Each train and test excerpt was converted in a single i-vector following a Total Variability modelling as in [35]. NIST SRE10 data belonging to MIXER 4, 5 and 6 corpora provides a challenging dataset to speaker recognition including a larger number of speakers involved, multilingual material and recording conditions. Further details can be found in [36].

4.2. Performance Evaluation

The performance of the system was assessed following the NIST SRE10 evaluation protocol [34] and results are presented in terms of the Equal Error Rate (EER). A total number of 50.000 trials/comparisons (25.000 female and 25.000 male, involving over ∼1500 speakers per gender), were processed; obtaining Equal Error Rates of 4.39%, 1.86% for the female and male subset, respectively. The HC algorithm is evaluated at three operating points, namely: FAR = 0.1%, FAR = 0.05%, and FAR = 0.01%, which correspond to a low, medium, and high security application according to [37].

4.3. Security Evaluation

In order to generate the user accounts $\mathcal{C}$ to be attacked with the HC algorithm, we randomly chose 100 male and 100 female

FAR	Normal BF		Uniform BF		i-vector BF		i-vector LN BF	
	SR	Eff ($\times 10^{-4}$)	SR	Eff ($\times 10^{-4}$)	SR	Eff ($\times 10^{-4}$)	SR	Eff ($\times 10^{-4}$)
0.10%	0%	-	0%	-	21%	1.1800	29%	0.9251
0.05%	0%	-	0%	-	10%	1.2426	16%	0.9360
0.01%	0%	-	0%	-	2%	1.0330	1%	0.5456

Table 1: SR and Eff for the four Brute Force (BF) attacks carried out at the operating points tested for the **male subset**.

FAR	Normal BF		Uniform BF		i-vector BF		i-vector LN BF	
	SR	Eff ($\times 10^{-4}$)	SR	Eff ($\times 10^{-4}$)	SR	Eff ($\times 10^{-4}$)	SR	Eff ($\times 10^{-4}$)
0.10%	0%	-	0%	-	19%	1.3043	47%	1.2194
0.05%	0%	-	0%	-	9%	1.1739	26%	0.9698
0.01%	0%	-	0%	-	0%	-	0%	-

Table 2: SR and Eff for the four Brute Force attacks (BF) carried out at the operating points tested for the **female subset**.

FAR	Male				Female			
	Uphill-Simplex HC		i-vector LN BF		Uphill-Simplex HC		i-vector LN BF	
	SR	Eff ($\times 10^{-4}$)	SR	Eff ($\times 10^{-4}$)	SR	Eff ($\times 10^{-4}$)	SR	Eff ($\times 10^{-4}$)
0.10%	100%	1.1350	29%	0.9251	100%	1.2268	47%	1.2194
0.05%	100%	1.1231	16%	0.9360	100%	1.2169	26%	0.9698
0.01%	100%	1.0997	1%	0.5456	100%	1.1669	0%	-

Table 3: Eff and SR at the operating points tested, compared to those obtained by the most efficient BF attack tested (i-vector LN BF).

i-vectors from the NIST SRE10 training set. The performance of the attack will be evaluated in terms of the Success Rate and Efficiency, defined as in [38].

Success Rate (SR): it is the expected probability that the attack breaks a given account. It is computed as the ratio between the number of broken accounts (A_B) and the total number of accounts attacked ($A_T = 170$): $\text{SR} = A_B/A_T$, indicating how dangerous the attack is: the higher the SR, the bigger the threat.

Efficiency (Eff): it is computed as the inverse of the average number of matchings needed by the attack to break an account. It is defined as $\text{Eff} = 1/\left(\sum_{t=1}^{A_B} n_t/A_B\right)$, where n_t is the number of matchings computed to bypass each of the broken accounts, thus giving an estimation of how easy it is for the attack to break into the system in terms of speed: the higher the Eff, the faster the attack.

5. Results

The goal of the experiments is threefold: i) apply standard protocols and metrics for the vulnerability evaluation of biometric recognition systems to speaker verification, ii) evaluate the performance of the proposed HC attack under a new scenario (speaker verification), and iii) analyse the impact of a standard trait-independent countermeasure against indirect attacks, in the framework of speaker verification.

5.1. Analysis of Different Operating Points

For completeness and also as baseline result with which to compare the performance of our attacking scheme, a brute force attack approach (i.e., an exhaustive search through a very large number of i-vectors) was also carried out. For this purpose, 20,000 i-vectors were randomly sampled from four different distributions, each of them constituting one brute force attack:

- **Normal Brute Force**: i-vectors are sampled from a normal distribution $\mathcal{N}(0, 1)$.
- **Uniform Brute Force**: i-vectors are sampled from a uniform distribution $\mathcal{U}(-0.5, 0.5)$.
- **i-vector Brute Force**: each parameter of a development pool of real i-vectors is modelled with a normal distribution $\mathcal{N}(\mu_j, \sigma_j)$, with $j = 1, \ldots, 600$. Then, the synthetic i-vectors for the brute force attack are sampled from those normal distributions.
- **i-vector with Length Normalization Brute Force**: analogous to the i-vector BF attack, but the distributions model i-vectors after Length Normalization [32].

As the tested system is working at operating points where, on average, one *real* i-vector in 1,000 (FAR = 0.1%) to 10,000 (FAR = 0.01%) would produce a false positive, it seems that 20,000 may be a reasonable amount of *synthetic* i-vectors to find one that is assigned to a given real identity. Therefore, those synthetic i-vectors were matched to the users of the database until one of the synthetic samples produced a score greater than the corresponding δ. The number of comparisons needed by the brute force strategy to achieve a false positive is M, being w_M the first i-vector that produced the winning score.

Results of all of the brute force attacks described are presented in Tables 1 (for the male subset) and 2 (for the female subset). We can observe that for the two schemes that require no previous information about the i-vectors (Normal BF and Uniform BF), no user accounts are broken. However, when we have an initial pool of i-vectors for modelling the normals means and standard deviations, as many as 47% of the user accounts are broken in the best case (i.e., the recognition system operating at FAR = 0.1% over the female subset). The SR of this attack (i-vector LN BF) decreases for more secure operating points, being only 1% for the male subset and 0% for the female one.

In Table 3, the results achieved by the attack proposed in the present article are compared to those of the most efficient brute force method (i-vector LN BF) in terms of the SR and Eff. We can observe that the SR is 100% (i.e., all accounts are broken) at all the operating points for the HC scheme, while it is only 1% for the BF attack at FAR = 0.01%. Similarly, for the most secure operating point, the Eff of the HC approach is twice that of the the BF scheme, meaning that we need half as many

comparisons in order to retrieve an i-vector similar enough to the original one according to the recognition system.

Therefore, not only the i-vectors are approximated with a considerably lower number of comparisons for the most secure operating points by the Uphill Simplex-based approach, but it also guarantees success in the attack, in contrast to the brute force schemes.

It should also be noted that the Eff does not vary greatly between different operating points for the proposed algorithm: for the male subset, it ranges between 1.1350×10^{-4} and 1.0997×10^{-4} (i.e., on average, we need between 8800 and 9100 comparisons to break an account); and for the female subset, it ranges between 1.2268×10^{-4} and 1.1669×10^{-4} (i.e., on average, we need between 8150 and 8550 comparisons).

It should be highlighted that the configuration for the Uphill-Simplex algorithm is the same one already used to attack signature and face recognition systems [30, 29]. Therefore, the ability of the attack to break different and independent sets of features extracted from completely different biometric traits has been further proved in the present study.

5.2. Countermeasuring the Attack: Score Quantization

The results presented in Sect. 5.1 have shown the vulnerability of the speaker recognition system considered in the experiments against the proposed HC algorithm: the SR is as high as 100% at all the operating points tested and the Eff is considerably higher than that of the best brute force attack considered (which, in addition, requires a development set of real i-vectors). As a consequence, we need to incorporate some attack protection method that increases the system's robustness against this threat. In this section we analyse the performance of score quantization as a way to countermeasure the attack.

Score quantization has been proposed as an effective biometric-based approach to reduce the effects of HC attacks by quantizing the score so that the HC algorithm does not get the necessary positive feedback to iteratively increase the similarity measure. The BioAPI consortium [28] recommends that biometric algorithms emit only quantized matching scores in order to prevent eventual HC attacks.

Here, we will consider the speaker recognition system operating at a high security operating point (FAR = 0.01%), and we will assume the same configuration used in the vulnerability assessment experiments for the HC algorithm.

In order to select the appropriate quantization step according to the trade-off that should be met in terms of its impact on the system performance (ideally as small as possible) and on the attack performance (as big as possible), we try several Quantization Steps (QS). The findings are summarized in Tables 4 and 5. PI is the percentage of iterations out of the total performed in the attack that produced a Positive Increase in the matching score (i.e., the score increase was higher than the QS). As we may observe, the EER increases drastically for QS = 2×10^{-1} for the male subset and for QS = 10^{-2} for the female subset: the performance of the system is not acceptable at these points. Therefore, we will consider two different QSs for each subset, where the EER increases about 0.5% and 5%, respectively: 10^{-2} and 10^{-1} for the males, 5×10^{-4} and 10^{-3} for the females.

Results are presented in Table 6. As it may be observed, the effects of the score quantization are not equivalent for the male and female subsets. While in the first case the SR decreases form 100% to 75%, and the Eff increases almost 5 times, in the second case the SR remains constant at 100% and the Eff barely decreases. The behaviour of the system under quantized scores scenario for the male subset suggests that the proposed countermeasure is effective for the strongest accounts. However, the attack is still able to break the weakest ones with a very low number of comparisons (Eff is considerably higher).

QS	10^{-6}	10^{-4}	10^{-3}	10^{-2}	10^{-1}	2×10^{-1}
PI (%)	7.11	7.11	1.55	0	0	0
EER (%)	1.86	1.86	1.86	2.21	6.73	100

Table 4: PI of the iterations and EER of the system for different quantization steps (QS) of the matching score for the **male subset**.

QS	10^{-6}	10^{-4}	5×10^{-4}	10^{-3}	10^{-2}	10^{-1}
PI (%)	4.99	4.99	1.04	0.37	0	0
EER (%)	4.39	4.39	5.40	8.65	92.87	100

Table 5: PI of the iterations and EER of the system for different quantization steps (QS) of the matching score for the **female subset**.

	Male		Female	
QS	10^{-2}	10^{-1}	5×10^{-4}	10^{-3}
SR	78%	73%	100%	100%
Eff ($\times 10^{-4}$)	1.5026	4.8166	1.1786	1.1853

Table 6: Performance (in terms of SR and Eff) of the HC attack against the system for different QSs.

6. Conclusions

The robustness of a PLDA speaker verification system against a HC attack based on the Uphill Simplex algorithm was studied following a standard protocol and metrics already used for the vulnerabilities evaluation of other biometric security systems against indirect attacks. The performance of the proposed attack as compared to several brute force schemes, clearly outperforming all of them and showing the lack of robustness of the tested system against the proposed attacking method.

The attacking algorithm considered had already been successfully tested against on-line signature [30] and face [29] verification systems. These new experimental results have therefore confirmed that the proposed HC attack can be applied to break biometric recognition systems working on different traits as long as they use fixed length feature vectors of real numbers.

As a possible way to minimize the effect of the attack, score quantization was studied. Although it considerably reduced the SR of the attack for the male subset, the proposed attacking scheme proved its robustness to this type of countermeasure.

We believe that studies such as the one presented here, using publicly available data and following standard evaluation protocols, will help bring some insight into the difficult problem of biometric security evaluation and to further develop the rapidly evolving speaker recognition technology.

7. Acknowledgements

This work has been partially supported by projects Contexts (S2009/TIC-1485) from CAM, Bio-Shield (TEC2012-34881) and CMC-V2 (TEC2012-37585-C02-01) from Spanish MINECO, TABULA RASA (FP7-ICT-257289) and BEAT (FP7-SEC-284989) from EU, and *Cátedra UAM-Telefónica*.

8. References

[1] A. K. Jain, A. Ross, and S. Pankanti, "Biometrics: a tool for information security," *IEEE Trans. on Information Forensics and Security*, vol. 1, no. 2, pp. 125–143, 2006.

[2] A. Mansfield and J. Wayman, "Best practices in testing and reporting performance of biometric devices," CESG Biometrics Working Group, Tech. Rep., August 2002, (http://www.cesg.gov.uk/).

[3] J. Lindberg and M. Blomberg, "Vulnerability in speaker verification - a study of technical impostor techniques," in *Proc. ECSCT*, vol. 3, 1999, pp. 1211–1214.

[4] T. Masuko, T. Hitotsumatsu, K. Tokuda, and T. Kobayashi, "On the security of hmm-based speaker verification systems against imposture using synthetic speech," in *Proc. Eurospeech*, vol. 3, 1999, pp. 1223–1226.

[5] T. Masuko, K. Tokuda, and T. Kobayashi, "Imposture using synthetic speech against speaker verification based on spectrum and pitch," in *Proc. ICSLP*, vol. 2, 2000, pp. 302–305.

[6] P. L. De Leon, V. R. Apsingekar, M. Pucher, and J. Yamagishi, "Revisiting the security of speaker verification systems against imposture using synthetic speech," in *Proc. ICASSP*, 2010, pp. 1798–1801.

[7] F. Alegre, R. Vipperla, N. Evans, and B. Fauve, "On the vulnerability of automatic speaker recognition to spoofing attacks with artificial signals," in *Proc. EUSIPCO*, 2012, pp. 36–40.

[8] P. Perrot, G. Aversano, R. Blouet, M. Charbit, and G. Chollet, "Voice forgery using alisp: Indexation in a client memory," in *Proc. ICASSP*, vol. 1, 2005, pp. 17–20.

[9] D. Matrouf, J.-F. Bonastre, and C. Fredouille, "Effect of speech transformation on impostor acceptance," in *Proc. ICASSP*, vol. 1, 2006.

[10] J.-F. Bonastre, D. Matrouf, and C. Fredouille, "Artificial impostor voice transformation effects on false acceptance rates," in *Proc. Interspeech*, 2007.

[11] Q. Jin, A. R. Toth, A. W. Black, and T. Schultz, "Is voice transformation a threat to speaker identification?" in *Proc. ICASSP*, 2008, pp. 4845–4848.

[12] T. Kinnunen, Z.-Z. Wu, K. A. Lee, F. Sedlak, E. S. Chng, and H. Li, "Vulnerability of speaker verification systems against voice conversion spoofing attacks: the case of telephone speech," in *Proc. ICASSP*, 2012, pp. 4401–4404.

[13] M. Blomberg, D. Elenius, and E. Zetterholm, "Speaker verification scores and accoustic analysis of a professional impersonator," in *Proc. FONETIK*, 2004.

[14] M. Farrús, M. Wagner, J. Anguita, and J. Hernando, "How vulnerable are prosodic features to professional imitators?" in *The Speaker and Language Recognition Workshop (Odyssey)*, 2008.

[15] S. Prince, P. Li, Y. Fu, U. Mohammed, and J. H. Elder, "Probabilistic models for inference about identity." *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 1, pp. 144–157, 2012.

[16] C. Soutar, R. Gilroy, and A. Stoianov, "Biometric system performance and security," in *Proc. IEEE AIAT*, 1999.

[17] A. Adler, "Sample images can be independently restored from face recognition templates," in *Proc. CCECE*, vol. 2, 2003, pp. 1163–1166.

[18] U. Uludag and A. K. Jain, "Attacks on biometric systems: a case study in fingerprints," in *Proc. SPIE-IE*, vol. 5306, no. 4, 2004, pp. 622–633.

[19] M. Martinez-Diaz, J. Fierrez, J. Galbally, and J. Ortega-Garcia, "An evaluation of indirect attacks and countermeasures in fingerprint verification systems," *Pattern Recognition Letters*, vol. 32, pp. 1643–1651, 2011.

[20] N. K. Ratha, J. H. Connell, and R. M. Bolle, "Enhancing security and privacy in biometrics-based authentication sytems," *IBM Systems Journal*, vol. 40, pp. 614–634, 2001.

[21] J. A. Markowitz, "Anti-spoof techniques for voice," biometric Consortium.

[22] P. L. De Leon, I. Hernaez, I. Saratxaga, M. Pucher, and J. Yamagishi, "Detection of synthetic speech for the problem of imposture," in *Proc. ICASSP*, 2011, pp. 4844–4847.

[23] F. Alegre, R. Vipperla, and N. Evans, "Spoofing countermeasures for the protection of automatic speaker recognition from attacks with artificial signals," in *Proc. Interspeech*, 2012.

[24] T. Satoh, T. Masuko, T. Kobayashi, and K. Tokuda, "Imposture using synthetic speech against speaker verification based on spectrum and pitch," in *Proc. Eurospeech*, vol. 2, 2001, pp. 302–305.

[25] C. C. Chibelushi, F. Deravi, and J. S. D. Mason, "A review of speech-based bimodal recognition," *IEEE Trans. on Multimedia*, vol. 4, no. 1, pp. 23–37, 2002.

[26] E. A. Rua, C. G. Mateo, H. Bredin, and G. Chollet, "Aliveness detection using coupled hidden markov models," in *Proc. Spanish Workshop on Biometrics*, 2007.

[27] Q. Jin, A. R. Toth, T. Schultza, and A. W. Black, "Voice covergin: Speaker de-identification by voice transformation," in *Proc. ICASSP*, 2009, pp. 3909–3912.

[28] BioAPI Consortium, "BioAPI specification (version 1.1)," March 2001, www.bioapi.org/Downloads/BioAPI

[29] M. Gomez-Barrero, J. Galbally, J. Fierrez, and J. Ortega-Garcia, "Face verification put to test: A hill-climbing attack based on the uphill-simplex algorithm," in *Proc. International Conference on Biometrics (ICB)*, 2012, submitted.

[30] ——, "Hill-climbing attack based on the uphill simplex algorithm and its application to signature verification," in *Proc. European Workshop on Biometrics and Identity Management (BioID)*. LNCS-6583, 2011, pp. 83–94.

[31] P. Kenny, "Bayesian Speaker Verification with Heavy-Tailed Priors," in *Odyssey: The Speaker and Language Recognition Workshop, Brno, Czech Republic,*, June 28 - July 1 2010.

[32] D. Garcia-Romero and C. Y. Espy-Wilson, "Analysis of I-Vector Length Normalization in Speaker Recognition Systems," in *INTERSPEECH*, 2011, pp. 249–252.

[33] M. Gomez-Barrero, J. Galbally, P. Tome-Gonzalez, and J. Fierrez, "On the vulnerability of iris-based systems to software attacks based on genetic algorithms," in *Proc. International Conference on Biometrics (ICB)*, 2012, submitted.

[34] National Institute of Standards and a. o. Technology, "The NIST Year 2010 Speaker Recognition Evaluation Plan," http://www.nist.gov/itl/iad/mig/upload/NIST_SRE10_evalplan-r6.pdf, 2010.

[35] N. Dehak, P. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet, "Front-End Factor Analysis for Speaker Verification," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 19, no. 4, pp. 788 – 798, February 2011.

[36] C. Cieri, L. Corson, D. Graff, and K. Walker, "Resources for New Research Directions in Speaker Recognition: The Mixer 3, 4 and 5 Corpora," Proc. Interspeech, 2007.

[37] 2001, ANSI X9.84-2001, Biometric Information Management and Security.

[38] J. Galbally, "Vulnerabilities and attack protection in security systems based on biometric recognition," Ph.D. dissertation, Universidad Autonoma de Madrid, 2009.