Improving Speaker Verification Performance in Presence of Spoofing Attacks Using Out-of-Domain Spoofed Data

Sarkar, Achintya Kumar; Sahidullah, Md; Tan, Zheng-Hua; Kinnunen, Tomi

Published in:
Interspeech 2017

DOI (link to publication from Publisher):
10.21437/Interspeech.2017-1758

Publication date:
2017

Document Version
Publisher's PDF, also known as Version of record

INTERSPEECH 2017
August 20–24, 2017, Stockholm, Sweden

Improving Speaker Verification Performance in Presence of Spoofing Attacks Using Out-of-Domain Spoofed Data

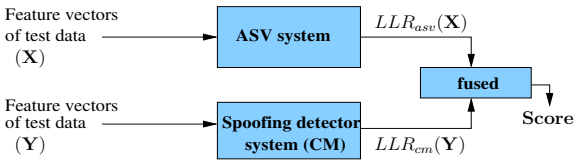
Achintya Kr. Sarkar[1], *Md. Sahidullah*[2], *Zheng-Hua Tan*[1] and *Tomi Kinnunen*[2]

[1]Department of Electronic Systems, Aalborg University, Denmark
[2]Speech and Image Processing Unit, School of Computing, University of Eastern Finland, Finland
akc@es.aau.dk, sahid@cs.uef.fi, zt@es.aau.dk, tkinnu@cs.uef.fi

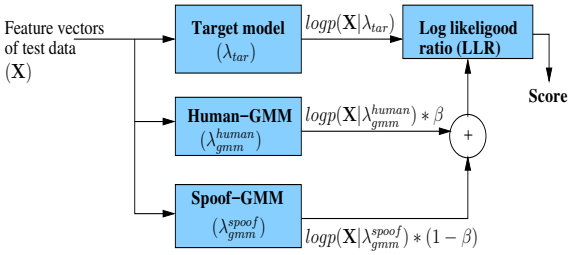
Abstract

Automatic speaker verification (ASV) systems are vulnerable to spoofing attacks using speech generated by voice conversion and speech synthesis techniques. Commonly, a countermeasure (CM) system is integrated with an ASV system for improved protection against spoofing attacks. But integration of the two systems is challenging and often leads to increased false rejection rates. Furthermore, the performance of CM severely degrades if in-domain development data are unavailable. In this study, therefore, we propose a solution that uses two separate background models – one from human speech and another from spoofed data. During test, the ASV score for an input utterance is computed as the difference of the log-likelihood against the target model and the combination of the log-likelihoods against two background models. Evaluation experiments are conducted using the joint ASV and CM protocol of ASVspoof 2015 corpus consisting of text-independent ASV tasks with short utterances. Our proposed system reduces error rates in the presence of spoofing attacks by using out-of-domain spoofed data for system development, while maintaining the performance for zero-effort imposter attacks compared to the baseline system.

Index Terms: Speaker verification, Spoofing, UBM, Cross-corpora



(a) ***Baseline***: *ASV system with spoofing detector*



(b) ***Proposed***: *Improved ASV system*

Figure 1: *Our baseline involves linear fusion of ASV and spoofing countermeasure that may use different features* $\mathbf{X}$ *and* $\mathbf{Y}$. *The proposed approach uses* shared *features with two background models whose contributions are traded off with parameter* β, *to tackle both zero-effort impostor and spoofing attacks.*

1. Introduction

In recent years, significant progress has been made in *automatic speaker verification* (ASV) technology, which now finds many real-world authentication applications, *e.g.* in physical and logical access control systems. But recent studies [1] show that *spoofing attacks* using artificial speech, generated by voice conversion (VC) or speech synthesis (SS) techniques, severely compromise the security of ASV systems regardless of the applied paradigms; degradations have been reported for *Gaussian mixture model – universal background model* (GMM-UBM) [2, 3, 4], *Hidden markov model* (HMM) [5] and *i-vector* [6, 7] based ASV systems, to mention a few.

In order to deal with spoofing attacks, an ASV system typically uses a *spoofing detector* module as a *countermeasure* (CM). Here, the task of a CM is to detect whether a speech signal is uttered by a human or is an artificial signal generated by VC or SS algorithms. The CM score is then combined with that of the ASV system to make a joint decision on a test speech signal. During the past few years, a number of CMs have been proposed to discriminate synthetic speech from real human voice [8, 9]. Mostly, these CMs are implemented as additional modules to ASV systems (as in Fig. 1a), demanding careful, potentially challenging, design considerations, such as the selection or optimization of front-end features that might differ from those used in the ASV system. Moreover, most spoofing detectors do not generalize well to unseen conditions including unknown types of attacks, data and environments, which collectively define a *domain*. For instance, a recent study [10] reported considerably increased error rates for various countermeasures when evaluated on another corpus (out-of-domain data).

On the basis of these observations, this work aims to provide a simpler and generalized solution to the artificial speech spoofing detection problem by enhancing ASV tolerance against spoofing attacks without a dedicated additional spoofing detector. Our idea is to enhance a conventional GMM-based ASV system so that it inherently rejects a spoofed speech trial as an impostor. A related prior work [7] with a similar goal tackled the problem using an i-vector approach to handle artificial speech spoofing attacks in long utterances using *probabilistic linear discriminant analysis* (PLDA) model trained on natural

This work is partly supported by the OCTAVE Project (#647850), funded by the Research European Agency (REA) of the European Commission, in its framework programme Horizon 2020. The views expressed in this paper are those of the authors and do not engage any official position of the European Commission. The work of Sarkar and Tan is also partly supported by the iSocioBot project, funded by the Danish Council for Independent Research - Technology and Production Sciences (#1335-00162).

http://dx.doi.org/10.21437/Interspeech.2017-1758

and artificial speech data. In contrast, we focus on the GMM-UBM framework better suited for short utterances [11, 12]. Our idea (Fig. 1b) is straightforward: in addition to the 'conventional' UBM, trained on human speech, we use an additional UBM trained on spoofed speech. We formulate and extensively experiment with various flavors of this key idea.

We conduct ASV experiments on the ASVspoof 2015 corpus [13] consisting of human speech and spoofed speech generated by different VC and SS methods, and address in particular the issue of out-of-domain training. To this end, we use the IDIAP-AVspoof database [14] as our out-of-domain training data and evaluate the countermeasures on ASVspoof 2015. The IDIAP-AVspoof database is considered out-of-domain as it differs from the ASVspoof 2015 data in terms of both human speech recording settings and spoofing methods. Our study reveals that ASV performance in the presence of spoofing attacks can be considerably boosted by appropriate data engineering without building a separate explicit spoofing detector. In fact, the proposed solution outperforms the combination of an ASV system and a spoofing detector.

2. ASV in Presence of Spoofing Attacks

2.1. Conventional GMM-UBM based ASV Systems

Previous studies [11, 12] indicate that traditional GMM-UBM [15] can outperform i-vector based speaker verification with short utterances. Therefore, earlier work on ASV with spoofed speech [2, 3, 4, 16] used GMM-UBM based systems for performance evaluation. We also use this technique where speaker models are adapted from a UBM with *maximum-a-posteriori* (MAP) adaptation using target speaker's training data [15]. During test, log-likelihood ratio (LLR) for a test utterance $\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_L\}$ is computed using the target model λ_{tar} and the background model $\lambda_{\text{ubm}}^{\text{asv}}$ as follows:

$$\text{LLR}_{\text{asv}}(\mathbf{X}) = \frac{1}{L}\big\{ \log p(\mathbf{X}|\lambda_{\text{tar}}) - \log p(\mathbf{X}|\lambda_{\text{ubm}}^{\text{asv}})\big\} \quad (1)$$

The LLR value is used for decision making. Typically, the background model, $\lambda_{\text{ubm}}^{\text{asv}}$, is trained with real human speech.

2.2. GMM-UBM system fused with a spoofing detector

To improve the performance of ASV systems in presence of spoofing attacks, a standalone spoofing detector countermeasure, or CM for short, is *integrated* with ASV. The task of the CM is to detect the non-human or spoofed speech as *impostor*. In literature [8, 17], two types of integration are considered. The first one is decision fusion where a test utterance is accepted only when it is jointly approved by both ASV and CM systems. In the second approach, recognition scores from the two systems are combined to obtain a final score for decision making. We adopt the former approach illustrated in Figure 1a, requiring only a single decision threshold. We use a two-class GMM-based method [18, 19] using two separate GMMs [20] trained independently with real human and spoofed data. In test, for a given test utterance with CM features as $\mathbf{Y} = \{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_L\}$, LLR value is calculated between the human and spoof speech GMM as,

$$\text{LLR}_{\text{cm}}(\mathbf{Y}) = \frac{1}{L}\big\{ \log p(\mathbf{Y}|\lambda_{\text{gmm}}^{\text{human}}) - \log p(\mathbf{Y}|\lambda_{\text{gmm}}^{\text{spoof}})\big\} \quad (2)$$

Here $\text{LLR}_{\text{cm}}(\mathbf{Y})$ is the countermeasure score used for detecting whether an utterance is spoofed or human speech irrespective of the target speaker. In general, different front-end features are used for spoofing detection and ASV [8]. The output score for a given test utterance in the combined ASV with CM system can be expressed as,

$$\alpha \times \text{LLR}_{asv}(\mathbf{X}) + (1-\alpha) \times \text{LLR}_{cm}(\mathbf{Y}), \quad (3)$$

which, using Eqs. (1) and (2), can be rewritten as,

$$= \frac{1}{L}\big[\, \alpha \log\ p(\mathbf{X}|\lambda_{\text{tar}}) - \{\alpha\ \log\ p(\mathbf{X}|\lambda_{\text{ubm}}^{\text{asv}}) \quad (4)$$
$$-(1-\alpha) \log\ p(\mathbf{Y}|\lambda_{\text{gmm}}^{\text{human}})\} - (1-\alpha)\ \log\ p(\mathbf{Y}|\lambda_{\text{gmm}}^{\text{spoof}})\big]$$

Here, $\alpha \in [0, 1]$ is a fusion weight. We observe that

- The special cases $\alpha = 0$ and $\alpha = 1$ correspond to, respectively, standalone CM and baseline ASV without any CM protection.
- In the special case of *shared* features for ASV and CM ($\mathbf{X} = \mathbf{Y}$) and shared human data used for ASV background and CM training ($\lambda_{\text{ubm}}^{\text{asv}} = \lambda_{\text{gmm}}^{\text{human}}$), Eq. (4) becomes,

$$\frac{1}{L}\big[\, \alpha \log\ p(\mathbf{X}|\lambda_{\text{tar}}) - (2\alpha - 1) \log\ p(\mathbf{X}|\lambda_{\text{gmm}}^{\text{human}}) \quad (5)$$
$$-(1-\alpha)\ \log\ p(\mathbf{X}|\lambda_{\text{gmm}}^{\text{spoof}})\big]$$

The joint system can be viewed as a speaker verification system with two background models with different weights for human and spoof UBMs and here the target speaker model is estimated from human data.

2.3. Proposed ASV system

Inspired by Eq. (5), we propose a *two-UBM* based ASV system, where one UBM is trained using human and another using spoofed speech. In enrollment, target speaker models are still MAP-adapted from the human UBM, $\lambda_{\text{gmm}}^{\text{human}}$, but the test LLR score computation involves combination of both UBMs scores. A control parameter β trades off the two UBM scores, leading to the following LLR score:

$$\text{LLR}(\mathbf{X}) = \frac{1}{L}\big\{ \log p(\mathbf{X}|\lambda_{\text{tar}}) - \quad (6)$$
$$\big[\, \beta \times \log p(\mathbf{X}|\lambda_{\text{gmm}}^{\text{human}}) + (1-\beta) \times \log p(\mathbf{X}|\lambda_{\text{gmm}}^{\text{spoof}})\big]\,\big\}$$

This system can be viewed as a joint spoofing and human background model for ASV that uses a shared feature space, instead of the approach combining two separate subsystems (an ASV system and a CM system that use different features as presented in Sec.2.2). For $\beta = 1$, it reduces to the standalone ASV system as in Eq. (1), while for the other extreme $\beta = 0$, it acts as a two-GMM based ASV system using the spoof UBM for background normalization:

$$\text{LLR}(\mathbf{X}) = \frac{1}{L}\big\{ \log p(\mathbf{X}|\lambda_{\text{tar}}) - \log p(\mathbf{X}|\lambda_{\text{gmm}}^{\text{spoof}})\big\} \quad (7)$$

The interesting values, however, are $0 < \beta < 1$. The proposed method always holds the concept of speaker verification by inherently accounting for the two negative hypotheses together in standalone ASV frameworks and will be able to reject more spoofing impostors than the conventional standalone ASV. We consider two alternative approaches for choosing this additional control parameter. In **System 1**, β is computed 'on-the-fly' for a given input, based on the likelihood of the test feature vectors scored against human and spoof UBMs,

$$\beta(\mathbf{X}) = \frac{p(\mathbf{X}|\lambda_{\text{gmm}}^{\text{human}})}{p(\mathbf{X}|\lambda_{\text{gmm}}^{\text{human}}) + p(\mathbf{X}|\lambda_{\text{gmm}}^{\text{spoof}})} \quad (8)$$

Table 1: *Interpretation of the control parameters for baseline and proposed integrated system.*

Method	Control parameter	Interpretation of the task
Baseline	$\alpha = 0$	Speaker verification suitable for spoofed impostors only.
	$\alpha = 1$	Speaker verification suitable for zero-effort impostors only.
	$0 < \alpha < 1$	Speaker verification suitable for both impostors but constrained control over the trade-offs
Proposed	$\beta = 0$	Speaker verification suitable for spoofed imposters and to some extent for zero-effort impostors.
	$\beta = 1$	Speaker verification suitable for zero-effort impostors.
	$0 < \beta < 1$	Speaker verification suitable for both impostors but with flexible control to trade-off of the two cases.

Hence, $\beta(\mathbf{X})$ emphasizes the score of the particular UBM according to the test data. In contrast, in **System 2**, we optimize β using a disjoint development set by linear grid search over $[0, 1]$ and fix this optimized value on future data.

Table 1 summarizes the functionality of the baseline and the proposed ASV systems for different value of their respective control parameters.

3. Experimental setup

Experiments are conducted on the joint ASV and CM protocol of the ASVspoof 2015 corpus [8, 13] consisting of 81 target speakers (35 male and 46 female). Each speaker has five utterances for enrollment. The joint protocol has two types of non-target trials: *zero-effort human impostors* (Z) and *spoof impostors* (S) (speech generated by SS and VC). Table 2 shows the trial statistics where G stands for human (genuine) trials.

Table 2: *Number of trials for ASV experiments on development (dev) and evaluation (eval) sets for ASVspoof2015.*

	Male			Female		
	G	Z	S	G	Z	S
Dev.	1498	4275	21375	1999	5700	28500
Eval.	4053	8000	80000	5351	10400	104000

The **baseline** ASV system is implemented with gender-dependent UBM of 512 Gaussians trained on the data from the IDIAP-AVSpoof database [14]. The target models are enrolled using MAP adaptation process with relevance factor of 3. The spoofing detector is also trained on the same corpora, i.e., IDIAP-AVSpoof. In total, it consists of 10887 male and 4661 female speech files for human; and 47000 male and 8255 female spoofed files. For the **proposed** ASV, we train the UBMs separately on human and spoofed speech. Since the main experiments are conducted on ASVspoof 2015 and spoofing detector is trained on AVSpoof, our evaluation follows the cross-copora evaluation strategy for countermeasures [10].

For ASV, we use 57-dimensional MFCCs (19 static+Δ+$\Delta\Delta$) computed from speech frames using 20ms Hamming window with 10ms overlap. We have performed RASTA [21] filtering, energy-based speech activity detection (SAD) and cepstral mean and variance normalization (CMVN). For the spoofing detector, we use 40-dimensional MFCCs (20 Δ+$\Delta\Delta$) as used in [10]. For the proposed ASV system, we also use the same 57-dimensional MFCCs as baseline ASV.

4. Results and Discussion

4.1. Baseline ASV performance under spoofing attacks

Table 3 shows the ASV performance under spoofing attacks of different types (human, spoof). As expected, EERs of ASV are generally much higher under spoof impostor attacks than under zero-effort impostors. EER of spoof impostors is higher in the *real speech UBM* based ASV system than the other. This is expected as this UBM has no information about the attributes of spoofed data and therefore is unable to reject spoofing impostors. *Mix1* (a UBM trained from pooled human and spoof data features) and *Mix2* (a UBM trained by first training human and spoof GMMs, each with 256 Gaussians, followed by merging the Gaussians and mixing weight re-normalization) cases show lower EERs specifically for the spoof impostors in contrast to the *real speech UBM* system, as their negative hypothesis, i.e. UBM has been built using spoofed data and hence captured some attributes of the spoofed speech. The performance of *Mix1* and *Mix2* systems is similar in general, as might be expected. In the remaining experiments, we consider *Mix1* as a baseline.

Table 3: *Effect of different datasets in building UBM on ASV performance (%EER) of development set of ASVspoof 2015 database.* Real*: human speech,* Mix1*: feature pooling,* Mix2*: Gaussian pooling to create a UBM. See text for details.*

UBM	Trn. data		Male		Female	
	Human	Spoof	Z	S	Z	S
Real	✓	×	6.99	36.99	11.29	31.81
Mix1	✓	✓	7.27	32.14	10.42	29.54
Mix2	✓	✓	6.80	33.87	10.49	29.86

4.2. Proposed ASV systems under spoofing attacks

We first study the performance of the proposed methods for different values of β on the development set. For simplicity, we present the ASV performance of **System 2** for different values of β (for male speakers) in Fig. 2.

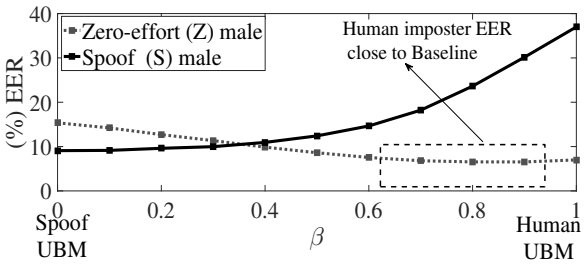


Figure 2: *Effect of β on ASV performance (%EER) for zero-effort and spoofing conditions with the proposed* ***System 2*** *on* **development set** *of ASVspoof2015.*

By increasing the value of β, which decreases the contribution of the spoof UBM, the EER of the real human zero-effort impostors (Z) decreases while the EER of the spoofing impostors (S) increases. The range $\beta \in [0.7, 0.9]$ yields closer performance for the Z impostor to the baseline ASV system. At the same time, EER of the spoof impostors reduces remarkably. From now onwards, we use $\beta = 0.7$ for the rest of the experiments with proposed **System 2**.

Table 4 compares the ASV performance on the evaluation set using the optimal value $\beta = 0.7$ obtained from the devel-

Table 4: *Performance comparison (in terms of %EER) of the proposed method with baseline for ASV under spoofing in ASVspoof 2015 database.*

System	Male		Female		Avg.
	Z	S	Z	S	EER
Development set:					
Baseline	7.27	32.14	10.42	29.54	19.84
Combined ASV	6.47	27.23	9.85	26.26	**17.45**
Prop. System 1	7.81	28.10	12.40	27.26	**18.89**
Prop. System 2 $\beta = 0.0$	15.36	9.08	25.56	20.61	**17.65**
Prop. System 2 $\beta = 0.5$	8.61	12.41	14.95	18.04	**13.50**
Prop. System 2 $\beta = 0.7$	6.80	18.21	11.36	19.95	**14.08**
Prop. System 2 $\beta = 1.0$	6.99	36.99	11.29	31.81	21.77
Evaluation set:					
Baseline	9.18	32.44	8.65	24.76	18.75
Combined ASV	9.03	28.05	7.10	20.18	**16.09**
Prop. System 1	9.67	28.81	9.70	21.95	**17.53**
Prop. System 2 $\beta = 0.0$	17.56	12.80	23.14	16.52	**17.50**
Prop. System 2 $\beta = 0.5$	10.65	15.32	12.28	14.07	**13.08**
Prop. System 2 $\beta = 0.7$	8.41	20.20	8.51	15.30	**13.10**
Prop. System 2 $\beta = 1.0$	8.47	36.68	9.28	26.68	20.27

opment set as well as for other two boundary cases for $\beta = 0$ and $\beta = 1$. In addition, we show results for a combined ASV system with a linear combination of two separate ASV systems – one trained with human speech as UBM and another with spoofed speech as UBM. The motivation to use this system is to observe whether the additional speaker model with respect to spoof GMM further helps the ASV under spoofing attacks or not. Similar to **System 2**, we determine the linear fusion weight using the development set.

From Table 4, we can see that the proposed method 1, 2 shows lower/comparable EER values for human impostor (Z) with compared to the baseline. However, the proposed methods shows much EER reduction for the spoof. Similar phenomena is observed on evaluation set. The error rate in **System 1** for spoof impostor is quite higher than the **System 2**. It could be due to the better optimization of β using data driven approach on development in compared to likelihood proportion of two UBMs. Overall observations indicate that the proposed method is very useful for the ASV system in real-life specially when system has no priori knowledge about the test spoof (i.e out-domain data set).

4.3. Comparison of the baseline with the best proposed method with/without an out-of-domain spoofing detector

In this section, we compare the performance of the baseline with the best proposed ASV system (**System 2** for $\beta = 0.7$) with or without a spoofing detector under spoofing attacks on ASVspoof 2015. We show the comparative performance in Fig. 3 for different values of fusion weight α (for simplicity in male speakers) on the development set. From Fig. 3, we observe that with lower value of α, i.e. the joint system gets more contribution from CM than ASV. As a result, EER value for the human impostor increases, and vice versa for the spoofed impostor. In Table 5, we show the performance of the baseline and the proposed **System 2** for different values of fusion weight on evaluation data. Fusion weight equal to 1.0 indicates that only the ASV system is used whereas a weight of 0.0 means the combined system acts as a spoofing detector only. We observe that the proposed system gives reasonable improvement for spoofed trials with $\alpha = 1.0$, i.e., even without using any spoofing detector. For fusion of ASV and CM, we have found an optimum value of α as 0.9 on the development set for both male and

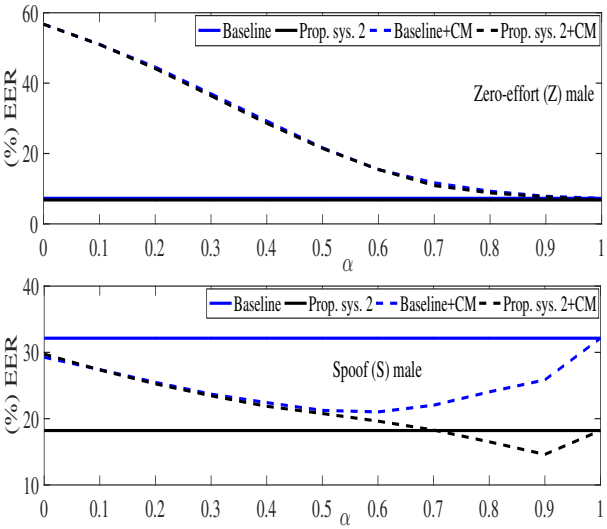


Figure 3: *Comparison performance (%EER) of standalone baseline ASV with the best proposed method with or without out-domain spoofing detector i.e. CM on development set of ASVspoof2015.*

Table 5: *Performance comparison (in terms of %EER) of the proposed ASV method with baseline ASV by combining CM on evaluation set of ASVspoof 2015 database. ASV and CM systems are combined using score fusion with fusion weight α as discussed in Section 2.2.*

ASV System	α	Male		Female		Avg.
		Z	S	Z	S	EER
Baseline	1.0	9.18	32.44	8.65	24.76	18.75
	0.0	52.75	24.62	51.14	27.30	38.95
	0.5	19.83	20.53	22.50	21.23	21.02
	0.9	9.47	27.83	9.00	20.29	16.64
Prop. System 2	1.0	**8.41**	20.20	8.51	15.30	13.10
	0.0	52.75	24.62	51.14	27.30	38.95
	0.5	19.56	19.99	20.91	20.82	20.32
	0.9	8.61	**17.91**	**8.50**	**14.37**	**12.34**

female speakers. This also gives best ASV performance with spoofed impostors on the evaluation set. The performance is also improved for human impostor trials in most cases for fused mode. Further investigations are required to study the generalization capability of the proposed ASV system in presence of more challenging spoofing attack such as replay.

5. Conclusion

Improving ASV performance in the presence of spoofing attacks is an open research problem, especially when matched (in-domain) data to train countermeasures is unavailable. To address this problem, we proposed a GMM-UBM based ASV system consisting of *two* UBMs, one trained using human speech and another using spoofed data. Our experiments on the ASVspoof 2015, using another out-of-domain data (IDIAP-AVspoof) for training, indicate that the proposed method is able to considerably reduce the EER for spoofing impostors compared to the baseline with or without a spoofing detector, *without compromising the performance under zero-effort spoofing*. The proposed method, presenting a simple alternative to dedicated countermeasures trained on custom features, holds promise.

6. References

[1] Z. Wu, N. Evans, T. Kinnunen, J. Yamagishi, F. Alegre, and H. Li, "Spoofing and Countermeasures for Speaker Verification: A Survey," *Speech Communication*, vol. 66, pp. 130–153, 2015.

[2] J.F. Bonastre, D. Matrouf, and C. Fredouille, "Artificial Impostor Voice Transformation Effects on False Acceptance Rates," in *Proc. of Interspeech*, 2007, pp. 2053–2056.

[3] F. Alegre, R. Vipperla, N. Evans, and B. Fauve, "On the Vulnerability of Automatic Speaker Recognition to Spoofing Attacks with Artificial Signals," in *Proc. of Eur. Conf. Speech Commun. and Tech. (Eurospeech)*, 2012, pp. 36–40.

[4] Z. Wu and H. Li, "Voice Conversion Versus Speaker Verification: An Overview," *APSIPA Transactions on Signal and Information Processing*, vol. 3(e17), 2014.

[5] J. Lindberg, M. Blomberg, et al., "Vulnerability in Speaker Verification - A Study of Technical Impostor Techniques," in *Proc. of European Conference on Speech Communication and Technology*, 1999, pp. 1211–1214.

[6] E. Khoury, T. Kinnunen, A. Sizov, Z. Wu, and S. Marcel, "Introducing I-vectors for Joint Anti-Spoofing and Speaker Verification," in *Proc. of Interspeech*, 2014, pp. 61–65.

[7] A. Sizov, E. Khoury, T. Kinnunen, Z. Wu, and S. Marcel, "Joint speaker verification and antispoofing in the i -vector space," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 4, pp. 821–832, 2015.

[8] M. Sahidullah, H. Delgado, M. Todisco, H. Yu, T. Kinnunen, N. Evans, and Z.-H. Tan, "Integrated spoofing countermeasures and automatic speaker verification: an evaluation on asvspoof 2015," *Interspeech 2016*, pp. 1700–1704, 2016.

[9] H. Yu, Z.-H. Tan, Z. Ma, and J. Guo, "DNN Filter Bank Cepstral Coefficients for Spoofing Detection," *IEEE Access*, 2017.

[10] P. Korshunov and S. Marcel, "Cross-Database Evaluation of Audio-Based Spoofing Detection Systems," in *Proc. of Interspeech*, 2016, pp. 1705–1709.

[11] A. Larcher, K. A. Lee, B. Ma, and H. Li, "Text-dependent Speaker Verification: Classifiers, Databases and RSR2015," *Speech Communication*, vol. 60, pp. 56–77, 2014.

[12] H. Delgado, M. Todisco, M. Sahidullah, A. K. Sarkar, N. Evans, T. Kinnunen, and Z.-H. Tan, "Further Optimisations of Constant Q Cepstral Processing for Integrated Utterance and Text-dependent Speaker Verification," in *Proc. of Spoken Language Technology Workshop (SLT)*, 2016, pp. 179–185.

[13] Z. Wu, T. Kinnunen, N. Evans, J. Yamagishi, C. Hanilci, M. Sahidullah, and A. Sizov, "ASVspoof 2015: the First Automatic Speaker Verification Spoofing and Countermeasures challenge," in *Proc. of Interspeech*, 2015.

[14] S. K. Ergunay, E. Khoury, A. Lazaridis, and S. Marcel, "On the Vulnerability of Speaker Verification to Realistic Voice Spoofing," in *Proc. of Int. Conf. on Biometrics: Theory, Applications and Systems (BTAS)*, 2015, pp. 1–6.

[15] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker Verification using Adapted Gaussian Mixture Models," *Digital Signal Processing*, vol. 10, pp. 19–41, 2000.

[16] T. Kinnunen, 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. of IEEE Int. Conf. Acoust. Speech Signal Processing (ICASSP)*, 2012, pp. 4401–4404.

[17] I. Chingovska, A.R.d. Anjos, and S. Marcel, "Biometrics Evaluation Under Spoofing Attacks," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 12, pp. 2264–2276, 2014.

[18] M. Sahidullah, T. Kinnunen, and C. Hanilci, "A Comparison of Features for Synthetic Speech Detection," in *Proc. of Interspeech*, 2015.

[19] H. Yu, A. Sarkar, D. A. L. Thomsen, Z.-H. Tan, Z. Ma, and J. Guo, "Effect of Multi-condition Training and Speech Enhancement Methods on Spoofing Detection," in *Proc. of International Workshop on Sensing, Processing and Learning for Intelligent Machines (SPLINE)*, 2016, pp. 1–5.

[20] D. A. Reynolds, "Speaker Identification and Verification using Gaussian Mixture Speaker Models," *Speech Communication*, vol. 17, pp. 91–108, 1995.

[21] H. Hermanksy and N. Morgan, "RASTA Processing of Speech," *IEEE Trans. on Speech and Audio Processing*, vol. 2, pp. 578–589, 1994.