2017

Generalization of spoofing countermeasures: A case study with ASVspoof 2015 and BTAS 2016 corpora

Paul, Dipjyoti

Institute of Electrical and Electronics Engineers (IEEE)

GENERALIZATION OF SPOOFING COUNTERMEASURES: A CASE STUDY WITH ASVSPOOF 2015 AND BTAS 2016 CORPORA

Dipjyoti Paul[1]*, Md Sahidullah*[2]*, Goutam Saha*[1]

[1]Department of E & ECE, Indian Institute of Technology Kharagpur, Kharagpur, India
[2]School of Computing, University of Eastern Finland, Joensuu, Finland
e-mail: dipjyotipaul@ece.iitkgp.ernet.in, sahid@cs.uef.fi, gsaha@ece.iitkgp.ernet.in

ABSTRACT

Voice-based biometric systems are highly prone to spoofing attacks. Recently, various countermeasures have been developed for detecting different kinds of attacks such as replay, speech synthesis (SS) and voice conversion (VC). Most of the existing studies are conducted with a specific training set defined by the evaluation protocol. However, for realistic scenarios, selecting appropriate training data is an open challenge for the system administrator. Motivated by this practical concern, this work investigates the generalization capability of spoofing countermeasures in restricted training conditions where speech from a broad attack types are left out in the training database. We demonstrate that different spoofing types have considerably different generalization capabilities. For this study, we analyze the performance using two kinds of features, mel-frequency cepstral coefficients (MFCCs) which are considered as baseline and recently proposed constant Q cepstral coefficients (CQCCs). The experiments are conducted with standard Gaussian mixture model - maximum likelihood (GMM-ML) classifier on two recently released spoofing corpora: ASVspoof 2015 and BTAS 2016 that includes cross-corpora performance analysis. Feature-level analysis suggests that static and dynamic coefficients of spectral features, both are important for detecting spoofing attacks in the real-life condition.

Index Terms— Spoofing Attack, Replay Attack, ASVspoof 2015, BTAS 2016, Generalized countermeasure.

1. INTRODUCTION

Spoofing attacks imitate a person's identity in order to gain illegitimate access to sensitive or protected resources. Nowadays, significant advancement in speech technology related to SS and VC techniques poses threat to speech-based biometric systems like *automatic speaker verification* (ASV) systems [1]. *Replay attacks* are another form of spoofing attack, where an adversary tries to attack a system using pre-recorded speech accumulated from target speakers [2]. Due to the availability of high-quality, low-cost recording and playback devices, replay attacks are also a serious threat to the voice biometric systems. Several replay spoofing detection approaches such as fixed pass-phrase method, spectral ratio

and modulation index were proposed in [3–5]. A study on cross database evaluation was demonstrated in [6].

To detect SS and VC attacks, diverse range of feature extraction methods such as *mel-frequency cepstral coefficients* (MFCCs) cepstral feature [7], phase features [8–10], a combination of both amplitude and phase feature [11], prosodic features [12] were reported. A concise experimental review of spoofing detection was presented in [13]. While, MFCCs are considered as the standard feature extraction techniques in speech processing, *constant Q transform cepstral coefficients* (CQCCs) have shown best detection performance, especially for unknown attacks in ASVspoof 2015 corpus [14]. However, it was not implemented for replay attack detection.

Techniques to generate voice converted speech and synthetic speech, made a rapid progress in recent times. Notable among them are *joint density-Gaussian mixture model* (JD-GMM) [15], *line spectrum pairs* (LSP) [16], MARY *text-to-speech synthesis* (MARY-TTS) [17], *hidden Markov model* (HMM) based TTS [18] etc. It is not practically possible to anticipate the kind of SS and VC attack all the time to include those types of speeches in the training database. At the same time, it is expected that detection performance will degrade if similar kinds of data are unavailable in the training corpus. The previous studies on spoofing detection do not focus on *attack dependency* which is the central theme of this work. There are some results generated in recent spoofing challenges with *unknown attack* types but no exhaustive study is done that can lead to *generalization* ability of certain training schemes over other for a range of unknown attacks.

In this work, we did a systematic study of attack dependency to discover corresponding generalization ability. We demonstrate the result using conventional MFCCs and newly proposed CQCCs features on GMM-maximum likelihood (GMM-ML) framework. It is found that GMM-ML as a classifier is better suited for spoofing detection task [19]. We have experimented on two recent databases: ASVspoof 2015, developed as a part of *Automatic Speaker Verification Spoofing and Countermeasure Challenge* [20] and BTAS 2016 corpus in *Speaker Anti-spoofing Competition* [21]. BTAS 2016 introduces more realistic replay attacks compared to ASVspoof
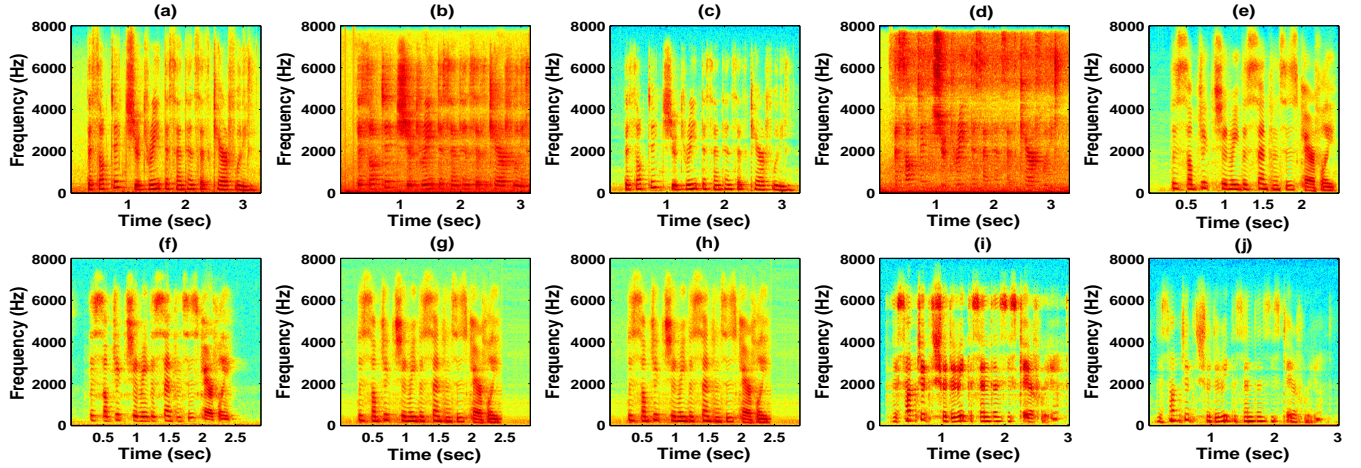
Fig. 1: Spectrogram of (a) genuine and replay speech signals for same sentence *"The subject should read the sentences carefully"*. The replayed signals are generated by using techniques based on (b) replay laptop, (c) replay laptop high quality, (d) replay phone, (e) SS, (f) replay SS, (g) replay SS high quality, (h) VC, (i) replay VC and (j) replay VC high quality.

database. Our study shows the generalization ability of one countermeasure over the other.

2. GENERALIZATION FRAMEWORK

Figure 1 illustrates the spectral characteristics of spoofed signals for diverse attacks. *Generalized countermeasure* refers to the ability to overcome the attack dependency in the detection process. This dependency signifies the types of attacks that are best represented by a similar pattern in the attack space. It involves a prior knowledge of attack type, which is not a realistic assumption in all the cases. Therefore, the countermeasure system needs to be robust enough to detect an attack even though that type of attack data is not used for training the model. Fig. 2 describes the functional block diagram of a generalized countermeasure framework where we find which kind of training has greater generalization ability.
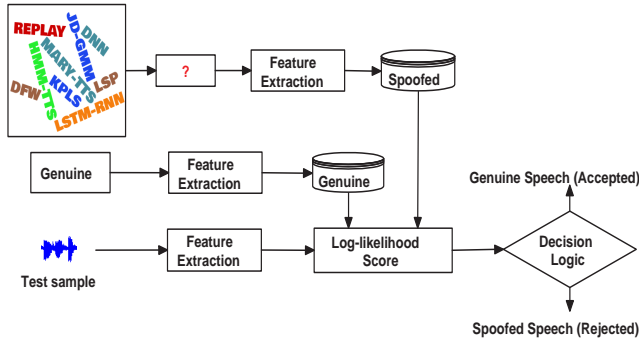

Fig. 2: A speech-based countermeasure system.

Initially, we train the models using all types of replay (i.e., genuine, SS and VC samples) and synthetic (SS and VC samples) attacks. Then, we study the impact when one type of spoofing data is not used for modeling the spoofed data.

3. EXPERIMENTAL SETUP

3.1. Database Description

ASVspoof 2015: ASVspoof database is created to assess ten different types of SS and VC synthetic speech samples

namely S1 to S10 [20]. It includes both the known attacks (S1-S5) and the unknown attacks (S6-S10).

BTAS 2016: BTAS database contains genuine and different kinds of replay attacks where genuine, SS and VC speech samples were played back using high-quality devices. Two new replay unknown attacks (R9 and R10) are introduced in the evaluation data to make it more challenging. The statistics regarding types of attacks and the number of utterances for each dataset are presented in Table 1.

Table 1: Number of utterances in BTAS 2016 database. LP: laptop, HQ: high quality speaker, PH1: Samsung Galaxy S4 phone, PH2: iPhone 3GS and PH3 is iPhone 6S.

Types			Training	Development	Evaluation
Genuine			4973	4995	5576
Replay	Replay LP LP	R1	700	700	800
	Replay LP HQ LP	R2	700	700	800
	Replay PH1 LP	R3	700	700	800
	Replay PH2 LP	R4	700	700	800
	Replay PH2 PH3	R9	-	-	800
	Replay LP PH2 PH3	R10	-	-	800
SS	SS LP LP	R5	490	490	560
	SS LP HQ LP	R6	490	490	560
VC	VC LP LP	R7	17400	17400	19500
	VC LP HQ LP	R8	17400	17400	19500

3.2. Feature Extraction Techniques

Mel-frequency cepstral coefficients (MFCCs): MFCC [22] feature utilizes mel-scale based triangular filter bank. The power spectrum is integrated using overlapping band-pass filters in the triangular filterbank. We use the configuration reported in [13].

Constant Q cepstral coefficients (CQCCs): The constant Q transform (CQT) gives a higher frequency resolution in lower frequencies and a greater temporal resolution in the higher frequency region. A spline interpolation method is applied to resample the geometric frequency scale into a uniform linear scale in order to apply linearly spaced DCT coefficients for CQCC cepstral feature computation [14].

CQCC feature is implemented with maximum frequency ($f_{max} = 4Khz$) and minimum frequency of ($f_{min} = 15Hz$). The number of bins per octave is assigned to 96.

Table 2: Performance (in % of EER) for MFCC and CQCC features on BTAS 2016 development data. The corresponding class of attacks that are not considered in the training are highlighted.

Features	Train Replay	Train SS	Train VC	Type	R1	R2	R3	R4	R5	R6	R7	R8	Average Replay	Average SS	Average VC	Average All
MFCC	✓	✓	✓	Static	0.14	0.61	0.09	0.00	0.00	0.84	0.00	0.02	0.21	0.42	0.01	0.21
				Static+$\Delta\Delta^2$	0.34	2.74	0.00	0.00	0.00	0.57	0.00	0.01	0.77	0.29	0.01	0.46
				$\Delta\Delta^2$	19.50	41.35	28.62	28.67	1.33	0.86	0.00	0.01	29.54	1.10	0.01	15.04
	×	✓	✓	Static	4.91	5.92	30.33	24.56	0.00	1.07	0.01	0.01	16.43	0.54	0.01	**8.35**
				Static+$\Delta\Delta^2$	4.73	6.98	30.51	26.29	0.00	0.65	0.00	0.00	17.13	0.33	0.01	8.65
				$\Delta\Delta^2$	25.44	44.39	34.34	35.52	1.33	1.10	0.00	0.01	34.92	1.23	0.01	17.77
	✓	×	✓	Static	0.27	0.63	0.00	0.00	0.00	1.93	0.00	0.02	0.23	0.97	0.01	0.36
				Static+$\Delta\Delta^2$	0.31	3.50	0.00	0.00	0.00	2.07	0.00	0.01	0.95	1.04	0.01	0.74
				$\Delta\Delta^2$	19.63	40.88	27.35	27.99	2.39	1.62	0.00	0.01	28.96	2.01	0.01	14.98
	✓	✓	×	Static	0.00	0.30	0.00	0.00	0.00	0.04	0.32	2.13	0.15	0.02	1.23	0.35
				Static+$\Delta\Delta^2$	0.04	0.08	0.00	0.00	0.00	0.02	0.55	0.85	0.03	0.01	0.70	0.19
				$\Delta\Delta^2$	2.44	24.05	5.80	2.69	0.14	0.85	3.60	6.30	8.75	0.50	4.95	5.73
CQCC	✓	✓	✓	Static	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.00**
				Static+$\Delta\Delta^2$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
				$\Delta\Delta^2$	20.04	6.43	25.22	41.05	1.10	0.11	0.09	0.18	23.19	0.61	0.14	11.78
	×	✓	✓	Static	1.27	0.00	47.71	40.84	0.00	0.00	0.00	0.00	22.46	0.00	0.00	11.23
				Static+$\Delta\Delta^2$	8.25	0.00	49.18	44.75	0.00	0.00	0.00	0.00	25.55	0.00	0.00	12.77
				$\Delta\Delta^2$	25.07	9.36	33.72	43.77	1.03	0.23	0.09	0.19	27.98	0.63	0.14	14.18
	✓	×	✓	Static	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.00**
				Static+$\Delta\Delta^2$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
				$\Delta\Delta^2$	20.31	6.48	25.54	40.93	2.26	0.20	0.08	0.18	23.32	1.23	0.13	12.00
	✓	✓	×	Static	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	**0.00**
				Static+$\Delta\Delta^2$	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
				$\Delta\Delta^2$	10.52	2.71	18.08	29.93	0.41	0.30	2.46	2.89	15.31	0.36	2.68	8.37

Table 3: Performance (in % of EER) for MFCC and CQCC (static) features on BTAS 2016 evaluation data. The corresponding class of attacks that are not considered in the training systems are highlighted.

Features	Train Replay	Train SS	Train VC	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	Average Replay	Average SS	Average VC	Average All
MFCC	✓	✓	✓	1.11	8.26	0.01	0.14	0.06	3.67	0.03	1.68	17.87	10.85	6.37	1.87	0.86	4.37
	×	✓	✓	6.61	16.12	32.89	24.38	0.14	5.34	0.06	2.88	30.50	40.89	25.23	2.74	1.47	15.98
	✓	×	✓	1.42	8.41	0.02	0.11	0.73	5.80	0.04	1.84	18.21	10.23	6.40	3.27	0.94	4.68
	✓	✓	×	0.22	2.31	0.00	0.00	0.00	0.16	3.06	4.90	19.47	6.59	4.77	0.08	3.98	3.67
CQCC	✓	✓	✓	0.00	0.04	0.00	0.00	0.00	0.00	0.00	0.00	7.56	0.00	1.27	0.00	0.00	**0.76**
	×	✓	✓	7.20	0.43	48.79	43.96	0.00	0.00	0.00	0.01	10.02	32.76	23.86	0.00	0.10	**14.32**
	✓	×	✓	0.00	0.00	0.00	0.00	0.19	0.00	0.00	0.00	13.08	0.03	2.19	0.10	0.00	**1.33**
	✓	✓	×	0.00	0.00	0.00	0.00	0.00	0.00	0.25	0.01	12.19	0.03	2.04	0.00	0.13	**1.25**

Speech activity detector (SAD) is not employed as non-speech frames could be helpful for spoofing detection.

3.3. Classifier and Performance Evaluation

We employ GMM-ML classifier for spoofing detection. Two target models $\boldsymbol{\lambda}_n$ and $\boldsymbol{\lambda}_s$ are created from natural and spoofed speech data respectively [23]. The log-likelihood score is calculated as, $\Lambda(\mathbf{X}) = \mathcal{L}(\mathbf{X}|\boldsymbol{\lambda}_n) - \mathcal{L}(\mathbf{X}|\boldsymbol{\lambda}_s)$, where $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_T\}$ is the feature matrix of the test utterance, T is the number of frames and $\mathcal{L}(\mathbf{X}|\boldsymbol{\lambda})$ is the average log-likelihood of $\mathbf{X}$ given GMM model $\boldsymbol{\lambda}$. We train GMMs with 10 iterations of expectation-maximization (EM) algorithm and 512 mixture components.

Equal error rate (EER) is used as the performance metric to evaluate spoofing attack detection. We use BOSARIS toolkit [24] to calculate the EER using receiver operating characteristics convex hull (ROCCH) method.

4. RESULTS AND ANALYSIS

4.1. BTAS 2016

We first conduct an experiment on BTAS 2016 replay spoofing development dataset to investigate the effects of different training data. The aim is to learn the system's ability to detect spoofed signals generated by various spoofing algorithms that are not incorporated in the training phase. Overall perfor-

mance evaluation results on eight replay attacks (R1-R8), obtained using conventional MFCC and proposed CQCC feature based countermeasures are reported in Table 2. Due to attack dependency, the performance degrades drastically when one attack type is excluded from training the models and when the system is confronted with the similar type of attack in the system assessment process. We also observe that including direct replay in training helps for both SS and VC but not vice versa. This can be justified by the fact that SS and VC attacks are different whereas replay attacks have high similarity with genuine speech in terms of frequency components and formant trajectories [1]. Consequently, the replay speech characteristics of natural signal cannot be captured properly when they are eliminated from training the models. Interestingly, the static spectral features lead to promising recognition accuracy as opposed to their dynamic counterparts. This is in contrast to previous studies [13, 14].

We perform further experiments for only static features on the evaluation dataset. The overall and individual results are reported in Table 3. CQCC feature yields superior result in all training conditions. This probably can be explained by the fact that CQCC feature provides higher resolution in lower and higher frequency regions that reflects better human perception system. Thus, they contribute better ability to capture

Table 4: Same as Table 2 but for ASVspoof 2015 evaluation.

| Features | Train | | Type | S1 | S2 | S3 | S4 | S5 | S6 | S7 | S8 | S9 | S10 | Average | | |
	SS	VC												SS	VC	All
MFCC	✓	✓	Static	1.54	7.54	0.00	0.00	6.80	5.33	2.25	0.04	2.12	26.66	8.89	3.66	5.23
			$\Delta\Delta^2$	0.09	1.46	0.00	0.00	0.36	0.30	0.02	0.03	0.02	19.45	6.48	0.33	2.17
	×	✓	Static	1.64	7.35	0.32	0.35	5.93	5.07	2.47	0.07	2.67	30.68	10.45	3.60	5.66
			$\Delta\Delta^2$	0.00	1.07	0.30	0.28	0.25	0.25	0.01	0.17	0.17	22.08	7.55	0.25	2.44
	✓	×	Static	19.46	35.82	0.00	0.00	35.16	34.41	26.98	0.26	24.54	8.97	2.99	25.23	18.56
			$\Delta\Delta^2$	40.96	29.02	0.00	0.00	11.21	11.39	2.18	0.19	2.57	26.23	8.74	13.93	12.38
CQCC	✓	✓	Static	0.02	0.44	0.00	0.00	1.54	1.04	0.09	0.15	0.13	19.11	6.37	0.49	2.25
			$\Delta\Delta^2$	0.02	0.31	0.01	0.03	0.27	0.25	0.12	2.29	0.15	0.94	0.33	0.49	**0.44**
	×	✓	Static	0.03	1.43	0.10	0.07	1.43	1.19	0.08	0.59	0.14	21.77	7.31	0.70	2.68
			$\Delta\Delta^2$	0.01	0.08	4.17	3.86	0.08	0.12	0.07	5.93	0.12	0.72	2.92	0.92	**1.52**
	✓	×	Static	2.68	16.24	0.00	0.00	25.74	22.09	10.56	0.05	12.22	6.05	2.02	12.80	9.56
			$\Delta\Delta^2$	26.59	7.86	0.01	0.03	7.86	7.85	1.90	2.09	2.96	35.20	11.75	8.16	**9.24**

Table 5: Cross corpora evaluation performance (in % of EER) on BTAS 2016 evaluation data trained using ASVspoof 2015 training dataset.

| Features | Train | | Type | R1 | R2 | R3 | R4 | R5 | R6 | R7 | R8 | R9 | R10 | Average | | | |
	SS	VC												Replay	SS	VC	All
MFCC	✓	✓	Static	46.91	50.00	50.00	50.00	46.68	49.94	42.47	49.28	46.05	50.00	48.83	48.31	45.88	48.13
			$\Delta\Delta^2$	49.93	34.39	49.88	49.82	34.31	4.92	3.97	4.80	44.62	49.87	46.42	19.62	4.39	**32.65**
	×	✓	Static	46.54	50.00	50.00	50.00	46.53	49.97	42.77	49.44	45.90	49.16	48.60	48.25	46.11	48.03
			$\Delta\Delta^2$	49.94	35.39	49.61	49.19	39.74	7.42	3.20	3.29	47.89	49.85	46.98	23.56	3.25	**33.55**
	✓	×	Static	50.00	49.98	43.23	47.96	50.00	49.99	50	46.14	50	49.97	48.52	50	48.07	48.73
			$\Delta\Delta^2$	50.00	47.42	49.97	50.00	20.03	3.26	14.32	10.67	50.00	49.98	49.56	11.65	12.50	**34.57**
CQCC	✓	✓	Static	48.50	49.96	38.68	44.14	41.39	49.88	49.94	49.16	49.96	45.21	46.08	45.64	49.55	46.68
			$\Delta\Delta^2$	47.67	36.43	49.68	50	19.42	17.29	18.11	19.97	46.64	49.08	46.58	18.36	19.04	35.43
	×	✓	Static	48.38	49.96	42.89	45.88	41.74	49.88	49.96	49.95	49.96	45.03	47.05	45.81	49.96	47.36
			$\Delta\Delta^2$	47.83	40.07	49.57	49.94	30.61	31.82	27.10	28.22	42.81	48.44	46.44	31.22	27.66	39.64
	✓	×	Static	49.99	49.92	40.11	49.99	45.22	49.96	49.99	47.36	50.00	50.00	48.34	47.59	48.68	48.25
			$\Delta\Delta^2$	44.44	46.12	42.62	49.47	9.87	13.80	21.89	23.71	49.95	49.96	47.09	11.84	22.80	35.18

replay characteristics while the models are trained by entire or a specific type of attacks. Furthermore, the pattern in EER values represents similar nature when features from unknown spoofing classes appear in the evaluation phase. It is worthwhile to mention that overall performance is compromised throughout all generalization systems for such unknown attacks (R9-R10). Comparing CQCC feature with MFCC feature, the CQCC feature outperforms other systems reported in [21] with an average EER of 0.76 %.

4.2. ASVspoof 2015

The results of generalized systems on ASVspoof 2015 synthetic spoofing database are reported in Table 4. We train the countermeasure with both and either of the SS and VC attacks. In this study, we find that the dynamic coefficients provide superior performance in detecting synthetic spoofed signals. The results show a large amount of deterioration in performance when a particular attack is not considered in training. Although both MFCC and CQCC features give poor performance, CQCC feature leads to better performance across all cases of generalization scenarios. It is also interesting that for a particular case of generalization (where SS type attack is only used for training), static features give higher recognition accuracy than dynamic features for S1 and S10 attacks. We also notice that best performance for a specific attack is obtained if data from the specific attack type is used in training.

4.3. Cross-corpora Evaluation

The goal of this study has been to check cross-corpora vulnerability in a similar attack dependency framework where ASVspoof 2015 synthetic data is used to model the countermeasure and system performance is evaluated on BTAS 2016

replay evaluation dataset. The cross-corpora evaluations are shown in Table 5. The overall performance is poor as SS and VC data of BTAS 2016 test set consist of replayed version of SS and VC attacks as oppose to the ASVspoof database. An interesting observation is that although replay attacks show poor recognition accuracy, dynamic features convey more distinct information in case of SS and VC attacks. It seems reasonable given that SS and VC based spoofed data are better modeled through dynamic characteristics. This, in turn, enhances the recognition accuracy while detecting replay version of VC and SS spoofed samples. Conventional MFCC feature proves to be more efficient in cross-corpora evaluation, but the performance of unknown attacks is poor for both features.

5. CONCLUSIONS

This work presents first analysis of spoofing countermeasures for attack dependency and generalization. A detailed study on BTAS 2016 with extensive experiments reveals that direct replay data have better generalization capability than SS and VC-based replayed data. Results on ASVspoof 2015 demonstrates that VC spoofed data in training can better represent the attack space. The cross-corpora evaluation performance is very poor due to lack of suitable data in training. Our study on both the databases also indicates that both static and dynamic parts of spectral features are useful for detecting spoofing attacks in generalized sense.

7. 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. Lindberg, M. Blomberg *et al.*, "Vulnerability in speaker verification-a study of technical impostor techniques." in *EUROSPECH*, 1999.

[3] W. Shang and M. Stevenson, "Score normalization in playback attack detection," in *ICASSP*. IEEE, 2010, pp. 1678–1681.

[4] J. Villalba and E. Lleida, "Detecting replay attacks from far-field recordings on speaker verification systems," in *Biometrics and ID Management*. Springer, 2011, pp. 274–285.

[5] ——, "Preventing replay attacks on speaker verification systems," in *Security Technology (ICCST), 2011 IEEE International Carnahan Conference on*. IEEE, 2011, pp. 1–8.

[6] P. Korshunov and S. Marcel, "Cross-database evaluation of audio-based spoofing detection systems," in *INTERSPEECH*, 2016.

[7] Z. Wu, C. E. Siong, and H. Li, "Detecting converted speech and natural speech for anti-spoofing attack in speaker recognition." in *INTERSPEECH*, 2012.

[8] P. L. De Leon, M. Pucher, J. Yamagishi, I. Hernaez, and I. Saratxaga, "Evaluation of speaker verification security and detection of HMM-based synthetic speech," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 20, no. 8, pp. 2280–2290, 2012.

[9] J. Sanchez, I. Saratxaga, I. Hernaez, E. Navas, and D. Erro, "The AHOLAB RPS SSD spoofing challenge 2015 submission," in *INTERSPEECH*, 2015, pp. 2042–2046.

[10] M. J. Alam, P. Kenny, G. Bhattacharya, and T. Stafylakis, "Development of CRIM system for the automatic speaker verification spoofing and countermeasures challenge 2015," in *INTERSPEECH*, 2015.

[11] X. Xiao, X. Tian, S. Du, H. Xu, E. S. Chng, and H. Li, "Spoofing speech detection using high dimensional magnitude and phase features: The NTU approach for ASVspoof 2015 challenge," in *INTERSPEECH*, 2015.

[12] P. L. De Leon, B. Stewart, and J. Yamagishi, "Synthetic speech discrimination using pitch pattern statistics derived from image analysis." in *INTERSPEECH*, 2012.

[13] M. Sahidullah, T. Kinnunen, and C. Hanilçi, "A comparison of features for synthetic speech detection," in *INTERSPEECH*, 2015.

[14] M. Todisco, H. Delgado, and N. Evans, "A new feature for automatic speaker verification anti-spoofing: Constant Q cepstral coefficients," in *Speaker Odyssey Workshop, Bilbao, Spain*, 2016.

[15] T. Toda, A. W. Black, and K. Tokuda, "Voice conversion based on maximum-likelihood estimation of spectral parameter trajectory," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 15, no. 8, pp. 2222–2235, 2007.

[16] D. Saito, K. Yamamoto, N. Minematsu, and K. Hirose, "One-to-many voice conversion based on tensor representation of speaker space." in *INTERSPEECH*, 2011, pp. 653–656.

[17] M. Schröder and J. Trouvain, "The german text-to-speech synthesis system mary: A tool for research, development and teaching," *International Journal of Speech Technology*, vol. 6, no. 4, pp. 365–377, 2003.

[18] J. Yamagishi, T. Kobayashi, Y. Nakano, K. Ogata, and J. Isogai, "Analysis of speaker adaptation algorithms for hmm-based speech synthesis and a constrained smaplr adaptation algorithm," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 17, no. 1, pp. 66–83, 2009.

[19] C. Hanilçi, T. Kinnunen, M. Sahidullah, and A. Sizov, "Classifiers for synthetic speech detection: A comparison," in *INTERSPEECH*, 2015.

[20] Z. Wu, T. Kinnunen, N. Evans, J. Yamagishi, C. Hanilçi, M. Sahidullah, and A. Sizov, "ASVspoof 2015: the first automatic speaker verification spoofing and countermeasures challenge," *INTERSPEECH*, 2015.

[21] P. Korshunov, S. Marcel, H. Muckenhirn, A. Gonçalves, A. S. Mello, R. V. Violato, F. Simoes, M. Neto, M. de Assis Angeloni, J. Stuchi, H. Dinkel, N. Chen, Y. Qian, D. Paul, G. Saha, and M. Sahidullah, "Overview of BTAS 2016 speaker anti-spoofing competition," in *Biometrics: Theory, Applications and Systems (BTAS), IEEE*, 2016.

[22] S. B. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," *Acoustics, Speech and Signal Processing, IEEE Transactions on*, vol. 28, no. 4, pp. 357–366, 1980.

[23] D. Paul, M. Pal, and G. Saha, "Novel speech features for improved detection of spoofing attacks," in *INDICON IEEE*, 2015.

[24] N. Brümmer and E. de Villiers, "The BOSARIS toolkit: Theory, algorithms and code for surviving the new DCF," *arXiv preprint arXiv:1304.2865*, 2013.