INTERSPEECH 2016
September 8–12, 2016, San Francisco, USA

Integrated Spoofing Countermeasures and Automatic Speaker Verification: an Evaluation on ASVspoof 2015

Md Sahidullah[1], *Héctor Delgado*[2], *Massimiliano Todisco*[2]
Hong Yu[3], *Tomi Kinnunen*[1], *Nicholas Evans*[2] *and Zheng-Hua Tan*[3]

[1]Speech and Image Processing Unit, School of Computing, University of Eastern Finland, Finland
[2]Dept. of Digital Security, EURECOM, France
[3]Signal and Information Processing, Dept. of Electronic Systems, Aalborg University, Denmark

sahid@cs.uef.fi, delgado@eurecom.fr, todisco@eurecom.fr, hy@es.aau.dk
tkinnu@cs.joensuu.fi, evans@eurecom.fr, zt@es.aau.dk

Abstract

It is well known that automatic speaker verification (ASV) systems can be vulnerable to spoofing. The community has responded to the threat by developing dedicated countermeasures aimed at detecting spoofing attacks. Progress in this area has accelerated over recent years, partly as a result of the first standard evaluation, ASVspoof 2015, which focused on spoofing detection in isolation from ASV. This paper investigates the integration of state-of-the-art spoofing countermeasures in combination with ASV. Two general strategies to countermeasure integration are reported: cascaded and parallel. The paper reports the first comparative evaluation of each approach performed with the ASVspoof 2015 corpus. Results indicate that, even in the case of varying spoofing attack algorithms, ASV performance remains robust when protected with a diverse set of integrated countermeasures.

Index Terms: Automatic speaker recognition, spoofing, countermeasures, presentation attack detection.

1. Introduction

It has long been suspected that automatic speaker verification (ASV) systems can be vulnerable to spoofing [1], also referred to as presentation attacks [2]. Spoofing refers to the fraudulent manipulation of an ASV system with specially crafted speech data in order to provoke false alarms. The Interspeech 2013 special session on Spoofing and Countermeasures for Automatic Speaker Verification [3] was organized to stimulate the collaboration needed for the collection of standard datasets and the definition of protocols and metrics for future research.

The first *Automatic Speaker Verification Spoofing and Countermeasures Challenge* (ASVspoof) [4] followed soon after in 2015. This first evaluation aimed to promote the development of generalized countermeasures [5], namely countermeasures with the potential to detect varying and unforeseen spoofing attacks; the ASVspoof 2015 evaluation dataset contained spoofing attacks generated with 10 different speech synthesis and voice conversion spoofing algorithms. Being the first evaluation of its kind, the evaluation focused on spoofing detection in isolation from ASV.

Evaluation results [4] showed considerable variation in spoofing detection performance. Many systems obtained good performance for some spoofing conditions, but relatively poor performance for others, most notably the S10 condition for which no similar training material was provided in the development set. These results suggest that a bank of fused countermeasures may prove beneficial.

While the broader picture was encouraging, with some exceptionally low error rates being achieved for some conditions, even small spoofing detection errors may yet have significant impacts on ASV performance. The integration of spoofing countermeasures with ASV was foreseen at the time as a future goal [6].

The contributions of this paper are thus two-fold. First, we report a new study of fused, state-of-the-art spoofing detection systems evaluated on the ASVspoof 2015 dataset. This work is performed with a host of different countermeasures developed by three different research groups. The second contribution relates to a study of different ASV and countermeasure integration strategies. The manner in which the two tasks should best be combined has attracted only modest attention to date [7]. The work reported in this paper is the first reported for the standard ASVspoof dataset.

2. ASV and countermeasure integration

Spoofing countermeasures (CMs) are expected to improve the reliability of biometric systems by preventing fraudulent access. It is however impossible to gauge the impact of CMs unless they are evaluated when integrated with a biometric system [8]. A diverse body of research reports the combination of CMs in the context of many different biometric modalities, especially for fingerprint and face verification [9, 10, 11, 12].

While they have a common goal of preventing fraudulent access, ASV and CM systems have specific objectives. They are illustrated in Table 1. While the ASV system should reject a zero-effort impostor (the speakers differ), the CM should detect a valid trial (which is genuine human speech). The problem of ASV and CM integration is somewhat different to conventional fusion which typically involves two systems with identical objectives.

Since genuine trials should be accepted by both systems and since either ASV or CM systems could cause the rejection of impostor or spoofed trials, a simple cascaded combination of ASV and CMs provides a straightforward solution. This approach is illustrated in Fig. 1(a). The cascaded approach was reported in [13] which describes a countermeasure to protect ASV from synthetic speech spoofing attacks. A similar approach was reported in [14] for the protection of ASV from voice conversion spoofing attacks. The cascaded system illustrated in Fig. 1

Copyright © 2016 ISCA

http://dx.doi.org/10.21437/Interspeech.2016-1280

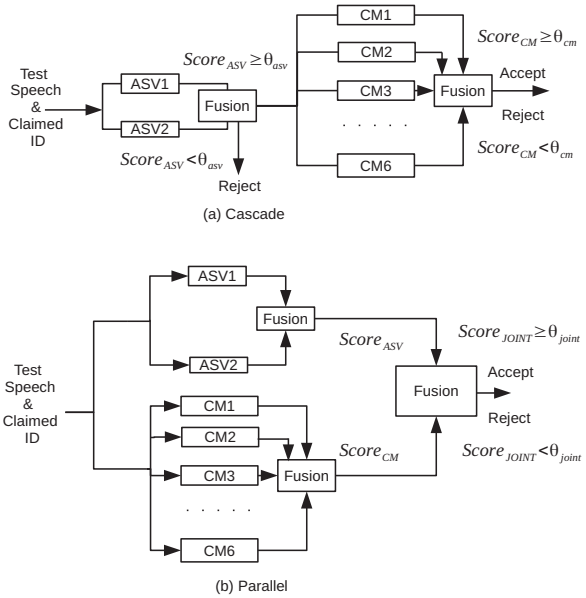

Figure 1: *Block diagrams showing integration of ASV with CM: (a) cascade and (b) parallel.*

Table 1: *Definition of positive and negative trials for ASV and CM tasks.*

Task	Genuine	Zero-effort impostor	Spoofing impostor
Stand-alone ASV	+1	-1	-1
Stand-alone CM	+1	+1	-1
ASV with CM	+1	-1	-1

will only accept trials which produce an ASV score ($\mathrm{Score}_{\mathrm{ASV}}$) greater than or equal to the recognition threshold (θ_{asv}). Only accepted trials are then processed by the CM where trials with a CM score ($\mathrm{Score}_{\mathrm{CM}}$) greater that the CM threshold (θ_{cm}) are then finally accepted. The final decision is thus obtained by ANDing the ASV and CM decisions.

The cascaded approach is, however, not the only solution. A parallel approach such as that illustrated in Fig. 1(b) can be used to combine ASV and CM scores in order to obtain the decision for a given trial based on a single threshold (θ_{joint}). An approach similar to this was used in [15]. This paper reports the performance of different ASV and CM systems which are first combined separately and subsequently fused according to one of the approaches illustrated in Fig. 1. Until now, no such comparison has been reported in the literature.

3. System descriptions

The ASV and CM systems used for experiments reported in this paper have been developed by different partners of the EU H2020 OCTAVE project[1]. There are two different ASV systems (ASV1-2) and six different CM systems (CM1-CM6).

3.1. CM systems

All CM systems employ the same back-end which is a simple two-class classifier based on Gaussian mixture models (GMM). A GMM is trained for each class, namely genuine and spoofed speech by means of the *expectation-maximization* (EM) algorithm. Scores are the log-likelihood ratio given the two models. The GMM-based approach often outperforms more sophisticated classification methods for the ASVspoof 2015 dataset, e.g. [22, 18].

The six different CM systems differ only in their front-ends. A summary of the different systems is shown in Table 2. All are recent and at the state-of-the-art as judged by results generated using the ASVspoof 2015 database [16, 18]. Common to all is the use of cepstral processing and, importantly, the absence of static features; dynamic features give better performance. Delta and delta-delta coefficients are all extracted in the usual way. No speech activity detection is used since non-speech frames tend also to contain information useful for spoofing detection at least on ASVspoof 2015. Details of each feature set and the common back-end are given below.

CM1: Conventional *mel-frequency cepstral coefficients* (MFCCs) are extracted using a bank of 20, triangular-shaped filters positioned on a mel scale and the application of the discrete cosine transform (DCT) to the filterbank log energies.

CM2: *Inverted-mel frequency cepstral coefficients* (IMFCCs) are extracted in similar fashion to MFCCs except that filters are placed on an inverted-mel scale [23] (See [16] for an illustration).

CM3: *Linear frequency cepstral coefficients* (LFCCs) are computed in the same way as MFCCs except that filters are positioned on a linear scale.

CM4 Originally reported in [17], *cochlear filter cepstral coefficients* (CFCCs) have shown promise in detecting spoofed speech. CFCCs model the physiological elements of the human hearing system, namely the cochlea, the inner hair cells and the nerve spike density. The work reported here uses the same configuration as [24] for the extraction of 12 static coefficients.

CM5: Originally reported in [18], *constant Q cepstral coefficients* (CQCCs) are based on the constant Q transform (CQT) [25] popular in the study of music processing. The constant Q transform is a time-frequency analysis tool which employs a variable time-frequency resolution, providing greater frequency resolution for lower frequencies and greater time resolution for higher frequencies. First, the power spectrum is computed with the CQT. Second, cepstral analysis is performed by first linearising the frequency scale and then by computing the DCT in the usual way to derive a set of 19 coefficients.

CM6: *Gammatone frequency cepstral coefficients* (GFCCs) are based upon Gammatone filters derived from psychophysical observations of the auditory periphery [19]. The filterbank is a standard model of cochlear filtering which emulates the characteristics of the human basilar membrane. Filters and filter bandwidths are positioned according to the equivalent rectangular bandwidth (ERB) scale. In contrast to the standard approach, the gain of each filter is set to give equal emphasis to each of 128 bands. The DCT is applied in the usual way to log filterbank outputs to produce a set of 20 coefficients.

3.2. ASV systems

ASV configurations are also summarised in Table 2. Details of each are provided in the following.

ASV1: Our first ASV system is based on the MFCC features and a Gaussian mixture model – universal background model (GMM-UBM) architecture [20]. MFCCs are the same as those used for CM1. For ASV, however, static coefficients are retained hence a feature dimension of 60. For training

[1] http://www.octave-project.eu/

Table 2: *Summary of the countersmeasures (CM) and speaker verification systems used for experiments reported in this paper.*

Task	Name	Feature (Dim.)	Classifier	Development Data
Countermeasures Systems	CM1 [16]	MFCC-40	GMM-ML Mixtures: 512	ASVspoof 2015 (Train Set)
	CM2 [16]	IMFCC-40		
	CM3 [16]	LFCC-40		
	CM4 [17]	CFCC-40		
	CM5 [18]	CQCC-40		
	CM6 [19]	GFCC-40		
Speaker Verification	ASV1 [20]	MFCC-60	GMM-UBM	TIMIT, RSR2015
	ASV2 [21]	MFCC-60	i-Vector	TIMIT, RSR2015

target models, first a gender-dependent UBM of 512 components is trained with the speech data from the TIMIT [26] and RSR2015 [27] corpora. Target models are created using maximum-a-posteriori (MAP) adaptation with a relevance factor of 3. Scores are the log-likelihood ratio computed between the target model and the UBM.

ASV2: The second is an i-vector system in which GMM super-vectors are projected into a low-dimensional space referred to as the total variability space [21]. i-vectors are computed using the same MFCCs as used for ASV1, from the Baum-Welch statistics and the total variability matrix $\mathbf{T}$. Gender-dependent UBMs are learned using the same TIMIT and RSR2015 databases. Whereas we use the full TIMIT data consisting of 630 speakers (438 male and 192 female) we use a subset of 10 different sentences for 300 speakers (157 male and 143 female) in the RSR2015 database. This gives 5950 sentences for male speakers and 3350 sentences for female speakers which are used as development data. The $\mathbf{T}$-matrix is estimated with the same data. The i-vector dimension is set to 300. Since each target has five different sentences for enrolment, extracted i-vectors are averaged to derive a single training i-vector per speaker. Finally, the score is given by the cosine similarity between the length normalized training and testing i-vectors. Note that using probabilistic linear discriminant analysis (PLDA) could be helpful for scoring [28], but we have not implemented it due to the unavailability of suitable development data such as WSJCAM [4].

4. Experimental setup

Described here is the database used for all experimental work and the evaluation metric.

4.1. Database description

All experiments are conducted with the ASVspoof 2015, a publicly available corpus[2] and supports both the study of ASV and spoofing CMs. The database is summarized in Table 3. It contains both genuine human as well as spoofed speech generated using 10 different voice conversion and speech synthesis methods. A subset of five algorithms are used to generate spoofed speech contained in both development and evaluation subsets and are thus referred to as *known* (K) attacks. The evaluation subset also contains spoofed speech generated with the other five spoofing algorithms and are thus referred to as *unknown* (U) attacks. Full details of the ASVspoof 2015 database are available in [4].

[2] http://datashare.is.ed.ac.uk/handle/10283/853

Table 3: *Database description of ASVspoof 2015 database for joint ASV and CM experiments.*

Trial Type	Male		Female	
	Dev	Eval	Dev	Eval
Genuine	1498	4053	1999	5351
Imposter	4275	8000	5700	10400
Spoofed (K)	21375	40000	28500	52000
Spoofed (U)	-	40000	-	52000

4.2. Evaluation metric

As per the ASVspoof 2015 evaluation plan [6], CM performance is assessed in terms of the *equal error rate* (EER), here calculated using the BOSARIS toolkit[3] and the so-called *receiver operating characteristics convex hull* (ROCCH) method. EERs are reported for the development and evaluation subsets and separately for unknown and known spoofing attacks.

ASV performance is assessed in terms of the *false rejection rate* (FRR) and the *false acceptance rate* (FAR). With this evaluation involving two types of negative class, namely *zero-effort impostor* and *spoofing impostor*, the FAR is reported separately for each. The FARs for two types of spoofing attacks, known and unknown, are also computed separately. The FAR for zero-effort impostors is referred to as FAR(Z), whereas that for spoofing impostors is referred to as FAR(K) in the case of known attacks and FAR(U) in the case of unknown attacks. FARs and FRRs are calculated with EER thresholds obtained from the gender-dependent development subsets and where EERs are computed from genuine trials and zero-effort impostors only.

5. Results and discussion

Performance is first reported for CM and ASV is isolation, then when integrated.

5.1. Countermeasure performance

Comparative CM results are illustrated in Table 4 for the development set and the known (K) and unknown (U) subsets of the evaluation set. All the systems perform well and the performance of CM5, which uses the recently proposed CQCC features [18], is the best among the six. Even then, performance for the unknown spoofing attacks is poorer than for known attacks.

Also illustrated in the last row of Table 4 are results for a logistic regression based fusion of scores produced by all six

[3] https://sites.google.com/site/bosaristoolkit/

Table 4: *Stand-alone spoofing detection performance (in terms of % EER) for the ASVSpoof 2015 database.*

System	Male			Female		
	Dev	Eval(K)	Eval(U)	Dev	Eval(K)	Eval(U)
CM1	0.54	0.53	1.58	0.25	0.23	4.21
CM2	0.12	0.12	0.96	0.19	0.25	2.91
CM3	0.03	0.06	0.76	0.21	0.16	2.43
CM4	1.41	1.10	1.26	0.74	0.61	1.75
CM5	0.01	**0.02**	0.41	0.03	0.03	1.34
CM6	0.10	0.11	0.61	0.09	0.06	1.95
Fused	**0.00**	**0.02**	**0.16**	**0.00**	**0.01**	**0.80**

Table 5: *Stand-alone ASV performance in terms of % of FRR and FAR for the ASVspoof 2015 database. FAR(Z): FAR for zero-effort impostor, FAR(K): FAR for spoofed known attack impostor, FAR(U): FAR for spoofed unknown attack impostor.*

System	Eval Metric	Male		Female	
		Dev	Eval	Dev	Eval
ASV1	FRR	5.67	7.85	7.60	7.77
	FAR(Z)	5.67	6.61	7.60	6.51
	FAR(K)	59.89	59.80	34.50	30.81
	FAR(U)	-	39.48	-	33.42
ASV2	FRR	10.62	15.96	14.22	10.22
	FAR(Z)	10.62	9.13	14.22	13.12
	FAR(K)	59.63	55.39	45.19	46.10
	FAR(U)	-	50.77	-	50.58
Fused	FRR	5.34	7.38	7.45	7.21
	FAR(Z)	5.34	6.24	7.45	6.26
	FAR(K)	60.76	59.15	34.99	31.68
	FAR(U)	-	39.76	-	34.03

CMs. CM fusion delivers universally improved or equivalent performance to the single best system. Of particular note, the improvement is greatest for the unknown attacks, thereby showing the benefit of a bank of diverse CMs for spoofing detection.

5.2. Speaker verification performance

Performance for ASV1 and ASV2 systems is shown in Table 5. Since the decision thresholds are computed on the development data with zero-effort impostors, the FRR is equal to the FAR(Z) in this case. The same thresholds are used on the evaluation set where the FRR and FAR(Z) then differ. Both FAR(K) and FAR(U) are considerably higher than the FAR(Z).

ASV1 outperforms ASV2; it is not uncommon for a basic back-end approach to give better results in the case of short duration ASV on clean data. This may also be due to the lack of suitable development data as used in [7]. Even so, performance once again generally improves with score fusion, although improvements are modest and not always consistent.

5.3. Integrated performance

Attention now turns to the integration of ASV and CMs. Results for cascaded and parallel approaches are given in Tables 6 and Table 7, respectively. In both cases, the results are presented for ASV1 combined with CM5 (ASV1-CM5) and for fused ASV and CMs (Fused-Fused). When subjected to spoofing, the performance for integrated ASV and CMs is considerably better than the performance of stand-alone ASV. For the integrated systems, the FRR and FAR(Z) are almost the same as for the stand-alone approach. However, FAR(K) and FAR(U) are significantly reduced. The FAR(U) is lower for the parallel integration of ASV and CM systems, though the FRR and FAR(Z) is considerably worse. Fusion is once again universally beneficial.

Table 6: *Performance for cascaded ASV and CM in terms of % of FRR and FAR for the ASVspoof 2015 database. FAR(Z): FAR for zero-effort impostor, FAR(K): FAR for spoofed known attack impostor, FAR(U): FAR for spoofed unknown attack impostor.*

	Eval Metric	Male		Female	
		Dev	Eval	Dev	Eval
ASV1-CM5	FRR	5.81	7.85	7.65	7.79
	FAR(Z)	5.66	6.61	7.61	6.51
	FAR(K)	0.00	0.02	0.01	0.01
	FAR(U)	-	0.96	-	2.60
Fused-Fused	FRR	5.47	7.40	7.50	7.29
	FAR(Z)	5.33	6.24	7.44	6.26
	FAR(K)	0.00	0.00	0.00	0.00
	FAR(U)	-	0.34	-	1.75

Table 7: *Performance for parallel ASV and CM in terms of % of FRR and FAR for the ASVspoof 2015 database. FAR(Z): FAR for zero-effort impostor, FAR(K): FAR for spoofed known attack impostor, FAR(U): FAR for spoofed unknown attack impostor.*

	Eval Metric	Male		Female	
		Dev	Eval	Dev	Eval
ASV1-CM5	FRR	20.83	27.69	18.95	20.65
	FAR(Z)	20.83	20.05	18.95	16.42
	FAR(K)	0.00	0.00	0.00	0.00
	FAR(U)	-	0.20	-	0.87
Fused-Fused	FRR	14.49	14.53	15.35	14.80
	FAR(Z)	14.49	18.65	15.35	14.73
	FAR(K)	0.00	0.00	0.00	0.00
	FAR(U)	-	0.27	-	1.50

6. Conclusions

This paper reports the first comparative study of different countermeasures and different approaches to their integration with automatic speaker verification using ASVspoof 2015 database. Countermeasure fusion is shown to offer the greatest potential to detect spoofing attacks especially in the face of unknown spoofing attacks – the only real scenario. The cascaded integration of ASV and CMs greatly reduces the FAR whereas the FRR relatively unaffected. On the other hand, while performance in the absence of spoofing deteriorates, the parallel integration of ASV and CMs gives better performance when the ASV system is subjected to spoofing attacks. The best performance in all cases is delivered through fusion which also increases resilience to unknown spoofing attacks. In future, investigation will be made on speaker-dependent techniques to tackle spoofing attacks.

7. Acknowledgements

The paper reflects some results from 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.

8. 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, no. 0, pp. 130 – 153, 2015.

[2] ISO/IEC, "Information technology—Biometric presentation attack detection—Part 1: Framework," International Organization for Standardization, Geneva, Switzerland, ISO/IEC 30107-1:2016, 2016.

[3] N. Evans, T. Kinnunen, and J. Yamagishi, "Spoofing and countermeasures for automatic speaker verification," in *INTERSPEECH*, 2013, pp. 925–929.

[4] 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," in *Proc. of INTERSPEECH*, 2015.

[5] F. Alegre, A. Amehraye, and N. Evans, "A one-class classification approach to generalised speaker verification spoofing countermeasures using local binary patterns," in *Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on*. IEEE, 2013, pp. 1–8.

[6] Z. Wu, T. Kinnunen, N. Evans, and J. Yamagishi, "ASVspoof 2015: Automatic speaker verification spoofing and countermeasures challenge evaluation plan," 2014.

[7] Z. Wu, P. L. D. Leon, C. Demiroglu, A. Khodabakhsh, S. King, Z. H. Ling, D. Saito, B. Stewart, T. Toda, M. Wester, and J. Yamagishi, "Anti-spoofing for text-independent speaker verification: An initial database, comparison of countermeasures, and human performance," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 4, pp. 768–783, 2016.

[8] A. Hadid, N. Evans, S. Marcel, and J. Fierrez, "Biometrics systems under spoofing attack: An evaluation methodology and lessons learned," *IEEE Signal Processing Magazine*, vol. 32, no. 5, pp. 20–30, 2015.

[9] B. Biggio, Z. Akhtar, G. Fumera, G. L. Marcialis, and F. Roli, "Security evaluation of biometric authentication systems under real spoofing attacks," *IET Biometrics*, vol. 1, no. 1, pp. 11–24, 2012.

[10] E. Marasco, P. Johnson, C. Sansone, and S. Schuckers, *Multiple Classifier Systems: 10th International Workshop, MCS 2011, Naples, Italy, June 15-17, 2011. Proceedings*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, ch. Increase the Security of Multibiometric Systems by Incorporating a Spoofing Detection Algorithm in the Fusion Mechanism, pp. 309–318.

[11] I. Chingovska, A. Anjos, and S. Marcel, "Anti-spoofing in action: Joint operation with a verification system," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2013 IEEE Conference on*, 2013, pp. 98–104.

[12] ——, "Biometrics evaluation under spoofing attacks," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 12, pp. 2264–2276, 2014.

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

[14] F. Alegre, A. Amehraye, and N. Evans, "Spoofing countermeasures to protect automatic speaker verification from voice conversion," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, 2013, pp. 3068–3072.

[15] 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.

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

[17] Q. Li and Y. Huang, "An auditory-based feature extraction algorithm for robust speaker identification under mismatched conditions," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 6, pp. 1791–1801, 2011.

[18] 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.

[19] A. Adiga, M. Magimai, and C. S. Seelamantula, "Gammatone wavelet cepstral coefficients for robust speech recognition," in *TENCON 2013 - 2013 IEEE Region 10 Conference (31194)*, 2013, pp. 1–4.

[20] D. Reynolds, T. Quatieri, and R. Dunn, "Speaker verification using adapted Gaussian mixture models," *Digital Signal Processing*, vol. 10, no. 1-3, pp. 19–41, 2000.

[21] 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, May 2011.

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

[23] S. Chakroborty, A. Roy, and G. Saha, "Improved closed set text-independent speaker identification by combining MFCC with evidence from flipped filter banks," *International Journal of Signal Processing*, vol. 4, no. 2, pp. 114–122, 2007.

[24] T. B. Patel and H. A. Patil, "Combining evidences from mel cepstral, cochlear filter cepstral and instantaneous frequency features for detection of natural vs. spoofed speech," in *INTERSPEECH*, Dresden, Germany, 2015, pp. 2062–2066.

[25] J. Brown, "Calculation of a constant q spectral transform," *Journal of the Acoustical Society of America*, vol. 89, no. 1, pp. 425–434, January 1991.

[26] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, N. L. Dahlgren, and V. Zue, "TIMIT acoustic-phonetic continuous speech corpus," *Linguistic Data Consortium, Philadelphia*, 1993.

[27] A. Larcher, K. Lee, P. L. S. Martinez, T. H. Nguyen, B. Ma, and H. Li, "Extended RSR2015 for text-dependent speaker verification over VHF channel," in *INTERSPEECH*, Singapore, 2014, pp. 1322–1326.

[28] S. Prince and J. Elder, "Probabilistic linear discriminant analysis for inferences about identity," in *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*, 2007, pp. 1–8.