The ASVspoof 2017 Challenge: Assessing the Limits of Replay Spoofing Attack Detection

Tomi Kinnunen[1], *Md Sahidullah*[1], *Héctor Delgado*[2], *Massimiliano Todisco*[2], *Nicholas Evans*[2], *Junichi Yamagishi*[3,4], *Kong Aik Lee*[5]

[1]University of Eastern Finland, Finland – [2]EURECOM, France
[3]National Institute of Informatics, Japan – [4]University of Edinburgh, UK
[5]Institute for Infocomm Research, Singapore

info@asvspoof.org

Abstract

The ASVspoof initiative was created to promote the development of countermeasures which aim to protect automatic speaker verification (ASV) from spoofing attacks. The first community-led, common evaluation held in 2015 focused on countermeasures for speech synthesis and voice conversion spoofing attacks. Arguably, however, it is replay attacks which pose the greatest threat. Such attacks involve the replay of recordings collected from enrolled speakers in order to provoke false alarms and can be mounted with greater ease using everyday consumer devices. ASVspoof 2017, the second in the series, hence focused on the development of replay attack countermeasures. This paper describes the database, protocols and initial findings. The evaluation entailed highly heterogeneous acoustic recording and replay conditions which increased the equal error rate (EER) of a baseline ASV system from 1.76% to 30.71%. Submissions were received from 49 research teams, 20 of which improved upon a baseline replay spoofing detector EER of 24.65%, in terms of replay/non-replay discrimination. While largely successful, the evaluation indicates that the quest for countermeasures which are resilient in the face of variable replay attacks remains very much alive.

Index Terms: automatic speaker verification, spoofing, countermeasure, replay attacks, ASVspoof

1. Introduction

Automatic speaker verification (ASV) [1, 2, 3] technology is used in a growing range of applications that require not only robustness to changes in the acoustic environment, but also resilience to intentional circumvention, known as *spoofing* [4]. Among other possible attack vectors, *replay attacks* are a key concern; they can be performed with ease and the threat they pose to ASV reliability has been confirmed by independent researchers [5, 6, 7]. Replay attacks are mounted using recordings of a target speaker's voice which are replayed to an ASV system in the place of genuine speech. An example is the use of a smart-device to replay a recording of a target speaker's voice to unlock a smartphone which uses ASV access control.

Spoofing *countermeasures* have consequently been developed to protect ASV systems from replay attacks. The literature shows three general strategies. *Prompted-phrase* ASV, e.g. randomised digit sequences [8], and utterance verification [9, 10] offers some protection, although multiple recordings can be remixed to produce a replay attack which matches the prompted phrase. *Copy detection* [11, 12], or *audio fingerprinting*, can also be used to detect recordings of genuine enrollment utterances or previous access attempts although this approach calls for the maintenance of a dynamically growing database. This study addresses a third strategy which aims to detect replay attacks using only the acoustic characteristics of a given utterance. Arguably, this solution has broader utility; this includes any ASV approach/system and both forms of replay attack outlined above.

The detection of replay attacks using acoustic characterisation is potentially problematic, however. The difficulty relates to the unpredictable variation in the quality of a replay attack. Recordings, perhaps collected surreptitiously, may contain significant additive or convolutional noise. The detection of replay attacks may then boil down to a ambient or channel noise classification problem. In contrast, recordings made with high-quality hardware in benign acoustic environments may be close to indistinguishable from genuine speech signals. At the limit, bit-to-bit digital copies of genuine audio recording, perhaps injected into an ASV system post-microphone, would be indistinguishable using *any* method. The question then is, what are the practical limits of replay attack detection?

The search for an answer to this fundamental question is the focus of the ASVspoof 2017 challenge[1]. ASVspoof 2017 follows two special sessions on spoofing and countermeasures for automatic speaker verification at INTERSPEECH 2013 [13] and 2015 [14] which formed the first evaluation, ASVspoof 2015 [15]. The first evaluation promoted the development of generalised countermeasures capable of protecting ASV from diverse text-to-speech (TTS) and voice conversion (VC) spoofing attacks [16]. While the mounting of these attack currently requires substantial expertise, replay attacks are accessible to the layperson using widely available consumer devices for audio recording and replaying. ASVspoof 2017 therefore promoted the development of replay attack countermeasures.

Previous attempts to assess the threat of replay spoofing attacks typically involved a modest number of evaluation conditions, e.g. [7, 5, 6, 17]. Some studies, e.g. [5, 6] report close-to-perfect recognition accuracy, albeit in the case of relatively homogeneous acoustic conditions. Other work [7] suggests that performance may degrade in more practical scenarios where the acoustic conditions can vary greatly. The primary technical goals of ASVspoof 2017 are therefore (i) to assess the practical limitations of replay attack detection and (ii) to promote the development of countermeasures with potential to detect replay spoofing attacks 'in the wild', namely in highly-varying acoustic conditions.

In identical fashion to the 2015 edition, ASVspoof 2017 focuses on standalone spoofing attack detection (here, replay

[1]http://www.asvspoof.org/

Table 1: *Statistics of the ASVspoof 2017 Corpus.*

Subset	# Speakers	# Replay sessions	#Utterances	
			Non-replay	Replay
Training	10	6	1508	1508
Devel.	8	10	760	950
Eval.	24	163	1298	12922
Total	42	179	3566	15380

attacks), i.e. spoofing detection in isolation from ASV. However, so that the initiative is at least aligned to ASV research and in contrast to the first edition, ASVspoof 2017 uses the recent text-dependent *RedDots* [18] data as the base corpus [19]. One additional change to the 2015 edition, made in order to encourage wider participation, is the provision of a baseline spoofing classifier [20]. This strategy appears to have had a positive impact; at the time of writing, the organisers have received 113 requests for the development set, while a total of 49 primary system scores were submitted for the evaluation set.

2. ASVspoof 2017 Corpus

The ASVspoof 2017 Corpus originates from the *RedDots* corpus[2] which was collected by volunteers from across the globe (mostly ASV researchers) using Android smartphones. **Non-replayed** utterances are a subset of the original RedDots recordings whereas **replayed** recordings are replayed and recaptured versions. Replayed utterances hence correspond to a 'stolen voice' replay attack scenario where the attacker has access to a digital copy of an original target speaker utterance which is then replayed through transducers of varying quality.

A total of 56% of replay files were collected by four participants of the EU Horizon 2020-funded OCTAVE project[3], (see [19]), while the remaining 44% were collected by other contributors. Replay recordings were collected from the replaying and re-recording of concatenated RedDots utterances with heterogeneous devices and acoustic environments. Non-replayed evaluation data was supplemented with utterances collected from 7 new speakers.

The ASVspoof 2017 Corpus is partitioned into three subsets: **training**, **development** and **evaluation**. Details of each are presented in Table 1. The first two subsets were provided to participants for the design of replay detectors (countermeasures), while re-partitioning of the training and development subsets was permitted. Metadata consisting of replay/non-replay ground-truth labels, in addition to speaker ID, phrase ID, and replay configuration details were provided for the training and development subsets. Only audio data and phrase ID were provided for the evaluation set for which participants were required to submit scores. Results were then determined by the organisers and returned to participants.

All three subsets are disjoint in terms of speakers. They are also somewhat disjoint in terms of data collection sites. The training subset was collected at a single site. The development subset was collected at the same site in addition to two more sites. Finally, the evaluation subset was collected at the same three sites and supplemented with additional data from two new sites. Nonetheless, even data from the same site was collected by different people using different recording and replaying devices and in different acoustic environments. The evaluation subset contains data collected from 163 replay sessions in 112 unique replay configurations[4]. Data heterogeneity has proven essential to the development of reliable spoofing countermeasures [21, 22, 23].

2.1. Evaluation conditions

The ASVspoof 2017 corpus comprises six evaluation conditions containing a disjoint set of replay trials (and a shared set of non-replay trials). Since the original RedDots source data and its replay recordings were collected in diverse conditions, the data exhibits multiple, concurrent variations (*e.g.* recording device quality, room dimensions, reverberation and vocal effort in the original recordings). The isolation or marginalisation of such variation is particularly challenging. Hence, the six conditions for the ASVspoof 2017 challenge were defined post-evaluation using clustering of well-ranked system scores.

So as to marginalise variation at the utterance level, *i.e.* to focus on differences in replay configurations rather than differences relating to individual utterances, clustering was applied to scores averaged across individual replay environments. The clustering process was performed as follows:

1. System scores for all submissions which out-performed the baseline were linearly fused using the Bosaris[5] toolkit to obtain a high-performance ensemble classifier.
2. Fused scores were then averaged across all replay trials corresponding to the same replay session (sharing the replay environment, playback and recording devices).
3. Averaged scores were then clustered using k-means to obtain a non-uniform partitioning of the score axis.
4. Resulting score clusters were then ordered according to increasing average fused score.

This procedure can be applied to cluster results into a number of different replay conditions. Those with a lower average fused score represent replay conditions which are generally easier to detect than replay conditions with a higher average fused score. We found empirically that clustering into 6 different replay conditions gave the most consistent and intuitive results. Condition **C1** represents replay trials with significant background noise or channel distortion which are generally among the most easily detected. Condition **C6** represents high-quality replay trials which are comparatively more difficult to detect. In the experimental part, we provide a quantitative analysis of the conditions in terms of signal quality measures and error rates.

2.2. Evaluation metrics

In line with the ASVspoof 2015 challenge, the 2017 edition concentrates on stand-alone spoofing detection without ASV integration. The task requires the assignment to a set of audio files a score which reflects the relative strength of two competing hypotheses, namely that the trial is non-replayed (genuine) or replayed (spoofed) speech. Higher scores are assumed to favor the non-replay/genuine hypothesis. The primary metric is the

[2] https://sites.google.com/site/thereddotsproject/

[3] https://www.octave-project.eu/

[4] A **replay configuration** refers to a unique (room, replay device, recording device) combination while a **session** refers to replay of specific source files, some of which share the same replay configuration.

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

Table 2: *Summary of the trials for text-dependent speaker verification experiment.*

Trial Type	Development	Evaluation
Genuine	742	1106
Zero-effort Spoof	5186	18624
Replay Spoof	940	11711

Table 3: *Text-dependent ASV performance (% of EER_{ASV}) with GMM-UBM system.*

Imposter Type	Development	Evaluation
Zero-effort Spoof	3.50	1.76
Replay Spoof	41.96	30.71

equal error rate (EER). Let $P_{\text{fa}}(\theta)$ and $P_{\text{miss}}(\theta)$ be the false alarm and miss rates at threshold θ defined according to:

$$\begin{aligned} P_{\text{fa}}(\theta) &= \frac{\#\{\text{replay trials with score} > \theta\}}{\#\{\text{Total replay trials}\}} \\ P_{\text{miss}}(\theta) &= \frac{\#\{\text{non-replay trials with score} \leq \theta\}}{\#\{\text{Total non-replay trials}\}}, \end{aligned}$$

so that $P_{\text{fa}}(\theta)$ and $P_{\text{miss}}(\theta)$ are, respectively, monotonically decreasing and increasing functions of θ. The EER corresponds to the threshold θ_{EER} at which the two detection error rates are (approximately) equal. It is estimated using the convex hull method available in the Bosaris toolkit. In contrast to the ASVspoof 2015 Challenge, the EER is computed from scores pooled across all the trial segments instead of condition averaging. The rationale is to promote the development of replay attack detectors yielding scores that are more consistent across variable spoofing conditions; see also [24, Table 12] and [15, Fig. 6].

3. Impact of replay to ASV accuracy

The vulnerability of ASV systems to replay spoofing attacks has been confirmed previously by independent teams [5, 6, 25]. This section reports the effect of ASVspoof 2017 replay spoofing attacks on a classical Gaussian mixture model with universal background model (GMM-UBM) [26] ASV system. This has been shown [27] to deliver competitive performance for RedDots data consisting of short-duration utterances. The ASV system uses mel-frequency cepstral coefficient (MFCC) features and a 512-component UBM trained using RSR2015 [28] and TIMIT[6] databases. Phrase-dependent target speaker models are created from RedDots enrollment data. The evaluation protocol involves a number of genuine trials and then either zero-effort impostor or replay spoofing attack trials. The number of each are shown in Table 2. Table 3 shows the degradation in ASV performance when zero-effort impostors are replaced with replay spoofing attacks. The baseline EER for speaker discrimination is seen to increase substantially, lending support to the development of replay attack countermeasures.

4. ASVspoof 2017 Challenge results

4.1. Overview

A total of 49 submissions were received. A summary of results for primary systems only is illustrated in Table 4[7] with that for the baseline replay/non-replay detector[8]. It uses *constant Q cepstral coefficient* (CQCC) features [20] with a Gaussian mixture model (GMM) back-end. CQCCs, based on a perceptually motivated time-frequency transform [29], have been widely adopted as a baseline to detect different types of spoofing attacks. Performance is shown for two baseline variants trained using either combined training and development data (**B01**) or training data alone (**B02**). The use of pooled data naturally results in better performance. There is substantial variation in EERs with 20 of the 49 submissions achieving better performance than the B01 baseline. This observation indicates the difficulty of the challenge and stresses the importance of avoiding over-fitting. The top-performing S01 system achieves an encouraging EER of 6.73 %.

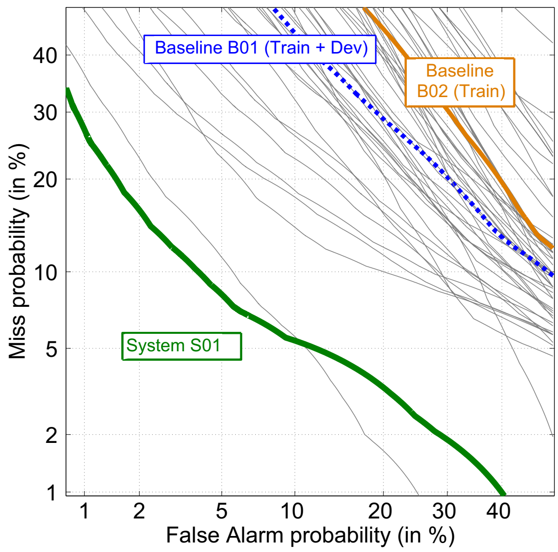

Figure 1: *Detection error trade-off (DET) graph for the replay-nonreplay detection task, illustrating the submitted 49 primary systems along with the two baseline system variants.*

Table 4: *Results of the baseline systems (B1 & B2) and primary systems (S01-S48, D01) submitted to ASVspoof 2017 challenge, in terms of EER_{RD} (%).*

ID	EER	ID	EER	ID	EER	ID	EER
S01	6.73	S14	22.04	S26	26.98	S38	31.59
S02	12.39	S15	22.23	S27	27.32	S39	31.76
S03	14.31	S16	22.41	S28	27.39	S40	32.59
S04	14.93	S17	23.11	S29	27.45	S41	34.78
S05	16.35	S18	23.19	S30	28.26	S42	35.57
S06	17.62	S19	23.53	S31	28.27	S43	36.05
S07	18.07	S20	23.85	S32	28.29	S44	37.20
S08	18.33	**B01**	**24.65**	S33	28.96	S45	38.15
S09	20.20	S21	24.66	S34	30.01	S46	38.51
S10	20.27	S22	25.10	**B02**	**30.17**	S47	39.06
S11	21.31	S23	25.19	S35	30.72	S48	45.82
S12	21.48	S24	26.21	S36	31.02	D01	7.39
S13	21.99	S25	26.51	S37	31.38	Avg.	25.91

[6] https://catalog.ldc.upenn.edu/ldc93s1
[7] D01 signifies a late submission.
[8] http://www.asvspoof.org/data2017/baseline_CM.zip

Table 5: *Evaluation conditions information: number of trials, mean and standard deviation of the signal-to-noise ratio (SNR) and cepstral distance (CSD), and quality of playback and recording devices (L=low, M=medium, H=high).*

Category	C1	C2	C3	C4	C5	C6
# Replay trials	183	1483	1178	5227	4475	376
SNR μ	28.77	44.81	29.60	34.93	36.05	39.86
SNR σ	7.61	11.22	8.50	8.44	9.59	10.70
CSD μ	0.91	0.82	0.64	0.61	0.45	0.26
CSD σ	0.26	0.12	0.18	0.23	0.18	0.10
Pl. quality	L	L	L/M	L/M	M/H	H
Rec. quality	L/M	L/M	L/M	M/H	M/H	H

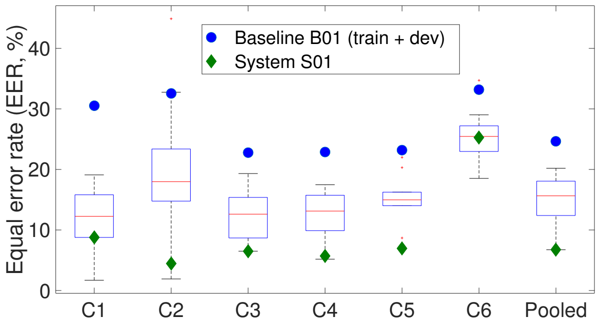

Figure 2: *A boxplot of the top-10 primary systems submitted to ASVspoof 2017 challenge, in terms of EER$_{RD}$ (%), broken down according to the evaluation conditions defined in subsection 2.1. **Pooled** corresponds to the challenge primary metric computed by pooling all the top-10 system detection scores across the six evaluation conditions.*

4.2. Condition analysis

Table 5 characterizes the six evaluation conditions derived from submission scores using the procedure detailed above. Illustrated are the mean and standard deviation of two standard quality measures. The *signal-to-noise ratio* (SNR) reflects the level of background noise and is estimated using the NIST STNR tool[9], while *cepstral distance* (CSD) is a two-sided estimate of the distortion between replay utterances and corresponding source recordings. It reflects average distances of frames of 25ms with 15ms overlap, computed as the Euclidean distance between the long-term average cepstra of replay and corresponding non-replay recordings without the DC coefficient c_0. Low CSD values indicate a high-quality replay sample, i.e. little distortion.

Table 5 shows an almost-consistent correlation between increasing SNR and difficulty. The one exception, namely condition C2, was found to exhibit low background noise but substantial spectral distortion stemming from the use of low quality replay (a netbook) and recording (webcam microphone) devices. Table 5 further indicates that the difficulty of each condition is entirely correlated with the CSD measurements; replay configurations which introduce greater distortion are easier to detect. This is easy to understand remembering that additional noise, reverberation or other distortions induced by low-quality replay attacks push such samples further away from authentic non-replay data. The last two rows of Table 5 illustrate the general quality of the recording/replay devices which characterise each condition. As expected, replay attacks mounted with low (L) and medium (M) quality devices are more easily detected that those mounted with high (H) quality devices.

[9] https://www.nist.gov/sites/default/files/documents/itl/iad/mig/spqa_2-3-sphere_2-5.tgz

4.3. System analysis

Fig. 2 illustrates, independently for each of the 6 conditions besides pooled scores, the performance of the top-10 ranked submissions and the B01 baseline system. Replay detection performance for category C6 is consistently the worst. Performance for the condition C1 is often not the easiest but, for many cases, the variation in performance across C1-C5 is substantial. System S01 is the best performing for 4 of the 6 conditions.

5. Conclusions

The ASVspoof 2017 challenge was highly successful with more than 100 development data requests and nearly 50 challenge entries. The second edition of the challenge is new in several respects; besides new data for the most common spoofing attack replay, speech signals are collected 'in the wild' from a large number of recording conditions. The focus is given for text-dependent ASV scenario where short pass-phrases are used for speaker authentication. This paper summarized the challenge corpus, task, preliminary evaluation results, and categorization of the evaluation data for further analysis.

The best results show an overall detection EER of 6.73% and the average EER of all the primary submission is 25.91%. The detection of replay attack seems to be more difficult than synthetic (artificial) speech detection task. Generalization of detectors remains a challenging open problem when the test data has significant variations within it.

Beyond the challenge evaluation metric with pooled EER, blind categorization of the trials according to the difficulty level needs further investigation. We expect that the challenge data and its evaluation results will help anti-spoofing and ASV researchers in exploring new ideas for developing advanced solutions for fraudulent free authentication.

6. Acknowledgements

The authors would like to express their gratitude for all the ASVspoof 2017 challenge participants. 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. The work also funded in part by Academy of Finland (grant #288558).

7. References

[1] D. Reynolds and R. Rose, "Robust text-independent speaker identification using Gaussian mixture speaker models," *IEEE Trans. on Speech and Audio Processing*, vol. 3, pp. 72–83, January 1995.

[2] T. Kinnunen and H. Li, "An overview of text-independent speaker recognition: from features to supervectors," *Speech Communication*, vol. 52, no. 1, pp. 12–40, January 2010.

[3] J. Hansen and T. Hasan, "Speaker recognition by machines and humans: A tutorial review," *IEEE Signal Proc. Mag.*, vol. 32, no. 6, pp. 74–99, 2015.

[4] Z. Wu, N. Evans, T. Kinnunen, J. Yamagishi, F. Alegre, and H. Li, "Spoofing and countermeasures for speaker verification: A survey," *Speech Comm.*, vol. 66, pp. 130–153, 2015.

[5] Z. Wu, S. Gao, E. Chng, and H. Li, "A study on replay attack and anti-spoofing for text-dependent speaker verification," in *Proc. APSIPA*, 2014, pp. 1–5.

[6] J. Galka, M. Grzywacz, and R. Samborski, "Playback attack detection for text-dependent speaker verification over telephone channels," *Speech Comm.*, vol. 67, pp. 143–153, 2015.

[7] F. Alegre, A. Janicki, and N. Evans, "Re-assessing the threat of replay spoofing attacks against automatic speaker verification," in *Proc. 13th International Conference of the Biometrics Special Interest Group (BIOSIG)*, 2014, pp. 157–168.

[8] T. Stafylakis, M. J. Alam, and P. Kenny, "Text-dependent speaker recognition with random digit strings," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 7, pp. 1194–1203, July 2016.

[9] Q. Li, B.-H. Juang, and C.-H. Lee, "Automatic verbal information verification for user authentication," *IEEE Transactions on Speech and Audio Processing*, vol. 8, no. 5, pp. 585–596, Sep 2000.

[10] T. Kinnunen, M. Sahidullah, I. Kukanov, H. Delgado, M. Todisco, A. Sarkar, N. B. Thomsen, V. Hautamäki, N. Evans, and Z.-H. Tan, "Utterance verification for text-dependent speaker recognition: a comparative assessment using the RedDots corpus," *Proc. INTERSPEECH*, 2016.

[11] C. Ouali, P. Dumouchel, and V. Gupta, "A robust audio fingerprinting method for content-based copy detection," in *Proc. 12th International Workshop on Content-Based Multimedia Indexing (CBMI)*, June 2014, pp. 1–6.

[12] M. Malekesmaeili and R. Ward, "A local fingerprinting approach for audio copy detection," *Signal Processing*, vol. 98, pp. 308 – 321, 2014.

[13] N. Evans, T. Kinnunen, and J. Yamagishi, "Spoofing and countermeasures for automatic speaker verification," in *Proc. INTERSPEECH*, Lyon, France, 2013.

[14] 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. INTERSPEECH*, Dresden, Germany, September 2015, pp. 2037–2041.

[15] Z. Wu, J. Yamagishi, T. Kinnunen, C. Hanilci, M. Sahidullah, A. Sizov, N. Evans, M. Todisco, and H. Delgado, "ASVspoof: the automatic speaker verification spoofing and countermeasures challenge," *IEEE Journal of Selected Topics in Signal Processing*, vol. PP, no. 99, pp. 1–1, 2017.

[16] Z. Wu, P. L. D. Leon, C. Demiroglu, A. Khodabakhsh, S. King, Z. 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 Trans. Audio, Speech & Language Processing*, vol. 24, no. 4, pp. 768–783, 2016.

[17] P. Korshunov *et al.*, "Overview of BTAS 2016 speaker anti-spoofing competition," in *Proc. IEEE BTAS*, 2016, pp. 1–6.

[18] K. Lee, A. Larcher, G. Wang, P. Kenny, N. Brümmer, D. A. van Leeuwen, H. Aronowitz, M. Kockmann, C. Vaquero, B. Ma, H. Li, T. Stafylakis, M. J. Alam, A. Swart, and J. Perez, "The RedDots data collection for speaker recognition," in *Proc. INTERSPEECH*, 2015, pp. 2996–3000.

[19] T. Kinnunen, M. Sahidullah, M. Falcone, L. Costantini, R. G. Hautamäki, D. Thomsen, A. Sarkar, Z.-H. Tan, H. Delgado, M. Todisco, N. Evans, V. Hautamäki, and K. A. Lee, "Reddots replayed: A new replay spoofing attack corpus for text-dependent speaker verification research," in *Proc. ICASSP*, New Orleans, USA, 2017.

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

[21] P. Korshunov and S. Marcel, "Cross-database evaluation of audio-based spoofing detection systems," in *Proc. INTERSPEECH*, San Francisco, USA, 2016, pp. 1705–1709.

[22] M. Todisco, H. Delgado, and N. Evans, "Constant Q cepstral coefficients: A spoofing countermeasure for automatic speaker verification," *Computer Speech & Language*, pp. –, 2017.

[23] D. Paul, M. Sahidullah, and G. Saha, "Generalization of spoofing countermeasures: A case study with ASVspoof 2015 and BTAS 2016 corpora," in *Proc. ICASSP*, 2016.

[24] Y. Qian, N. Chen, and K. Yu, "Deep features for automatic spoofing detection," *Speech Communication*, vol. 85, pp. 43–52, 2016.

[25] S. Ergunay, E. Khoury, A. Lazaridis, and S. Marcel, "On the vulnerability of speaker verification to realistic voice spoofing," in *Proc. IEEE BTAS*, Arlington, USA, Sept. 2015, pp. 1–6.

[26] 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.

[27] H. Delgado, M. Todisco, M. Sahidullah, A. Sarkar, N. Evans, T. Kinnunen, and Z.-H. Tan, "Further optimisations of constant Q cepstral processing for integrated utterance verification and text-dependent speaker verification," in *Proc. IEEE Spoken Language Technology Workshop (SLT)*, 2016, pp. 179–185.

[28] A. Larcher, K. Lee, B. Ma, and H. Li, "Text-dependent speaker verification: Classifiers, databases and RSR2015," *Speech Comm.*, vol. 60, pp. 56–77, 2014.

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