An improved normalized Gain-based score normalization technique for the spoof detection algorithm

Original Scientific Paper

Ankita Chadha
School of Computer Science,
Taylor's University,
Subang Jaya, Selangor, Malaysia 47500
chadhaankitanaresh@sd.taylors.edu.my

Azween Abdullah
School of Computer Science,
Taylor's University,
Subang Jaya, Selangor, Malaysia 47500
azween.abdullah@taylors.edu.my

Lorita Angeline
School of Computer Science,
Taylor's University,
Subang Jaya, Selangor, Malaysia 47500
lorita.angeline@taylors.edu.my

Abstract *– A spoof detection algorithm supports the speaker verification system to examine the false claims by an imposter through careful analysis of input test speech. The scores are employed to categorize the genuine and spoofed samples effectively. Under the mismatch conditions, the false acceptance ratio increases and can be reduced by appropriate score normalization techniques. In this article, we are using the normalized Discounted Cumulative Gain (nDCG) norm derived from ranking the speaker's log-likelihood scores. The proposed scoring technique smoothens the decaying process due to logarithm with an added advantage from the ranking. The baseline spoof detection system employs Constant Q-Cepstral Co-efficient (CQCC) as the base features with a Gaussian Mixture Model (GMM) based classifier. The scores are computed using the ASVspoof 2019 dataset for normalized and without normalization conditions. The baseline techniques including the Zero normalization (Z-norm) and Test normalization (T-norm) are also considered. The proposed technique is found to perform better in terms of improved Equal Error Rate (EER) of 0.35 as against 0.43 for baseline system (no normalization) wrt to synthetic attacks using development data. Similarly, improvements are seen in the case of replay attack with EER of 7.83 for nDCG-norm and 9.87 with no normalization (no-norm). Furthermore, the tandem-Detection Cost Function (t-DCF) scores for synthetic attack are 0.015 for no-norm and 0.010 for proposed normalization. Additionally, for the replay attack the t-DCF scores are 0.195 for no-norm and 0.17 proposed normalization. The system performance is satisfactory when evaluated using evaluation data with EER of 8.96 for nDCG-norm as against 9.57 with no-norm for synthetic attacks while the EER of 9.79 for nDCG-norm as against 11.04 with no-norm for replay attacks. Supporting the EER, the t-DCF for nDCG-norm is 0.1989 and for no-norm is 0.2636 for synthetic attacks; while in case of replay attacks, the t-DCF is 0.2284 for the nDCG-norm and 0.2454 for no-norm. The proposed scoring technique is found to increase spoof detection accuracy and overall accuracy of speaker verification system.*

Keywords*: spoof detection, speaker verification, score normalization, replay attack, voice conversion, speech processing.*

1. INTRODUCTION

The voice of a speaker is a unique way of identifying an individual and signifies various traits of the speaker such as his pitch, pauses, breathiness, and vocal tract length. The authentication based on the voice has gained importance to secure our biometric systems such as phone banking, person identification, voice command devices, voice assistants, and many more [1]. These applications require Automatic Speaker Verification (ASV) to detect enrolled and unknown speakers [2]. The spoof detection algorithm intends to detect the imposter attacks

on the ASV system. Hence, the aim of the spoof detection algorithm is to accurately classify the incoming speech sample as spoofed or genuine speech. These spoofing attacks may be categorized as Logical Access (LA) and Physical Access (PA) [3]. The development of a spoof detection algorithm includes feature representation, model training, and decision making as the major steps. While ASV also implicates score normalization for obtaining standardized scores which is indeed a crucial step in decision making [4]. Hence, considering score normalization for spoof detection is equally important for improvising the detection scores similar to the ASV framework [5]. In absence of score normalization, the variations are seen in the distribution of genuine and spoofed scores for more than one model. This happens for every speaker enrolled during the training. This leads to difficulty in choosing a unique threshold for all the enrolled speaker models. Moreover, a single enrolled speaker is likely to have variation in test utterance distribution due to changed environmental conditions such as acoustic variations, recording environment, language, and gender variations. Thus, developing a scoring technique that overcomes the mismatched conditions observed in the test speech is the essential for contributing to accurate detection of unknown test speech.

The elementary speaker verification is shown in Fig. 1 with two major sub-tasks: training and testing phase. During the training phase, the feature extraction represents the enrolled samples for various types of attacks along with the genuine speech utterances. The commonly used renowned features for spoof detection in ASV framework are Linear Prediction Residual [6], Glottal Flow parameters (GFP) [7], CQCC [3], Line Frequency Cepstral Coefficient (LFCC) [3], Phase based features like Modified Group Delay (MGD) [8] and Deep features [9]. These features have shown significant improvement in the EER. The speaker-specific features are then trained using appropriate machine learning algorithms such as GMM [3], Support Vector Machines (SVM) [10], and deep learning models like Recurrent Neural Networks (RNN) [11], Convolution Neural Network (CNN) [12], [13], Residual Networks [14] etc. During the testing, the unknown utterances are classified as genuine or spoofed speech using the target and imposter models.

2. RELATED WORK

The testing phase may include normalizing the scores by comparing the claimant score to the trained model score. The score distributions from the imposter and genuine speakers are normalized to improve the overall accuracy of the detection system under mismatched conditions [4]. The dissimilarities in scores are observed due to intra-speaker and inter-speaker variations [15]. This work focuses on using a unique score normalization technique for improving the EER and t-DCF of a spoof detection system.

The score normalization process works on a similar principle as that of the basis function in the wavelets, where we scale up or down and shift the score distributions according to individual speaker models for tuning the threshold to a single value. It is also widely used in other speech applications such as speaker recognition [16], [17] and outbreak classification [18]. The main work began in speaker recognition where Z-norm was employed to select speech segments [19]. Since, Z-norm does not consider handset variations, its variant, Handset normalization (H-norm) was also proposed [4]. Following this, the T-norm was used which was based on the test speech signals [20]. Other kinds of normalizations employed include ZT-norm [21], [22], HT-Norm [20], Cellular normalization (C-norm) [23], Symmetric normalization (S-norm) [16] and Distance normalization (D-norm) [16]. In [16], adaptive score normalization has also been proposed for speaker recognition and is found to perform equally well as the S-norm using NIST 2016 dataset. Although the importance of score normalization has proven to be evident in improving the accuracy of ASV, research in spoof detection is scarce [5]. Table 1 shows the research done in score normalization based on the dataset and area of application.

The study of these various normalization schemes yields three important conclusions as highlighted below:

- For score normalization, if prior knowledge about the speech samples is available then it may prove to be beneficial for that scoring technique. To elaborate on this, consider handset, language, and gender-related information to be available; then H-norm and HT-norm may help in boosting the performance of the speaker verification and recognition system. But HT norm also requires high computational time [15]. On the contrary, the knowledge about the handset, language and gender is less likely to be known for an unknown test utterance.
- Some scoring techniques are based on speaker or imposter-centric approaches where cohorts are chosen either closer to the target speaker or imposter speaker. The selection of imposters plays a vital role in such normalization schemes [16]. Still, the prior knowledge about the imposter is rarely known and the cohort selection is incomplete without imposter information.
- The normalization techniques are generally based on the estimation of mean and variance for score distributions which include Z-norm and T-norm [24].

So far, there is no specific set of rules for selecting a score normalization technique and the dataset used for scoring is not uniform due to its application in various domains (as seen from Table 1). Based on the authors' knowledge, there is no scoring technique based on cumulative gain using the rank of the speaker and has not been used for spoof detection task. Hence in this work, the nDCG-norm is used to regulate the loglikelihood scores which show promising results with a reduction in EER and t-DCF scores. The nDCG-norm is based on the goodness of ranking as well as the cumulative accumulation of relevance of the scores.

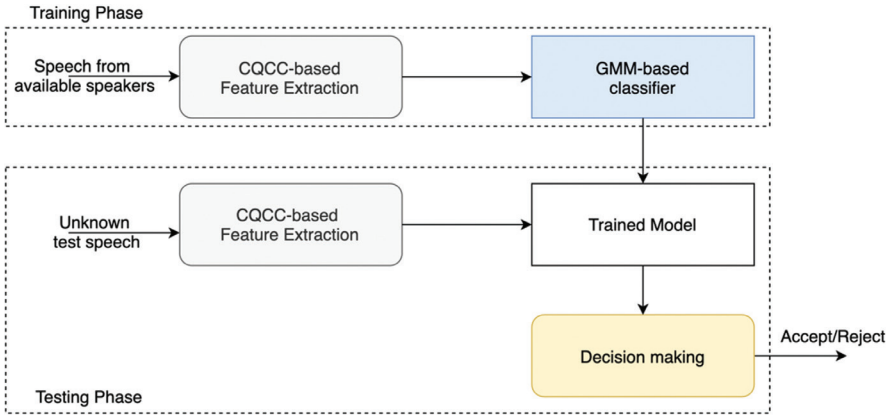


Fig. 1 Generic Speaker verification system.

Table 1. Research work in score normalization wrt datasets and various speech domains.

Speech Application	Author	Year	Normalization	Datasets
Speaker Verification	Auckenthaler et. al [20] 2000	T-norm	NIST 1997	
	Castro et. al [25]	2006	Kullback-Leiber – T-norm	NIST 2005
	Kenny et. al [21]	2008	T-norm, Z-norm, ZT-norm	NIST 2006
	Villalba et. al [26]	2011	ZTnorm	NIST SRE 2008
	Kinnunen et. al [27]	2012	ZT norm	NIST 2006
	Kons et. al [22]	2013	ZT norm	WF corpus
	Alegre et. al [28]	2014	T-norm	NIST 2005 and NIST 2006
	Khemiri et. al [29]	2016	T-norm	RSR 2015
	Li and Wang [30]	2016	Cohort scores	CSLT- DSDB
	Tong et. al [31]	2020	Adaptive (A) scoring	CH Data, Voxceleb2 and FFSV 2020
	Sahidullah et. al [32]	2020	AS-norm	SdSv challenge dataset
	Zhao et. al [33]	2021	S-norm	Voices 2019
Speaker Recognition	Matejka et. al [16]	2017	S-norm	NIST 2016
	Swart and Brummer [17]	2017	Generative scoring	RSR 2015
Spoof detection - Replay Attack	Shang, Stevenson [5]	2010	Mean, standard deviation	Custom made

Thus, the objectives of this work are three-fold:

I. Exploring nDCG-norm for computing normalized scores for LA and PA attacks.

II. Investigating the performance of nDCG-norm using EER, t-DCF, and Detection Error Tradeoff (DET) curve for the spoof detection task.

III. Comparing the proposed score normalization technique with baseline no normalization, state-of-the-art Z-norm, and T-norm-based scoring algorithms.

This article is structured as follows: Section 3 describes the Baseline Techniques and Section 4 includes Proposed Score Normalizing technique respectively. Section 5 discusses the Experimental setup and results of this work. Lastly, the conclusion of this work can be found in Section 6.

3. BASELINE TECHNIQUES

The effect of score normalization is visible on the decision accuracy. Although the literature in score normalization is two decades old, its progress is sluggish with lack of work done in spoof detection domain. Hence, there is a vivid need to explore score normalization for spoof detection task as well. The objective of the score normalizing technique is to decrease intra-speaker variations which leads to better accuracy, score calibration, and improved threshold selection. This section elaborates the baseline CQCC -GMM detection along with baseline score normalization techniques including Z-norm and T-norm as shown in Fig. 2.

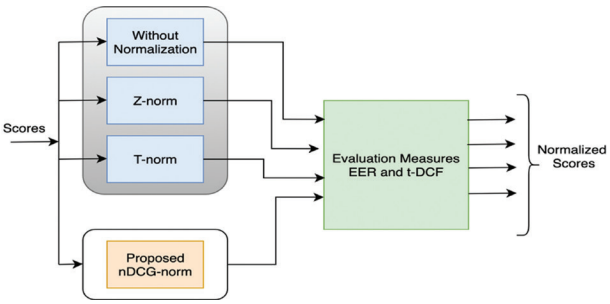


Fig. 2. Score Normalization Process – no normalization, Z-norm, T-norm and Proposed nDCG norm algorithm.

3.1 CQCC-GMM SPOOF DETECTION

The CQCC features were introduced after the ASVspoof 2015 challenge to detect the S10 attack (an attack found to be difficult to detect in the ASVspoo2015 challenge) and were also the state-of-the-art features for the ASVspoof 2019 challenge [34]. These features are based on CQT rather than discrete Fourier Transform as they promote temporal content present at higher frequencies. This is an important criterion for distinguishing genuine speech from spoofed speech.

The state-of-the-art GMM classifier is commonly used in spoof detection scheme due to its ability to perform well and capture generality in the data [35]. The task of a GMM classifier or detector is to categorize the input unknown test sample as genuine or spoofed. This is done by computing log-likelihood scores from the individual trained model – genuine speech (Θ_{gen}) and spoofed speech model (Θ_{spoof}). Hence, while testing the unknown test speech (s), the difference in log-likelihood (l) can be computed using equation (1).

$$Score\ (s) = \ell(\Theta_{gen}(s)) - \ell(\Theta_{\text{spoof}}(s)) \tag{1}$$

3.2 SCORE NORMALIZATION

The general scores are resultant of enrolled (r) and test speech which is denoted as score(r,s). The likelihood of a speaker model Θ (speaker model consists of mixture weights) with the extracted feature set $Y=\{y_1, y_2, ..., y_m\}$ where m is the number of utterances, is given in equation (2).

$$\text{score} = \mathcal{L}(Y|\Theta) = \sum_{m=1}^{M} \log p(y_m|\Theta) \tag{2}$$

3.2.1 Zero Normalization (Z-norm)

The most reliable and simplest form of normalization that is based on the estimation of mean and variance for the genuine or target speaker distribution is Z-norm [29]. The important highlight of the Z-norm is that it doesn't need to perform online permutations during the training process. The trained speaker model is compared to the subset of enrolled samples following which mean μ_r and variance δ_r are estimated. The Z-norm score $C_{z\text{-}norm}$ normalized can be computed as shown in equation (3)

$$C_{z-norm} = \frac{\log p(Y|\Theta) - \mu_r}{\delta_r} \tag{3}$$

3.2.2 Test-Normalization (T-norm)

The T-norm is based on a similar principle to Z-norm except for the imposter score distribution [16]. This arrangement boosts the accurate distribution of cohort samples because of the variance and can be computed as shown in equation (4)

$$C_{t-norm} = \frac{\log p(Y|\Theta) - \mu_s}{\delta_s} \tag{4}$$

Where, μ_s and δ_s are mean and variance of imposter cohort score distribution.

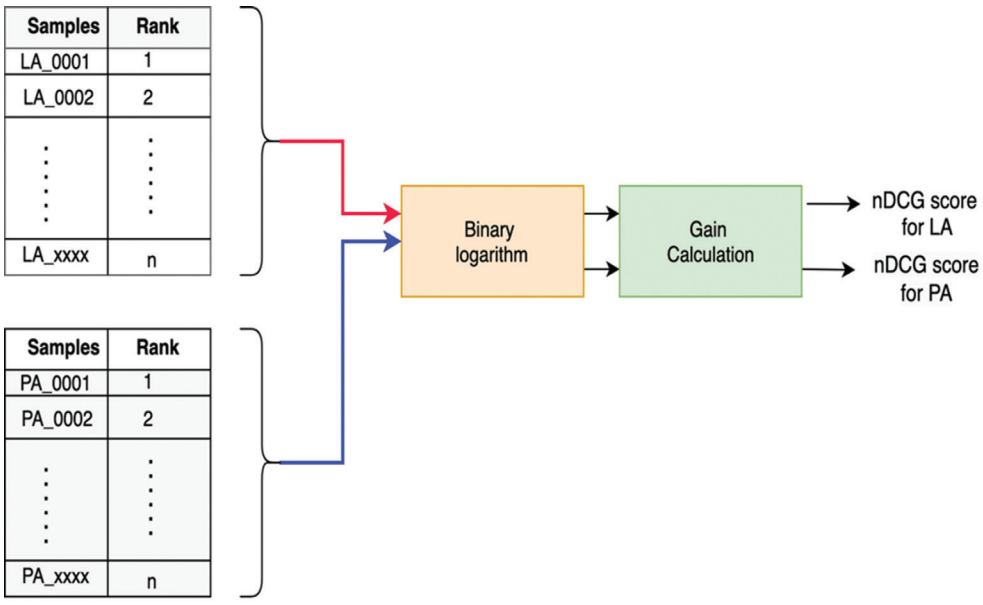


Fig. 3 Steps to compute proposed nDCG normalization.

4. PROPOSED SCORE NORMALIZING TECHNIQUES

The experiments conducted in the score normalization have revealed the difficulty of various parameters that need to be considered before selecting an appropriate scoring technique. Some of these parameters are the number of speakers, the number of utterances, language dependency, handset reliability, challenges of pseudo imposters, speaker dependency, gender variations, and test data dependency which influence the performance of the scoring technique. Hence, there is a need for a more reliable scoring technique that consid-

ers variations between the training and testing phase for both known as well as unknown speakers. In this work, we propose an nDCG-norm that works on a similar principle to reduce false acceptance ratios and cumulate the score relevance through the ranking of the speaker samples. To elaborate further, Fig. 3 depicts the computation steps for the proposed scoring technique.

The scores for LA and PA attacks are normalized separately. The scores are ranked based on their degree of score value and then scaled using the binary logarithm. The nDCG-norm can be calculated as shown in equations (5) and (6) [36].

$$DCG = \sum_{i=1}^{K} \frac{N_i}{log_2(i+1)} \quad (5)$$

$$nDCG = \frac{DCG}{IDCG} \quad (6)$$

Where K is the number of test samples, N_i is the rank of ith sample and inverse DCG is the reverse order rank DCG of the score distribution. The nDCG-norm does not require cohort score and hence, the difficulty of choosing the cohort data is averted in contrast to state-of-the-art Z-norm and T-norm. Furthermore, the nDCG-norm when used as a part of spoof detection framework may reduce EER subsequently. So to evaluate its performance, the t-DCF score and DET are also employed.

5. EXPERIMENTAL SETUP AND RESULTS

The baseline ASV for developing a spoof detection system is the CQCC-GMM algorithm whose log-likelihood scores are considered for normalization. The 30 coefficient CQCC includes the delta and double delta coefficients and GMM is used as a two-class classifier with 512 components [37]. To evaluate the proposed and baseline normalization techniques, we used the ASV spoof 2019 dataset [3] which includes all three attacks including voice converted speech, text-to-speech (TTS) [36], and replay speech. For objective evaluation, the EER [39] and t-DCF are used to measure the performance of the score normalization techniques along with the DET curve [15] on the test dataset. The corpus and results are elaborated below in sub-sections.

5.2 ASV SPOOF 2019 CORPUS

The ASV spoof 2019 [3] corpus is adapted from the VCTK dataset [40] [41] which comprises of a separate data for LA and PA attacks. The LA dataset has synthetic speech while the PA dataset includes the replay speech. The corpus is split into three parts: training subset with 20 speakers (12 Female, 8 Male), development subset with 10 speakers (6 Female, 4 Male), and unknown speaker-based test data with nearly 48 speakers (27 Female, 21 Male). In this work, the baseline spoof detection system is trained using a training and development subset of the data and evaluated using unknown test data.

5.3 EXPERIMENTAL RESULTS

The spoof detection algorithm needs to be evaluated for measuring its performance and susceptibility to various attacks. This is possible through objective measures including EER and t-DCF functions. The EER measures the ratio of false positives to the false negatives and its value must be as low as possible. The t-DCF is the most important metric that calculates the error between the speaker verification system and its counter-measure or spoof detection system.

Table 2. EER and t-DCF for LA and PA attack for Baseline and Proposed normalization schemes.

Type of Attack	Type of Scoring	Development Dataset		Evaluation Dataset	
		EER	t-DCF	EER	t-DCF
LA	Baseline with no normalization	0.4311	0.01564	9.57	0.2366
	Baseline Z-norm	0.4302	0.01298	9.32	0.2298
	Baseline T-norm	0.4299	0.01267	9.15	0.2207
	Proposed nDCG-norm	0.3571	0.01037	8.96	0.1989
PA	Baseline with no normalization	9.87	0.1953	11.04	0.2454
	Baseline Z-norm	9.51	0.1921	10.87	0.2395
	Baseline T-norm	8.76	0.1865	10.66	0.239
	Proposed nDCG-norm	7.83	0.1782	9.79	0.2284

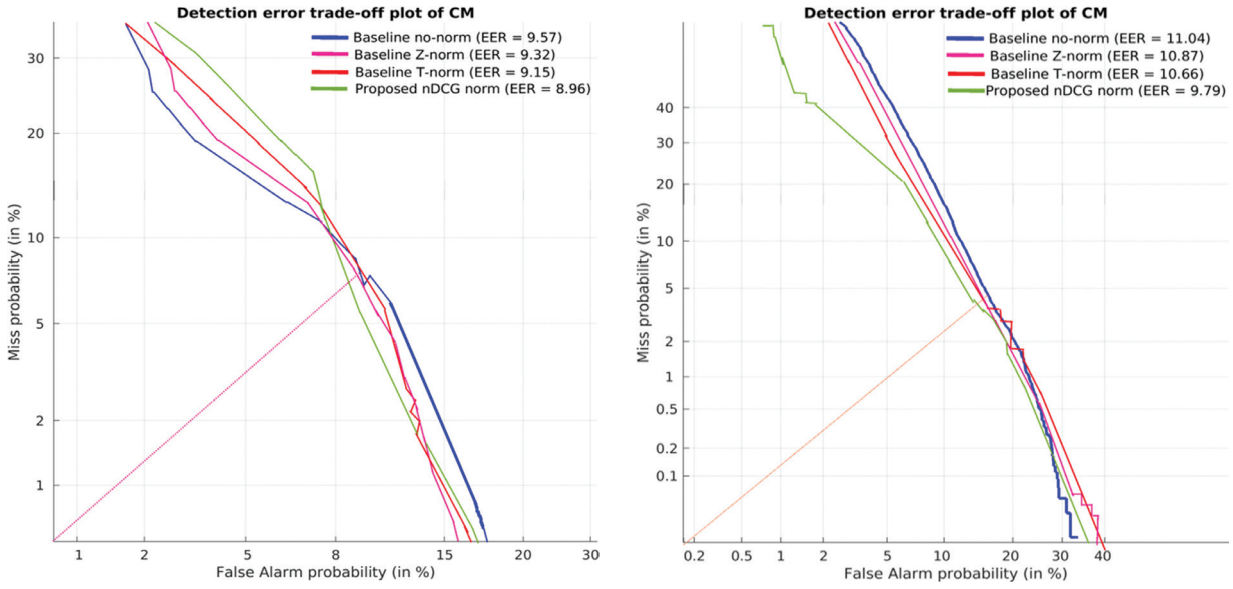
Its value must lie between 0 and 1 where 0 implies error-free between verification and counter-measure while 1 means no further improvement can be seen in spoof detection [38].

Table 2 shows EER and t-DCF scores for LA and PA attacks using development and evaluation dataset. The baseline scores include non-normalized scores, Z-norm, and T-norm scores as against the proposed nDCG-based scoring technique.

The Z-norm and T-norm scores show negligible improvements in the decision accuracy of the spoofing algorithm. This might be due to the lower number of speakers in the training data. On the contrary, the nDCG-norm does not dependent on the number of

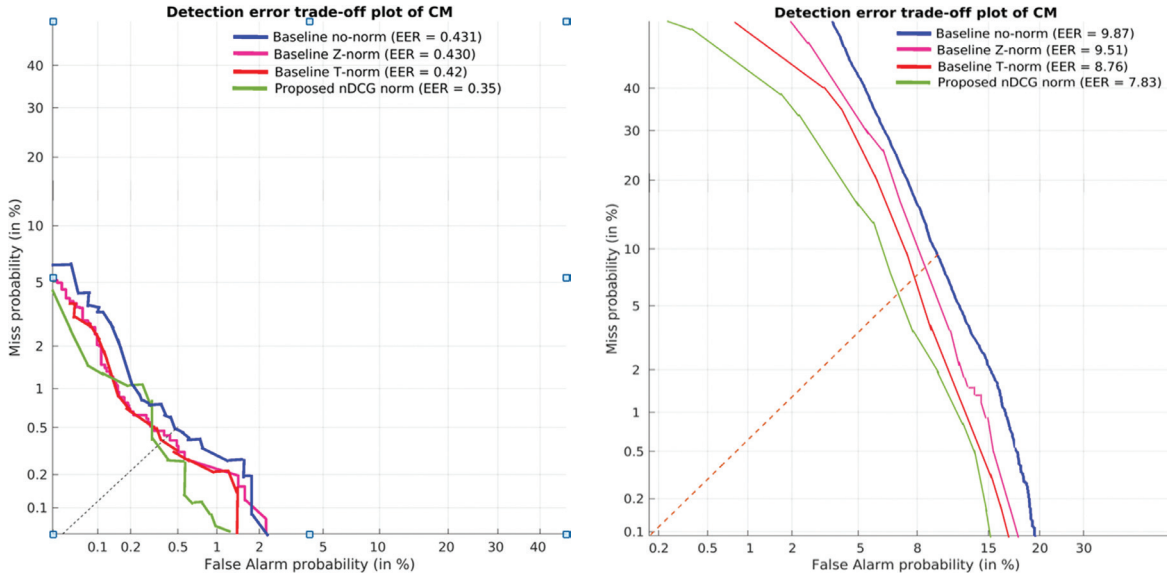
speakers. During the development stage for LA attacks, the EER for nDCG-norm is 0.35 in contrast to 0.43 with no normalization while t-DCF is 0.0103 for nDCG-norm and 0.015 for no-norm. Similarly, during the evaluation stage for LA attacks, the EER for nDCG-norm is 8.96 as against 9.57 with no-norm while t-DCF is 0.1989 for nDCG-norm and 0.2366 with no-norm. Furthermore, during development stage for PA attacks, the EER for nDCG-norm is 7.83 in contrast to 9.87 with no-norm while the t-DCF scores are 0.1782 for nDCG-norm and 0.1953 for no-norm. Similarly, during the evaluation stage for PA attacks, the EER for nDCG-norm is 9.79 as against 11.04 with no-norm while t-DCF is 0.2284 for nDCG-norm and 0.2454 with no-norm.

The nDCG-norm performs better than the system with no normalization and the baseline scoring techniques. The main reason for improvements is due to no involvement of cohort in nDCG computation; simply, the speaker-based ranking and gain computations are carried out. The selection of cohort is laborious and involves no ground rule but surely depends on the number of spoofed speakers that sound more like the genuine speakers individually. To support the above objective measures, the DET plots are used to show the relation between False-Acceptance Ratio (FAR) and False Rejection Ratio (FRR) as shown in Fig. 4 and Fig. 5 for LA and PA attacks using development and evaluation dataset respectively.



(a) (b)

Fig. 4. DET plot for Baseline no norm, Z-norm, T-norm and Proposed nDCG norm spoof detection system using evaluation data based on – (a) LA attacks (b) PA attacks.



(a) (b)

Fig. 5. DET plot for Baseline no norm, Z-norm, T-norm and Proposed nDCG norm spoof detection system using development data based on – (a) LA attacks (b) PA attacks.

In case of LA attack (Fig. 4(a) and Fig. 5(a)), the DET curves for no normalization, Z-norm and T-norm show slight variation in slope and operating point of the system i.e. the EER. The nDCG-norm shows significant improvement in lowering the false positives as compared to the other three baseline techniques. While on the other hand, for PA attack (Fig. 4(b) and Fig. 5(b)), the slope for no-norm and Z-norm are similar, with minute variation in slope is observed for T-norm. The nDCG-norm has an improved slope implying reduced false positives ad increased in true values. Overall EER and t-DCG scores are reduced for proposed normalization as against Z-norm and T-norm score. Moreover, the normalization of scores is proven to influence the accuracy of the spoof detection system than with no normalization. It is also rightful to state that the overall ASV performance is thus improved.

6. CONCLUSION

The task of spoof detection is challenging yet crucial for stimulating secure environments for imposter-resistant networks including the ASV framework. The score normalization is not a compulsory but necessary step in improving the decision accuracy of the ASV. In this work, a unique score normalization technique is proposed for the spoof detection task. The proposed nDCG-norm is found to perform equally well in contrast to state-of-the-art normalization schemes. Moreover, the EER and t-DCF for all the baseline techniques are higher than the proposed scoring technique including LA and PA attacks. In the case of LA attacks, the nDCG-norm achieved an EER of 0.35 and t-DCF of 0.01 which is superior to the EER of 0.43 and t-DCF of 0.015 for the baseline technique with no normalization during the development stage. Further nDCG-norm achieved an EER of 8.96 and t-DCF of 0.198 which is superior to the EER of 9.57 and t-DCF of 0.236 for the baseline technique with no normalization during the evaluation stage. Additionally, considering PA attacks, the EER is 9.87 and t-DCF is 0.19 for no-norm in the development stage, with no major variations observed for Z-norm and T-norm; while a significant reduction in EER of 7.83 and t-DCF of 0.17 are observed for nDCG-norm. Similarly, during the evaluation stage, the EER is 11.04 and t-DCF is 0.245 for no-norm, whereas improved EER of 9.79 and t-DCF of 0.228 are obtained for proposed nDCG-norm.

The overall objective of improving accuracy by reducing the false positives is achieved by the proposed score normalization technique. Moreover, the simplicity of extraction of nDCG-norm and lower computation complexity makes it potentially viable in the post-processing stage of the spoof detection algorithm. In the future, this work can be extended for feature normalization and investigating an alternative for rank selection.

7. REFERENCES

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

[2] M. Sahidullah, H. Delgado, M. Todisco, T. Kinnunen, N. Evans, J. Yamagishi, K.-A. Lee, "Introduction to Voice Presentation Attack Detection and Recent Advances", Advances in Computer Vision and Pattern Recognition, 2019, pp. 321–361.

[3] J. Yamagishi et al. "ASVspoof 2019: Automatic Speaker Verification Spoofing and Countermeasures Challenge Evaluation Plan", 2019, https://www.asvspoof.org/asvspoof2019/asvspoof2019_evaluation_plan.pdf (accessed: 2022)

[4] D. A. Reynolds, "Comparison of background normalization methods for text-independent speaker verification", Proceedings of the 5th European Conference on Speech Communication and Techology, 22-25 September 1997.

[5] W. Shang, M. Stevenson, "Score normalization in playback attack detection", Proceedings of the International Conference on Acoustics, Speech and Signal Processing, Dallas, TX, USA, 14-19 March 2010, pp. 1678–1681.

[6] K. Phapatanaburi, L. Wang, S. Nakagawa, M. Iwahashi, "Replay Attack Detection Using Linear Prediction Analysis-Based Relative Phase Features", IEEE Access, Vol. 7, 2019, pp. 183614–183625.

[7] A. Chadha, A. Abdullah, L. Angeline, "A Unique Glottal Flow Parameters based Features for Anti-spoofing Countermeasures in Automatic Speaker Verification", International Journal of Advanced Computer Science and Applications, Vol. 12, No. 8, 2021, pp. 827–835.

[8] I. Saratxaga, J. Sanchez, Z. Wu, I. Hernaez, E. Navas, "Synthetic speech detection using phase information", Speech Communication, Vol. 81, Jul. 2016, pp. 30–41.

[9] Y. Qian, N. Chen, K. Yu, "Deep features for automatic spoofing detection", Speech Communication, Vol. 85, 2016.

[10] B. Nasersharif, M. Yazdani, "Evolutionary fusion of classifiers trained on linear prediction based fea-

tures for replay attack detection", Expert Systems, Vol. 38, No. 3, 2021.

[11] A. Gómez Alanís, A. M. Peinado, J. A. Gonzalez, A. Gomez, "A Deep Identity Representation for Noise Robust Spoofing Detection", Proceedings of Interspeech, Hyderabad, India, 2-6 September 2018, pp. 676–680.

[12] I. Himawan, S. Madikeri, P. Motlicek, M. Cernak, S. Sridharan, C. Fookes, "Voice Presentation Attack Detection Using Convolutional Neural Networks", Advances in Computer Vision and Pattern Recognition, 2019, pp. 391–415.

[13] H. Tak, J. Patino, M. Todisco, A. Nautsch, N. Evans, A. Larcher, "End-to-End anti-spoofing with RawNet2", Proceedings of Interspeech, 2021, pp. 6369–6373.

[14] Y. Zhang, F. Jiang, Z. Duan, "One-class Learning Towards Synthetic Voice Spoofing Detection", IEEE Signal Processing Letters, Vol. 28, 2020, pp. 937–941.

[15] F. Bimbot et al. "A Tutorial on Text-Independent Speaker Verification", EURASIP Journal on Advances in Signal Processing, Vol. 2004, No. 4, 2004, p. 101962.

[16] P. Matějka, O. Novotný, O. Plchot, L. Burget, M. D. Sánchez, J. Černocký, "Analysis of Score Normalization in Multilingual Speaker Recognition", Proceedings of Interspeech, Stockholm, Sweden, 20-24 August 2017, pp. 1567–1571.

[17] A. Swart, N. Brümmer, "A Generative Model for Score Normalization in Speaker Recognition", Proceedings of Interspeech, Stockholm, Sweden, 20-24 August 2017, pp. 1477–1481.

[18] Z. Mustaffa, Y. Yusof, "A comparison of normalization techniques in predicting dengue outbreak", Proceedings of the International Conference on Business and Economics Research, Kuala Lumpur, Malaysia, Vol. 1, 2011.

[19] K.-P. Li, J. E. Porter, "Normalizations and selection of speech segments for speaker recognition scoring", Proceedings of the International Conference on Acoustics, Speech, Signal Processing, New York, NY, USA, 11-14 April 1988, pp. 595–598.

[20] R. Auckenthaler, M. Carey, H. Lloyd-Thomas, "Score normalization for text-independent speaker verification systems", Digital Signal Processing: A Review Journal, Vol. 10, No. 1, 2000.

[21] P. Kenny, P. Ouellet, N. Dehak, V. Gupta, P. Dumouchel, "A study of interspeaker variability in speaker verification", IEEE Transactions on Audio, Speech and Language Processing, Vol. 16, No. 5, 2008.

[22] Z. Kons, H. Aronowitz, "Voice transformation-based spoofing of text-dependent speaker verification systems", Proceedings of Interspeech, Lyon, France, 25-29 August 2013, pp. 945–949.

[23] D. A. Reynolds, "Channel robust speaker verification via feature mapping", Proceedings of the International Conference on Acoustics, Speech, Signal Processing, Vol. 2, Hong Kong, China, 6-10 April 2003, pp. II-53–6.

[24] A. Mittal, M. Dua, "Automatic speaker verification systems and spoof detection techniques: review and analysis", International Journal of Speech Technology, 2021.

[25] D. Ramos-Castro, J. Fierrez-Aguilar, J. Gonzalez-Rodriguez, J. Ortega-Garcia, "Speaker verification using speaker- and test-dependent fast score normalization", Pattern Recognition Letters, Vol. 28, No. 1, 2007, pp. 90–98.

[26] J. Villalba, E. Lleida, "Detecting replay attacks from far-field recordings on speaker verification systems", Lecture Notes in Computer Science, Vol. 6583, 2011.

[27] T. Kinnunen, Z. Wu, K. A. Lee, F. Sedlak, E. S. Chng, H. Li, "Vulnerability of speaker verification systems against voice conversion spoofing attacks: The case of telephone speech", IEEE International Conference on Acoustics, Speech and Signal Processing, Kyoto, Japan, 25-30 March 2012.

[28] F. Alegre, A. Janicki, N. Evans, "Re-assessing the threat of replay spoofing attacks against automatic speaker verification", Lecture Notes in Informatics, Vol. P-230, 2014.

[29] H. Khemiri, D. Petrovska-Delacretaz, "Cohort selection for text-dependent speaker verification score normalization", Proceedings of the 2nd International Conference on Advanced Technologies for Signal and Image Processing, Monastir, Tunisia, 21-23 March 2016, pp. 689–692.

[30] L. Li, D. Wang, C. Zhang, T. F. Zheng, "Improving Short Utterance Speaker Recognition by Modeling Speech Unit Classes", IEEE/ACM Transactions on Audio, Speech, Language Processing, Vol. 24, No. 6, 2016, pp. 1129–1139.

[31] Y. Tong, W. Xue, S. Huang, L. Fan, C. Zhang, G. Ding, X. He, "The JD AI Speaker Verification System for the FFSVC 2020 Challenge", Proceedings of Interspeech, Shanghai, China, 25-29 October 2020, pp. 3476–3480.

[32] M. Sahidullah et al. "UIAI System for Short-Duration Speaker Verification Challenge 2020", Proceedings of the IEEE Spoken Language Technology Workshop, Shenzhen, China, 19-22 January 2021, pp. 323–329.

[33] L. Zhao, M.-W. Mak, "Channel Interdependence Enhanced Speaker Embeddings for Far-Field Speaker Verification", Proceedings of the 12th International Symposium on Chinese Spoken Language Processing, Hong Kong, China, 24-27 January 2021, pp. 1–5.

[34] H. Tak, J. Patino, A. Nautsch, N. Evans, M. Todisco, "An explainability study of the constant Q cepstral coefficient spoofing countermeasure for automatic speaker verification", Proceedings of the The Speaker and Language Recognition Workshop, Tokyo, Japan, 1-5 November 2020, pp. 333–340.

[35] M. G. Kumar, S. R. Kumar, M. S. Saranya, B. Bharathi, H. A. Murthy, "Spoof Detection Using Time-Delay Shallow Neural Network and Feature Switching", Proceedings of the Automatic Speech Recognition and Understanding Workshop, Singapore, 14-18 Dcember 2019, pp. 1011–1017.

[36] Y. Wang, L. Wang, Y. Li, D. He, W. Chen, T. Y. Liu, "A theoretical analysis of NDCG ranking measures", Journal of Machine Learning Research, Vol. 30, 2013.

[37] G. Lavrentyeva, S. Novoselov, A. Tseren, M. Volkova, A. Gorlanov, A. Kozlov, "STC Antispoofing Systems for the ASVspoof2019 Challenge", Proceedings of the Annual Conference of the International Speech Communication Association, Graz, Austria, September 2019, pp. 1033–1037.

[38] S. Mo, H. Wang, P. Ren, T.-C. Chi, "Automatic Speech Verification Spoofing Detection", arXiv:2012.08095, 2020.

[39] B. Chettri, D. Stoller, V. Morfi, M. A. M. Ramírez, E. Benetos, B. L. Sturm, "Ensemble Models for Spoofing Detection in Automatic Speaker Verification", Proceedings of the Annual Conference of the International Speech Communication Association, September 2019, pp. 1018–1022.

[40] C. Veaux, J. Yamagishi, K. MacDonald, "CSTR VCTK Corpus: English Multi-speaker Corpus for CSTR Voice Cloning Toolkit", The Centre for Speech Technology Research, 2016, https://datashare.ed.ac.uk/handle/10283/3443 (accessed: 2022)

[41] J. Yamagishi et al. "ASVspoof 2021 : accelerating progress in spoofed and deepfake speech detection", Proceedings of the ASVspoof 2021 Workshop - Automatic Speaker Verification and Spoofing Coutermeasures Challenge, September 2021.