ASVspoof 2015: the first automatic speaker verification spoofing and countermeasures challenge

Zhizheng Wu, Tomi Kinnunen, Nicholas Evans, Junichi Yamagishi

November 20, 2015

Mounting evidence has exposed the potential vulnerability of biometric authentication systems to spoofing[1] [1]. In response, there has been a movement in the academic community over the last two decades to develop spoofing countermeasures. The research is now relatively mature, with several competitive evaluations having been organised for various modalities including face [2], fingerprint [3] and iris [4] recognition.

The first significant action within the speaker recognition community involved the organisation of a Special Session at Interspeech 2013 entitled 'Spoofing and Countermeasures for Automatic Speaker Verification' [5]. An IEEE SLTC newsletter article [6] released in conjunction with that initiative set out the importance to research progress of standard datasets, protocols and metrics.

The authors of this article subsequently embarked upon the preparation of the first standard evaluation for automatic speaker verification (ASV). This came to fruition in the form of ASVspoof: the Automatic Speaker Verification Spoofing and Countermeasures Challenge [7]. The first edition, ASVspoof 2015 [8], was held as a special session at Interspeech 2015. This article outlines the challenge, participation, results and the organisers' plans for a second edition.

ASVspoof

In 2013 there were no standard datasets, protocols or metrics to gauge and tackle the threat of spoofing to ASV systems. ASVspoof was conceived to fill the void through the provision of an evaluation platform which allowed, for

[1] Spoofing attacks are also known as *presentation attacks* in ISO nomenclature

the first time, results generated by different research teams to be compared reliably.

Whereas it is the impact of spoofing and countermeasures on ASV performance that is of the greatest interest, ASVspoof focused exclusively on spoofing detection, that is to say, detection in isolation from ASV. This choice was adopted mainly for simplicity; decoupling the evaluation of standalone spoofing detection from the complexities of ASV avoids certain complications related to what would otherwise involve the joint evaluation of two, coupled systems [9]. This strategy furthermore helped to maximise participation which required no prerequisite expertise in ASV.

The first edition of ASVspoof focused on the detection of speech synthesis and voice conversion spoofing attacks. Furthermore the evaluation aimed to promote the development of generalised countermeasures, namely those with greater potential to detect unforeseen spoofing attacks. Generalised countermeasures are expected to have greater relevance in the practical context in which the nature of a spoofing attack can never be known with certainty.

Dataset

The ASVspoof 2015 dataset is based upon the Spoofing and Anti-Spoofing (SAS) corpus [10] and contains speech collected from 106 speakers. The dataset comprises three distinct, non-overlapping subsets for training, development and evaluation. The training and development subsets contain both genuine speech and spoofed speech. All examples of the latter are generated artificially using one of five different, well-known speech synthesis or voice conversion spoofing algorithms and one of two different vocoders.

The evaluation subset also includes genuine speech and almost 200k spoofed speech utterances, this time generated with one of 10 different spoofing algorithms. They include the same five algorithms used to generated the training and development subsets (referred to as known attacks) and five additional algorithms (referred to as unknown attacks not used elsewhere). In this sense, better performance is expected in the case that countermeasures generalise well to unknown spoofing attacks.

Impact upon speaker verification

With the focus of ASVspoof 2015 on spoofing detection alone, it was important to verify first that the spoofing attacks are indeed successful in manipulating an ASV system. Figure 1 illustrates detection error trade-off

(DET) profiles for a standard, baseline iVector-PLDA system and the same system subjected to each of the 10 different spoofing attacks (S1-S10) included in the ASVspoof 2015 dataset. These results show that they all pose a genuine threat to ASV performance, with the baseline equal error rate of 2% increasing to between 3% and 44%.

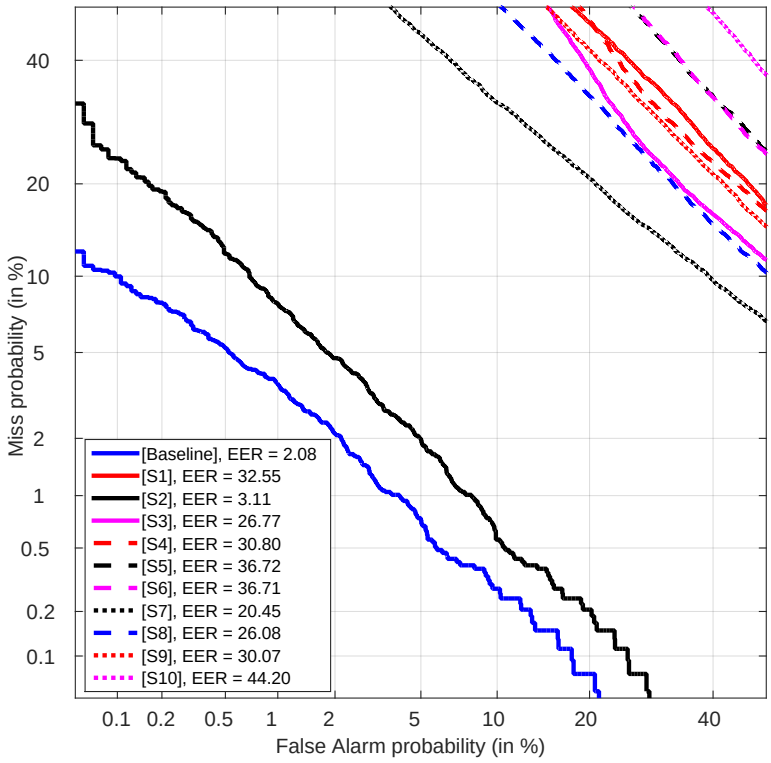

Figure 1: DET profiles of ASV performance for the iVector-PLDA baseline and the same system subjected to each of the 10 different spoofing attacks (S1-S10).

Results

The organisers received 43 submissions from 16 participants. Figure 2 shows a Tukey boxplot of the 16 primary submission results for the evaluation set, sub-divided across known and unknown spoofing attacks. The left, red box shows that the known attacks, for which training and development data are provided, are detected relatively easily. The right, green box shows

considerably higher error rates are achieved for the unknown attacks, for which no training or development data are not provided.

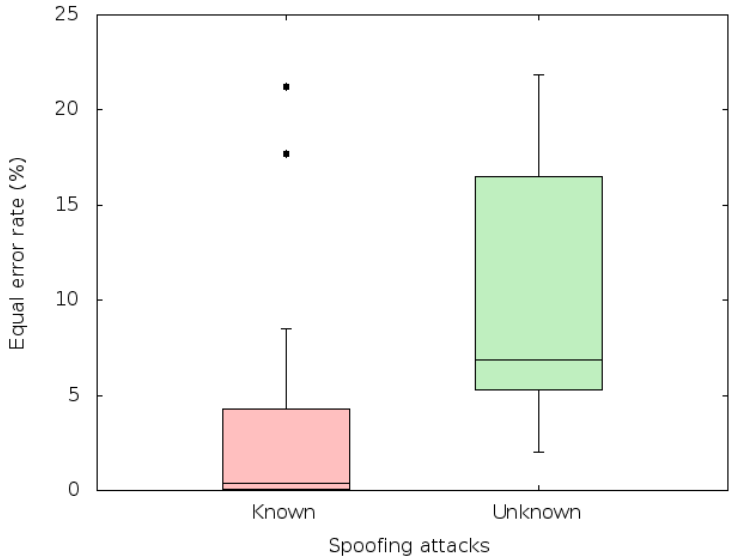

Figure 2: Tukey boxplots of spoofing detection performance. The plot illustrates the distribution of equal error rates for the 16 primary submissions to ASVspoof 2015 and for known and unknown spoofing attacks.

The system that achieved the best results for unknown spoofing attacks [11] achieved an equal error rate of 2% cf. 0.5% for known attacks. Even if these error rates are low, that for unknown spoofing attacks is still four times higher than that for known attacks. Furthermore, even small spoofing detection error rates can impact significantly on the false alarm rate of an ensuing ASV system. The problem of spoofing detection is thus very much far from being solved.

The future

The ASVspoof special session at Interspeech 2015 included a lively open feedback and discussion session. Ideas for future consideration include a broader variety of spoofing attacks including impersonation and replay, the consideration of text-dependent speaker recognition, noise and channel effects, a comparison to human liveness detection, speaker dependent countermeasures and a study of the data quantities used to effect spoofing attacks.

There was also a suggestion to develop more appropriate metrics for the study of spoofing and countermeasures.

More general feedback questioned the focus on spoofing as only one example of how ASV technology may be compromised and also suggested that greater insight into these threats might be gained from collaboration with the forensic and other related communities, e.g. those of voice conversion and speech synthesis. All of these issues will contribute to the authors' plans for a future, second edition of ASVspoof, tentatively planned for 2017.

Further information

Further information, including the referenced evaluation plan, summary and survey papers, in addition to copies of all ASVspoof participants' papers can be found on the ASVspoof website:

http://www.spoofingchallenge.org

The ASVspoof database can be downloaded here:

http://dx.doi.org/10.7488/ds/298

The organisers can be contacted by email:

asvspoof2015@spoofingchallenge.org

Acknowledgements

The organisers wish to thank the following for their help in creating the ASVspoof 2015 database: Daisuke Saito (University of Tokyo), Tomoki Toda (Nara Institute of Science and Technology), Ali Khodabakhsh (Ozyegin University), Cenk Demiroglu (Ozyegin University) and Zhen-Hua Ling (University of Science and Technology of China). They also wish to thank Cemal Hanilçi, Md Sahidullah and Aleksandr Sizov (University of Eastern Finland) for their contribution to the design and validation of the ASVspoof 2015 protocols.

References

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

[2] M. M. Chakka, A. Anjos, S. Marcel, R. Tronci, D. Muntoni, G. Fadda, M. Pili, N. Sirena, G. Murgia, M. Ristori, F. Roli, J. Yan, D. Yi, Z. Lei, Z. Zhang, S. Z. Li, W. R. Schwartz, A. Rocha, H. Pedrini, J. Lorenzo-Navarro, M. Castrillon-Santana, J. Maatta, A. Hadid, and M. Pietikainen, "Competition on counter measures to 2-D facial spoofing attacks," in *Biometrics (IJCB), 2011 International Joint Conference on*, Oct 2011, pp. 1–6.

[3] G. Marcialis, A. Lewicke, B. Tan, P. Coli, D. Grimberg, A. Congiu, A. Tidu, F. Roli, and S. Schuckers, "First International Fingerprint Liveness Detection Competition – LivDet 2009," in *Image Analysis and Processing ICIAP 2009*, P. Foggia, C. Sansone, and M. Vento, Eds., vol. 5716 of *Lecture Notes in Computer Science*, pp. 12–23. Springer Berlin Heidelberg, 2009.

[4] D. Yambay, J. S. Doyle, K. W. Bowyer, A. Czajka, and S. Schuckers, "LivDet-Iris 2013 – Iris Liveness Detection Competition 2013," in *Biometrics (IJCB), 2014 IEEE International Joint Conference on*, Sept 2014, pp. 1–8.

[5] N. Evans, T. Kinnunen, and J. Yamagishi, "Spoofing and countermeasures for automatic speaker verification," in *INTERSPEECH 2013, 14th Annual Conference of the International Speech Communication Association*, Lyon, France, Aug. 2013.

[6] N. Evans, J. Yamagishi, and T. Kinnunen, "Spoofing and countermeasures for speaker verification: a need for standard corpora, protocols and metrics," *IEEE Signal Processing Society Speech and Language Technical Committee Newsletter*, May 2013.

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

[8] Z. Wu, T. Kinnunen, N. Evans, J. Yamagishi, C. Hanilci, M. Sahidullah, and A. Sizov, "ASVspoof 2015: the first automatic speaker verification spoofing and countermeasures challenge," in *INTERSPEECH 2015, 16th Annual Conference of the International Speech Communication Association*, Dresden, Germany, Sept. 2015.

[9] 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, pp. 20–30, Sept. 2015.

[10] Zhizheng Wu, Ali Khodabakhsh, Cenk Demiroglu, Junichi Yamagishi, Daisuke Saito, Tomoki Toda, and Simon King, "SAS: A speaker verification spoofing database containing diverse attacks," in *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, 2015.

[11] 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 2015, 16th Annual Conference of the International Speech Communication Association*, Dresden, Germany, Sept. 2015.

Zhizheng Wu is with the Centre for Speech Technology Research (CSTR) at the University of Edinburgh, United Kingdom. His current research interests include speech synthesis, voice conversion and anti-spoofing. Email: zhizheng.wu@ed.ac.uk

Tomi H. Kinnunen is with University of Eastern Finland (UEF). His current research interest include speaker verification, robust feature extraction and voice conversion. Email: tkinnu@cs.uef.fi.

Nicholas Evans is with EURECOM, Sophia Antipolis, France. His interests include speaker diarization and recognition and acoustic signal processing. Email: evans@eurecom.fr.

Junichi Yamagishi is with National Institute of Informatics, Japan and with University of Edinburgh, UK. His interests include speaker synthesis and speaker adaptation. Email: jyamagis@inf.ed.ac.uk