Prague, July 10–13, 2022

Work-in-Progress, Student paper

Can DeepFake voices steal high-profile identities?

Bence Mark Halpern[123]*and Finnian Kelly*[4]

[1]*University of Amsterdam, Amsterdam, The Netherlands*
[2]*Netherlands Cancer Institute, Amsterdam, The Netherlands*
[3]*Delft University of Technology, Delft, The Netherlands*
b.m.halpern@uva.nl
[4]*Oxford Wave Research, Oxford, UK*
finnian@oxfordwaveresearch.com

Computer-generated synthetic voices are increasingly growing indistinguishable from human voices. While these high-quality synthetic voices open new horizons for the entertainment industry, they can be also used with malicious intent. Examples of the latter include obtaining unauthorised access to bank accounts using fake-voice biometrics (Wang et al., 2020), or rapidly spreading disinformation via deepfake videos of political leaders (Wakefield, 2022; Lorenzu-Trueba, 2018; Thomas, 2020). As the amount of data required to build convincing synthetic voices decreases, it is becoming increasingly important to develop automatic tools that can reliably detect malicious usage of this technology.

Celebrity voices are a great example of a scenario where large amounts of data is available to build a synthetic voice. In this paper, we consider the evaluation of a Deep Neural Network (Dilated ResNet) based spoofing detector (Halpern et al., 2020) with a celebrity deepfake speech corpus. The corpus, collected for the present study from various online sources, consists of one deepfake recording and one genuine recording for each of 30 celebrities. We note that this data is uncontrolled, with varying levels of noise, compression, and other artefacts.

The evaluation of the corpus resulted in a spoof detection Equal Error Rate (EER) of 16.7%. Speaker-wise, all except two of the 30 speakers, namely Bill Clinton and Winston Churchill, correctly produced higher detection scores for their genuine recording than for their deepfake recording. We hypothesise that older genuine recordings, and that of Winston Churchill in particular, may contain artefacts resulting from post-hoc speech enhancement, which influence the detector.

We further consider the use of the evaluation scores from the celebrity deepfake corpus to calculate a genuine/spoof likelihood ratio (LR) for a questioned sample from a new speaker. Using the probability densities of genuine and spoof evaluation scores to represent genuine and spoof hypotheses respectively, we calculate a genuine/spoof LR for a Zelenskyy deepfake (Wakefield, 2022), as shown in Fig. 1. We additionally calculate the LR for a genuine recording of Zelenskyy (one with a similar SNR to the deepfake). Converting the detector score to an LR in this way, by considering the competing genuine and spoof hypotheses given relevant data, produces a result that can be directly interpreted. In the present example, the LR for the deepfake recording is less than one (0.18) and the LR for the genuine recording is greater than one (6.3). These LRs therefore provide correct support in both deepfake and genuine cases.

Ongoing work is investigating the influence of environmental noise, recording devices, compression, as well as the speech duration, on the performance of deepfake detection.

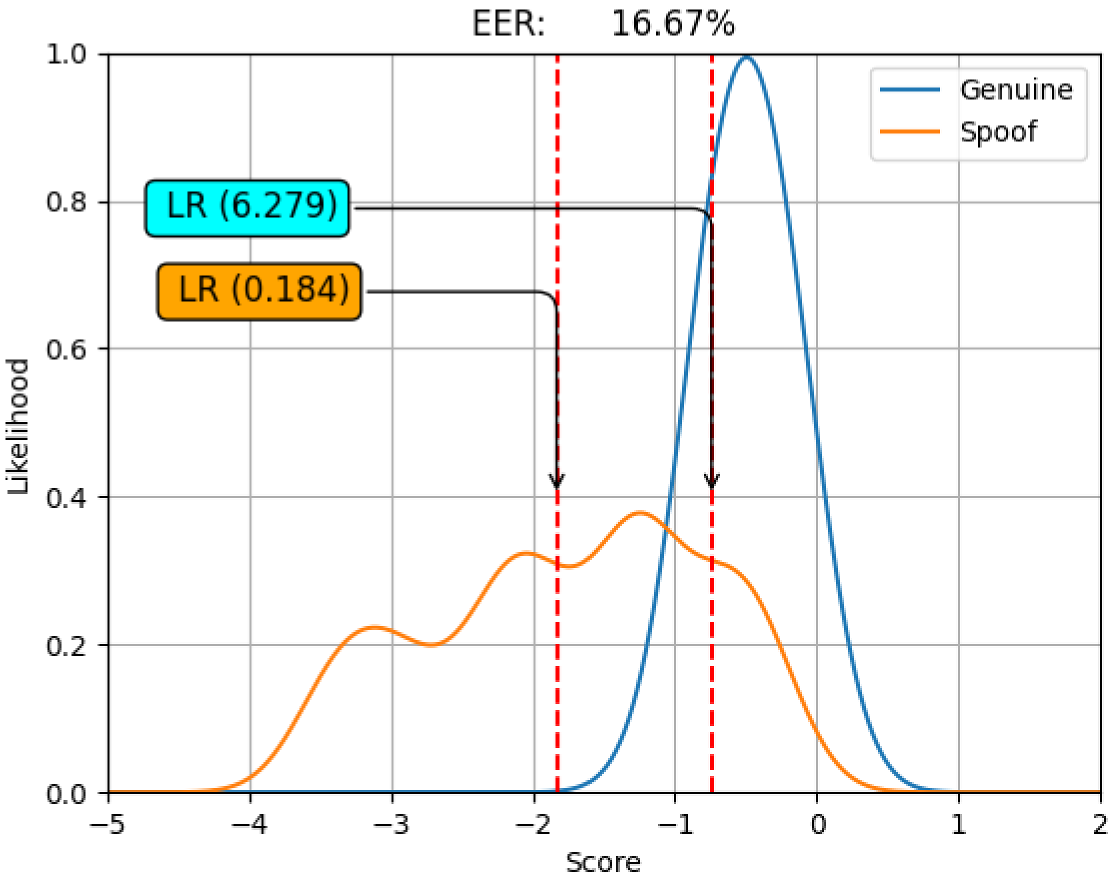

Figure 1 Probability density curves for the genuine and spoof evaluation scores obtained from the celebrity deepfake speech corpus (EER of 16.67%). The orange box indicates the genuine/spoof LR for the Zelenskyy deepfake (0.184) and the blue box indicates the genuine/spoof LR for the Zelenskyy genuine recording (6.279).

References

Halpern, B. M., Kelly, F., van Son, R., & Alexander, A. (2020). Residual networks for resisting noise: analysis of an embeddings-based spoofing countermeasure. *Speaker Odyssey 2020.*

Lorenzo-Trueba, J., Fang, F., Wang, X., Echizen, I., Yamagishi, J., & Kinnunen, T. (2018). Can we steal your vocal identity from the Internet?: Initial investigation of cloning Obama's voice using GAN, WaveNet and low-quality found data. *arXiv preprint arXiv:1803.00860.*

Thomas, D.: Deepfakes: A threat to democracy or just a bit of fun? (2020), https://www.bbc.com/news/business-51204954 (Date of access: 2022. 05. 20)

Wakefield, J.: Deepfake presidents used in Russia-Ukraine war (2022), https://www.bbc.com/news/technology-60780142 (Date of access: 2022. 05. 20)

Wang, X., Yamagishi, J., Todisco, M., Delgado, H., Nautsch, A., Evans, N., ... & Ling, Z. H. (2020). ASVspoof 2019: A large-scale public database of synthesized, converted and replayed speech. *Computer Speech & Language*, *64*, 101114.