Robust Audio Anti-Spoofing System Based on Low-Frequency Sub-band Information (Paper ID: 151)

Menglu Li, Xiao-Ping Zhang

Department of Electrical, Computer and Biomedical Engineering
Toronto Metropolitan University, Toronto, Ontario, Canada

INTRODUCTION

Deepfake audio is the use of Deep Learning algorithms, to manipulate speech content or even create fake sound voices. Deepfake Audio detection is getting more crucial as human ears are being fooled easily by fake audios in their daily lives, since the fake audio does keep linguistic features, such as breathing or emotion.
It is essential to provide a handful of tools to identify fabricated audio, and to prevent Deepfake audio spreading out misinformation or threatening voice-abled authentication systems.

Our Contributions:

1. We identify the **discriminative information** presented in the low-frequency sub-band of power spectrograms, especially for detecting TTS-generated speeches.
2. Based on the finding, we **propose a robust anti-spoofing system with only 57K parameters**, which outperforms all official baselines of ASVspoof2021 Deepfake Challenge.

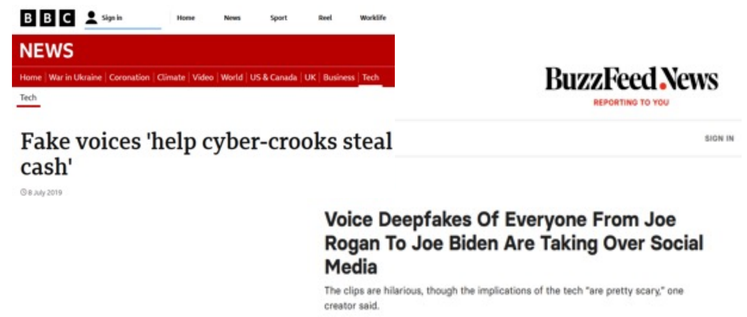

Figure 1. News snippets about Deepfake audio

PROPOSED MODEL

Our assumption:

The non-speech parts such as silence segments in the low-frequency sub-band should contain more essential discriminate information than the voice parts.

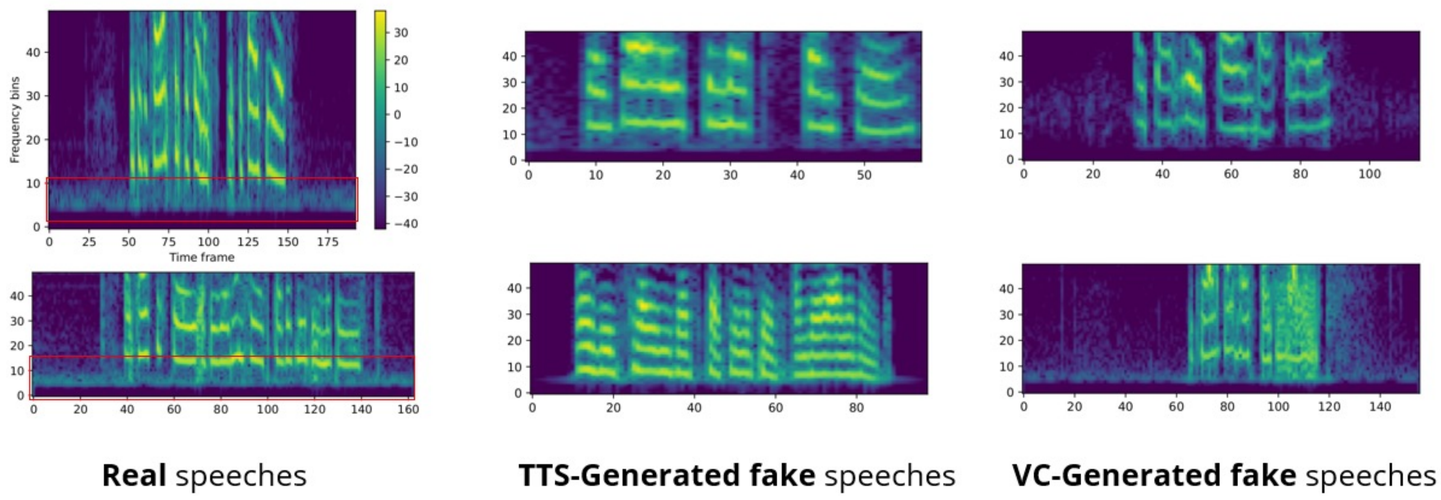

Figure 2. Illustration of the spectrogram of low-frequency sub-band (first 50-bins) for both bonafide and spoofing speeches.

- We can observe that there is a continuous noise band between the first 10 frequency bins **in the real speeches** throughout the entire time frame. This noise band occurs in all real speech samples in the dataset, which should be caused by electronic recording devices.
- For **the TTS-generated speech**, there is no such continuous noise band at the low-frequency range.
- For **the VC-based fake audio**, since they are generated directly using real recording speeches, the electronic noise may or may not be retained depending on the particular architecture of the VC algorithm.

Following by this idea, we design our anti-spoofing system as seen in **Figure 3**. We only utilize the first 50 frequency bins of the power spectrogram as input.

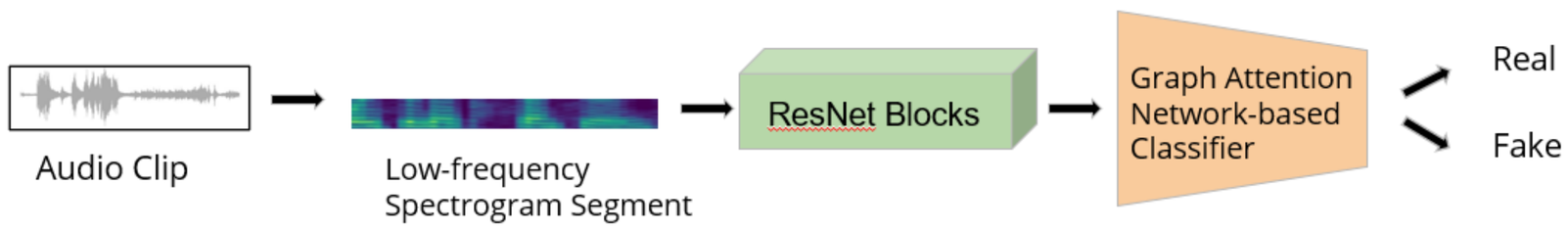

Figure 3. The proposed model

DATASET

Our model is evaluated on **Deepfake Speech (DF) track of ASVspoof2021** Challenge dataset.

- The largest and the most diverse dataset
- 14,869 real speeches and 519,059 synthesized speeches
- Contains more than hundreds of spoofing algorithms across different codecs.

EXPERIMENT & RESULTS

Table 1. Performance on the ASVspoof2021 DF evaluation set

	EER [%]	Parameters
Low-Frequency system (Ours)	**22.13**	**57K**
Baseline01: RawNet2 [1]	22.38	25000K
RawGAT [2]	22.47	440K
Baseline02: LFCC-LCNN [3]	23.48	-
Res2Net [4]	24.47	923K
Baseline03: LFCC-GMM[3]	25.25	-
Baseline04: CQCC-GMM [3]	25.56	-

CONCLUSION

- We propose a robust GAT-based Deepfake audio detector, utilizing a low-frequency band as input, to detect fake speech across various conditions. Because of its good robustness and usage efficiency, our model, our model has gained a higher possibility of being applied in real-world use cases, such as automated Deepfake screening.
- We propose that the non-speech parts contain more identification features to make the detection judgment, especially for detecting TTS-generated speech. This finding provides a direction for the potential design of future anti-spoofing systems.

REFERENCE

[1] H. Tak, J. Patino, M. Todisco, A. Nautsch, N. Evans, and A. Larcher, "End-to-end anti-spoofing with Rawnet2," ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2021.
[2] H. Tak, J.-w. Jung, J. Patino, M. Kamble, M. Todisco, and N. Evans, "End-to-end spectro-temporal graph attention networks for speaker verification anti-spoofing and speech deepfake detection," 2021 Edition of the Automatic Speaker Verification and Spoofing Countermeasures Challenge, 2021.
[3] J. Yamagishi, X. Wang, M. Todisco, M. Sahidullah, J. Patino, A. Nautsch, X. Liu, K. A. Lee, T. Kinnunen, N. Evans, et al., "Asvspoof 2021: Accelerating progress in spoofed and deepfake speech detection," 2021 Edition of the Automatic Speaker Verification and Spoofing Countermeasures Challenge, 2021.
[4] X. Li, N. Li, C. Weng, X. Liu, D. Su, D. Yu, and H. Meng, "Replay and synthetic speech detection with res2net architecture," ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2021, pp. 6354–6358.
[5] "Fake voices 'help cyber-crooks steal cash'," BBC News, 08-Jul-2019. [Online]. Available: https://www.bbc.com/news/technology-48908736.
[6] M. Li, X-P. Zhang, "Robust Audio Anti-spoofing Detection System based on Low-frequency Sub-band Information," *To appear* in 2023 IEEE Workshop on Application of Signal Processing to Audio and Acoustics (WASPAA), 2023.