IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.301

Volume 3, Issue 3, May 2023

Deep Fake Voice Synthesizer

D. Arvind Sai, D. Shiva Sai, C. Sathya Narayana, Mr. Christal Anand

Department of Computer Science and Engineering

Prathyusha Engineering College, Poonaamallee, Thiruvallur, Chennai, India

Abstract: *With the rapid-fire development of computer technology, voice technology has become an exploration hotspot in the field of deep literacy. Allowing computers to hear, see, speak and feel is the unborn development direction of mortal- computer commerce. Among them, voice will become the most promising mortal- computer commerce system in the future which has further advantages than other commerce styles. Voice cloning is one of the branches of voice technology, introducing a language modelling approach for textbook to speech conflation(TTS). Specifically, we train a neural codec language model using separate canons deduced from an out- the- shelf neural audio codec model, and regard TTS(Text to Speech) as a tentative language modelling task rather than nonstop signal retrogression as in former work. During the pre- training stage, we gauge the TTS training data to 1k hours of English speech which is hundreds of times larger than systems. Software emerges in- environment literacy capabilities and can be used to synthesise high- quality substantiated speech with only a 10- second enrolled recording or written textbook of an unseen speaker as an aural advice. trial results show that our software significantly outperforms the state- of- the- art zero- shot TTS system in terms of speech light heartedness and speaker similarity. In addition, we find our software could save the speaker's emotion and the aural terrain of the aural advisement in conflation and there are diligence similar as audio books, filmography(dubbing) and people with disabilities facing problems through audio. To break the problem of furnishing a large number of sample lines to reduplicate a voice and the long waiting time for voice cloning.*

Keywords: Voice Clone; A Few Samples; Real Time

REFERENCES

[1]. Aaron van den Oord, Sander Dieleman, Heiga Zen, et al. Wavenet: A Generative Model for Raw Audio[J/OL]. arXiv Preprint arXiv:1609.03499, 2011,

[2]. Sercan O. Arik, Mike Chrzanowski, Adam Coates, et al. Deep Voice:real-time neural text-to-speech[J/OL]. arXiv preprint arXiv:1702.07825, 2017.

[3]. Sercan Arik, Gregory Diamos, Andrew Gibiansky, et al. Deep Voice Multi-Speaker Neural Text-to-Speech. arXiv Preprint arXiv: 1705.08947, 2017.

[4]. Wei Ping, Kainan Peng, Andrew Gibiansky, et al. Deep Voice 3: Scaling Text-to-Speech with Convolutional Sequence Learning. arXiv Preprint arXiv: 1710.07654.

[5]. Yuxuan Wang, RJ Skerry-Ryan, Daisy Stanton, et al. Tacotron: Towards End-to-End Speech Synthesis. arXiv Preprint arXiv: 1703.10135, 2017.

[6]. Li Wan, Quan Wang, Alan Papir, and Ignacio Lopez Moreno. Generalised end-to-end loss for speaker verification. arXiv Preprint arXiv:1710.10467, 2017.

[7]. J. Shen, R. Pang, R. J. Weiss, et al. Natu- ral tts synthesis by conditioning wavenet on mel spectrogram predictions, in Proc. International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2018, pp. 4779– 4783.

[8]. N. Kalchbrenner, E. Elsen, K. Simonyan, et al. Efficient neural audio syn- thesis. arXiv Preprint arXiv:1802.08435, 2018.

[9]. W. B. Kleijn, F. SC Lim, A. Luebs, et al. Wavenet based low rate speech coding, in Proc. International Conference on Acous- tics, Speech, and Signal Processing (ICASSP), 2018, pp. 676– 680.

[10]. Xavi Gonzalvo, Siamak Tazari, Chun-an Chan, et al. Recent advances in google real-time hmm-driven unit selection synthesiser. Interspeech, 2016,pp. 2238–2242.

[11]. Z. Jin, A. Finkelstein, G. J. Mysore, and J. Lu, Fftnet: A real- time speaker-dependent neural vocoder, in Proc. Interna- tional Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2018, pp. 2251–2255.

[12]. S. Mehri, K. Kumar, I. Gulrajani, R. Kumar, S. Jain, J. Sotelo, Courville, and Y. Bengio, SampleRNN: An unconditional end-to- end neural audio generation model, arXiv:1612.07837, 2016.

[13]. Fu Qiang. Research on speech parameter representation and objective quality evaluation [D]. Xidian University 2001