

## Anonymize the speaker emotion by converting the expressive speech to natural utterance

Speech signal is a rich resource to disclose several speakers' states, such as their emotional and mood states, confidence and stress levels, physical condition, age, gender, and personal traits.

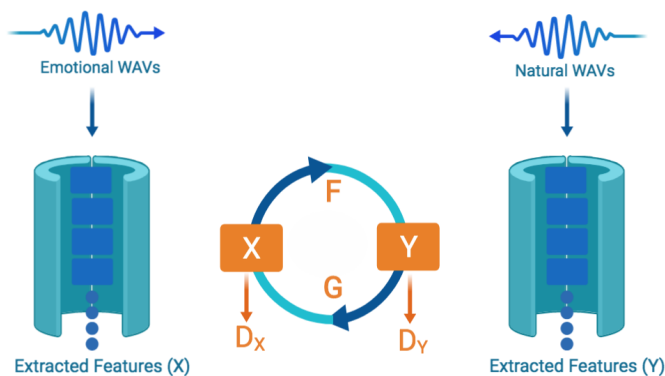


Figure1: Emotion Anonymization based on CycleGAN-VC

### Framework

The proposed method in (Figure1,3) is focused on what the content is in regardless how to say it.

### Experiment

- The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS)[1]
- WORLD: speech analysis, manipulation and synthesis[2]
- CycleGAN-VC: non-parallel voice conversion [3]

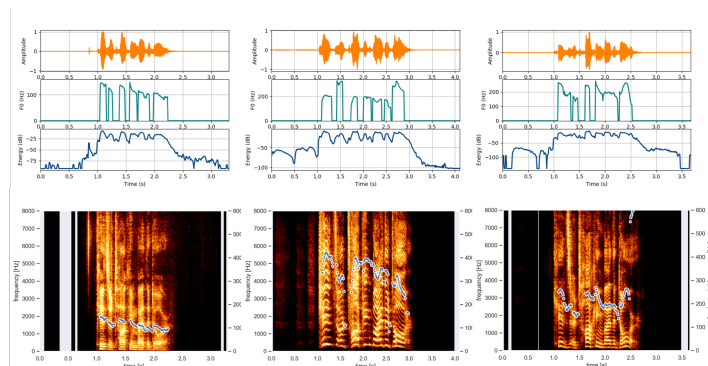


Figure2: Speech features analysis for different emotion categories which are natural, angry, happy respectively

### Following Steps

- Analyzing the prosody features (Figure2)
- Testing the quality of generated WAVs
- Measuring privacy with different Metrics
- Further experiments and techniques will be conducted

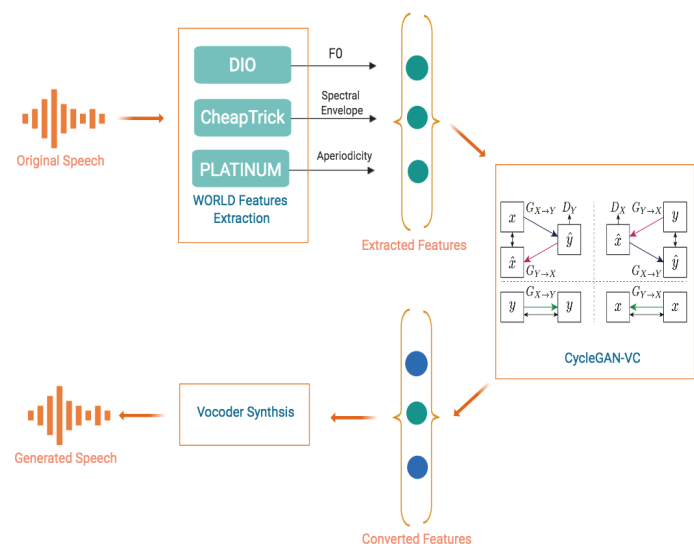
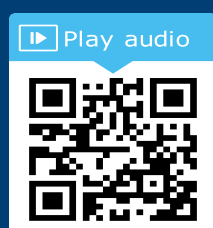


Figure3: Overview of the System Architecture [4]

### References

- [1] S. R. Livingstone and F. A. Russo, "The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVDESS)," 2018.
- [2] M. Morise, F. Yokomori, and K. Ozawa, "World: a vocoder-based high-quality speech synthesis system for real-time applications," 2016.
- [3] T. Kaneko, H. Kameoka, K. Tanaka, and N. Hojo, "Cyclegan-vc2: Improved cyclegan-based non-parallel voice conversion," 2019.
- [4] K. H. Kaneko, Takuhiro, "Parallel-data-free voice conversion using cycle-consistent adversarial networks," 2017.

Contact Information  
r.aloufi18@imperial.ac.uk



← Check the results

