Neural Text-to-Speech

Russel Shawn Dsouza

Electronics and Communications Engg.
National Institute of Technology Karnataka
Surathkal, India - 575025

October 9, 2019

Speech synthesis

Artificial production of human speech

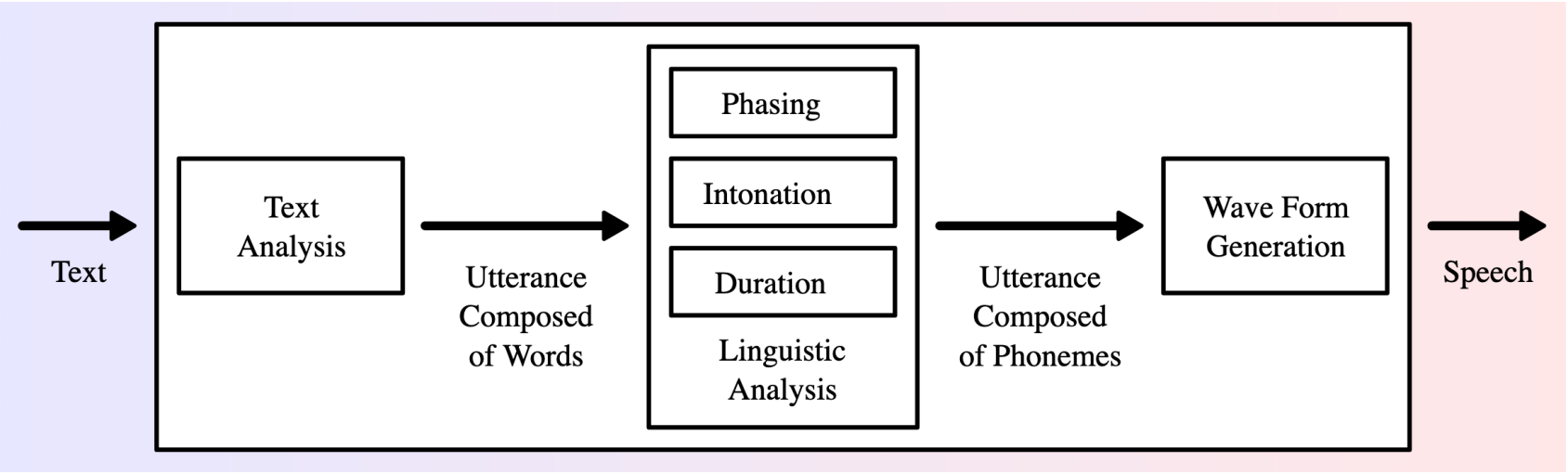

Figure: A typical text-to-speech system

History of speech synthesis

Concatenative

- Large database of human speech used
- Database has to be changed for every new style

Parametric

- Simulate human voice using a parametric physical model
- Easy to modify voice

Neural

- Generate human voice using neural networks
- More natural sounding output

Approaches in Neural text-to-speech

LSTM

WaveNet

WaveNet based

WaveNet

A deep neural network for generating raw audiowaveforms.

- Probabilistic
- Autoregressive
- Beats all previously known methods

Figure: Time domain representation of 1 second of generated speech

WaveNet: Architecture

WaveNet: Pros and Cons

FloWaveNet

FloWaveNet: Architecture

FloWaveNet: Training

FloWaveNet: Reported results

FloWaveNet: Improvements over WaveNet

Neural TTS: The future

Summary

Conclusion