

Developing text-to-speech methods using image information

Leyuan Sheng

Advisor: Pavlovskiy Evgeniy

December 6, 2017

Outline

- 1 Synthesis technology introduction
- 2 Speech synthesis application scenarios
- 3 A typical speech synthesis system flow chart
- 4 Synthetic speech evaluation criteria
- 5 Popular models of speech synthesis system
- 6 Use the image information

1. Synthesis Technology Introduction

What is Speech Synthesis ?

- **Automatic speech recognition (ASR)**

Speech (continuous time series) \longrightarrow Text (discrete symbol sequence)

- **Machine translation (MT)**

Text (discrete symbol sequence) \longrightarrow Text (discrete symbol sequence)

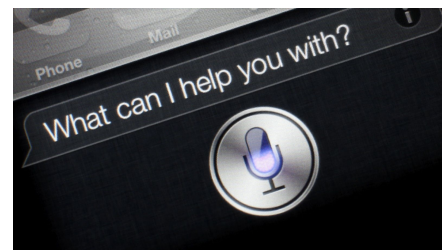
- **Text-to-speech synthesis (TTS)**

Text (discrete symbol sequence) \longrightarrow Speech (continuous time series)

2. Speech synthesis application scenarios

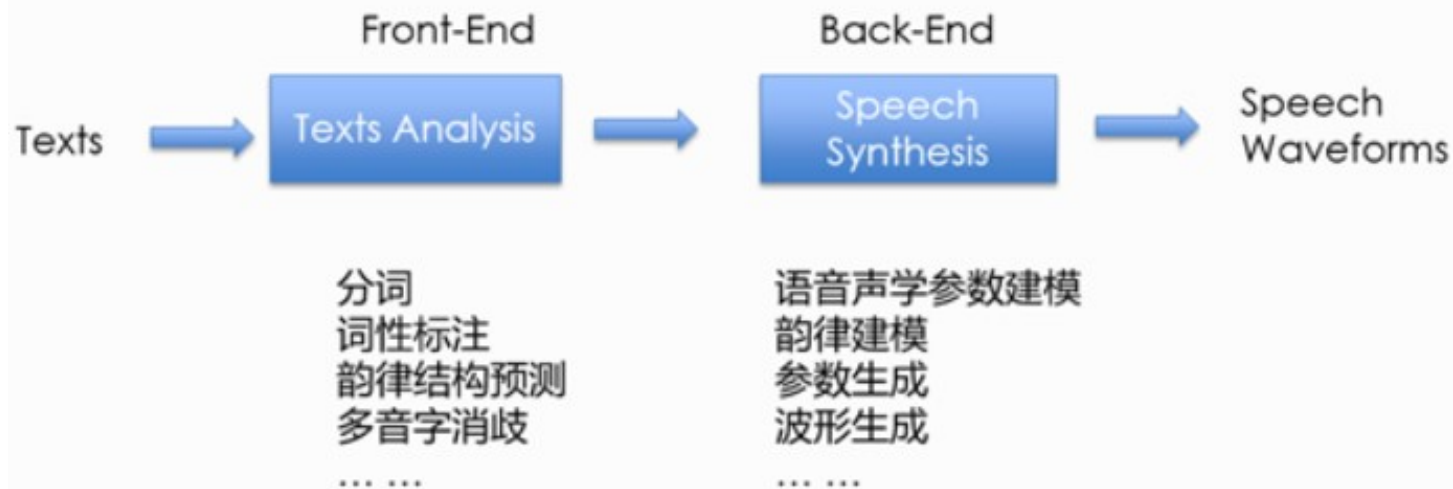
Why we need Speech Synthesis ?

- Why we need Speech Synthesis
 - More natural human-computer interaction
 - More convenient access to information
 - Release your hands
 - Safer driving experience
 -
- Application of speech synthesis
 - Smart assistant (Siri, Google Assistant, Cortana, ...)
 - Car voice navigation
 - Customer Service / Call Center
 - Smart speaker, TV, home (AliGenie, Alexa, Google home, ...)
 - Novel, news broadcast
 - Speech to speech translation
 -



3. A typical speech synthesis system flow chart

Components in a typical TTS system



4. Synthetic speech evaluation criteria

MOS(Mean Opinion Score):

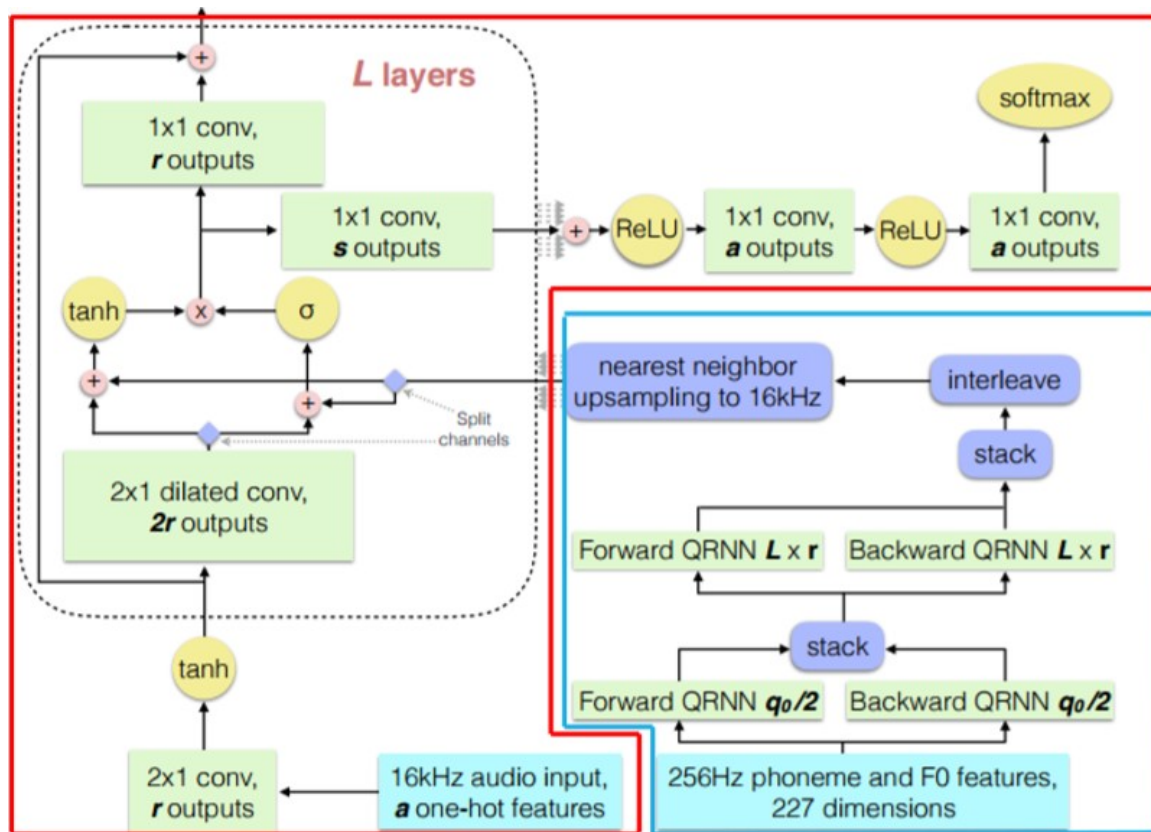
- 1). 1 — Bad
- 2). 2 — Poor
- 3). 3 — Fair
- 4). 4 — Good
- 5). 5 — Excellent

5. Popular models of speech synthesis system

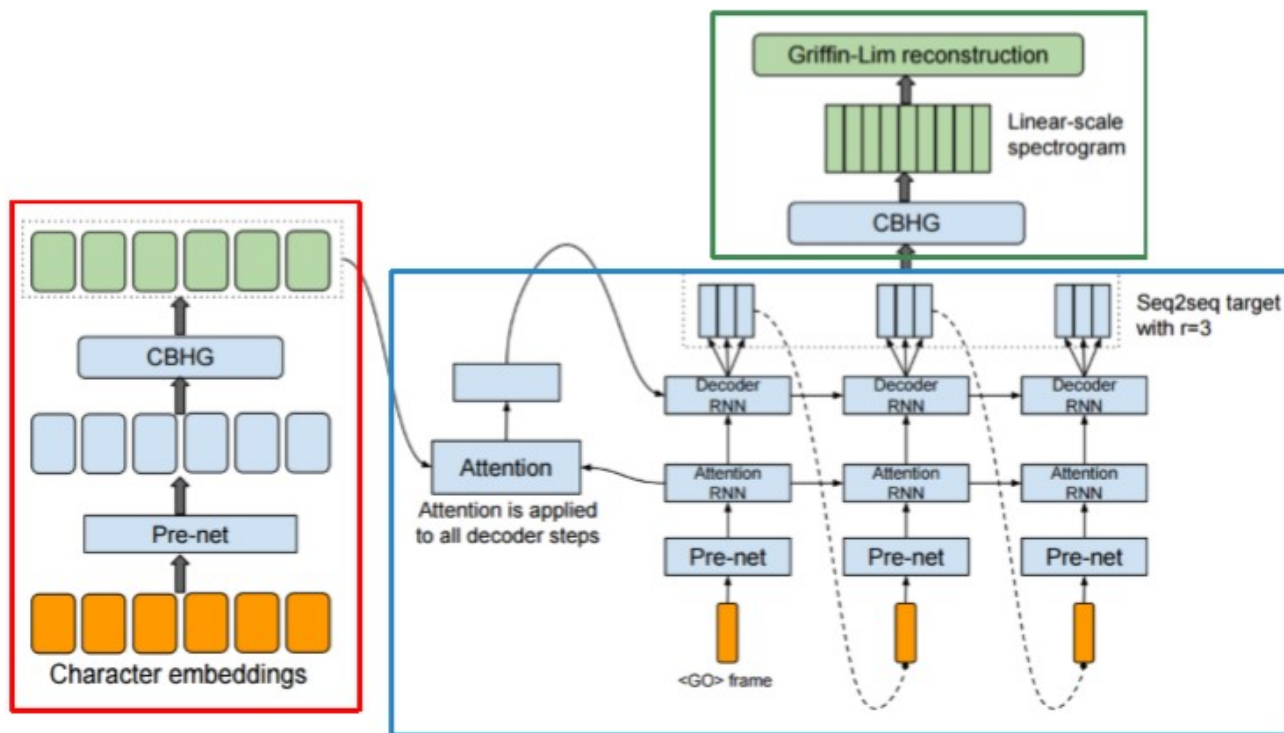
- **WaveNet** — **Deepmind (2016, 2017)**
Audio synthesis network with intermediate representation
- **Deep Voice** — **Baidu (2017)**
Composed end-to-end learning
- **Tacotron** — **Google (2017)**
Truly end-to-end learning

5.1 WaveNet

5.2 Deep Voice — Composed end-to-end learning



5.3 Tacotron — Truly end-to-end learning



6. What is the image information ?



图片标注与分类



天空、人、山

图片描述



一群人背着背包走在山间的小路上

6.1 How to use the image information

- 1 Text recognition
- 2 Image caption
- 3 Text recognition and Image caption
- 4 ... ??? Some fancy ideas

Thank you for your attention !