Deep Learning Audio

Lecture 1

Pavel Severilov

Moscow Institute of Physics and Technology

2022

Outline

Outline

Organisation

1. $\sim$ 7 lectures
2. 2 homeworks
3. Grade:
   - 50% homeworks + 50% one exam question
   - or 100% your project
4. Github course page:
   slides+videos+homeworks+materials+papers
   `https://github.com/severilov/2022-DL-Audio-Course`
5. Discussion: telegram chat (contact @severilov)

Outline

Voice Technologies: Applications

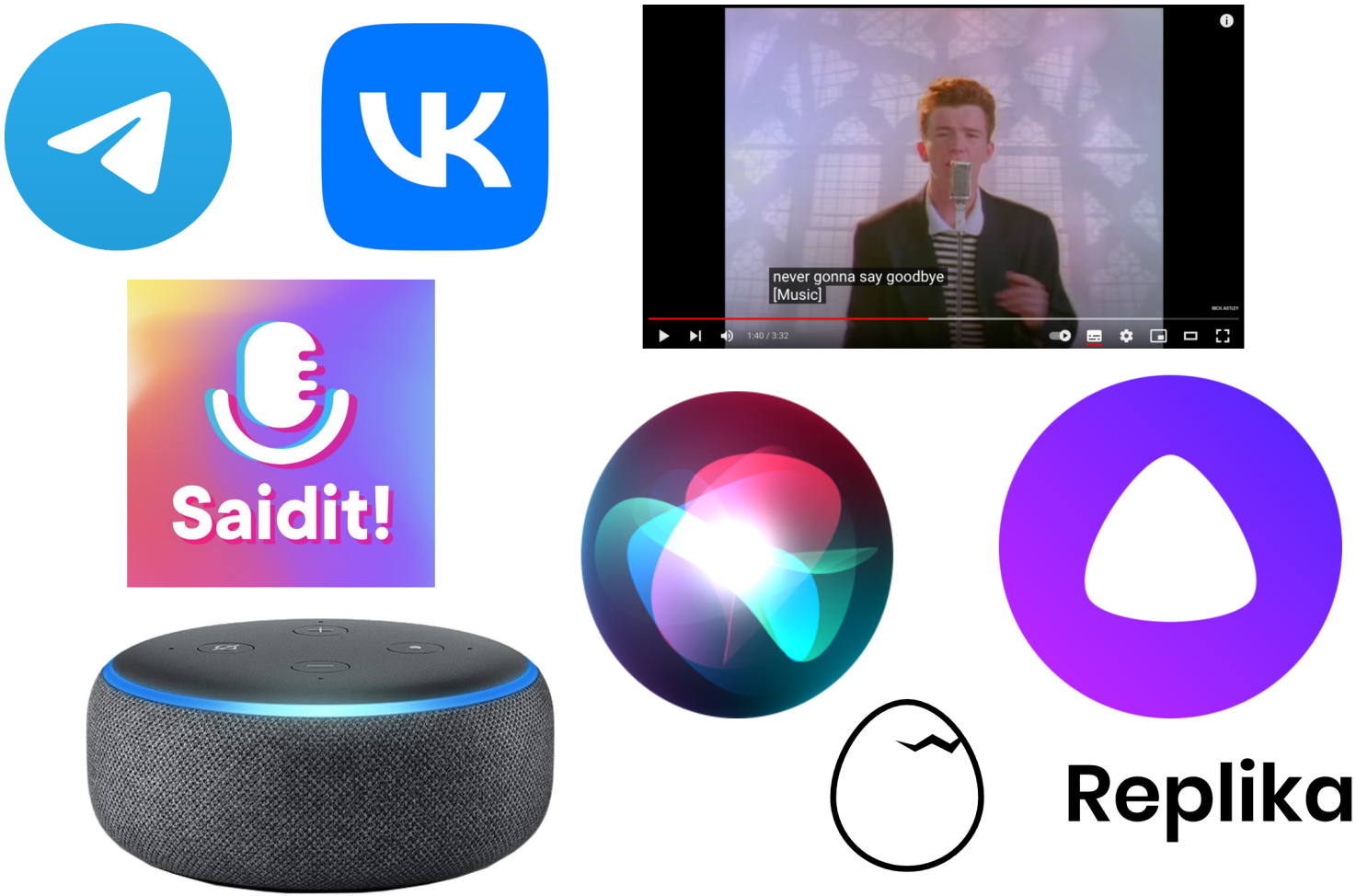
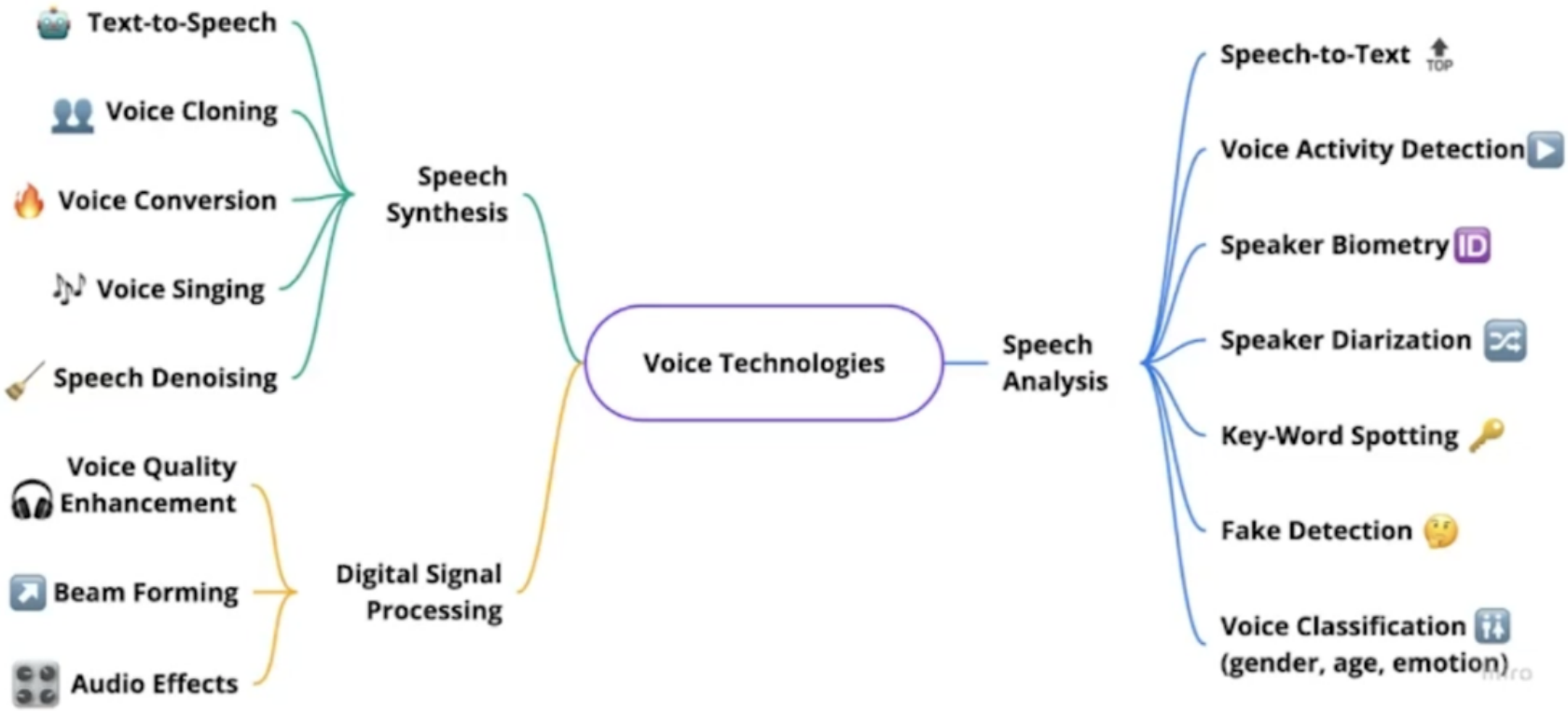
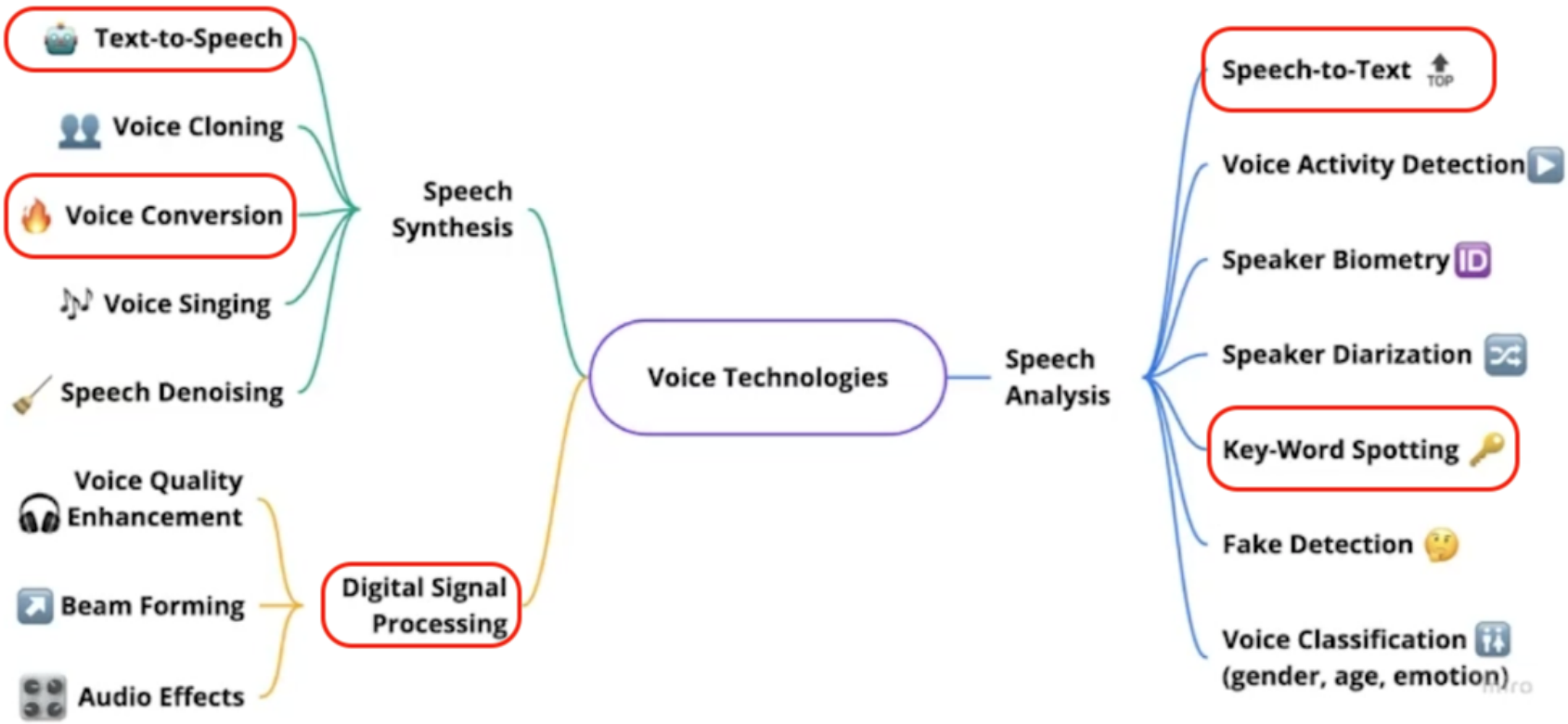


Figure: Siri, Amazon Alexa, Alisa, Replika, Telegram, VK, YouTube

Voice Technologies, Tasks: Mind Map



Voice Technologies, Tasks: Course

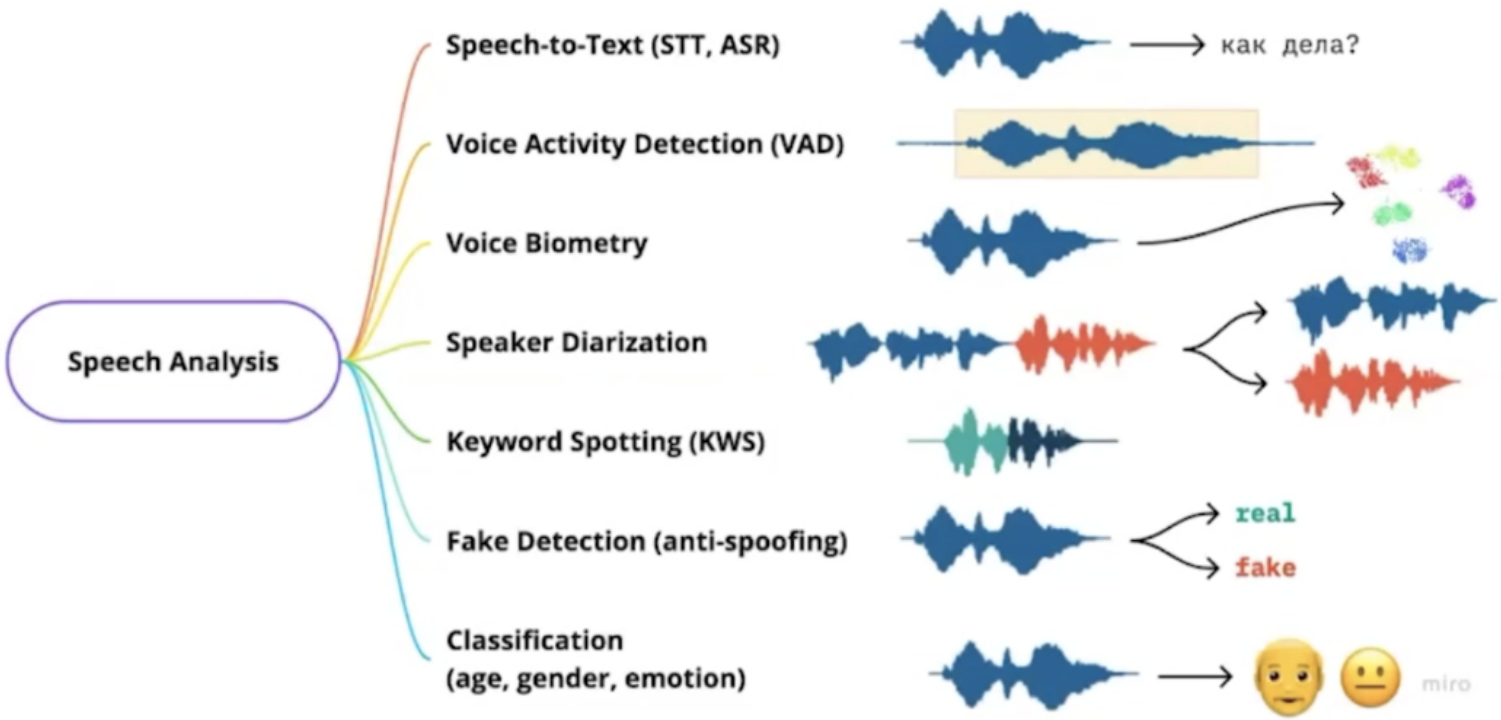


Outline

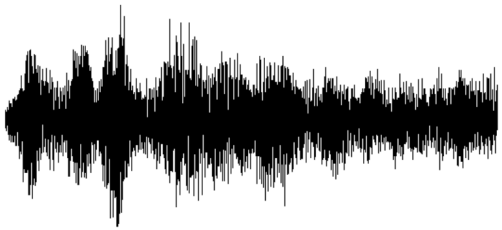
History of Speech Recognition

- **50's**: 1952, Bell Laboratories, "Audrey" system, could recognize single voice speking digits
- **60's**: 1961, IBM, "Shoebox", understood 16 words in English
- **70's**: DARPA, understood over 1000 words (Siri spin-out)
- **80's**: using HMM, understood several thousand words
- **90's**: became faster because of processors
- **00's-10's**: ML, DL, Big Data, GPUs

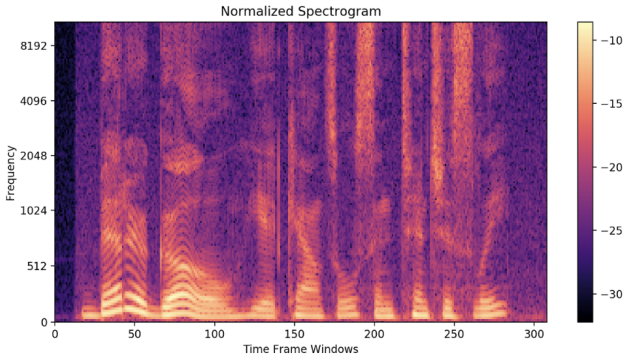
Speech Analysis Tasks: Mind Map



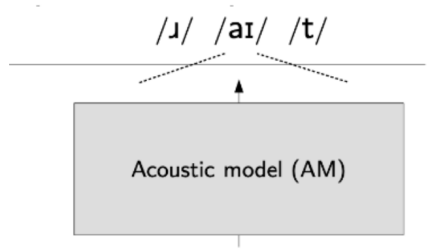
Speech Recognition & Deep Learning: Idea



Raw Audio



Feature: Spectrogram



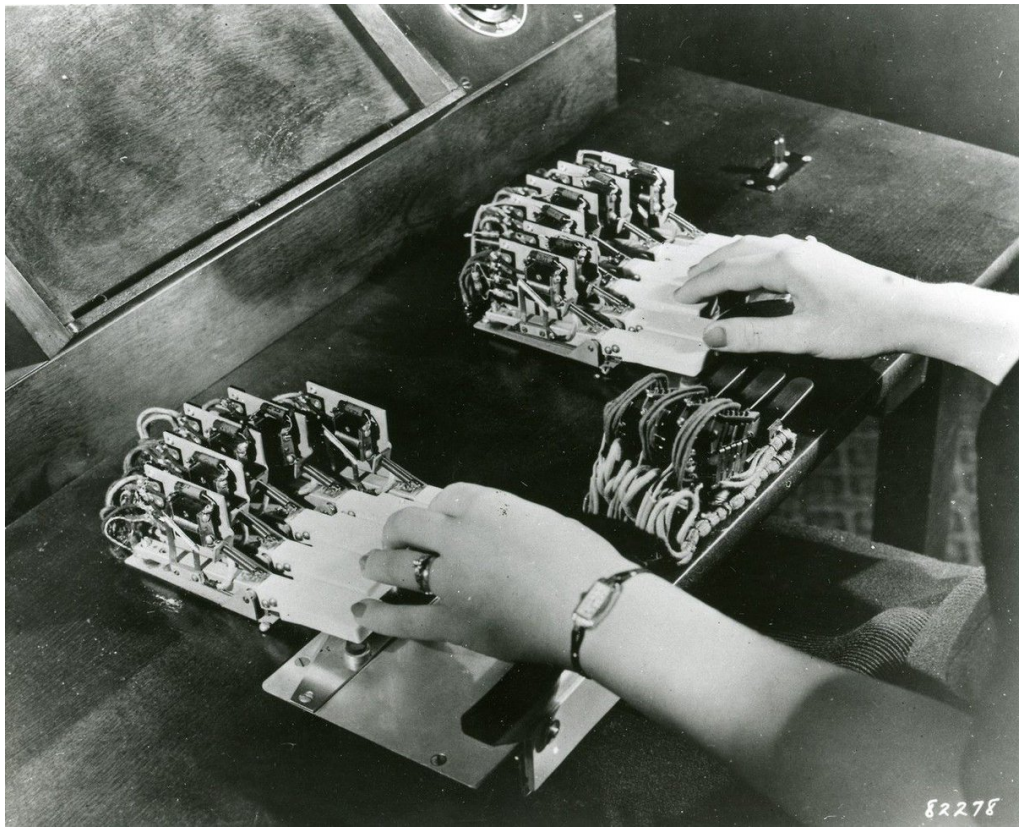
Acoustic model: phonemas

"right"

Language Model

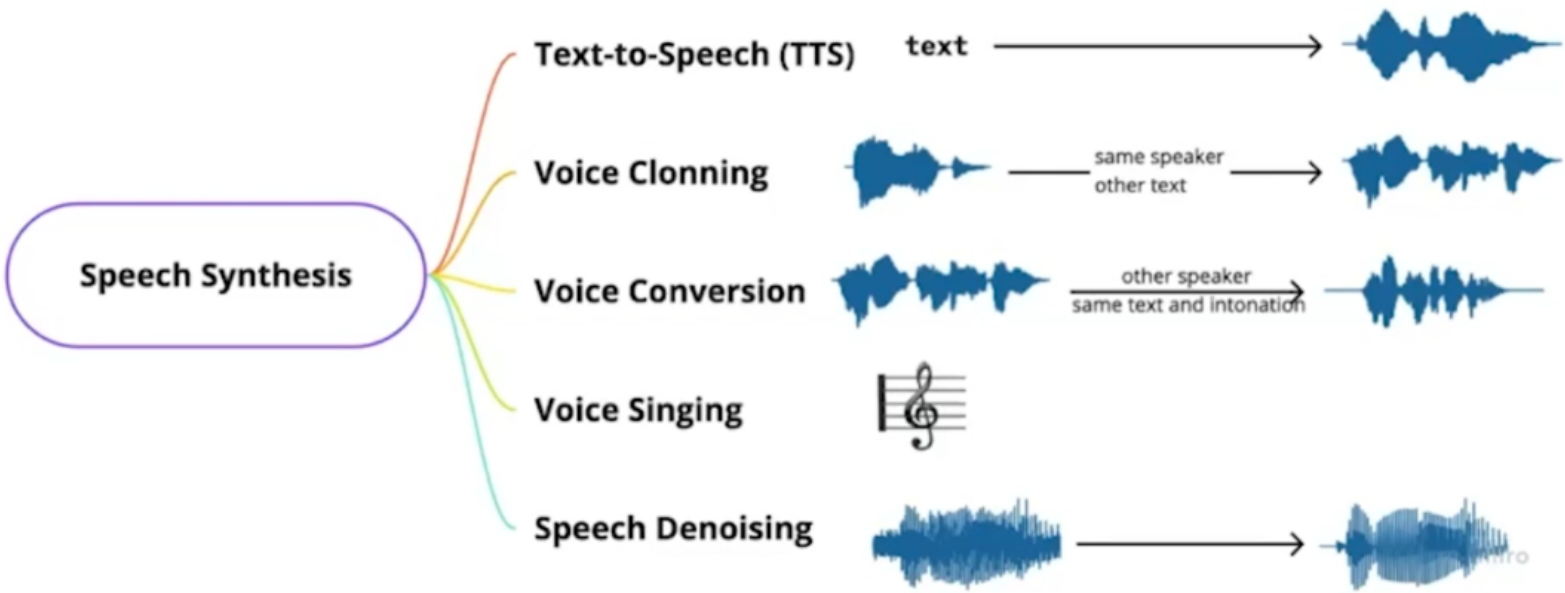
Outline

History of Speech Synthesis



- **30's**: 1939, Bell Laboratories, "Voder",
- **80's**: Format-based on rules, Atari/Sega
- **90's-00's**: Concatenative synthesis
- **10's**: ML, DL, Big Data, GPUs

Speech Synthesis Tasks: Mind Map



Speech Synthesis & Deep Learning: Idea

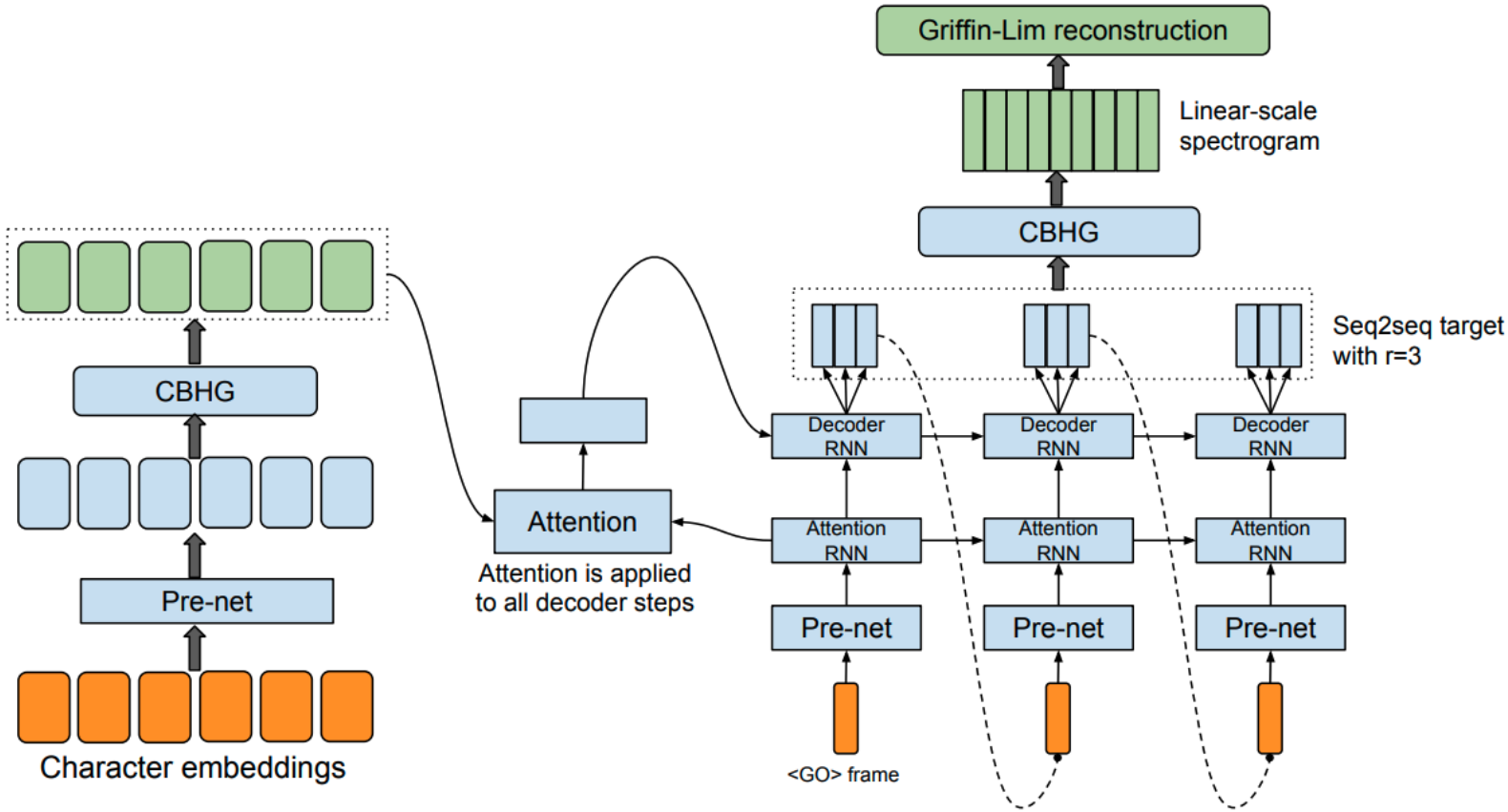


Figure: Example of Deep Learning approach to speech synthesis: encoder-decoder structure with reccurent parts

---

Wang, Yuxuan et al. "Tacotron: Towards End-to-End Speech Synthesis." INTERSPEECH (2017), Google Inc.