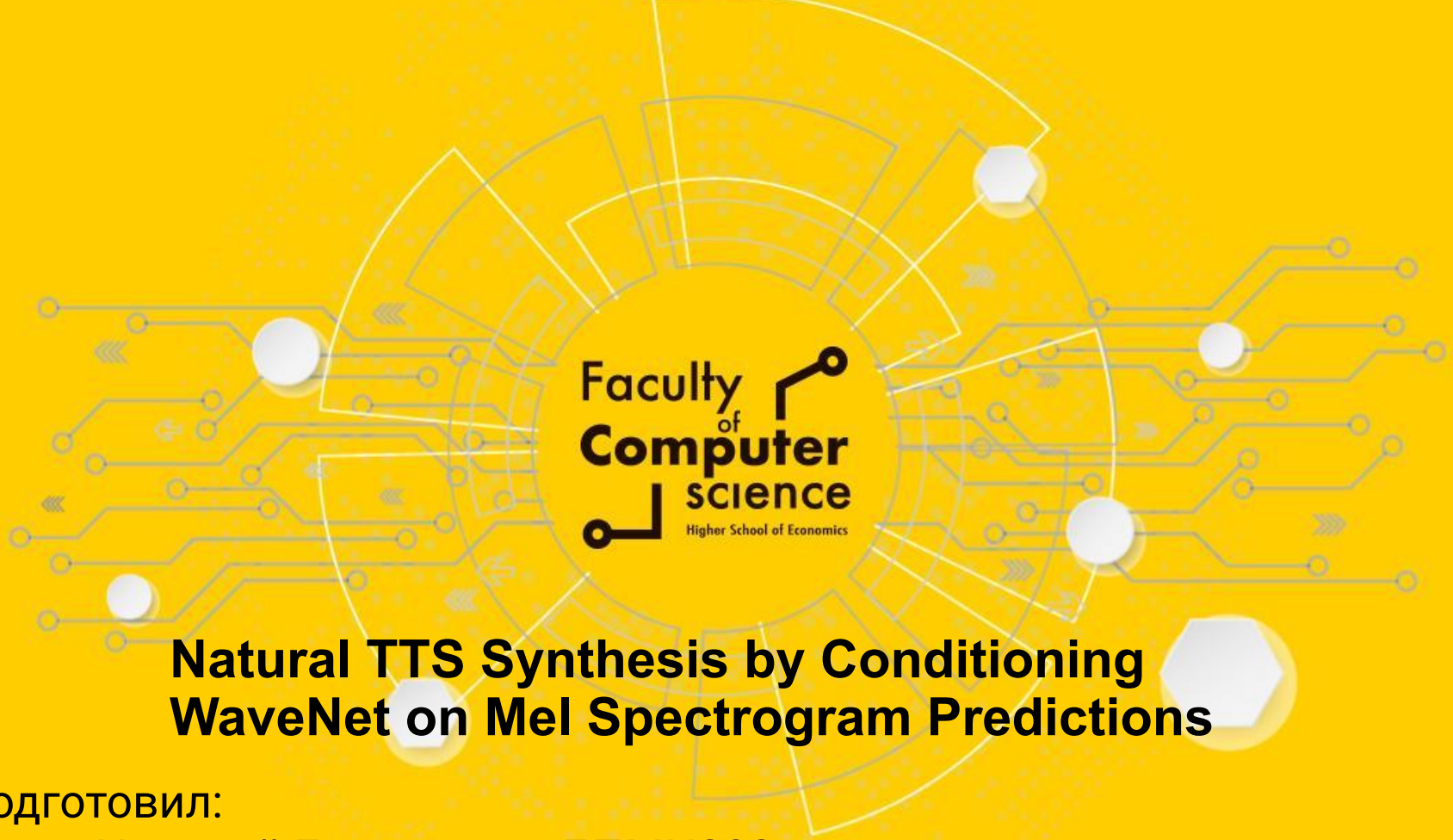


Natural TTS Synthesis by Conditioning WaveNet on Mel Spectrogram Predictions

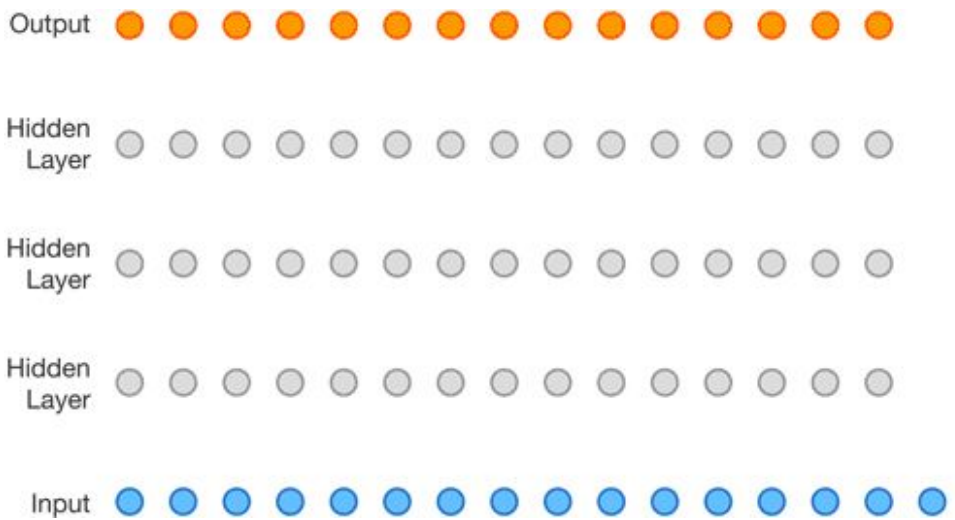
Подготовил:
Косса Николай Евгеньевич, БПМИ202

План

1. WaveNet
2. Tacotron
3. Спектрограммы
4. Tacotron 2:
    - архитектура
    - modified WaveNet vocoder
    - эксперименты

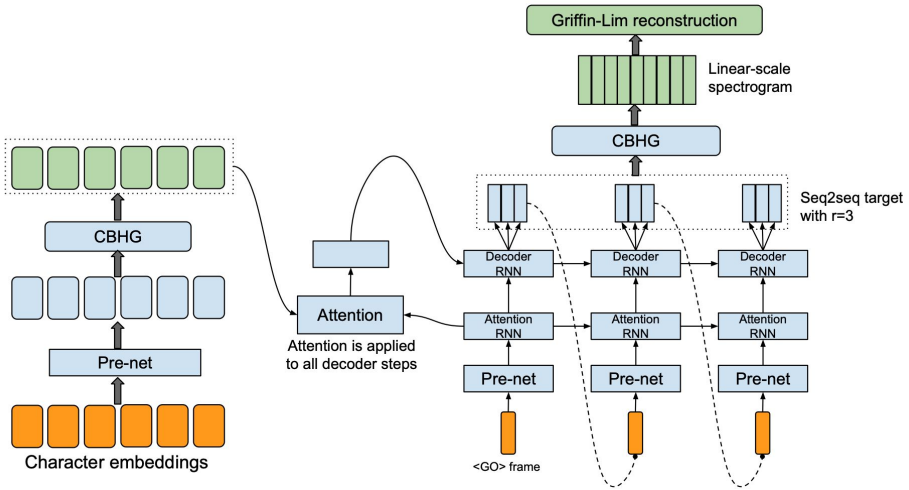
WaveNet

- Входные данные: лингвистические признаки, f0, длительность фонем.
- Средняя оценка MOS: 4.34.



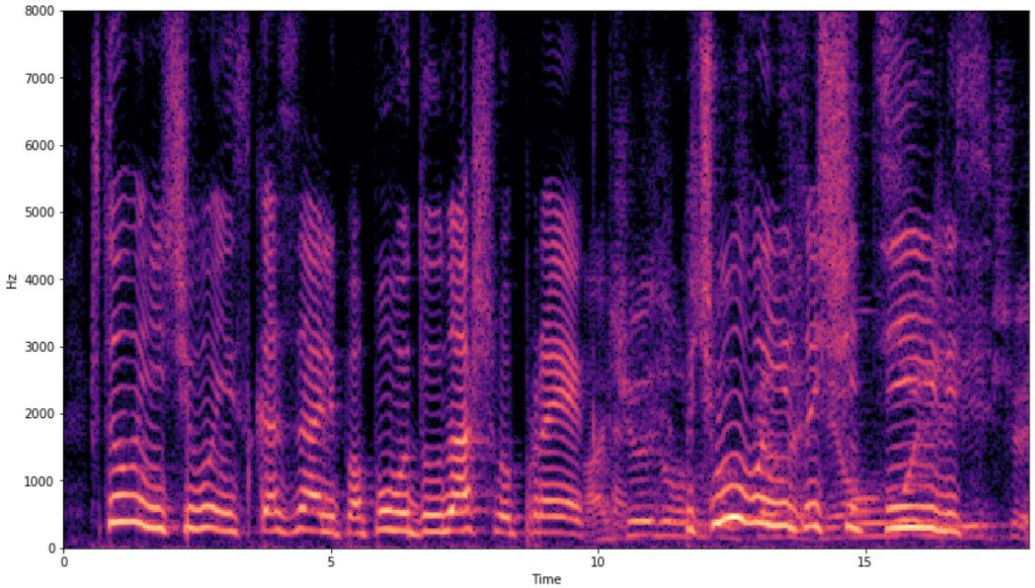
Tacotron

- Входные данные: текст.
- Выход: mel-спектрограмма.
- Алгоритм Гриффина-Лима.
- Средняя оценка MOS: 4.0.



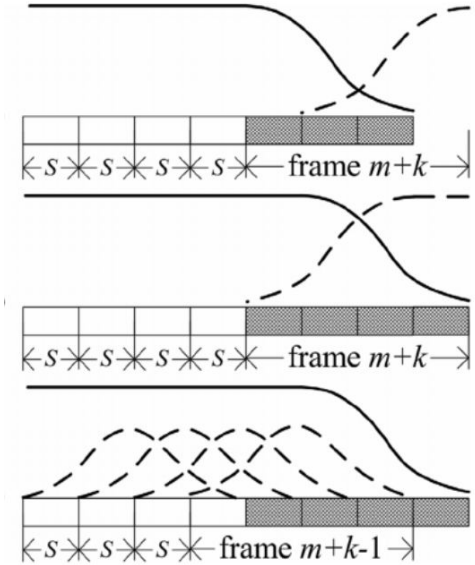
Спектрограммы

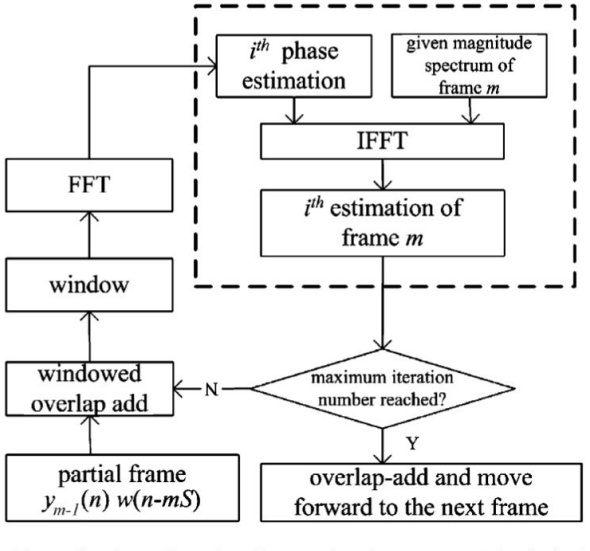
- Спектрограмма не содержит информации о фазе.
- Фаза зависит от времени.



Спектрограммы

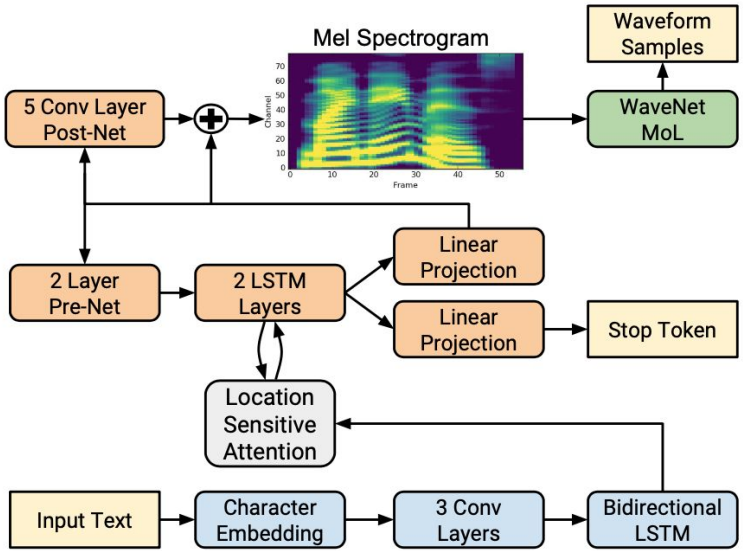
- Алгоритм Гриффина-Лима (GLA) пользуется избыточностью STFT.





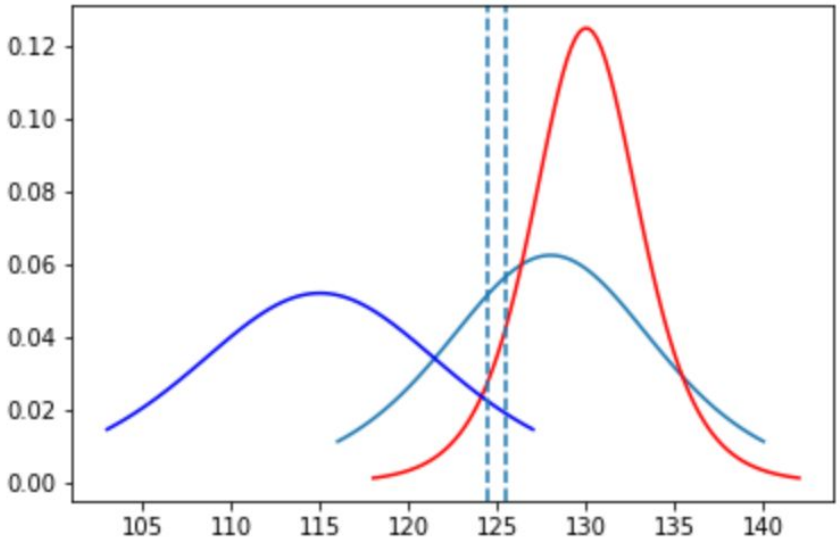
Tacotron 2

- В отличие от первой версии, tacotron 2 не нуждается в GLA.
- Аудио предсказывает модифицированная WaveNet.



Tacotron 2

- Вместо softmax используется 10-компонентная смесь логистических распределений (MoL).



Tacotron 2

- Качество итоговой модели выше baseline.

System	MOS
Parametric	3.492 ± 0.096
Tacotron (Griffin-Lim)	4.001 ± 0.087
Concatenative	4.166 ± 0.091
WaveNet (Linguistic)	4.341 ± 0.051
Ground truth	4.582 ± 0.053
Tacotron 2 (this paper)	$\mathbf{4.526 \pm 0.066}$

Tacotron 2

- Обучение на предсказанных спектрограммах лучше, чем на реальных.

	Synthesis	
Training	Predicted	Ground truth
Predicted	4.526 ± 0.066	4.449 ± 0.060
Ground truth	4.362 ± 0.066	4.522 ± 0.055

Tacotron 2

- WaveNet качественнее GLA.
- Mel-спектрограммы строго лучше линейных.

System	MOS
Tacotron 2 (Linear + G-L)	3.944 ± 0.091
Tacotron 2 (Linear + WaveNet)	4.510 ± 0.054
Tacotron 2 (Mel + WaveNet)	$\mathbf{4.526 \pm 0.066}$

Материалы

- https://arxiv.org/pdf/1712.05884.pdf
- https://arxiv.org/pdf/1609.03499v2.pdf
- an-explanation-of-discretized-logistic-mixture-likelihood
- https://wiki.aalto.fi

Вопросы?