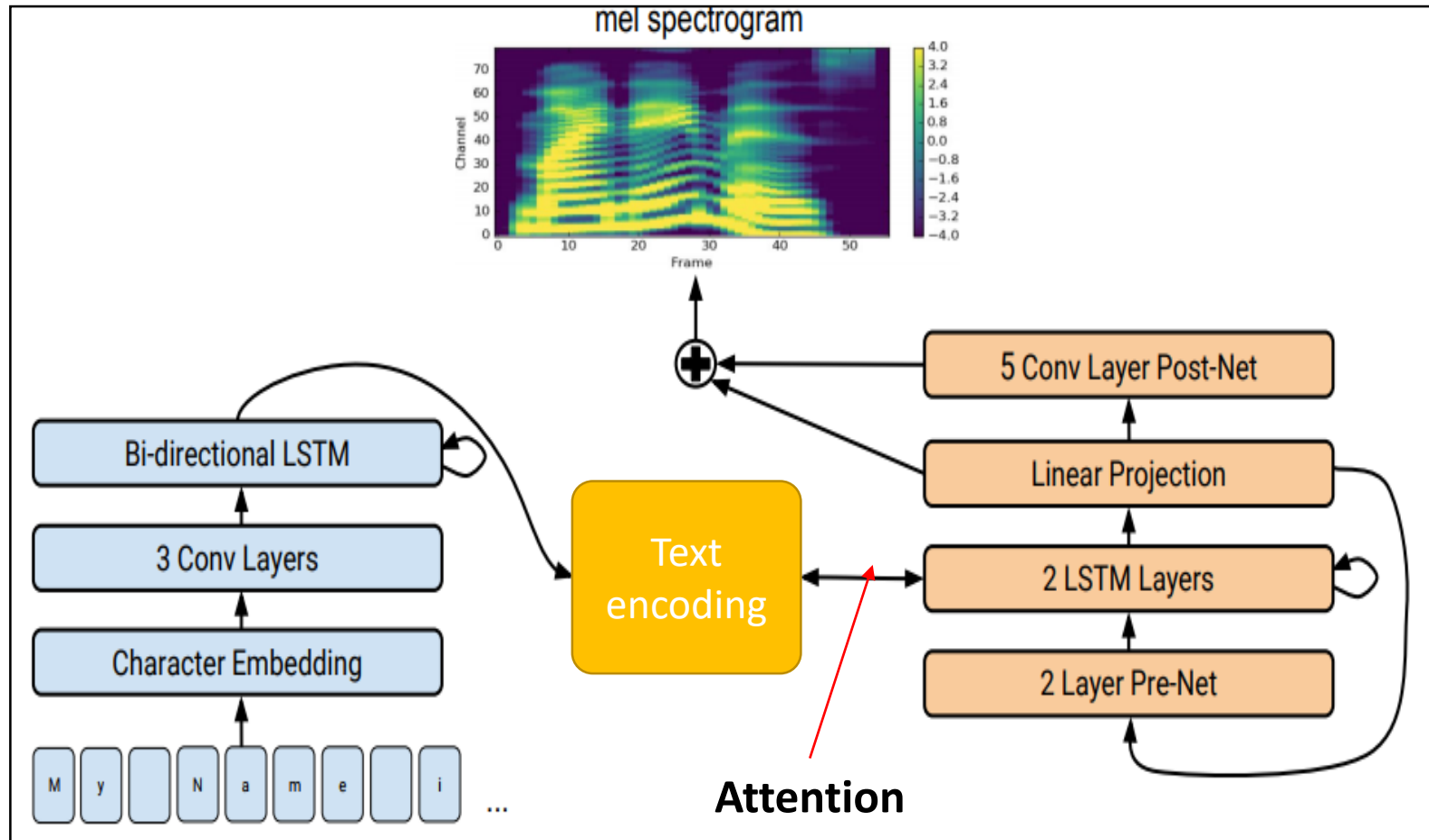


GraphTTS: graph-to-sequence modelling in neural text-to-speech

Yoon Hyung Lee

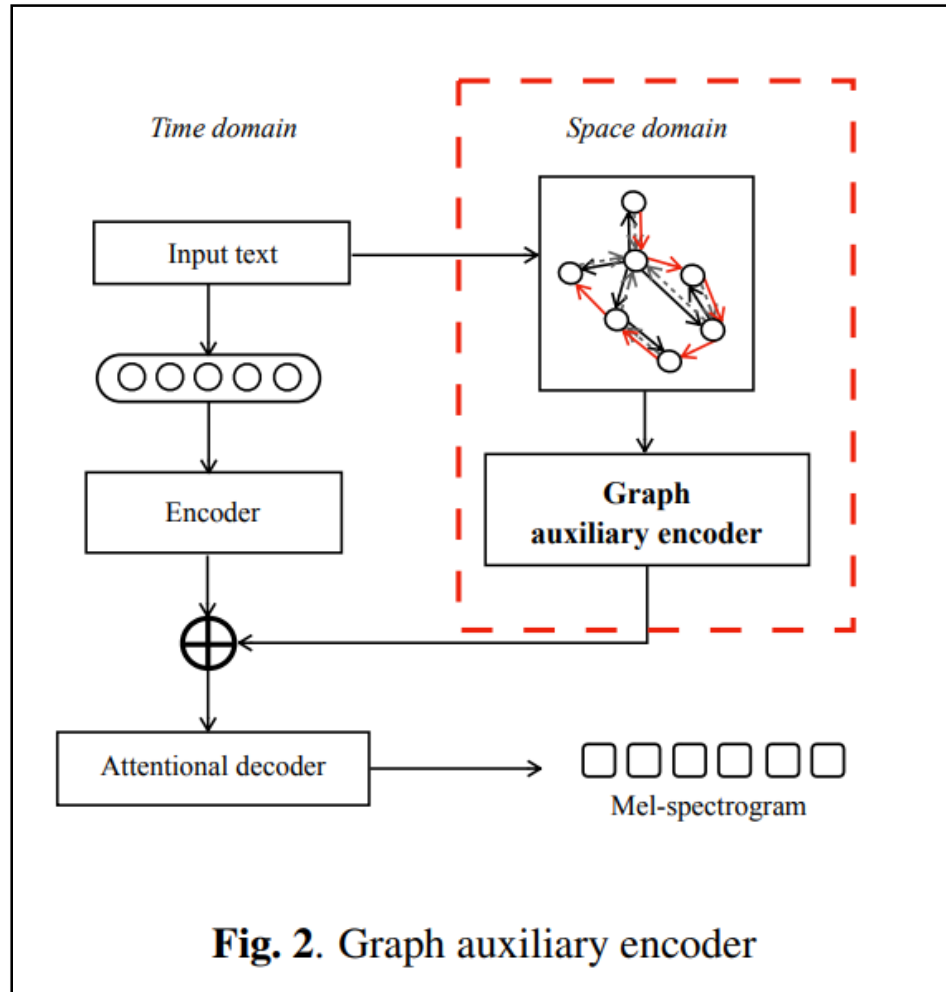
Seoul National University

Text-to-Speech



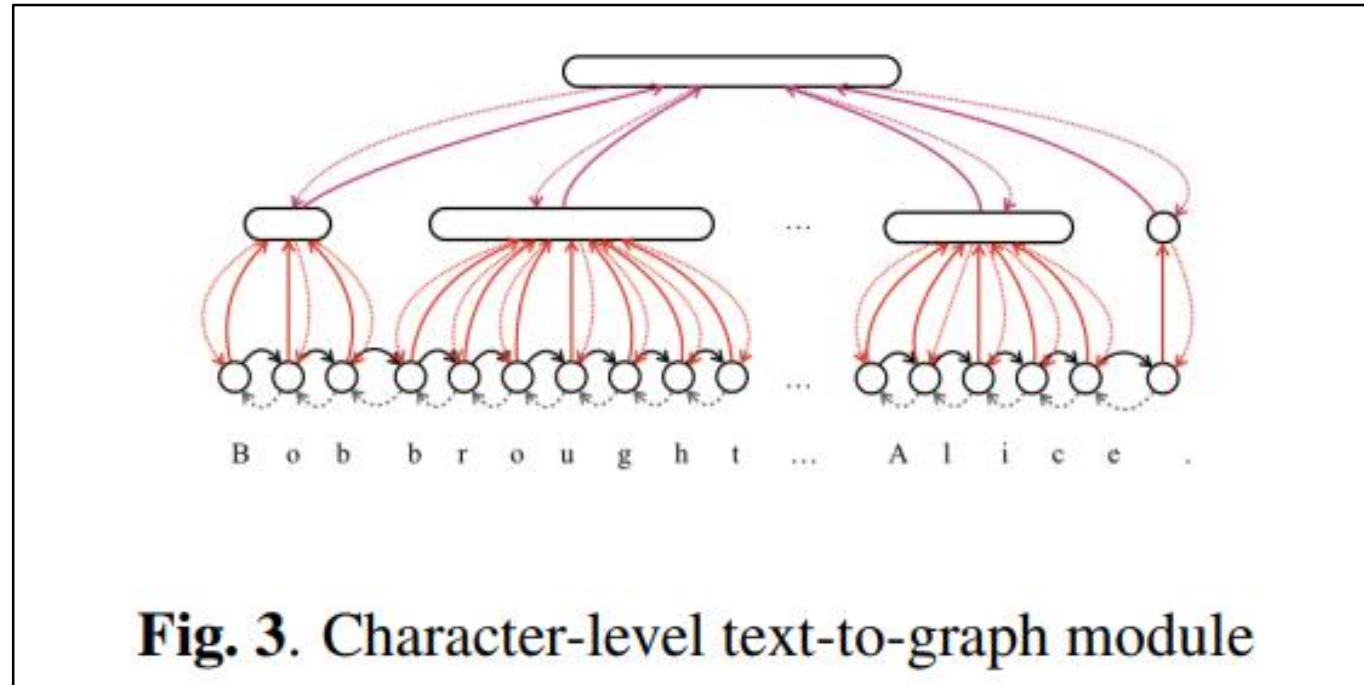
Sequence-to-sequence w/ attention

Graph Auxiliary Encoder (GAE)



- Alignment
 - base encoder (ex. BLSTM, ...)
- Prosody
 - graph auxiliary encoder
 - incorporate syntax information

Text-to-Graph conversion



Edge embedding

(E, 2, 3)

the # of edges

Source/target

the # of types

Experiments

- Dataset: 24-hour female (Eng/Chi) native speakers
- failure rate / prosody score

Metric	MOS	FR	PS
Tacotron2	4.17	0.2%	3.8
GraphTTS	4.39	5%	4.5
GAE	4.43	0.8%	4.35
Human recording	5.0	—	—