CONVOLUTIONAL SEQUENCE TO SEQUENCE MODEL WITH NON-SEQUENTIAL GREEDY DECODING FOR GRAPHEME TO PHONEME CONVERSION

Grapheme-to-phoneme (G2P) Task

- **G2P models** have been frequently employed in text-to-speech (TTS) and automatic speech recognition (ASR) systems.
- Convolutional sequence-to-sequence models have not been applied to G2P problem yet.

Non-sequential Greedy Decoding (NSGD)

- It is a good strategy to fill in the easiest blank first, referring to the hints given so far.
- According to this greedy strategy, the easy parts filled in earlier can be used as hints later.

How the proposed model works

- Our convolutional encoder-decoder model uses two inputs, source sequence \( S = (P_1, \ldots, P_N) \) and target sequence \( T = (T_1, \ldots, T_M) \) of tokens each.
- The model outputs prediction probabilities \( P = (P_1, \ldots, P_N) \).
- With NSGD, the model iteratively infers the most likely part among the candidate positions of \( T \) that are not inferred yet.
- NSGD initially starts with \( T_1 \), and updates \( T^{k+1} \) to \( T^{k+1} \) until the fully generated target sequence \( T^N \) is obtained.

Our Main Result

- We proposed a non-sequential greedy decoding method (NSGD) that generalizes traditional greedy decoding and two algorithms for inference and training.
- We also proposed a fully convolutional encoder-decoder model for NSGD.
- We were able to show the effectiveness of the proposed model and the decoding method by achieving the state-of-the-art performance on the G2P task.