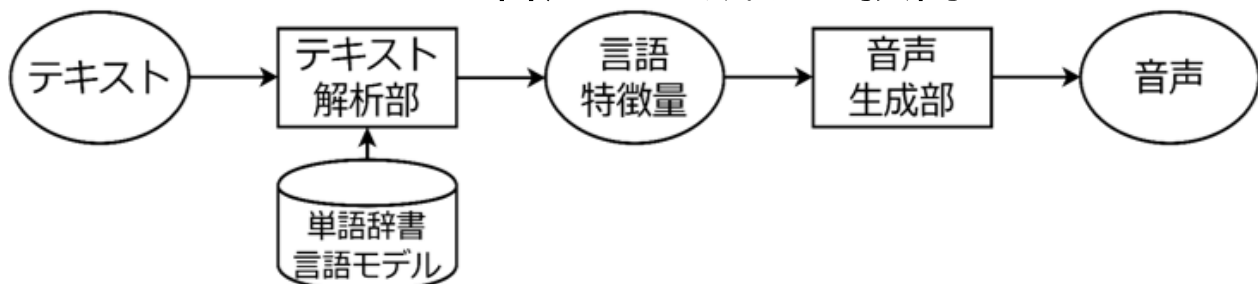


seq2seqモデルを用いた 日本語テキストからの読み仮名・韻律記号列推定の検討

M1 懸川直人(岡山大学 ヒューマンセントリック情報処理学研究室)

□ テキスト音声合成(TTS: text-to-speech)

- テキストから音声を生成する技術



□ 日本語はテキスト解析が困難

- 文字種の多さ, 複数の読み仮名など
- 人手だと音声・言語の専門的知識が必要

□ 機械翻訳に用いられる

Sequence-to-Sequence(seq2seq)モデルで高精度な推定を目指す

□ 入力: 漢字仮名混じりの日本語テキスト

□ 出力: 読み仮名・韻律記号のトークン列

その日に出場を決断した

ソノ ヒ' ニ @ シュ^ ツ ジョー オ ...

□ ... モデルで扱う単位 (トークン)

アクセント核	'
大・中ポーズ	.
小ポーズ	@
アクセント句境界	/
母音の無声化	^

□ 提案モデル

- Attentionつきseq2seqモデル
- Encoder: BiLSTM Decoder: LSTM

□ データセット

- ニュース記事からクロールしたテキスト
- 形態素解析・アクセント推定した結果を正解データとして扱う

□ 評価指標

$$\text{読み韻律正解精度}[\%] = 100 \times (N - S - D - I) / N$$

N : 総トークン数 S : 置換誤り数

D : 脱落誤り数 I : 挿入誤り数

□ ベースライン

- 提案モデルのEncoderをLSTMに変更したモデル

□ 実験結果

	読み	読み・韻律
BiLSTM Enc. (提案モデル)	84.3%	97.1%
LSTM Enc.	80.2%	96.8%

□ まとめ

- seq2seqモデルを用いて読み, 韻律それぞれを高精度に推定できることを確認

□ 今後の課題

- 提案モデルから得られた読み・韻律をもとに音声合成, その品質を調査