

# TTSとウェアラブルデバイスを用いた 曖昧な入力からの会話支援システムの検討

岩崎 茉理(M1 岡山大学 阿部研究室)



OKAYAMA UNIVERSITY

## 1. 研究背景・目的

- 発話困難者：
  - ✓ 舌の摘出や脳性麻痺などにより発話が困難な方
- Text-to-Speech(TTS):
  - ✓ テキストから音声を作成する技術
- 既存の会話支援
  - ✓ 手話 ⇒ **習熟コストが高い**
  - ✓ スマートフォン等による音声合成 ⇒ **画面を見て操作する必要がある**

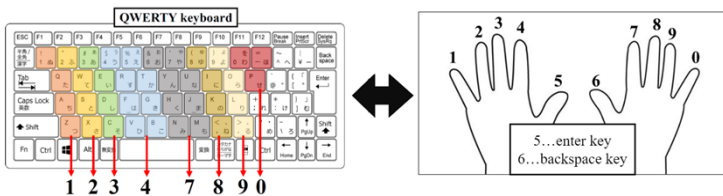


### 提案方式

TTSとウェアラブルデバイスを用いた  
入力に目視のいらない会話支援システム

## 2. 提案方式

### ① タッチタイピングに基づく入力方式



キーと指の対応関係に基づき文字列を推定

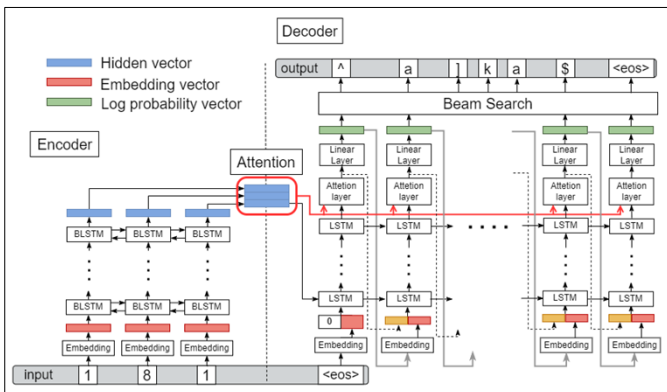
例) 私は大学生です  
21412871318418723833275  
↓  
音素・韻律記号列  
^w a [ t a s h i w a # d a  
[ i g a ] k u s e e d e s u \$

課題: 入力の曖昧性

121 ⇐ 朝 (asa)  
泡 (awa)

前後の文脈に考慮

### ② ニューラル機械翻訳(NMT)を用いた文章推定



➢ Attention機構で注意を向ける箇所を決定することで文脈を保持した推定が可能

➢ NMTのパラメータ

	Encoder	Decoder
RNN	Bi-LSTM	LSTM
レイヤ数	5	5
ユニット数	512→512	1層: 1024→512 2~5層: 512→512

## 3. 評価実験

### 学習条件

活性化関数	ReLU	損失関数	Softmax Cross Entropy
最適化手法	Adam	ミニバッチ	128文

客観評価実験: 4種のモデルでテストデータ(2000文)を翻訳し変換精度で評価

Model	出力シンボル	コーパス	データ数 (文)	
			Training	Validation
Ph_C1	音素・韻律	JESC	270万	2000
Ph_C2	音素・韻律	Yahoo	3200万	2000
Ph_C1+C2	音素・韻律	JESC & Yahoo	3470万	2000
Ja_C1+C2	仮名・漢字	JESC & Yahoo	3470万	2000

### 結果

Model	シンボル	文	評価指標
Ph_C1	84.02%	26.85%	① シンボル正解率 ✓ 1文字単位の正解率
Ph_C2	89.72%	43.65%	
Ph_C1+C2	91.48%	43.45%	② 文正解率 ✓ 文単位の正解率
Ja_C1+C2	84.02%	31.10%	

- データ数に比例してシンボル正解率が向上
- 音素・韻律記号列の方が推定に適している

### 主観評価実験

- 会話相手が合成音声をどの程度理解したか(伝達度)
- 提案方式を用いてお題あてのタスクを実施
  - ✓ 話し手: ① お題を決める (例: りんご)  
② 特徴をシステムに入力 (例: 赤い果物です)
  - ✓ 聞き手: 合成音声を聞き、お題を当てる

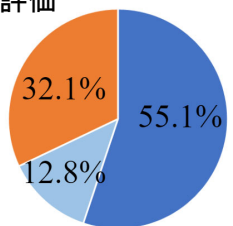
### 結果

✓ 推定精度(全76文)

シンボル	90.57%
文	50.00%

✓ 合成音声の評価

- 聞き取れる
- 推定できる
- 推定できない



- 伝達度(聞き取れる+推定できる)は **67.95%**
- 文正解率 < 伝達度  
⇒ **推定誤りがあっても意味を推測できる場合がある**

### 変換誤りを含むが推定できる例

入力文	その多くは著作権を無視しているそうです
推定文	^s o [ n # o ] o k u w a # g y o [ s a ] c k e n o # . . .

## 4. まとめと今後の課題

### まとめ

- TTSとウェアラブルデバイスを用いた会話支援システムの検討
- 最も良いモデルの変換精度は **91.48%**
- 合成音声の伝達度は **67.95%**

### 今後の課題

- ウェアラブルデバイスを用いた実験の実施
- タイプミスへの対処