

ASR・MT・TTSによる多言語音声翻訳システムの構築と評価

趙 澤毅† 林 実†
† 明星大学理工学部

1. はじめに

COVID-19 の収束に伴い、世界中からの訪問者が増え、多言語音声翻訳の需要が見込まれる。ここでは、音声認識(ASR)、機械翻訳(MT)、音声合成(TTS)技術を用い、多様な言語に対応できる自動音声翻訳システムの構築とその評価を行う。

2. 音声翻訳システム

自動音声翻訳システムは、ASR、MT、TTS によって構成され、自然に話された原言語の音声の音響信号をテキストに変換し、そのテキストを対象言語のテキストに翻訳し、翻訳された対象言語を自然な対象言語の音声に自動で生成するシステムである[1, 2]。

3 実験方法

3.1 実装

実験で実装した自動音声翻訳システムの構成を図 1 に示す。まず、PythonでStreamlitを用いWeb Graphical User Interface、多言語 ASR、多言語 MT および多言語 TTS の各システムを構築する。構築した各システムを組合せた音声翻訳実験では、音声をアップロードし、Speech-to-Text に送り、音声から特徴を抽出し、エンコーダ、デコーダ処理しテキストを出力する。音声認識で得られたテキストを設定した対象言語に、API より翻訳モデル、復号アルゴリズム処理し言語モデルを通して、テキスト変換する。変換されたテキストは Text-to-Speech により、フロントエンド処理し音響モデルとボコーダを通して音声を生成する。

3.2 評価方法

実装した自動音声翻訳システムの性能を定量化するため評価を行った。一連の自動音声翻訳プロセスにおいて、音声認識および音声翻訳結果をそれぞれ評価指標とした。音声認識における単語認識精度を式(1)に、単語正解率を式(2)に、音声翻訳結果における BLEU 値は式(3)にそれぞれ示す[3]。

$$\text{Correct} = \frac{\text{正解単語数}}{\text{全単語数}} \times 100\% \quad (1)$$

$$\text{Accuracy} = \frac{\text{正解単語数} - \text{挿入誤り}}{\text{全単語数}} \times 100\% \quad (2)$$

$$\text{BLEU} = \text{BP}_{\text{BLEU}} \cdot \exp \left(\sum_{n=1}^N w_n \log p_n \right) \quad (3)$$

$$p_n = \frac{\sum_i \text{翻訳文 } i \text{ と参照文 } i \text{ で一致した } n\text{-gram 数}}{\sum_i \text{翻訳文 } i \text{ 中の全 } n\text{-gram 数}}, \quad w_n = \frac{1}{N}$$

p_n は、評価コーパス全体について翻訳結果と参照訳を比較し、 n -gram の一致率を算出しているものである。

評価は、合成音声と自然音声の 2 種類の音声データを用い、本実装音声翻訳システムの出力と参照訳として Deepl の出力を比較し、BLEU 値により評価を行った。

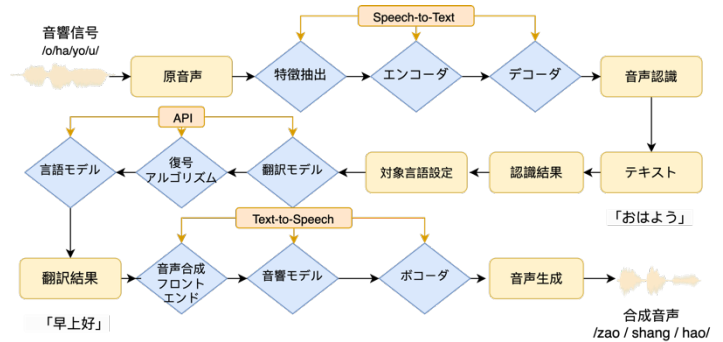


図 1. 実装した自動音声翻訳システムの構成

4. 実験結果

本実装システムの評価結果をそれぞれ表 1 に示す。

表 1. 本実装システムの評価結果

音声合成データから実験結果						
	データ数	BLEU 値	BLEU 評価	単語数	単語認識精度	単語正解率
CN-JP	30	0.5959	適切かつ流暢な翻訳	424	98.58%	98.58%
EN-JP	30	0.5196	適切かつ流暢な翻訳	230	97.24%	97.24%
JP-EN	30	0.6252	高品質な翻訳	371	98.65%	98.65%
JP-CN	30	0.5034	適切かつ流暢な翻訳	371	98.65%	98.65%
自然音声データから実験結果						
CN-JP	30	0.5196	適切かつ流暢な翻訳	329	98.48%	98.48%
EN-JP	30	0.4974	適度な品質の翻訳	267	96.25%	95.14%
JP-EN	30	0.5169	適切かつ流暢な翻訳	280	97.50%	97.50%
JP-CN	30	0.4631	適度な品質の翻訳	280	97.50%	97.50%

5. 考察

表 1 に示すように、音声認識精度に応じて翻訳の BLEU 値が高くなる傾向が見られることから、今後は、音声認識の精度向上を図り、自動音声翻訳システムの性能を高めることを目指す。

6. まとめ

多言語音声翻訳システムの構築と評価を行った結果、認識精度約 98%、BLEU 値約 0.5 の自動音声翻訳システムを実装できた。

参考文献

[1] 中村哲, "音声言語の自動翻訳," コロナ社, 2018.
 [2] 中村哲, "自動音声翻訳から自動音声通訳へ," 情報処理学会研究報告 音声言語情報処理 2019-SLP-12, pp.1-6, 2019.
 [3] 今村賢治, 隅田英一郎, 松本裕治, "機械翻訳自動評価指標の比較," 言語処理学会第 10 回年次大会発表論文集, pp.452-455, 2004.