

多言語音声翻訳システム VoiceTra の開発*

松田 繁樹, 木村 法幸, Paul Dixon, 大熊 英男, 柏岡 秀紀, 隅田 英一朗 (NICT)

1 はじめに

独立行政法人情報通信研究機構 (NICT) では、地球上の様々な言語を話す人々との円滑なコミュニケーションを実現する多言語音声翻訳技術を確立するため、実際の利用環境を想定した実証実験 [1][2] や、世界各国の研究機関との共同研究 [3][4] を通して、多言語音声翻訳システムの研究開発を行っている。本報告では、NICT が世界で初めて開発したスマートフォン向けネットワーク型多言語音声翻訳システム VoiceTra の概要及び、本システムを用いた大規模実証実験で得られた実利用音声ログに対する分析結果について述べる。

本実証実験は、多言語の自動音声翻訳システムにおける実利用音声データの分析及び、大量に収集された音声データを用いた音声翻訳システムの性能改善を目的として行った。本実証実験で用いた VoiceTra システムは、音声入力や翻訳結果の表示、合成音声の再生を行うクライアントアプリと、音声認識、言語翻訳、音声合成を行うサーバシステムから構成されている。クライアントアプリは、米国 Apple 社のスマートフォン iPhone 向けに開発し、2010 年 7 月末から AppStore にて無料で一般公開した。また、Android OS が導入されたスマートフォン向けにもクライアントアプリを開発し、2011 年 4 月から AndroidMarket にて無料で一般公開した。報道発表や、テレビ局等の報道機関からの取材及び番組上での紹介を通して、2012 年 12 月時点で約 70 万ダウンロード、約 1000 万アクセスを達成した。

本報告では、本実証実験で用いた VoiceTra システムの概要及び、収集された大規模音声データを人手で聴取し分析した結果について述べる。

2 多言語音声翻訳システム VoiceTra

2.1 VoiceTra の概要

VoiceTra は、ネットワーク型の多言語音声翻訳システムである。ユーザの発話した音声波形はネットワークを通して VoiceTra サーバへ送信され、処理負荷の大きな音声認識、言語翻訳、音声合成処理がサーバ側で行われる。その為、大規模な音声認識用モデルや言語翻訳用モデルの利用や、実利用音声データの収集及び、それら収集データを用いた音声認識用モデルや言語翻訳用モデルの更新を容易に行うことができる。

表 1 に VoiceTra で翻訳可能な 21 言語の内訳を示す。印の付いた日本語、英語、中国語、インドネシア語、ベトナム語、韓国語の 6 言語は音声認識と音声合成による入出力が可能である。その他の言

表 1. VoiceTra で翻訳可能な言語のリスト

日本語	マレー語
英語	ポルトガル語
中国語	ポルトガル語 (ブラジル)
インドネシア語	ロシア語
ベトナム語	スペイン語
韓国語	タガログ語
台湾華語	タイ語
フランス語	アラビア語
ドイツ語	オランダ語
ヒンディ語	デンマーク語
イタリア語	



図 1. VoiceTra の動作画面

語については、テキストによる入出力が可能である。

2.2 ユーザインターフェース

本システムは広く一般に VoiceTra クライアントアプリを利用して貰うため、所内にて必要な機能や使用感等のレビューを行い、直観的かつ分かり易いインターフェースの設計を心掛けた。

図 1 に VoiceTra の動作画面を示す。図の左側は日本語音声から英語音声への翻訳画面、右側は言語選択画面である。翻訳画面の上段は音声認識結果、下段は英語へ翻訳した結果である。翻訳結果は、テキストだけでなく文頭の再生ボタンをタップすることで英語の合成音声を再生することができる。中段は、英語の翻訳結果を日本語へ逆翻訳した結果であり、音声認識結果の意味と比較することで、翻訳結果の正しさをユーザ自身で確認することができる。音声入力は、ユーザが耳元に iPhone を近付けたタイミングで開始され、手元に戻したタイミングで終了する。この音声入力方法は、ユーザにとって自然な動作でありながら、強制的にマイクと口の間の距離を近付けることができ、結果として高い SNR を持つ音声を端末へ入力することができる。翻訳言語の方向は、画面上部の矢印をタップすることで即座に変更することができる。

所内レビューで得られた意見から、左下の時計のボタンをタップすることで、過去に翻訳した結果の表示、虫眼鏡のボタンをタップすることで文字の拡大を行う機能の追加を行った。また、会話する相手は VoiceTra の利用経験が無いことが想定されるため、日本語、英語、中国語 (簡体字, 繁体文) での操

* "Development of multi-lingual speech-to-speech translation system "VoiceTra" by Shigeki Matsuda, Noriyuki Kimura, Paul Dixon, Hideo Okuma, Hideki Kashioka, Ei-ichiro Sumita.

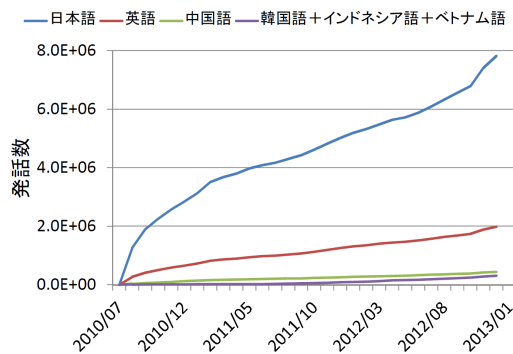


図 2. VoiceTra へのアクセス数

作方法の説明を表示できるようにした。言語選択画面は、当初は言語名のみ表示であったが、各言語の左側に人の顔のアイコンや鉛筆のアイコン、吹き出しのアイコンを付与することで、音声認識が可能な言語、テキスト入力のみ可能な言語、音声合成が可能な言語の判別が容易になるように修正した。

3 利用者からの反応への対応

実証実験期間中は、利用者からの質問や要望がサポート宛メールボックスに多数寄せられた。最も多かった質問は音声入力方法に関するもので、iPhoneを耳に近付けても振動しないと書いた内容であった。これは、iPhoneのバイブレーションの設定方法の説明不足から起った問題であった。この問題に対処するため、VoiceTraのサポートページとヘルプ画面にバイブレーションの設定方法についての説明を掲載することで対応した。また、音声入力方法が分かりづらい等の質問に対して、使いかたビデオの掲載や、SiriやGoogle音声検索で用いられているマイクボタンをタップすることによる音声入力機能を新たに追加した。これにより、サポート宛メールの件数は減少していった。

その他、利用履歴の削除方法に関する質問や、海外でのネットワークの利用料金に関する質問等に対する回答をサポートページに掲載するなどの対応を行った。クライアントアプリのバグの報告や、音声翻訳をもっと早くして欲しい等の要望、目の不自由な方からのVoiceOver機能(タップした場所の意味や目的を合成音で読み上げる機能)の追加についての要望に対してクライアントアプリのアップデートで対応を行った。

4. 実利用音声ログの分析

VoiceTraサーバへのアクセス数を図2に示す。図に示すように、実証実験の開始以来、順調にアクセス数が増加しており、2012年12月時点で約1054万発話が収集されている。言語の内訳は、日本語が74%、英語が19%、中国語が4%、韓国語が2%、インドネシア語とベトナム語は各々1%未満である。

国内からのアクセスは94.7%、海外からのアクセスが5.3%であった。国内から頻りに利用している上位100端末、海外から頻りに利用している上位50端末で発話された音声を手で聴取し利用目的を分析した結果を図3に示す。国内利用において最

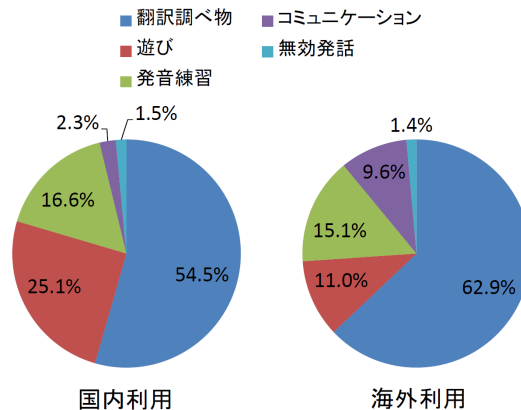


図 3. 国内と海外での利用目的

も多い利用目的は、「翻訳調べ物」の54.4%であった。相手がない状況で事前に翻訳内容を確認したり、単語の辞書引き等を行っている判断された発話である。次いで、「遊び」の25.1%であり、アニメの有名なセリフや歌詞など、遊びを目的とした発話である。「発音練習」の16.6%は、日本人が英語音声を発話し音声認識結果を確認している等の発話である。最後に、「コミュニケーション」の2.3%は、外国人との会話に利用している発話である。このように、国内からの利用において、音声翻訳としての実用的な利用の割合は、全体の56.7%(翻訳調べ物とコミュニケーション)であった。一方、海外からの利用では、「翻訳調べ物」の割合が63.0%に増加し、さらに外国人との会話に用いられた「コミュニケーション」の割合が9.6%と大幅に増加しており、全体として、72.6%が音声翻訳としての実用的な利用であった。

4. まとめ

本報告では、NICTにて世界で初めて開発したスマートフォン向けのネットワーク型多言語音声翻訳システムVoiceTraの概要及び、本システムを用いた実証実験で収集された音声ログの分析について述べた。VoiceTraクライアントアプリの開発において、所内レビューや利用者からの質問要望メールへの対応を行うことで、サポート宛メールの件数は減少していった。また、国内利用において全体の56.7%が音声翻訳としての実用的な利用であったのに対し、海外利用では、全体の72.6%の発話が実用的な利用であった。今後は、本実証実験で収集された1000万発話から成る実利用音声データを用いた、音声認識用モデル、翻訳用モデルの高精度な教師無し学習法に関する検討を行う予定である。

参考文献

- [1] 清水他, “携帯電話を用いた日中音声翻訳実証実験システムの評価,” 音講論(春), 3-Q-27, 2009.
- [2] 河井他, “H21年度全国音声翻訳実証実験の概要,” 音講論(秋), 3-9-6, 2010.
- [3] 中村他, “アジア言語音声翻訳コンソーシアム: A-STARについて,” 音講論(秋), 1-3-14, 2007.
- [4] C. Hori, et al., “Breaking Down the Language Barrier Among 23 Countries: Network-Based Speech Translation Communication Protocol based on ITU Standards,” New Breeze, Vol. 24, No. 4, 2012.