

### 3.5.1 ユニバーサルコミュニケーション研究所 音声コミュニケーション研究室

室長 河井 恒 ほか7名

#### グローバルコミュニケーション計画の実現に向けた音声認識、音声合成、音声対話の研究開発

##### 【概要】

本研究室では、人間にとって自然で簡便な情報伝達手段である音声によるコミュニケーションを用いた音声翻訳・音声対話システムの実現に向けて、音声の認識・合成・話処理の研究開発を行っている。特に、平成26年4月に総務省のグローバルコミュニケーション(GC)計画が開始されたことを受け、多言語音声認識技術、同技術強化の前提となる大規模な多言語音声コーパスの収集及び多言語音声合成技術の研究開発に注力している。

##### 【平成27年度の成果】

##### ●大規模な多言語音声コーパスを構築

音声コーパスの収集は、NICTが実証実験として無料公開を行っている音声翻訳アプリVoiceTraを使用した際にサーバーに記録される利用ログの音声データの発話内容を人手で書き起こす方法と、模擬会話による音声収録、の2種類の方法で行った。VoiceTra利用ログが豊富に存在する日本語と英語では、前者を中心に、それ以外の言語では、後者を中心に音声コーパスを構築した。VoiceTra利用ログは、発話様式、雑音環境が音声翻訳技術の実利用状況に近く、またコーパス構築に要するコストが模擬会話に比べて小さい反面、GC計画が2020年時点での利用形態として想定する、旅行、交通、ショッピング、防災、医療などの会話と比較すると、発話内容が想定場面に適合しているか否かが不明である。模擬会話は、発話内容を想定場面に合わせて設計可能である反面、発話様式が棒読みになりがちという欠点がある。

模擬会話の収録においては、医療、ショッピング等の各分野において、二者(例えば、医療分野であれば、患者やその家族と医者や看護師等の医療従事者)間で交わされる会話状況を想定し、各分野10パターン以上の対話シナリオを言語ごとに作成した。話者は対象言語を日常的に話す母語話者とし、男女比や年代、地域(方言)のバランスを考慮して募集し、対話シナリオに即した二者による模擬会話を収録した。収録には実際の音声翻訳アプリケーションを想定してスマートフォン上の専用の収録アプリケーションを用いた。

本年度は、VoiceTra利用ログに関しては、日本語、英語、中国語、韓国語の4言語を合わせて2,814時間の音声コーパスを構築した。模擬会話に関しては、中国語、韓国語、インドネシア語、ベトナム語、タイ語、ミャンマー語、ブラジルポルトガル語の7言語を合わせて4,268時間の音声コーパスを収録した。これらにより、前年度までに構築したものと合わせて、7,799時間の多言語音声コーパスを得た。図1に言語別・収集方法別の内訳を示す。

今後は、コーパス規模と認識精度の関係を実験的に明らかにすることにより、適正なコーパス規模を予測するとともに、模擬会話の収録方法を改良し、コスト削減を図る予定である。

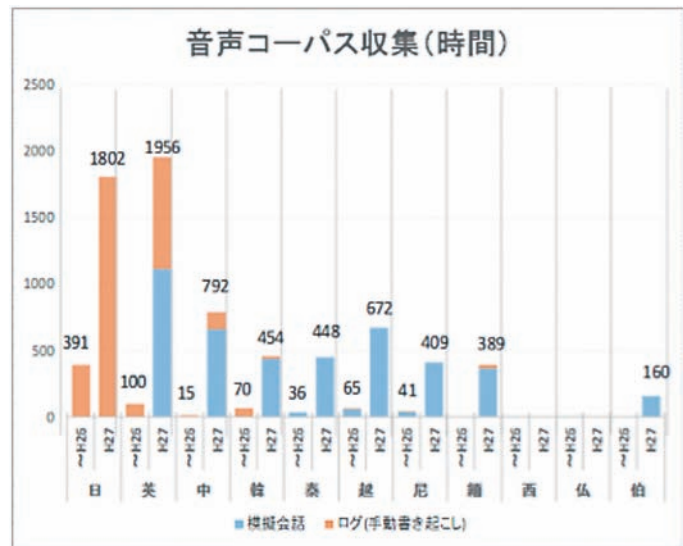


図1 多言語音声コーパスの言語別・収集方法別内訳

##### ●多言語音声認識・合成の研究開発を推進

本年度は、タイ語、ミャンマー語、ブラジルポルトガル語の音声認識システム及びミャンマー語、ブラジルポルトガル語、ロシア語の音声合成システムを新たに試作した。

### ●世界初のミャンマー語の音声翻訳システムを無料公開

ヤンゴンコンピュータ大学(USCY)より研究員を受け入れて、GC計画の対象である10言語の1つミャンマー語(ビルマ語)の自動音声認識・音声合成の研究開発を進めてきた。本年度はDNN(Deep Neural Network)に基づき3万9千語の語彙に対応した音声認識部を構築し、音声合成部とともにテキスト翻訳部と統合、多言語音声翻訳システムVoiceTraに組み入れ、2015年10月から無料公開した(図2)。ミャンマー語の公開音声翻訳システムは有料・無料を問わず世界初である(当研究室調べ)。ヤンゴンにてUSCYと共同で報道発表を実施した結果(写真1)、衛星放送を含む4つのテレビ局で取り上げられ、発表後4週間で6万発話を超える利用を数えるなど大きな反響を得た。年度末までに通算13万発話の利用があった。

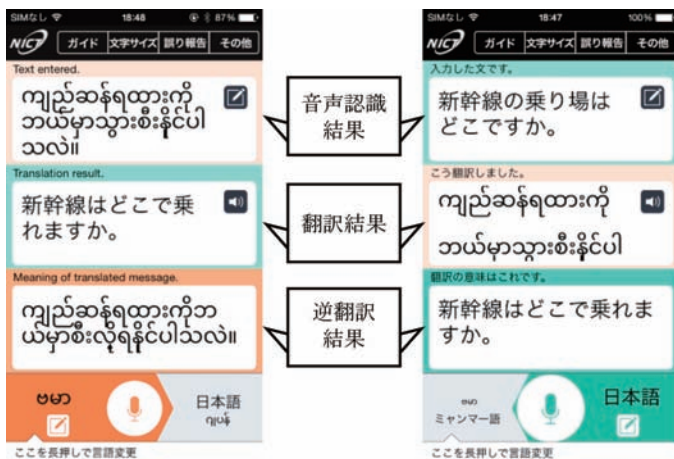


図2 ミャンマー語と日本語との相互音声翻訳の例  
(音声翻訳アプリVoiceTraの画面)



写真1 ミャンマー国内での記者会見の様相

### ●クロスリンガル音声対話システムを試作

異なる言葉話す人々が「言語の壁」を意識せず、同時に対話することが可能な情報案内システムの実現を目指して、クロスリンガル音声対話システムの研究開発を行った。クロスリンガル音声対話システムは、(1)言語と独立に話の流れを把握し、どの言語で話しかけられても一連の対話を続けることができる、(2)ユーザーが何語で話したのかを識別し、同じ言語で応答を返すことができる、(3)異なる言語のユーザーが話の流れを理解できるよう、対話内容を自動翻訳し、画面に表示する、という3つの特長を有している。

コンセプト検証のため、日英中韓の4言語で簡単な交通案内を行うプロトタイプシステムを試作し、紹介ビデオを<https://www.nict.go.jp/video/cross-lingual.html>で公開している。図3は、対話の例である。

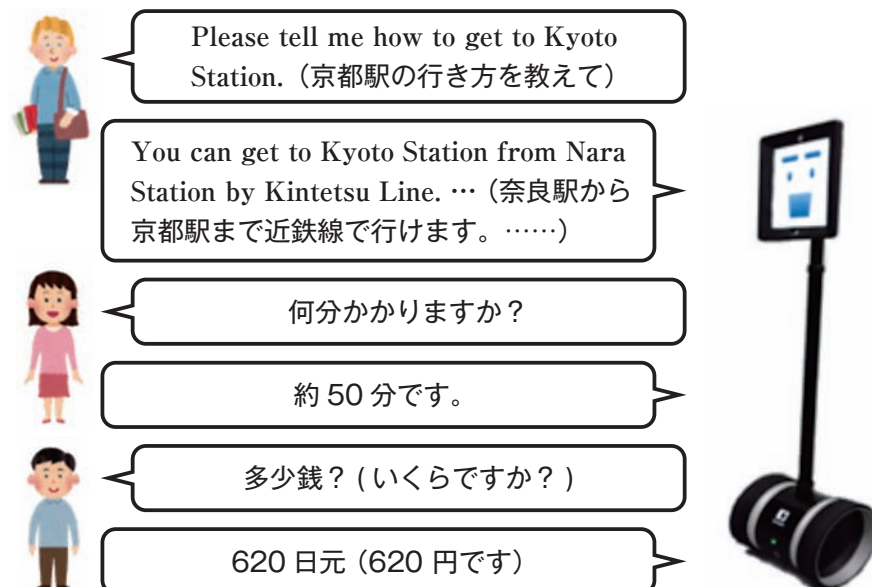


図3 日英中3言語の対話例