

# 対面業務の外国語会話を支援する同時通訳システム

Simultaneous Interpretation Technology Supporting Conversations in Foreign Languages for Face-to-Face Services

釜谷 聡史

坂本 明子

住田 一男

■ KAMATANI Satoshi

■ SAKAMOTO Akiko

■ SUMITA Kazuo

音声による外国語会話の機会が増加するに伴い、様々な場面で使える機械通訳システムへのニーズが高まっており、ユーザーの発声を逐次的に通訳するアプリケーションが提案されている。

東芝は、機械通訳における即時性を向上させるため、ユーザーの連続的な自由発声を受け付けて、同時並行で翻訳する同時通訳システムを開発した。このシステムは、ユーザーが対面で会話する場面、特に、これまで応用が広がっていなかった、顧客との円滑な会話を要する対面業務において利用が期待でき、現在、日本語と英語及び日本語と中国語の間の通訳が可能である。また、評価実験を行い、旅行案内と日常生活の場面を想定して設定した会話の目的を、平均約90%の解決率で達成できることを確認した。

With the increasing opportunities for conversation in foreign languages, demand has been growing for a simultaneous machine interpretation technology that can be used in many different situations.

Toshiba has developed a simultaneous interpretation system for continuous speech conversation taking place in various face-to-face services at stores, reception desks, counters at public offices, and so on. This system, is capable of both Japanese/English and Japanese/Chinese interpretation, supports smoother communication between speakers of different languages by processing their continuous spontaneous speech and incrementally outputting the interpretation results. As a result, a user can immediately understand what a conversational partner is saying. We have conducted field experiments and confirmed that a solved task ratio of approximately 90% is achieved for various tasks including buying souvenirs and asking for directions regarding a bus route.

## 1 まえがき

資本、情報、及び人々の交流や移動が地球規模で行われるようになり、人々が外国語に触れる機会が増えている。また、近年の音声言語処理技術の進展や計算機の進化により、旅行会話や日常会話の用途であれば、実用的な機械通訳アプリケーションが現れつつある<sup>(1)</sup>。

しかし、従来の機械通訳アプリケーションでは、ユーザーは、会話の相手に伝えたい内容をあらかじめ考えて、簡潔な文で発声する必要がある。また、全ての内容を話し終えるまで、その内容が会話の相手に伝わらないなど、円滑な会話が難しい。そのため、旅行案内所、小売店舗、及び役所の窓口など、円滑な会話を要するビジネスシーンでの活用が広がっていない。

そこで東芝は、前述のような顧客と対面して会話する業務(対面業務)を想定し、ユーザーの連続的な自由発声を受け付け、これを同時並行で翻訳する同時通訳システムを開発した。音声認識と機械翻訳を並行動作させ、ユーザーが連続的に発声する自由発声音声を処理し、訳すべき単位を検出して、順に翻訳する。これにより、ユーザーは発声の長さを気にする必要がなくなる。また、発声の途中から会話の相手に内容が順次

伝えられ、円滑な意志疎通が可能になる。

ここでは、開発した同時通訳システムの概要と、このシステムを介した会話の達成度を評価した実験の結果について述べる。

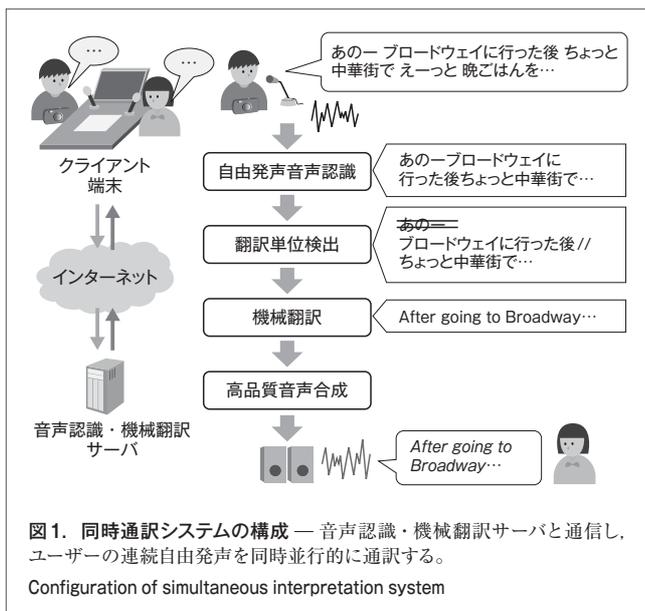
## 2 同時通訳システムの構成と処理技術

開発した同時通訳システムは、通訳機能を実現する音声認識・機械翻訳サーバと、ユーザーインタフェースを提供するクライアント端末から構成されている(図1)。サーバとクライアント端末は、一般のインターネット回線を介して接続され使用される。

システムは、次の処理によって、話者の発声に追従して連続的に通訳する、同時通訳を実現している。

- (1) フィラー<sup>(注1)</sup>を含む自由発声音声を、一定の無音区間ごとに確定して認識する。
- (2) 認識結果から翻訳するべき単位を検出する。
- (3) 翻訳単位ごとに前から順に翻訳する。
- (4) 必要に応じて、翻訳結果に基づき音声を合成する。

(注1) 「あー」など、会話の隙間を埋めるための、具体的な意味を担わない語句。



(1)と(4)の処理に用いている音声認識と音声合成の技術は、この特集に掲載されている別の論文でそれぞれ詳しく述べられているので、ここでは(2)と(3)の処理について述べる。

### 2.1 翻訳単位の検出

現在の音声認識技術の多くは、所定時間以上継続する無音区間を検出すると、それ以前の音声系列を一つの単位と見なして処理し、これに対応する文字列を出力する。しかし、無音区間の存在は発声のくぎりの手がかりとはなるが、必ずしも翻訳に適した単位のくぎりとはならない。また、無音区間の長さは話し方や話の状況によって変化し、あらかじめ適切な時間を設定しておくことが難しい。

更に、人間の自由な発声には、「ええと」や「あー」のようなフィルラーが含まれる。このような語句は通訳の必要がないばかりか、解析の妨げになることがある。

これらの問題を解決するため、音声認識が出力した結果を解析し、翻訳すべき単位を検出する。翻訳単位の検出は、翻訳不要句の除去と翻訳単位の推定という二つの段階で行う(図2)。

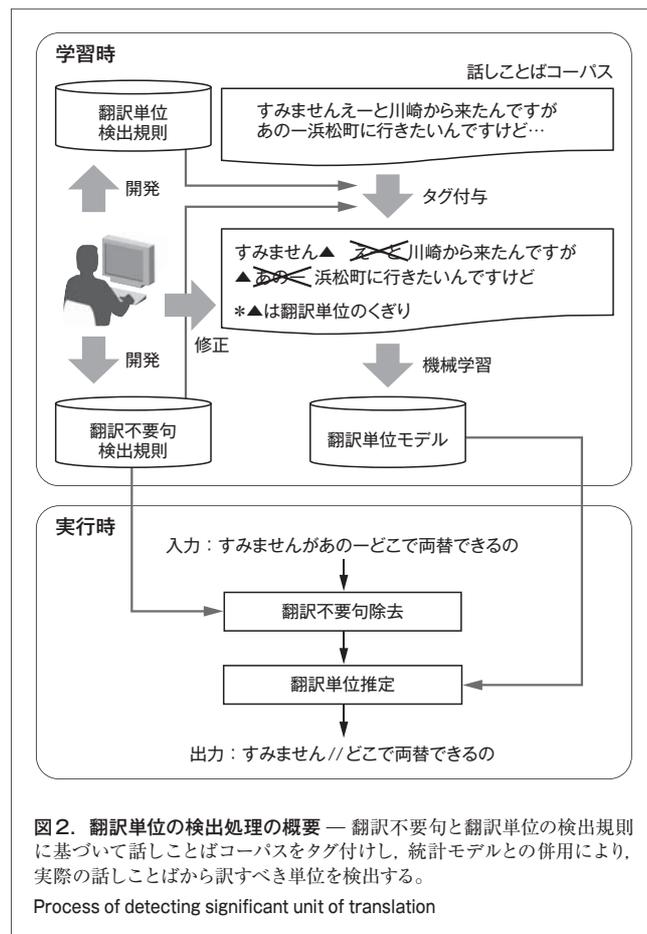
まず、音声認識結果からフィルラーなどの翻訳不要句を除く。ここでは、解析結果に付与されている品詞や表層の系列情報に基づく、単純なパターンマッチの手法を使用した。

次に、音声認識が出力した文字列中に、複数の文あるいはは分割して訳出できる単位があれば、その開始位置を推定し翻訳単位とする。翻訳単位の検出はラベリング問題として扱い、学習のためのコーパスを作成し、機械学習により翻訳単位モデルを構築して推定する。

### 2.2 ハイブリッド機械翻訳エンジン

当社は、性質の異なる2種類の翻訳方式を開発している。

一つは、人間が作成した文法や辞書を参照して翻訳する規

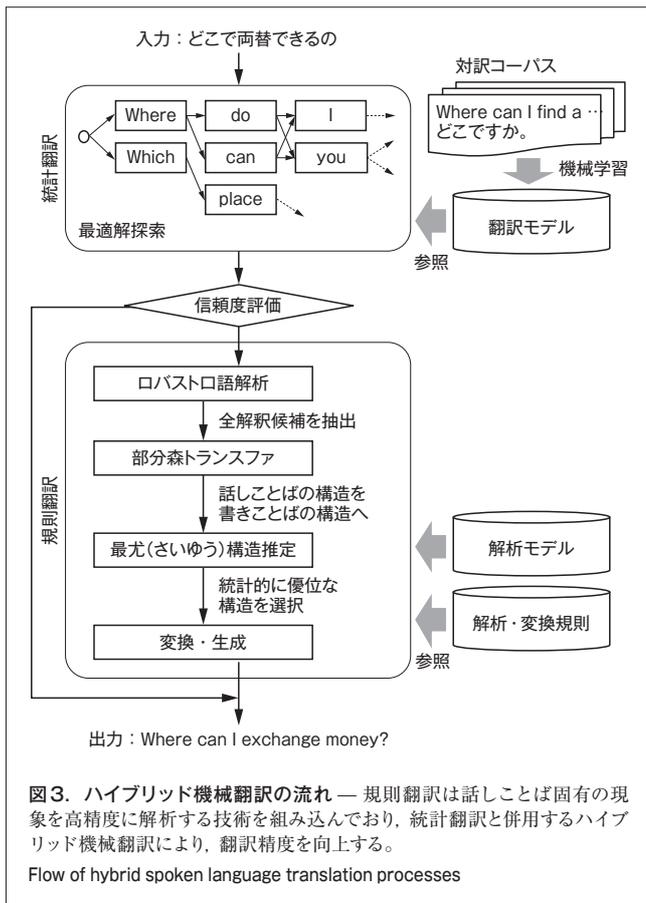


則翻訳である。単語を処理の単位とし、言語間の文法構造や単語の変換、及び目的言語における単語の並びなどを記述した規則を組み合わせることで適用し、訳文を作る。そのため、ドメインを限らず、多様な入力に対して一定の品質を得られる反面、やや表現の硬い訳文になりやすい。

そこで、東芝ソリューション(株)が製品化している翻訳ソフトウェアThe 翻訳™シリーズの開発で培ってきた翻訳知識(規則)に加えて、独自の話しことば解析技術を導入することで、話しことばでも高精度に翻訳できる技術を開発した。話しことばは、書きことばと比較して断片的であり非文法性が高いため、従来の解析技術では正確な翻訳が難しいという問題があった。この問題を解決するため、解析過程で全ての解釈の曖昧性を考慮し、更に、話しことば固有の構文構造を書きことばの構造に変換する技術を開発した。

もう一つは、あらかじめ蓄積した翻訳用例から、言語間の変化を確率的に捉え翻訳モデル化して翻訳する、統計翻訳である<sup>(2)</sup>。この方式は、用例に近い表現の文は良質に翻訳できる反面、用例から外れると品質が低下する傾向にある。

これら二つの翻訳方式は相補的な関係にある。そこで、規則翻訳の結果と統計翻訳の結果を統計翻訳で計算した翻訳確率に基づいて切り替える、ハイブリッド方式の翻訳エンジン



を構築した<sup>(3)</sup>(図3)。

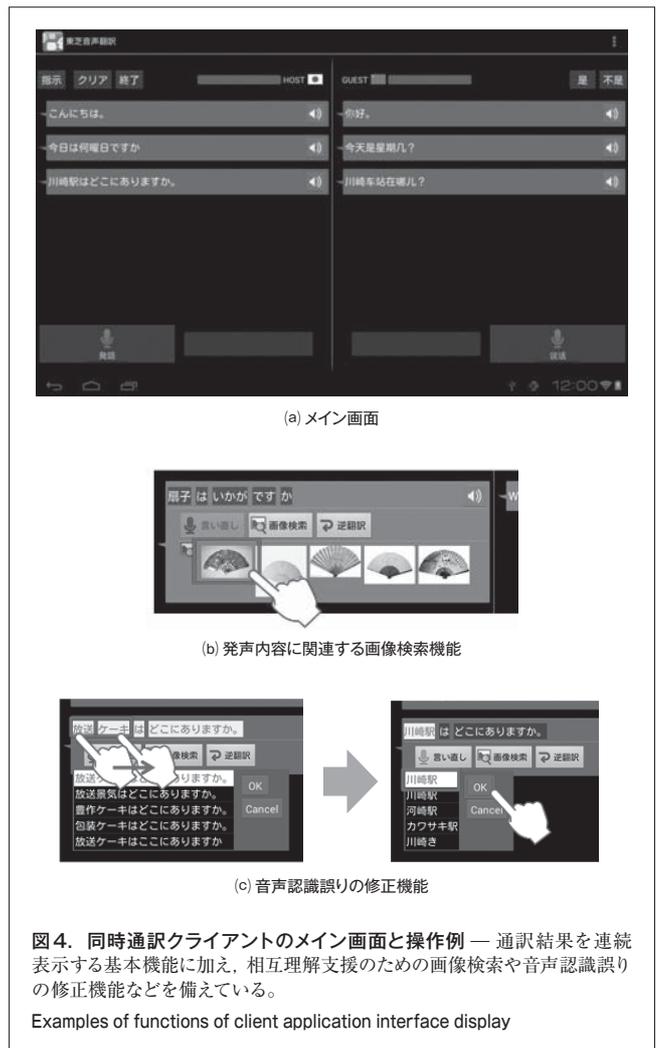
これにより、訳文の品質を高めながら、多様な入力に対する頑健性を得ることができる。翻訳方式の切替えは入力される翻訳単位ごとに行い、単位全体をいずれかの方式で翻訳する。

### 2.3 クライアントアプリケーション

ゲスト側(顧客)とホスト側(店員)が画面を共有して会話することを想定し、Android<sup>TM</sup>(注2)プラットフォームを搭載したタブレットで動作するクライアントアプリケーションを開発した(図4)。

クライアントは、ユーザーの連続発声を取り込んで、音声認識・機械翻訳サーバに送信し、通訳結果を受信する基本機能に加えて、特にホスト側の支援を目的として、頻出発声文の呼出し機能、入力文による画像検索機能、及び翻訳結果の再翻訳機能などを実装した。これらの機能は、初めてシステムを用いる外国人ゲストに教示するには難しいが、ホスト側の操作を見て使用することを考え、ゲスト側にも同様に実装した。ユーザーは、自身の発声を開始するときにボタンを押下するだけでよい。その後は、操作することなく連続的に発声することが可能であり、自身の発声を終了するとき、又は別のユーザーが発声を開始するときに再度ボタンを押下する。

(注2) Androidは、Google Inc.の商標又は登録商標。

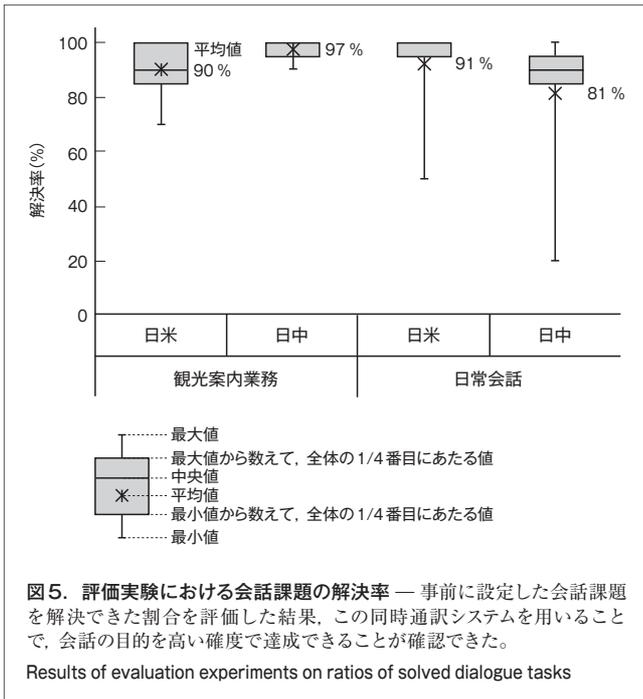


## 3 評価実験

開発した同時通訳システムを用いて、通訳会話の達成度を実験により評価した<sup>(4)</sup>。評価実験は、旅行案内と日常生活を対象とし、事前に設定した課題を、システムを介した会話によりどの程度解決できるかを測定した。

ゲスト役の米国人と中国人の被験者(各7名)に会話課題を記したシートを渡し、システムを介してホスト役の日本人被験者(7名)と1対1で会話してもらい、その課題の解決率を評価した。日本人被験者に対しては、ホスト役として機器の扱いに習熟した状況を作り出すため、評価の開始前に一通りの操作を説明し、機器に慣れる時間をとった。

評価実験結果である会話課題の解決率を、箱ひげ図として図5に示す。日米及び日中とも、旅行案内で90%、日常生活で80%を超える解決率であった。通訳の過程では音声認識誤りや機械翻訳誤りが含まれることもあるが、対面した状況であれば、このシステムを用いることで、会話の目的を高い確度で達成できることが確認できた。



前記の実験に加えて、通訳時間の削減率を評価する実験も行った。実験では、四つの文から成る連続発声について、その発声開始から、通訳結果である合成音声の出力が終了するまでの時間を測定し、評価した。文ごとに逐次翻訳する、当社の従来技術による音声翻訳アプリケーションのポケット通訳™が約36 sを要したのに対して、この同時通訳システムは約20 sと、約44 %の時間を削減できた。

## 4 あとがき

対面業務の外国語会話を支援する目的で開発した同時通訳システムについて述べた。

このシステムは、ユーザーの連続発声を受け付け、これを適切な単位にくぎって順次翻訳する。これにより、ユーザーは発声の長さや文を意識することなく話すことができ、より短時間での会話が可能になる。

今後、評価実験を通して得られた知見に基づいてシステムの改善を進め、通訳を介した意思疎通の時間を短縮することで、実際の場面において役だつシステムに進化させていく。

## 文献

- (1) 井阪岳彦 他. スマートフォン向け日中英音声翻訳システム. 東芝レビュー. 65, 8, 2010, p.48-51.
- (2) 王 海峰 他. 多くの言語対に対応する統計的機械翻訳システム. 東芝レビュー. 64, 10, 2009, p.66-67.
- (3) 知野哲朗 他. ハイブリッド機械翻訳技術による日中英音声翻訳システム. 東芝レビュー. 64, 2, 2009, p.26-29.
- (4) 釜谷聡史 他. 対面業務支援を想定した連続発話可能な同時通訳システムの開発と評価. 情報処理学会研究報告. 2013-SLP-95, 17, 2013, p.1-8.



**釜谷 聡史 KAMATANI Satoshi**

研究開発センター 知識メディアラボラトリー研究主務。  
自然言語処理の研究・開発に従事。情報処理学会、言語処理学会会員。

Knowledge Media Lab.



**坂本 明子 SAKAMOTO Akiko**

研究開発センター 知識メディアラボラトリー。  
自然言語、音声言語処理の研究・開発に従事。言語処理学会会員。

Knowledge Media Lab.



**住田 一男 SUMITA Kazuo, D.Eng.**

東芝リサーチ・コンサルティング(株) シニアフェロー、工博。  
音声翻訳の研究開発に従事。情報処理学会、電子情報通信学会、言語処理学会、人工知能学会会員。

Toshiba Research Consulting Corp.