

コンシューマ・システム論文

高速・高精度なコンパクト・スケーラブル 自動通訳ソフトウェアの開発と実用性評価

花沢 健^{1,a)} 奥村 明俊² 岡部 浩司¹ 安藤 真一¹

受付日 2011年12月16日, 採録日 2012年4月13日

概要: ある言語で話される音声を異なる言語へと変換する音声翻訳技術は、自動通訳技術とも呼ばれ、コミュニケーションを支援する技術として従来からニーズが高く、研究が行われている。今回、リアルタイムのコミュニケーションを阻害しないように、実用的な精度と速度を持つコンパクト・スケーラブル自動通訳ソフトウェアを構築し、携帯端末に実装した。統合的発話解釈評価手法を提案し、他の自動通訳ソフトウェアとの比較とともに実用性を評価、我々のソフトウェアが精度面・速度面で比較すると優位であり、かつ実用性が高いことを示した。

キーワード: 音声翻訳, 音声認識, 機械翻訳, 実用性評価

Development and Evaluation of the Fast and Accurate Compact-scalable Speech Translation Software

KEN HANAZAWA^{1,a)} AKITOSHI OKUMURA² KOJI OKABE¹ SHINICHI ANDO¹

Received: December 16, 2011, Accepted: April 13, 2012

Abstract: Speech translation technology is a key technology to assist cross-lingual communications. We have developed the Compact-scalable Speech Translation software which realizes both high speed and high accuracy on embedded devices such as mobile phones, to assist communication in the real world effectively. We also defined an evaluation method of the speech translation to assist cross-lingual communications in the real world. Evaluated by the method on a mobile phone, we could conclude that our speech translation software is effective enough and outperforms other present ones both on the speed and the accuracy.

Keywords: speech translation, speech recognition, machine translation, practical evaluation

1. はじめに

ある言語で話される音声を異なる言語へと変換する自動通訳技術（音声翻訳技術）は、コミュニケーションを支援する技術として従来から多くの研究が行われている [1], [2]. 自動通訳は、図 1 に例示するように観光・ビジネス・国際会議・海外生活など様々な活用場面が考えられ、ヨーロッパ・アジア言語を中心に、活動が活発になっている。評価型

の国際ワークショップも開催されている [3], [4]. 我々もこれまで、世界の様々な国の人々が、いつでも、どこでも、誰とでもコミュニケーションできる社会の実現を目指し、そ



図 1 自動通訳の活用場面

Fig. 1 Use scenes of speech translation.

¹ 日本電気株式会社 情報・メディアプロセッシング研究所
NEC Information and Media Processing Laboratories,
Kawasaki, Kanagawa 211-8666, Japan

² NEC 情報システムズ
NEC Informatic Systems, Ltd., Kawasaki, Kanagawa 213-
8511, Japan

a) k-hanazawa@cq.jp.nec.com



図 2 自動通訳ソフトウェアの利用イメージ

Fig. 2 Typical use image of speech translation.

のための重要な構成要素の1つとして自動通訳システムの研究開発を積極的に推進してきた [5], [6], [7], [8], [9], [10].

近年, 自動通訳技術のニーズの高まりと端末や通信網の発展を背景として, 特に携帯端末上で自動通訳を実現するサービス・ソフトウェアが多く登場している [11]. これらは, ユーザが相手と携帯端末を介してリアルタイムにコミュニケーションするという使い方を想定しており, 主に旅行会話を対象としている. 我々が開発を推進してきたシステムも同様である. 図 2 に, 自動通訳ソフトウェアの利用イメージを示す. しかし, 現行のサービス・ソフトウェアは, 異言語間のコミュニケーション・意思疎通支援という価値を十分に実現するほど実用的ではなく, 精度・速度が不十分である.

そこで我々は, 異言語間コミュニケーションでニーズが高い旅行会話を対象として, 実用的な精度を保ちつつ高速性を極めた自動通訳の実現を目指している. 今回, コンパクト・スケーラブル音声認識と語彙規則型機械翻訳から構成されるコンパクト・スケーラブル自動通訳により, 携帯端末上で高い精度と高速レスポンス性を実現し, 実用性が高いことを評価したので報告する. 実用的な精度は明確に定義されていないが, 相手に意味が伝わる発話の割合として, 最低限の精度は7割程度, 十分な精度は8割以上と考えられる [12]. 実用的な速度については, リアルタイムのコミュニケーションを阻害しない, リアルタイム比1倍程度(1倍が理論上最速値)の速度が望ましく, 実機上ではユーザが話し終わった直後に結果が出ていることが必要である. 実際に我々は, 空港やホテルの実現場で行った実証実験(国交省・総務省) [13], [14]を通じ, 7~8割程度の精度とリアルタイム比1倍近い速度があれば異言語間コミュニケーションを支援できることを確認している. また, 精度と速度を実験室で計算機シミュレーションによって評価しても, 実用的とはいえない. 実機上で, 実際に動作させた場合を可能な限り模擬した評価が必要である. 今回, 実用性を評価する評価手法についても提案する.

本稿では, まず2章において, 現行の自動通訳サービス・ソフトウェアの代表的な2つの実現形態であるスタンドアロンタイプとサーバ連携タイプの長所・短所を, 3章で自動通訳を構成する音声認識技術と機械翻訳技術の課題につい

て述べる. 次に4章で, コンパクト・スケーラブル自動通訳ソフトウェアの構築・試作について説明する. 次に5章において, 実用性を評価する統合的発話解釈評価手法を提案し, 6章では統合的発話解釈評価手法により, 試作したコンパクト・スケーラブル自動通訳ソフトウェアを, 入手可能な他の3つの自動通訳ソフトウェアとともに比較評価する. 7章では, 統合的発話解釈評価手法によって高い実用性を示すことができたことを考察する.

2. 自動通訳ソフトウェアのタイプ

現行の自動通訳ソフトウェアの実現方法は, 携帯端末単体で動作を完結させるスタンドアロンタイプと, 高速な通信網と大規模なサーバを活用して動作を行うサーバ連携タイプの2つに大別できる. それぞれの長所・短所は以下のとおりである.

1. スタンドアロンタイプ

携帯端末単体で動作が完結するために, 通信網を必要としない. このため, 通信を介することによるレスポンスの低下や利用可能エリアの制限がないという長所がある. 一方, CPUやメモリサイズといった計算リソースは, 大規模サーバに比べると携帯端末単体では限界があるという短所がある. このため, 特に音声認識や機械翻訳など大規模なリソースを必要とする処理は, 効率的に行うよう工夫しないと, 精度や速度の劣化を招く恐れがある.

2. サーバ連携タイプ

通信網の先にある大規模サーバを, 場合によっては複数同時に利用することができる. このため, 高精度化が容易という長所がある. また, サーバ側はユーザの手を煩わすことなくアップデートやメンテナンスが可能という長所もある. 一方, つねに通信網を介するため, 通信網の速度や通信環境の変化によりレスポンスが劣化するという短所がある. 近年, 無線通信網は高速化が進んでいるが, 通信を行うプロセス自体は原理的に遅延の原因となる. 通信環境という面では, 過疎地/途上国や高速移動中などそもそも安定した高速通信が困難な環境はまだ存在し, 通信を必要とすることが利用可能エリアの制限につながるという短所になる. また, 高額な通信利用料がユーザにとって負担になる場合もある.

今回我々は, 応答速度が速いというスタンドアロンタイプの長所を活かし, サーバ連携タイプに匹敵する精度を持つことを目指す.

3. 自動通訳を構成する技術

現行の自動通訳は, 音声認識技術と機械翻訳技術を主たる構成要素とする.

音声認識技術で現在主流であるのは, 統計モデルを利用した大語彙連続音声認識である. 音響的確からしさを与え

る音響モデルと、言語的確からしさを与える言語モデルとを、大規模な音声・テキストデータベースからそれぞれ学習し、これらを用いて入力音声に対する最適解を探索する。数万語規模の辞書を備えてその自由な組合せによる入力を許すことにより、多様な表現・言い回しを受理可能という長所を持つ。近年、大規模データベースの整備や学習手法の確立が進んだことで、音声認識技術の実用性は高まっている。しかし、大語彙かつ多様な表現を受理するために、また高い精度を達成するために、探索の規模が大きくなり、計算機リソースあるいは処理時間が必要であるという課題がある。

機械翻訳技術としては、ルールベース翻訳、用例ベース翻訳および統計翻訳の3つの手法が現在主流である。ルールベース翻訳は、専門家によって記述された文法と対訳辞書とを備え、入力文を解析して得られる構文および語句を変換する。用例ベース翻訳は、入力文に類似した文の対訳用例を活用する。統計翻訳は、語句の変換を行う翻訳モデルと、言語的確からしさを与える言語モデルとを、大規模なテキストデータベースからそれぞれ学習し、これらを用いて入力文に対する最適解を探索する。ルールベース翻訳は、文法記述を詳細化することで高精度化が可能という長所があるが、専門家による作業が必要となるために構築コストが大きいという短所を持つ。統計翻訳は、学習データさえ準備すれば構築が容易という長所がある。一方、精度が学習データ依存であり、特に翻訳モデルを学習する対訳データベースの不足から、高精度化が困難という短所を持つ。用例ベース翻訳は、両者の中間的な特徴を持つ。さらに、いずれの手法であっても入力文の解析および変換において多くの可能性を展開するため、メモリ消費量が大きいという課題がある。

今回我々は、統計モデルを利用した大語彙連続音声認識と、高精度化が可能なルールベース翻訳とを活用し、携帯端末上で高い精度と高速レスポンス性を実現するコンパクト・スケーラブル自動通訳ソフトウェアを独自開発する。

4. コンパクト・スケーラブル自動通訳ソフトウェア

リソースの限定された環境において、実用的な精度を保ちつつ高速性を極めるためには、処理時間を主たる課題とする音声認識処理において高速処理を実現し、メモリ消費量を主たる課題とする機械翻訳処理において省メモリ動作を実現する必要がある。このため、高速処理を実現するコンパクト・スケーラブル音声認識と、省メモリ動作を実現する語彙規則型機械翻訳とをそれぞれ独自開発し、コンパクト・スケーラブル自動通訳として統合する。図3に示すように、コンパクト・スケーラブル自動通訳ソフトウェアは、コンパクト・スケーラブル音声認識部と語彙規則型機械翻訳部、およびそれらを統合する通訳統合部からなる。

音声認識と機械翻訳を、日英それぞれ用意することで、日英・英日双方向の自動通訳を実現する。なお、通訳結果の提示方法としては音声とテキストの2種類がありうるが、図2に示すような利用シーンにおいてはテキストによる画面表示でも実用上十分であり、今回は音声合成を除いて音声認識と機械翻訳を主たる課題とする。したがって想定利用シーンは、音声を入力し、得られた通訳結果をテキストで相手に画面提示するというものになる。評価についても、音声入力から画面表示までの速度および表示されたテキストの精度を対象とする。

4.1 コンパクト・スケーラブル音声認識部

コンパクト・スケーラブル音声認識部の構成を図4に示す。コンパクト・スケーラブル音声認識部[15]は、図4に示すように、入力した音声进行分析する音声分析部、分析結果の特徴量と音響モデルとの距離を計算する距離計算部、距離計算結果から単語辞書と言語モデルを用いて探索を行う最適単語列探索部とからなる。各計算処理は固定小数点化を行うことで高速処理を実現している。なお、音声認識エンジン自体は言語非依存であり、モデルや辞書などのリソースを切り替えることで言語の切替えを可能としている。

4.1.1 コンパクト・スケーラブル音声認識エンジン

コンパクト・スケーラブル音声認識エンジンを構成する各モジュールの特徴を説明する。音声分析部は、入力音声

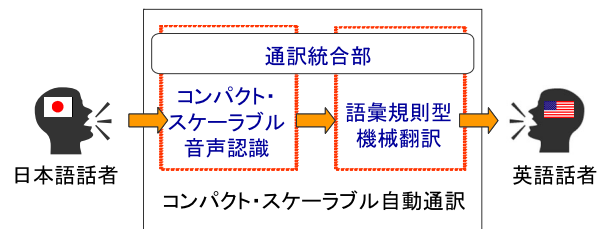


図3 コンパクト・スケーラブル自動通訳ソフトウェアの構成
Fig. 3 Overview of the compact-scalable speech translation software.

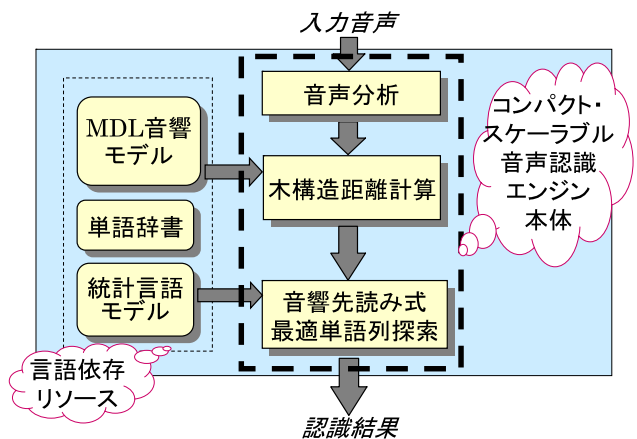


図4 コンパクト・スケーラブル音声認識部の構成
Fig. 4 Compact-scalable speech recognition.

波形を音声認識に適した特徴量系列であるケプストラムに分析する。

距離計算部では、コンパクト化のために、MDL 基準を用いた音響モデルの混合ガウス分布数の削減、ガウス分布の対角共分散行列の共有化を行う。MDL 基準を用いることで端末の使用リソースに合わせた、すなわちスケラブルな混合ガウス分布数の最適化が可能となる。高速化のために、木構造を利用した音響モデルの効率的な出力確率計算を行う。これらの工夫により、モデルサイズは 1/3、計算量は 1/10 以下と大幅なコンパクト化・高速化を達成しており [16]、コンパクト・スケラブル音声認識の高速化に重要な貢献をしている。

最適単語列探索部では、高速化のために音響先読みによる最適単語列探索とスコア計算結果の再利用による計算処理の削減を行う。コンパクト化のために使用メモリのガベージコレクションによる再利用を行う。さらに、精度劣化を防ぐために言語モデル先読み値の平滑化 [17] を行っている。

4.1.2 言語依存リソース

統計モデルである音響モデルと統計言語モデルは、大量の音声およびテキストコーパスから学習する。日本語は標準語、英語は北米英語を対象とした。音響モデルは、日英とも 600 時間以上の音声コーパスを用いて不特定話者かつ性別非依存の状態共有 triphone HMM を学習した。統計言語モデルは、日英とも数十万文規模の旅行会話テキストコーパスを構築し、単語 trigram モデルを学習した。単語辞書は、テキストコーパスに出現するものをベースに、頻度情報を利用することで、少ない語彙で広いコーパスカバー率が得られるよう工夫した。語彙サイズは日英とも 3 万語を超える。

4.2 語彙規則型機械翻訳部

語彙規則型機械翻訳部の構成を図 5 に示す。語彙規則型機械翻訳は、図 5 に示すように、原言語解析・言語変換・

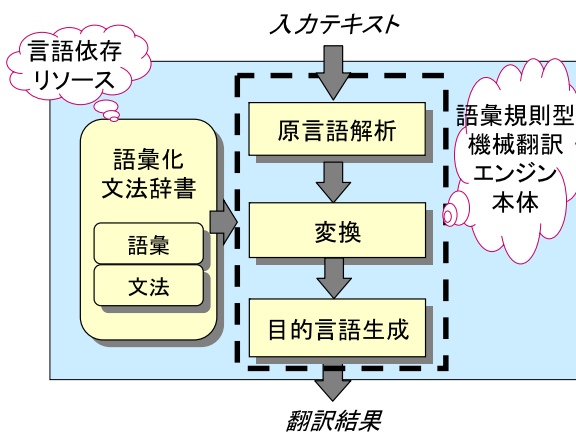


図 5 語彙規則型機械翻訳部の構成

Fig. 5 Lexicalized tree automata-based machine translation.

目的言語生成からなるルールベースの翻訳であり、語彙と文法を一体管理する語彙化文法辞書を有するのが特徴である。エンジン自体は言語非依存であり、語彙化文法辞書を切り替えることで言語の切替えが可能である。

4.2.1 語彙規則型機械翻訳エンジン

機械翻訳エンジンとしては、語彙規則型のコンパクト機械翻訳エンジン [18] を使用する。本エンジンは、文法知識が単語辞書中に局所化されている。メモリ上にすべての文法を保持する必要がなく、翻訳対象文に必要な文法のみ展開するだけでよいため、コンパクト化が容易となる。

4.2.2 語彙化文法辞書

語彙化文法辞書は、独自の語彙化ツリーオートマトン文法を持つ。語彙化ツリーオートマトン文法は、ツリー形式で記述された個々の文法を、適用順序を表現するオートマトン上に記述した文法形式である。各単語は、自分自身をヘッドとする文法の列を単語内に保持している。すべての単語がツリーオートマトンを持つが、オートマトンのレベルおよび構成要素であるツリー文法のレベルで共通部分を共有することにより、文法記述コストを削減し、実装のコンパクト化を行っている。

文法および語彙としては、旅行会話自動通訳向けに開発した翻訳言語知識ベースを搭載している。翻訳言語知識ベースは音声認識辞書と共通であり、日英方向・英日方向とも 3 万語規模の語彙を持つ。日英翻訳では、汎用の日英翻訳文法をベースとして、省略主語の推定、固定表現や熟語など単語の組合せに応じた訳し分け、口語的な文末表現への対応などの話し言葉対応を強化している。英日翻訳では、汎用の英日翻訳文法をベースとして、旅行場面で多く見られる依頼や質問などの表現の強化、口語的表現や定型表現への対応、丁寧表現の生成などを強化している。このように、語彙化文法を採用することにより日英翻訳・英日翻訳で独自の強化が可能であり、旅行会話の語彙に対して個別にきめ細かい文法を与えることができるため、コンパクト化しても高い精度を実現できる。

4.3 通訳統合部

通訳統合部は、コンパクト・スケラブル音声認識部と語彙規則型機械翻訳部の統合・制御をつかさどる。音声認識部から機械翻訳部への情報受け渡しにおいては、認識結果の単語表記に加えて、単語区切り、単語の読み情報、継続時間長の情報を渡す。これらの情報は、読みのあいまい性や統語的あいまい性の解消に利用しており、精度を高めている。

5. 統合的発話解釈評価手法

自動通訳の評価手法としては、これまでも様々な方法が検討・実行されている。

評価型の国際的なワークショップである GALE [5] や

IWSLT [6] においては、手法の有効性を比較評価することを目的として、定量的かつ客観的な精度評価を主に行っている。ここでは、統計翻訳の客観的な精度評価で用いられる BLEU など統計的尺度が用いられる。BLEU は、翻訳結果と参照訳（正解）との単語連鎖レベルの一致度を表したものである。参照訳との一致度が高ければ高いスコアが得られる。しかし統計的尺度では、必ずしも人間に意味が通じることの評価にはなっていない [19]。実利用場面において意味が通じかどうかを評価するには、人間の評価者が主観評価するのが直接的であり、参照訳との一致度が高ければ良いとは限らない。たとえば、英日翻訳において主格と目的格の助詞が入れ替わった場合、「太郎が花子に飲み物を頼んだ」と「太郎に花子が飲み物を頼んだ」では、一致度は高いが意味は大きく異なってしまう。

また、従来の評価では音声認識あるいは機械翻訳の単体評価を主に行ってきた。しかし、ささいな音声認識誤りであれば機械翻訳の精度に影響を与えない場合もあれば、個別には正しくても全体を通して誤りであることもある。自動通訳としての精度は個々の単体評価の組合せで導き出せるとは限らない。たとえば、英日通訳において、入力文「You can't bring a knife in.」に対して英語音声認識結果「You can bring a knife in.」と 1 単語だけ誤認識した場合、英日翻訳結果「あなたはナイフを持ち込める」は翻訳単体評価では正しい。このように、たった 1 単語の誤認識が致命的な誤りとなりうる。逆に、「Do you like an apple?」を「Do you like apples?」に誤認識したとしても、翻訳結果は意味が通じ、精度に影響を与えない場合もある。音声入力を機械翻訳した結果を直接評価する一貫評価を行う方が実利用場面での精度が反映される。

以上のように、これまでの評価手法には以下の問題があった。

1. 統計的尺度では、必ずしも人間に意味が通じることの評価にはなっていない。
2. 音声認識・機械翻訳エンジン単体の評価では、全体性能は見えにくい。

そこで今回、次の 2 つの方針により、上記 2 つの評価手法の問題を改善する。

1. 発話の意味が理解/解釈できるかどうかを主観評価する。
2. 音声入力から翻訳まで通しての一貫評価とする。

なお、自動通訳のユーザが操作に慣れていない場合を考

慮し、入力のやり直しを許さない 1 回の操作での評価を行う。本方針に基づいて実機上で精度と速度を評価する統合的発話解釈評価手法を提案する。本評価によって意味理解率と速度・リアルタイム比を評価する。

6. 統合的発話解釈評価手法による評価

携帯端末上の実装した自動通訳ソフトウェアの評価を行う。現在入手可能な 3 つの自動通訳ソフトウェアと比較を行い、実用上必要な精度と高速な処理の実現を確認する。携帯端末としては市販のスマートフォンを用いた。動作周波数は 1.0 GHz である。評価の公平性を保つため、各ソフトウェアへの音声入力はすべて同じ音声を用いることとし、スピーカ再生による模擬音声入力とする。なお、サーバ連携版では外部との通信が必須である。今回は国内の携帯電話回線（3G 回線）を用いた。

6.1 評価尺度

評価は、処理速度と通訳結果（音声認識および翻訳の結果）の精度とを対象とする。処理速度の評価には、音声入力の開始から翻訳結果の出力までの時間を計測し、それを入力した音声の長さで除算した値であるリアルタイム比を用いる。一般的に入力した音声の終端にまで到達しないと、すなわち音声入力終了しないと、最終的な音声認識および翻訳結果が出せないため、リアルタイム比 1.0 が最も高速である。精度の評価には、主観評価による意味理解率を用いる。主観評価は目的言語（英日方向であれば日本語）のネイティブ評価者が行う。意味理解率は、表 1 における 4 段階評価の c 評価以上のものを相手に意味が伝わったとし、その割合を求める。

6.2 評価方法

今回は、英日通訳を対象とする。

1. 音声データ

音声データは、いわゆる旅行会話集や旅行における対話文 [20] と同等の、旅行会話英語音声データを用いる。男女各 4 名ずつ少量の音声データをランダムに抽出した。合計 60 発声、音声の平均長は約 3 秒である。典型文例からなる旅行会話集と同等のデータを用いることで、旅行会話で最低限必要となるコミュニケーションの支援が可能となることを評価する。これらの評価データは、我々

表 1 翻訳精度の評価値と評価基準

Table 1 Evaluation value and criteria of translation.

評価値	評価基準
a	原文と照らし合わせて自然で適切
b	言い回しなどに不自然さはあるが統語的に正しく、原文の意味も正しく理解できる
c	文法的に誤りがあるなど b 評価にはならないが、原文の意味は伝わる
d	原文の意味が理解できない/誤解を生じる

表 2 主観評価結果

Table 2 Evaluation results.

評価値	Proposed	SystemA	SystemB	SystemC
a	35.0%	6.7%	10.0%	13.3%
b	23.3%	8.3%	8.3%	6.7%
c	10.0%	21.7%	25.0%	28.3%
d	31.7%	63.3%	56.7%	51.7%

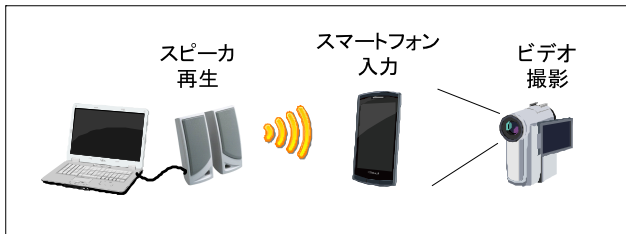


図 6 実験の構成

Fig. 6 Evaluation outline.

の自動通訳ソフトウェアでは学習に含んでおらず、オープンな評価である。一方、比較対象の他のソフトウェアにおいては確認が不可能であり、学習に含まれている可能性がある。

2. 比較対象ソフトウェア

比較対象の自動通訳ソフトウェアとしては、サーバ連携タイプを 2 種、スタンドアロンタイプを 1 種用いた。我々のコンパクト・スケーラブル自動通訳ソフトウェアとあわせて合計 4 種で比較する。スタンドアロンタイプ 1 種 (SystemA) と、サーバ連携タイプ 1 種 (SystemC) は旅行会話を対象としているが、もう 1 種のサーバ連携タイプ (SystemB) は旅行会話に限らず広く一般を対象としている。正確な確認は不可能であるが、他のソフトウェアにおいて、音声認識技術として大語彙連続音声認識を、機械翻訳技術としては SystemA はルールベース翻訳、サーバ連携タイプでは統計翻訳を採用していると推測される。すべてのソフトウェアは、同一の携帯端末上で順次動作させる。

3. 手順

図 6 に実験の模式図を示す。評価は、静かな居室環境でスピーカ再生による模擬音声入力により行った。入力音声データをスピーカ再生し、携帯端末 (スマートフォン) 上のソフトウェアに入力する。またその様子をビデオ撮影し、後からの精度・速度評価を可能にしている。ソフトウェアの操作は 1 発声ごとに人手で行い、その後撮影したビデオを確認しながら日本人評価者 1 名による英日通訳結果の主観評価と速度評価を行った。

図 7 に、我々のコンパクト・スケーラブル自動通訳ソフトウェアを評価した際の動作例を示す。この図では、英語音声の認識結果と英日翻訳の結果とが上下の順に表示されている動作例を示している。

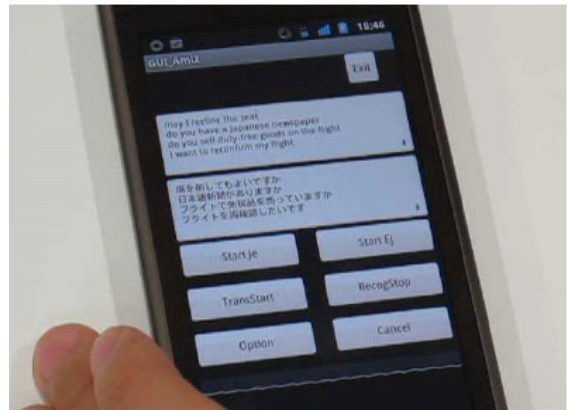


図 7 コンパクト・スケーラブル自動通訳ソフトウェア動作例

Fig. 7 Output examples of the compact-scalable speech translation software.

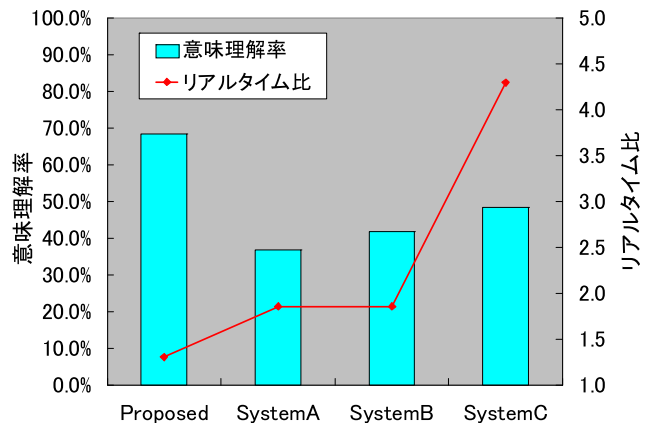


図 8 精度と速度の比較

Fig. 8 Comparison of accuracy and speed.

6.3 精度と速度の比較

4 種の自動通訳ソフトウェアによる主観評価結果を表 2 に示す。“Proposed” が我々のソフトウェアである。表 2 における主観評価結果の a, b, c の合計を意味理解率とする。

各ソフトウェアの意味理解率とリアルタイム比を図 8 に示す。図 8 は、意味理解率を左軸 (棒グラフ)、リアルタイム比を右軸 (折線グラフ) とした比較結果である。

1. 精度

精度は Proposed が優位である。大規模リソースを利用可能なサーバ連携タイプである SystemB や SystemC と比較しても精度が高い。サーバ連携タイプのこれら 2

つのソフトウェアは統計翻訳を採用していると推測され、Proposed ではルールベース翻訳による文法記述詳細化の効果が出ている。SystemA は Proposed と同じルールベース翻訳と推測されるが、語彙規則型機械翻訳により旅行会話の語彙に対して個別にきめ細かい文法記述を行った効果が出ている。また、Proposed がスタンドアロンタイプであるにもかかわらず精度が高い別の理由として、コンパクト・スケーラブル音声認識による精度劣化を抑えたコンパクトかつ高速な音声認識の実現があげられる。実用に十分な8割には届いていないものの、最低限の7割の精度が得られている。

SystemB と SystemC の精度差の原因としては、SystemC が対象を旅行会話に限定しており、今回の評価データに適していることがあげられる。SystemA の低精度の理由としては、コンパクト化・高速化において精度を犠牲にしている可能性があげられる。

2. 速度

速度は Proposed が優位である。理由として、スタンドアロンタイプであるためサーバ連携せずに通信遅延が発生しないこと、コンパクト・スケーラブル音声認識により高速な処理が可能となっていることがあげられる。

SystemB と SystemC はともにサーバ連携版であるにもかかわらず速度に大きな違いが出ている。この理由としては、サーバ側の処理能力あるいは通信プロトコルの違いなどが考えられる。またスタンドアロンタイプである SystemB とサーバ連携タイプである SystemC がほとんど同じ速度との結果が出ている。このことから、速度の劣化要因は単純にタイプの違いによる通信遅延だけではなく、アルゴリズム・実装上の工夫により高速化を図ることの重要性が分かる。

6.4 評価手法の比較

評価手法の比較のため、主観評価である意味理解率と、統計的尺度として広く用いられている BLEU との比較を行った結果を図 9 に示す。図 9 は、意味理解率を左軸（棒グラフ）、同じ英日通訳結果に対する BLEU の値を求めて右軸（折線グラフ）に示したものである。意味理解率の優劣と、BLEU の優劣とは、必ずしも一致しない。たとえば、参照訳「免税品の機内販売をしませんか」に対して英日通訳結果「フライトで免税品を売っていますか」が得られた事例がある。単語連鎖の一致度の低さから BLEU の値は低くなるが、意味は通じる。同様の現象が Proposed と SystemB において多く見られており、これが両者の意味理解率と BLEU との差分である。SystemC では逆に、参照訳「チェックインは何時までにしなければなりませんか」に対して英日通訳結果「何時までに宝石しなければなりませんか」のように、単語連鎖の一致度は高いが、致命的な誤認識のため意味は通じないという事例が見られた。こ

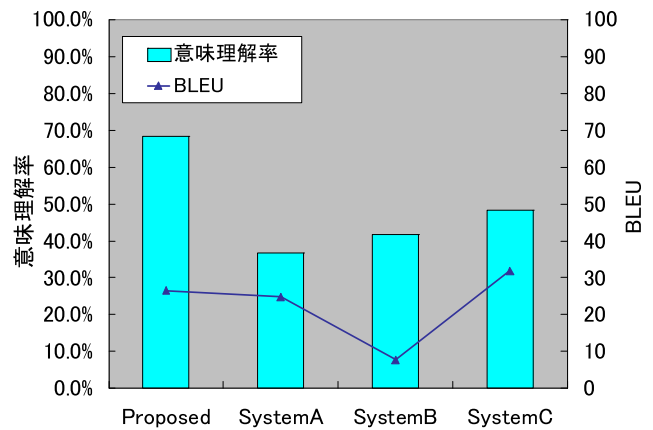


図 9 統計的尺度との比較

Fig. 9 Comparison with statistical criteria.

のように、相手に意味が通じるかどうかという基準では、BLEU に代表される統計的尺度よりも主観評価による意味理解率が適している。

また、単体評価と一貫評価の比較という観点から、音声認識の単体評価を行った。Proposed と SystemB では約 80%、SystemA と SystemC では 60% 台の精度を得た。音声認識での優劣と通訳の優劣は必ずしも一致していない。具体的には、入力文「Do you have a Japanese newspaper?」に対して英語音声認識結果「you have a Japanese newspaper」と 1 単語だけ誤認識し、英日翻訳結果「あなたは日本語の新聞を持っている」が得られた事例がある。音声認識単体ではわずかな誤りであり、翻訳単体では正しい結果である。しかし、通訳結果において入力文の意味は通じていない。このように、単体評価の組合せでは全体の精度を表せない事例が多く存在し、実利用場面での精度を表すには一貫評価が適している。

7. 考察

今回我々が構築・実装したコンパクト・スケーラブル自動通訳ソフトウェアは、統合的発話解釈評価手法の方針に基づいた実機上の精度・速度評価において、意味理解率で 7 割の精度と、リアルタイム比で 1 倍程度の速度を達成した。これにより、我々のソフトウェアの実用性が高いことを示した。他の現行のソフトウェアとの比較においては、精度・速度ともに優れているという結果が得られた。精度に関しては、他のソフトウェアが 4~5 割程度の意味理解率であるのに対し、我々のソフトウェアは約 7 割の意味理解率が得られており、10 回話のうち 7 回程度は相手に正しく伝わる。速度に関しては、他のソフトウェアが速いものでもリアルタイム比 2.0 程度であり、入力後に入力した音声とほぼ同じ長さの待ち時間が発生するのに対し、我々のソフトウェアは約 1.3 が得られており、ユーザが話し終わったほとんどその直後に結果が表示されるという速度を実現している。

評価手法の比較においては、統計的尺度である BLEU による客観評価結果と意味理解率による主観評価結果の優劣は必ずしも一致せず、事例による検証から、相手に意味が通じるかどうかという基準では主観評価が適していることを確認した。同様に、単体評価と一貫評価との比較においても、評価結果の優劣の不一致と事例による検証から、一貫評価が適していることを確認した。

今後の発展の方向性として、利用場面の拡大があげられる。今回の評価は、携帯端末を用いた利用場面での自動通訳の実用性を確認したものである。今後、作業中/業務中にハンズフリーで使用するなど、利用場面を拡大するためには、ウェアラブルコンピュータなど他のデバイスでの実現が望ましい。デバイスが異なると自動通訳に最適なユーザインタフェースが変わる可能性があるため、我々はウェアラブルコンピュータへの実装も試みている [21]。

8. まとめ

今回、異言語間のコミュニケーションを支援することを目的として、コンパクト・スケーラブル自動通訳ソフトウェアを構築、評価した。我々が実現を目指す自動通訳ソフトウェアは、リアルタイムのコミュニケーションを阻害しないように、実用的な精度と高速性を両立させたものである。

コンパクト・スケーラブル音声認識と語彙規則型機械翻訳によりコンパクト・スケーラブル自動通訳ソフトウェアを構築し、携帯端末に実装した。統合的発話解釈評価手法を提案し、旅行会話を対象として現行の自動通訳ソフトウェアとの比較評価を行ったところ、我々のソフトウェアが実用的であり、他のソフトウェアと比較して精度面・速度面で優位であることを確認した。また、実用性の評価手法として、統合的発話解釈評価手法が適していることを示した。今後、利用場面の拡大を狙い、異なるデバイスでの評価・ユーザインタフェースの工夫といった課題に取り組んでいきたい。

参考文献

- [1] 中村 哲, 隅田英一郎, 清水 徹: ここまできた音声翻訳技術, 情報処理, Vol.49, No.6, pp.606-610 (2008).
- [2] 奥村明俊: 携帯端末用多言語自動通訳システムの実用化に向けて, 情報処理, Vol.49, No.6, pp.611-616 (2008).
- [3] NIST: NIST Machine Translation Evaluation for GALE (online), available from <http://www.itl.nist.gov/iad/mig/tests/gale/> (accessed 2012-04-26).
- [4] Paul, M., Federico, M. and Stüker, S.: Overview of the IWSLT 2010 Evaluation Campaign, *International Workshop on Spoken Language Translation*, pp.3-27 (2010).
- [5] 奥村明俊, 服部浩明, 磯谷亮輔ほか: 携帯端末用多言語自動通訳システムの実用化技術に関する研究開発, フジサンケイビジネスアイ (オンライン), 入手先 <http://fbi-award.jp/sentan/jusyuu/2007/nec.pdf> (参照 2012-04-26).
- [6] 奥村明俊, 磯谷亮輔, 山端 潔ほか: 携帯端末など組み込み機器向け多言語自動音声翻訳システムの実用化技術の

- 開発, 情報処理, Vol.50, No.7, p.687 (2009).
- [7] Watanabe, T., Okumura, A., Sakai, S., et al.: An automatic interpretation system for travel conversation, *Proc. ICSLP-2000*, Vol.4, pp.444-447 (2000)
- [8] 山端 潔, 磯谷亮輔, 安藤真一ほか: PDA で動作する旅行会話向け日英双方向音声翻訳システム, 電子情報通信学会技術研究報告, 2002-NL-150-9 (2002).
- [9] 花沢 健, 荒川隆行, 岡部浩司ほか: 携帯電話試作機上で動作する旅行会話向け音声認識, 情報処理学会講演論文集, 2D-3 (2009).
- [10] 長田誠也, 花沢 健, 磯谷亮輔ほか: 携帯電話試作機上で動作する旅行会話向け自動通訳システムの開発, 情報処理学会講演論文集, 2D-2 (2009).
- [11] Google: Google translate (online), available from <http://translate.google.co.jp/> (accessed 2012-04-26).
- [12] 下郡信宏: 英語字幕による会議支援, 情報処理, Vol.51, No.1, pp.26-29 (2010).
- [13] NEC: PDA 用に開発した「日英/英日」音声通訳支援ソフトウェアの評価・実証実験を新東京国際空港で実施, NEC プレスリリース (オンライン), 入手先 <http://www.nec.co.jp/press/ja/0205/3002.html> (参照 2012-04-26).
- [14] NEC: 山梨県の観光地で多言語音声翻訳技術の実証実験を開始, NEC プレスリリース (オンライン), 入手先 <http://www.nec.co.jp/press/ja/1001/2002.html> (参照 2012-04-26).
- [15] 磯谷亮輔, 畑崎香一郎, 服部浩明ほか: 話し言葉認識に向けた基本技術と応用, 情報処理学会研究報告, 2005-NL-169 (2005).
- [16] 磯 健一, 磯谷亮輔, 畑崎香一郎ほか: 大語彙連続音声認識技術と応用, 日本音響学会春季講演論文集, 1-18-1 (2004).
- [17] 岡部浩司, 花沢 健, 磯谷亮輔ほか: 言語モデル先読み値の平滑化による探索誤りの改善, 日本音響学会講演論文集, 1-1-15 (2008).
- [18] 山端 潔, 安藤真一, 三村清美: 語彙化されたツリーオートマトンに基づく会話文翻訳システム, 言語処理学会第 6 回年次大会講演論文集, pp.264-267 (2000).
- [19] 辻井潤一, 越前谷博, 江原輝将ほか: 機械翻訳及び辞書構築に関する研究, 平成 20 年度 AAMT/Japio 特許翻訳研究会報告書, pp.2-13 (2009).
- [20] Takezawa, T.: Multilingual Spoken Language Corpus Development for Communication Research, *Computational Linguistics and Chinese Language Processing*, Vol.23, No.3, pp.303-324 (2007).
- [21] 花沢 健, 長田誠也, 後藤由希子: 携帯端末を活用した自動通訳実証実験, NEC 技報, Vol.63, No.1, pp.68-70 (2010).



花沢 健 (正会員)

1997 年東京工業大学大学院計算工学専攻修士課程修了。同年日本電気(株)入社。音声認識, 音声翻訳の研究開発に従事。現在, NEC 情報・メディアプロセッシング研究所主任研究員。日本音響学会, 人工知能学会各会員。



奥村 明俊 (正会員)

1986年京都大学大学院工学研究科修士課程修了。同年日本電気(株)入社。自然言語処理, 音声翻訳, メディア情報処理の研究開発に従事。1992~1994年南カリフォルニア大学客員研究員。工学博士。現在, (株) NEC 情報システムズ執行役員。言語処理学会, 人工知能学会各会員。



岡部 浩司

2007年東京大学大学院情報理工学系研究科修士課程修了。同年日本電気(株)入社。音声認識の研究開発に従事。現在, NEC 情報・メディアプロセッシング研究所勤務。日本音響学会会員。



安藤 真一 (正会員)

1992年大阪大学大学院基礎工学研究科物理系専攻修士課程修了。同年日本電気(株)入社。自然言語処理, 音声認識・合成の研究に従事。1995~1997年ATR 音声翻訳通信研究所。現在, NEC 情報・メディアプロセッシング研究所研究部長。言語処理学会, 人工知能学会各会員。