

携帯端末向け自動通訳の実用化に向けた 開発と評価

花沢健[†] 奥村明俊^{††}
岡部浩司[†] 安藤真一[†]

ある言語で話される音声を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、実用的な精度を保ちつつ高速性を極めたコンパクト・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定義し、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・速度面で比較優位であり、かつ実用性が高いことを示した。

Development and Evaluation of a Speech Translation System on Embedded Devices

Ken Hanazawa[†] Akitoshi Okumura^{††}
Kouji Okabe[†] Shinichi Ando[†]

Speech translation technology is a key technology to assist cross-lingual communications. We have developed the Compact-scheable 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.

1. はじめに

ある言語で話される音声を変換する音声翻訳技術（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、そのための重要な構成要素の一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[5-10]。



図1. 自動通訳の活用場面

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。

[†] 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

^{††} NEC 情報システムズ
NEC Informatic Systems, Ltd



図 2. 自動通訳ソフトウェアの利用イメージ

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、実用的とは言い難い。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、まず第2節において、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・短所を議論する。次に第3節で、我々が求めるリアルタイムの異言語間コミュニケーション支援という価値を実現するためのコンパクト・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。次に第4節において、実用性を評価する統合的発話解釈評価手法を提案し、第5節では通訳実用性評価手法により、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。第6節では、通訳実用性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、スタンドアロンタイプとサーバ連携タイプの2つに大別できる。すなわち、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である。それぞれの長所・短所について議論する。

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、常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、スタンドアロンタイプの長所である高速性を活かし、サーバ連携タイプに匹敵する精度を持つことを目指した、コンパクト・スケーラブル自動通訳ソフトウェアを実現した。

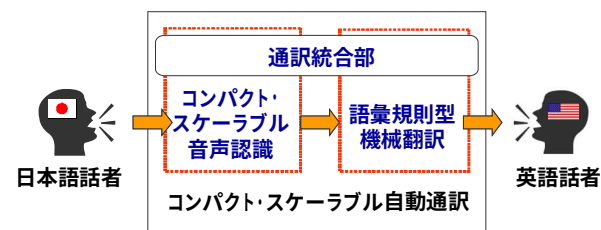


図 3. コンパクト・スケーラブル自動通訳ソフトウェアの構成

3. 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構築し、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成る。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種類があり得るが、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。

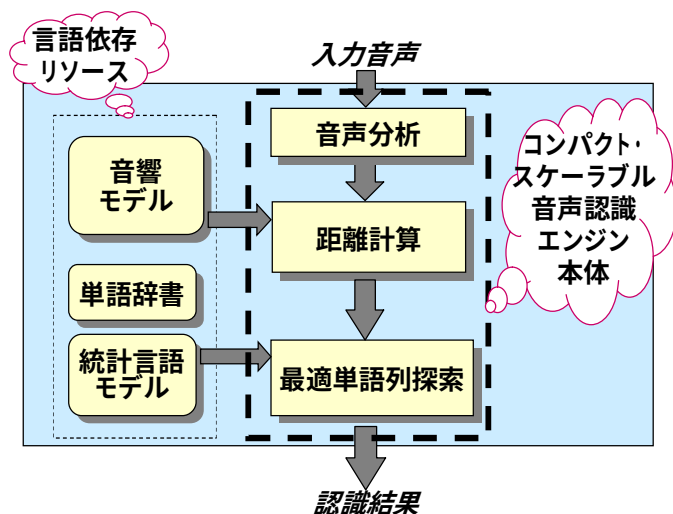


図4. コンパクト・スケーラブル音声認識部の構成

3.1 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。音声認識としては、現在主流である統計的モデルを用いた大語彙連続音声認識[15]を行う。コンパクト・スケーラブル音声認識部は、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的確からしさを与える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的確からしさを与える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成る。音響モデルとしては triphone HMM、言語モデルとしては統計言語モデルである単語 trigram を用いる。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り替えを可能としている。

3.1.1 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、すなわちスケー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・スケー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、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、高速化のためにスコア計算結果の再利用による計算処理の削減を行う。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。

3.1.2 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、大語彙連続音声認識である。

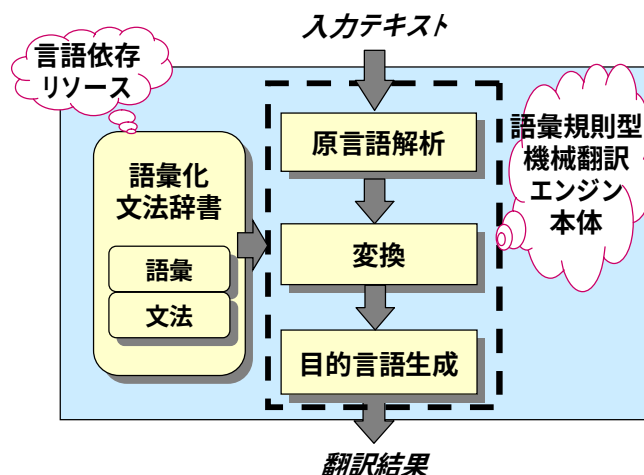


図5. 語彙規則型機械翻訳部の構成

3.2 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・目的言語生成から成るルールベースの翻訳モジュールであるが、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り替えが可能である。

3.2.1 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。

3.2.2 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み合わ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等の表現の強化、口語的表現や定型的表現への対応、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。

3.3 通訳統合部

通訳統合部は、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。

自動通訳の方向を、発話のターンに応じて高速に切り替えるためには、日本語・英語の音声認識エンジン、日英・英日の機械翻訳エンジンをすべて起動しておくか、高速に起動可能にする必要がある。今回は高速起動ではなく各エンジンをすべて起動しておく方式を採用した。このような構成とする場合にはメモリ割り当てが大変厳しくなるが、特に動作時に必要となるワークメモリを確保するために、各エンジンがスタティックに確保するメモリを削減するとともに、ワークメモリをこまめに解放して共有する等の制御を行っている。

3.4 エンジン単体評価

今回構築したコンパクト・スケーラブル自動通訳ソフトウェアを構成する各エンジン単体での性能を見るため、計算機シミュレーションによって評価した。

1. コンパクト・スケーラブル音声認識

日本語・英語の音声認識エンジンを評価した。評価データとしては、旅行会話の読み上げ発声データを用いた。雑音のない環境で収録した、男女各10名、合計約900発声である。音声認識の精度を一般的な評価尺度である単語正解精度（誤挿入をカウントする）で評価した結果、日本語音声認識においては94.4%の認識率が得られた。英語音声認識においては91.9%の認識率が得られた。日英とも9

割以上の認識率が得られている。

2. 語彙規則型機械翻訳

日英・英日の機械翻訳エンジンを評価した。評価データとしては旅行会話のテキストコーパスからランダムに500文を抽出し、主観評価により4段階評価した。主観評価における4段階の評価値と評価基準を表1に示す。今回は、相手に意味が伝わるための意味理解率を考え、4段階のうちc以上のものを意味が伝わる翻訳結果としてカウントした。評価の結果、日英翻訳では91%、英日翻訳で91%であった。日英・英日とも9割以上の意味理解率が得られた。

表 1. 翻訳精度の評価値と評価基準

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

4. 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。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る。

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が、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り、自動通訳としての精度は個々の単体評価の組み合わせで導き出せるとは限らない。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。

そこで我々は、小規模ではあるが実システムを実際に使い、主観評価かつ一貫評価によってコミュニケーション支援に直接影響を与える精度と速度を評価してきた[9-10]。しかしこれまでは、被験者が自動通訳の操作に慣れていることを前提として、

やり直しを許した複数回の試行により相手側に意味が通じかどうかを判定していた。ところが、ソフトウェアのユーザが所有者でない場合、必ずしも自動通訳の操作に慣れていることが前提とできず、一回目の失敗でコミュニケーションが破綻する危険がある。やり直しを許さない一回の試行での評価を行う方が、より実利用場面を考慮した評価と言える。

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
3. 複数回の試行による評価では、操作に不慣れなユーザによる試行を考慮していない

そこで今回、次の3つの方針により、上記3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
3. やり直しを許さない一回の試行での評価とする

本3つの方針に基づいて実機上で精度と速度・リアルタイム性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で7～8割以上の精度とリアルタイム比で1倍程度の速度を達成することで、真に実用的と見せる。

5. 統合的発話解釈評価手法による評価

携帯端末上に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。特に、スタンドアロンタイプとサーバ連携タイプとの両方を比較対象とすることで、速度面での優位性を議論する。

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1.0GHz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。

5.1 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、すなわち音声入力が終了しないと、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。

5.2 評価方法

今回は、ソフトウェアの所有者として日本人を想定し、操作に慣れていない相手側ユーザの入力として英日方向の通訳を評価する。やり直しを許さない一回の試行での評価である。

1. 音声データ

音声データとしては、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。典型文例から成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るため、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。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に実験の様子を表した模式図を示す。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、その後に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。

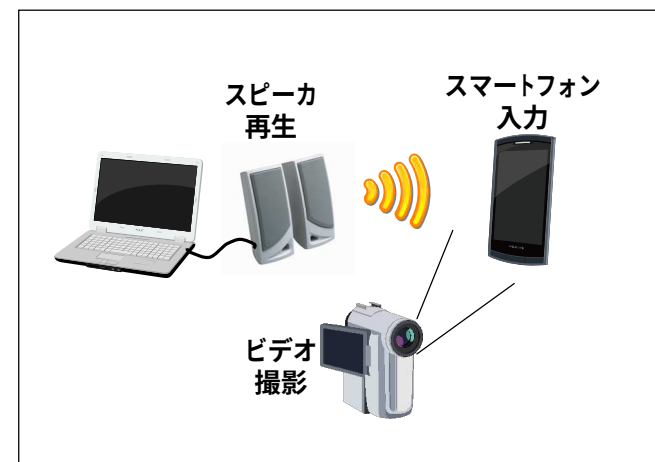


図6. 実験の構成

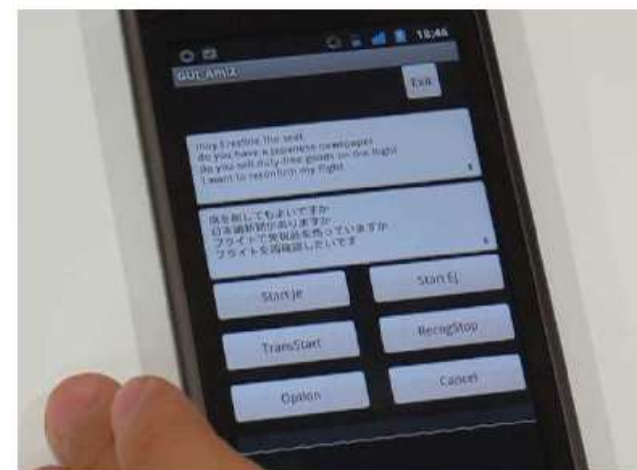


図7. コンパクト・スケーラブル自動通訳ソフトウェア動作例

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。画面上のユーザインタフェースは、マイク操作等のボタン類と、音声認識結果・機械翻訳結果を表示する画面とで構成されている。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。

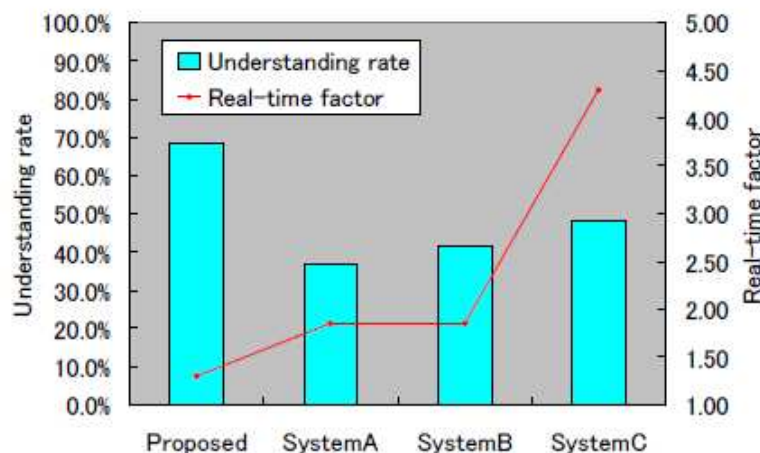


図8. 評価結果

5.3 評価結果

4種の自動通訳ソフトウェアによる比較評価の結果を図8に示す。図8は、意味理解率を左側(棒グラフ)、リアルタイム比を右側(折線グラフ)とした比較結果である。"Proposed"が我々のソフトウェアであり、"Proposed"と"SystemA"がスタンドアロンタイプ、"SystemB"と"SystemC"がサーバ連携タイプである。

1. 精度

精度は Proposed が他と比較して優れていることがわかる。高精度である理由としては、数万語規模の大規模辞書を音声認識・翻訳ともに備えていること、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チューニングが可能であること、が挙げられる。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く、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。

2. 速度

速度は Proposed が他と比較して優れていることがわかる。この理由としては、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、が挙げられる。SystemB と SystemC はともにサーバ連携版であるにも関わらず速度に大きな違いが出ており、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わかる。SystemA および SystemB ではリアルタイム比が 2.0 近いのに対し、Proposed では 1.3 程度である。すなわち、SystemA と SystemB では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が、Proposed では実際には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。

以上、統合的発話解釈評価手法により実機上で精度と速度を評価し、我々のコンパクト・スケーラブル自動通訳ソフトウェアが意味理解率で7割以上の精度とリアルタイム比で1倍程度の速度を達成、かつ他の現行ソフトウェアと比較して優位であることを示した。

6. 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3つ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が実用的であることを示した。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割～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。

今後の発展の方向性として、利用場面の拡大が挙げられる。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]。

7. 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、実用的な精度を保ちつつ高速性を極めたものである。

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、携帯端末の実機に実装した。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も精度面・速度面で優位であることを確認した。大規模サーバを利用可能なサーバ連携タイプと比較しても高精度であるし、他のスタンドアロンタイプと比較しても高速である。精度は実用上最低限の7割の精度を確保した。速度はリアルタイム比で約1.3を実現し、ユーザが話し終わった直後に結果が表示されているという圧倒的な高速性を実現した。

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) 中村他, “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, “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 Machine Translation Evaluation for GALE”, <http://www.itl.nist.gov/iad/mig/tests/gale/>
- 4) M. Paul, et al, “Overview of the IWSLT 2010 Evaluation Campaign”, International Workshop on Spoken Language Translation, pp.3-27, 2010.
- 5) 奥村他, “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, <http://fbi-award.jp/sentan/jusyou/2007/nec.pdf>, 2007.
- 6) 奥村他, “携帯端末など組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, pp.687, 2009.
- 7) T. Watanabe, et al, “An automatic interpretation system for travel conversation”, ICSLP-2000, vol.4, pp.444-447, 2000
- 8) 山端他, “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”, 情報研報, 2002-NL-150-9, 2002.

- 9) 花沢他, “携帯電話試作機上で動作する旅行会話向け音声認識”, 情報全国大会, 2D-3, 2009.
- 10) 長田他, “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”, 情報全国大会, 2D-2, 2009.
- 11) “Google translate”, <http://translate.google.co.jp/>
- 12) 下郡, “英語字幕による会議支援”, 情報処理, Vol.51, No.1, pp.26-29, 2010.
- 13) “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, <http://www.nec.co.jp/press/ja/0205/3002.html>, 2002.
- 14) “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, <http://www.nec.co.jp/press/ja/1001/2002.html>, 2010.
- 15) 磯谷他, “話し言葉認識に向けた基本技術と応用”, 情報研報, 2005-NL-169, 2005.
- 16) 磯他, “大語彙連続音声認識技術と応用”, 音講論集(秋), 1-1-15, 2008.
- 17) 岡部他, “言語モデル先読み値の平滑化による探索誤りの改善”, 音講論集(秋), 1-1-15, 2008.
- 18) 山端他, “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) 辻井他, “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) T. Takezawa, et al, “Multilingual Spoken Language Corpus Development for Communication Research”, Computational Linguistics and Chinese Language Processing, Vol. 23, No. 3, pp.303-324, 2007.
- 21) 花沢他, “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.