

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

釜谷聡史^{1,a)} 坂本明子¹ 渡辺奈夕子¹ 鈴木博和¹ 藤井寛子¹ 舘森三慶¹ 河村聡典¹
住田一男¹

概要: ユーザ同士が対面で会話する場面、特に、対面でサービスを行う対面業務を対象に、ユーザの連続自由発話を受け付け、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。日本語、英語、中国語で、ユーザは文の長さを意識することなく発話が可能であり、その発声に追従して通訳結果を出力することができる。本システムを用いて、事前に設定した対話課題を解く評価実験を行ったところ、約81%の課題を解決できることを確認した。

Evaluation of a Simultaneous Interpretation System for Continuous-Speech Conversation

Abstract: We developed a simultaneous interpretation system for face-to-face services at shops, front desks, windows, etc.. The system supports interpretation between Japanese and English, or Japanese and Chinese. It processes user's continuous and spontaneous speech, and incrementally outputs interpretation results. We conducted a field test of the system to evaluate the "solved task ratio" for tasks including buying souvenirs, asking a bus route. In the result, we achieved the solved task ratio of 81%.

1. はじめに

外国人と自由に意思疎通を図りたい。自動翻訳の研究開発は、計算機の黎明期から長く続けられている。音声言語技術の進展や計算機の進化に伴い、音声を入力として利用可能なサービスが身近になり、音声翻訳の技術も広く知られるようになった。

また、音声翻訳の精度も、旅行会話や、簡単な日常会話であれば、実用に耐え得る性能に達しつつある。一方で、ビジネスシーンでの活用は広まっていない。例えば、旅行会話に近い対話内容が予想される、旅行案内所、小売店舗、役所窓口などにおいて、外国語会話の翻訳ニーズがありながら、利用されてはいない。

前述のような、顧客と対面して行う業務においては、旅行者が自分の意思を伝えることを主な目的とする旅行会話と違い、より円滑なコミュニケーションが望まれる。しか

し、従来の音声翻訳では、ユーザは、一文あるいは数文を発話した後に一旦発話を止め、システムが出力する音声翻訳結果を対話相手に提示する必要がある。そのため、ユーザは、伝えたい内容をあらかじめ考えて、簡潔な文で発話する必要があったり、話者側が話し終えるまで対話相手に内容が伝わらなかつたりする等、円滑な対話を実現することが難しかった。

そこで我々は、ユーザの連続自由発話を受け付け、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。本システムでは、音声認識と機械翻訳とを並行動作させ、利用者が連続的に発話する自由発話音声进行处理し、訳すべき単位を検出して前から順に翻訳する。これにより、ユーザは発話の長さを気にすることなく発声することが可能となり、また、発話の途中段階であっても、順次対話相手に内容が伝わることから、制約の少ない対話が可能となる。

本稿では、開発した同時通訳システムについて2章に紹介し、同システムを用いた対話実験の概要を3章に述べる。実験結果とその分析を4章に報告して、5章に本稿をまとめる。

¹ 株式会社東芝 研究開発センター
Toshiba Corporation, Corporate Research & Development Center

^{a)} satoshi.kamatani@toshiba.co.jp

2. 同時通訳システム

図1に、開発した同時通訳システムの構成を示す。図1の左側部分に示すように、本システムは、音声翻訳を実現する音声認識・翻訳サーバと、ユーザインターフェイスを提供するクライアントからなり、インターネットを介して接続して使用する構成を採っている。

図1の右側部分は、我々の同時通訳システムの処理の流れを示している。我々のシステムは、1) フィラーを含む自由発話音声を一定の無音区間毎に確定して認識し、2) 認識結果から翻訳すべき単位を検出し、3) その単位毎に前から順に翻訳することで、利用者の連続発声に追従して翻訳する、同時通訳を実現している。また、必要に応じて、4) 翻訳結果を音声合成することができるように構成している。

以降、各処理モジュールについて、その概略を説明する。

2.1 音声認識

音声認識には、自社で開発した大語彙連続音声認識エンジンを使用した。今回の同時通訳システムは、店頭や窓口での対面業務への適用を想定しているため、雑音に強い音声認識が必要となる。

この音声認識エンジンでは、認識に用いる音声特徴量として、雑音環境でも音声の本質的な特徴を安定して表現し得る特徴量である帯域別平均時間ケプストラム特徴 [1] を用い、また、独自の音声区間検出方式 [2] を採用することで、雑音に強い音声認識を実現している。さらに、大語彙化にともなう認識性能の低下を防ぐため、音声を構成する一つ一つの音素の違いを際立たせる高精度音素識別技術 [3] を導入することで、高精度な大語彙連続音声認識を実現している。

言語モデル作成には、Web から自動収集した大規模テキストコーパスをベースとして、今回の使用状況を想定した旅行会話文コーパスをブレンドしたコーパスを用いた。認識語彙は、コーパス内での出現頻度の高い語彙を中心とし

た日本語 20 万単語、英語 3 万単語、中国語 3 万単語の基本辞書に加えて、3 章に示す実験条件に合わせて、千葉県観光に関する語彙（観光地名、施設名、交通機関名など）を強化した。

2.2 翻訳単位の検出

音声認識が出力する文字列は、必ずしも機械翻訳に適した単位とは限らない。音声認識では、所定時間以上の無音区間を検知すると、それ以前の音声系列を一つの単位として処理し、これに対応する認識文字列を出力する。複数の文が同じ認識文字列に含まれることもあるし、本来一文として解釈すべきものが、複数の認識文字列に分かれてしまうことがある。

そこで、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。

2.2.1 検出手法

翻訳単位の検出は、2 段階で行う。まず、音声認識結果から翻訳不要なフィ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・表層の系列情報に基づいて、除去には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、あるいは分割して訳すことができる単位があれば、その開始位置を推定することで、翻訳単位を検出する。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構築して、機械学習により推定モデルを構築する [4]。

学習には条件付き確率場 (CRF) を用い、その実装には CRF++ [5] を利用した。学習素性には、前後 2 形態素 (但し、中国語は文字) の表層情報のみを用いた。

2.2.2 学習データ

日本語は日本語話し言葉コーパス [6] の 14 万文を、英語は TED Talk の書き起こしデータである WIT³ [7] から得たコーパス 11 万文を、中国語は独自に構築した中国語旅行会話コーパス 40 万文を、それぞれ用いた。

これらのコーパスには、正解となる翻訳単位は付与されていない。そこで、英語と中国語は、句読点の位置を翻訳単位とした。一方、日本語は、概ね節の単位を、人間が文意を理解する際の単位として捉え [8]、これをパターンマッチングにより付与した。

2.2.3 検出精度

日本語 244 文、英語 1664 文、中国語 3648 文を、句読点を除いて 1 文字列として評価データを作成した。元の句点位置を正解とし、学習したモデルを用いて分割した時の精度を評価した (表 1)。表において、精度、再現率は下記の様に計算した。

$$\text{精度} = \frac{\text{翻訳単位の推定境界と、元の境界位置との一致数}}{\text{推定した翻訳単位の境界数}}$$

$$\text{再現率} = \frac{\text{翻訳単位の推定境界と、元の境界位置との一致数}}{\text{元コーパスの句点数}}$$

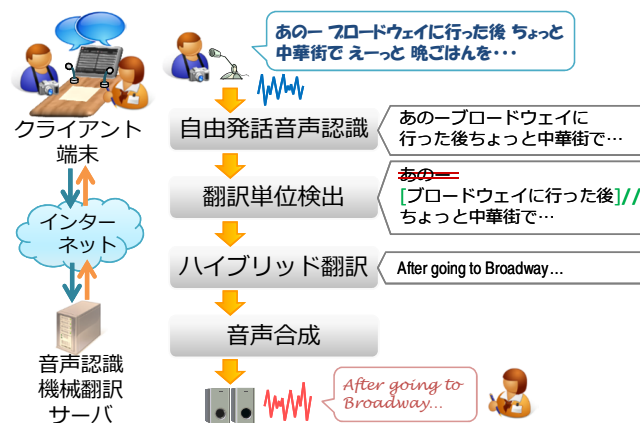


図 1 同時通訳システムの概要

表 1 翻訳単位検出精度

	精度	再現率	F 値
日本語	0.739	0.672	0.705
英語	0.720	0.809	0.763
中国語	0.679	0.685	0.681

2.3 機械翻訳

2.3.1 統語森駆動規則ベース翻訳方式

我々は、これまで、マニュアルや特許文書などの書き言葉を主対象とした規則ベース翻訳技術を開発してきた [9]。しかし、音声通訳のための規則翻訳では、書き言葉と違い、断片的で非文法的な話し言葉を適切に解析する必要がある。

そこで、我々は、発話の断片が入力されることを前提として、全ての統語的可能性を一括解析、最尤の各断片をまとめる節と、節間の関係を推定することで、話し言葉に対して高い翻訳性能を持つ統語森駆動規則ベース翻訳方式を開発した [10][11]。図 2 に、その処理の流れを示す。

統語森駆動規則ベース翻訳方式では、文法的尤度を伴って解析するよう拡張した一般化 LR 解析法により、全ての解釈を圧縮共有統語森（以下、統語森と呼ぶ）として導出、選好される構造を抽出して変換することで、入力に対する最尤構文候補からの翻訳を実現し、高精度化を図っている。また、訳文生成には、特に市場での成功例に富むトランスファ方式で使われる翻訳知識を活用し、高い品質とカヴァレッジを実現している。

2.3.2 ハイブリッド翻訳方式

旅行案内業務など、ドメインを限定した状況では、用例に基づく統計翻訳の方が良質な翻訳結果を得られる。一方、規則ベース翻訳は、やや硬い翻文になりやすい反面、ドメインを限らず、多様な入力に対して一定の翻訳結果を得られる。

そこで、統計翻訳エンジン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 [12][13]。入力原文を最初に統計翻訳エンジンに掛け、十分なスコアで翻訳できなかった場合に、規則ベース翻訳エ

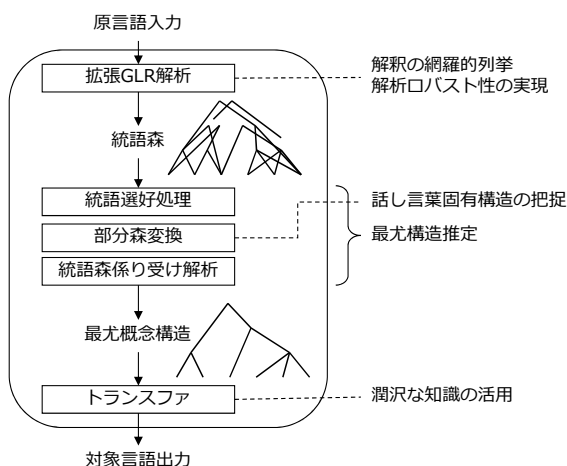


図 2 統語森駆動規則ベース翻訳方式の処理の流れ

ンジンで再翻訳することで、訳文の品質を高めつつ、多様な入力に対する頑健性を得る。翻訳エンジンの切り替えは入力される翻訳単位毎に行い、単位全体をいずれかのエンジンで翻訳する。

統計ベース翻訳には、我々が開発したフレーズベース統計翻訳エンジン [14] を用いた。日英・英日翻訳には、旅行会話を対象として独自に開発した対訳コーパス約 22 万対と、高度言語情報融合フォーラム ALAGIN *1が配布している、「日英翻訳エンジン学習・評価用対訳コーパス」約 2 万対を学習コーパスとして用いた。日中・中日翻訳には、同様に独自開発した日本語-中国語約 21 万対（全約 21 万対を独自に開発）を学習コーパスとした。

2.3.3 翻訳精度

旅行会話文 100 文をテキスト入力し、その翻訳結果を下記の 5 段階で目視評価した結果を図 3 に示す。

4. 正しくかつ流暢であり、問題点がない
3. 文法的に正しく意味も伝わるが、非流暢
2. 文法的な誤りがあるが、正しく意味が伝わる
1. 意味が分からない、間違った内容が伝わる
0. 翻訳結果が出力されない

図において、H はハイブリッド、R は規則ベース、S は統計ベースの各翻訳エンジンの性能を示している。いずれの翻訳方向、エンジンにおいても、評価値 0 の文はない。

日英・英日は、規則ベース翻訳エンジンの翻訳規則が、長期の蓄積により成熟しており、統計ベース翻訳エンジンとの相乗効果が高く出で、全体としての性能改善も見られる。一方、日中・中日は、翻訳規則の開発期間が比較として短く、基本的な規則しか網羅されていないことから、若干の品質改善は見られるものの、基本的には統計ベース翻訳の性能が濃く反映される結果となった。

対話を成り立たせる上で必要な翻訳品質として、評価値 2 以上を合格水準とすると、日英・英日で概ね 9 割、日中・中日で概ね 8 割の翻訳精度にある。

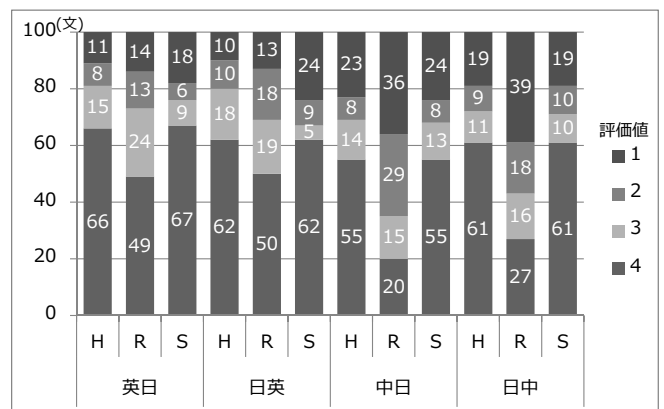


図 3 翻訳品質の目視評価結果

*1 高度言語情報融合フォーラム, 言語資源・音声資源サイト.
<https://alaginrc.nict.go.jp/index.php>

2.4 ユーザインターフェイス

客と店員が画面を共有して対話することを想定し、AndroidOSを搭載したタブレット端末で動作する、同時通訳クライアントアプリケーションを開発した。図4に、開発したユーザインターフェイスを示す。

話し手側のユーザは、対応する言語側の発話ボタンを押して発声を開始する。同じ話者の発話が連続している限りはボタンを押す必要はないが、再度同じボタンを押すことで明示的に通訳を終了させることも可能である。音声入力途中には、それまでの時点で得られている仮の認識結果(最尤系列)をグレーアウト状態で提示し、認識結果と翻訳結果が確定し次第、最終結果を提示する。

加えて、特に店員側を支援を目的として、下記の機能を実装した。

- 発話と類似する対訳用例を検索できる翻訳メモリ機能
- 単純なYES/NOの回答や、発声方法に関する指示、および、定型文をメニューから呼び出して提示する機能
- スワイプ操作により認識・翻訳結果を削除する機能

これらの機能は、初めてシステムを用いる外国人客に教示するには難しいが、店員側の操作を見て習得、使用することを考えて、双方に使用可能な機能として実装した。

図5に、日英翻訳の動作時の様子を示す。通訳(音声認識と翻訳)結果は、翻訳単位毎に表示される。通訳する必



図4 クライアント画面

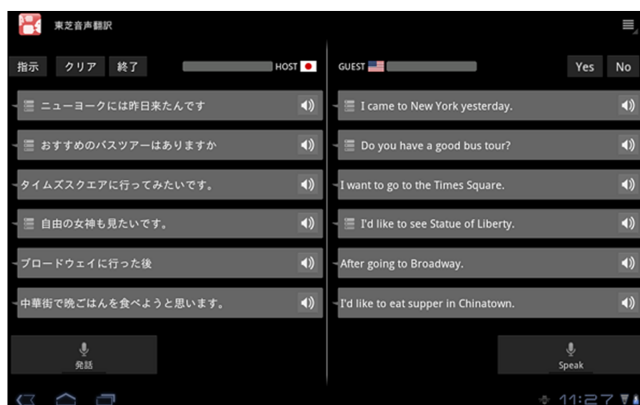


図5 日英同時通訳の動作時の様子

要のない部分は、通訳結果が確定すると削除されてしまうため、図には現れない。

3. 評価実験の設計

3.1 実験概要

開発した同時通訳システムの通訳性能を評価するため、システムが想定している対面業務に当たる旅行案内業務と日常会話とを対象とし、実環境において評価実験を行った。

実験は、事前に設定した課題を通訳機を介した対話を通して解決してもらう方式を採った。客役の米国人および中国人に課題を記したシートを渡して、通訳システムを介して店員役の日本人と会話してもらい、その課題の解決率を評価した。

3.2 実験環境

実験場所には、想定している使用場面に近い環境として、千葉市観光情報センター*2と、千葉市国際交流プラザ*3にご協力いただいた。観光情報センターでは旅行案内に関する対話を、国際交流プラザでは日常会話に関する対話を、それぞれ課題として設定した。

タブレット端末には、東芝 REGZA Tablet AT700を使用し、双方の被験者に画面が見えるようにテーブル上に設置した(図6)。音声入力には、モノラルマイクロホン2台をアダプターを用いて1入力に束ねたものを用い、タブレットのモノラル端子に接続した。翻訳結果提示に合わせて音声合成はせず、必要に応じてユーザがボタンを押すことで音声合成されるように設定した。サーバとの通信には、Wi-Fi接続に対応した、市販のLTEデータ通信端末*4を利用した。

3.3 被験者

外国人被験者には、米中ネイティブを各6名募集した。被験者の出身地、日本での居住年数は表2の通りである。

日本人被験者は、観光情報センターでは、実際にその場で業務に当たってる日本人職員2名に、国際交流プラザで



図6 実験機材の設置の様子(観光情報センター)

*2 http://www.chibacity-ta.or.jp/info/tourist_info_center/

*3 <http://www.ccia-chiba.or.jp/001kyokai/shisetu01.html>

*4 受信時最大 75Mbps / 送信時最大 25Mbps (規格値)

課題文	回答欄
Ask the way to get to a money exchange shop near here.	Place [] Did you complete this task? <input type="checkbox"/> Yes/ <input type="checkbox"/> No
Now you would like to know a bus routemap and its schedule in Chiba city. Ask how you can get these information.	Did you complete this task? <input type="checkbox"/> Yes/ <input type="checkbox"/> No
Ask the best souvenir of Chiba. Ask its features and how to get to a store where you can buy it.	Souvenir [] Did you complete this task? <input type="checkbox"/> Yes/ <input type="checkbox"/> No

図 7 課題シートの例 (米国人被験者用)

表 2 外国人被験者のプロフィール

	性別	年齢	出身地	居住年数
中国人	女	31	瀋陽	1
	男	42	成都	3
	男	28	北京	1
	女	31	成都	2
	女	25	上海	1.5
	女	40	北京	2
米国人	男	22	New York	1
	男	30	California	4
	女	25	Utah	3
	男	32	Pennsylvania	4
	男	40	California	4
	女	48	Florida	3

表 3 日本人被験者のプロフィール

所属	性別	年齢	出身地
観光情報センター	女	-	和歌山
	女	-	群馬
千葉大学	男	23	長崎
	男	23	栃木
	男	23	埼玉
	女	23	千葉
	女	21	千葉
	女	23	福島

は千葉大学生 6名に参加いただいた。被験者のプロフィールは表 3 の通りである。

3.4 対話課題

対話課題は、想定される対話の難易度を考慮して、1 から 2 ターンでの対話で解決できる簡単なものから、3 ターン以上または複数文を含む対話でなければ解決できない複雑なものまで、段階を設けて作成した。

観光情報センター向けには、同センターでの通常業務において想定される内容を選び、8 課題を作成した。難易度 (L1 から L4、数字が大きい方が難しい) 毎に例を示す。

- L1 ツアーの予約が可能か尋ねる
- L2 千葉駅からスポーツセンター駅までの運賃を尋ねる
- L3 千葉市のバス路線と運行時間を知る
- L4 訪問する観光スポットを相談して決める

国際交流プラザ向けには、日常会話を想定して、10 課題を作成した。難易度 (L1 から L5) 毎に例を示す。

- L1 曜日を尋ねる
- L2 ディズニーランドへの行き方を尋ねる
- L3 日本料理の好みを伝え、お店を探す
- L4 日本で困ったことを伝え、アドバイスを求める
- L5 相手と共通する関心事を見つける

3.5 実験手順

日本人被験者は、機器の操作に慣れている状況を想定し、事前に機器を提供して練習していただいた。一方、外国人被験者については、機器の操作方法を示す簡単な説明資料を準備し、通訳システムと共に配置するのみで、口頭での操作方法は説明しなかった。

外国人被験者には、簡単な実験の概要説明をした後、課題の書かれたカードを渡した。図 7 に提示した課題シートの例を示す。課題シートは、課題文と回答欄からなっている。被験者には、対話によって得られた情報と共に、課題解決の成否を回答欄に記入するよう指示した。更に実験後、日本人評価者が同回答内容と、対話ログを分析し、正しい内容が伝達されているか否かを判定した。

与える課題は、被験者毎に順序を変え、観光情報センターで 30 分、国際交流プラザで 60 分を対話時間の限度とし、これを越えた時点で、残りの課題は未達として評価した。実験中は、双方、自国の言語以外は理解できないものとし、表示された相手側言語から情報を得ないように指示した。観光情報センターでは、日常業務で使用している地図や資料などは使用を許可した。

4. 実験結果

4.1 タスク達成率

対話課題の解決率であるタスク達成率を定義して評価した (図 8)。被験者の自己申告による判定結果 (黒塗りグラフ) と、実験のログデータを元にした、採点者の判定結果 (白抜きグラフ) とを示している。両グラフの数値差が、被

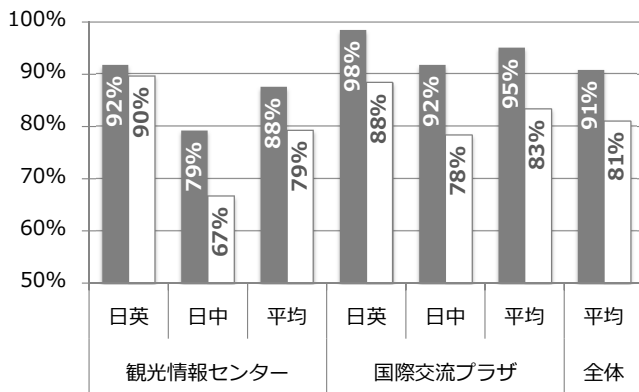


図 8 タスク達成率

験者間で発生した誤解の割合と考えることができる。

図 8 に示すように、観光情報センターにおけるタスク達成率よりも、国際交流プラザでのタスク達成率が高い。これは、旅行案内会話では被験者が課題の解決を明確に判定できるのに対し、日常会話では判定が曖昧になってしまうことが考えられる。加えて、具体的な地名、品名などが適切に伝わる必要がある旅行案内会話に比べて、言い回しを変えることによって語彙の不足を補える日常会話の方が、対話内容の工夫が可能であったためと考えられる。

言語別に見ると、日英に比べて、日中のタスク達成率が低い。また、被験者の自己申告による判定結果と、採点者の判定結果に見る誤解の程度も、日英に比べて日中が悪い傾向にある。この要因としては、音声認識や機械翻訳の精度が日英に比べて日中が低いことが直接表れている、と考えられる。

出身地や性別の違いによる、有意なタスク達成率の差は見られなかった。これには、タスク数が限られている上、システム上のトラブルなどが影響を与えていることも考えられる。また、これらの差は、音声認識結果や翻訳結果に表れると考えられ、その分析には時間を要することから、今後の課題としたい。

4.2 対話内容の分析

図 9 に「家族へのお土産を選ぶ」課題に対する英日対話の例を、図 10 に「外貨両替所を探す」課題に対する中日対話の例を示す。図では、一つの元発声に対して複数の翻訳結果が対応づけられているもの（例えば、6ej や 4cj）は、翻訳単位検出処理により、2 文以上に分割されたことを示している。

5ej や 1cj の発話に見られる誤認識や、7je や 8cj の発話のような誤訳も見られるが、文脈から意図を推定しながら、課題の解決に至った様子が分かる。連続的に発話を処理して全て提示するが故に、6cj のように本来対話相手に伝える必要のないものも伝わってしまうが、双方の反応を窺い、対話の流れを把握することで、必要な情報を拾いながら対

話を進行している。

ところで、音声認識結果に誤りが認められた場合、被験者は、1 単語毎、ゆっくり区切りながら発声する傾向が見られた。これは、音声認識過程で表示される途中結果を確認しつつ、それに合わせて発声を変えるためと考えられる。

しかし、このような断片的な発声では、言語モデルが有効に働かず、誤認識が引き起こされるばかりでなく、推定に必要な文脈知識も不足するため、全体として性能が更に低下する可能性が高い。また、認識結果が細かく分かれてしまうことにより、文としての解釈ができず、翻訳精度も低下する。

これは、認識の観点から見れば、認識の単位を区切る無音区間をどの程度とるかという問題であり翻訳単位の検出の観点から見れば、どの範囲の発話を連結させるかという問題である。これらは、予め決めておけることではなく、動的に決定する手法が必要となる。このような、本来まとめて解釈すべき文が複数に分かれて認識出力された場合のまとめ上げ処理は、今後検討が必要である。

4.3 通訳時間

各課題の、最初の発声から最後の発声までの時間を計測し、その平均をとったものを表 4 に示す。計測した時間には、機器やネットワークなどの不調に伴う遅延時間も含まれている。これを除いて計測することも考えられる。しかし、これらの時間を含めたものが真の通訳時間であると捉え、システム全体の評価として、これを含めた。

表 4 には、全ての対話時間を平均した単純平均と、最長と最短の対話時間を除いて平均した中間項平均と、課題が正しく達成できた（採点者の判定結果に基づく）対話の時間を平均したものを示している。

いずれの結果においても、概ね 2、3 分以内に課題を完了している。また、英日対話の方が、中日対話に比べて長い傾向が見られる。これは、課題内容の難しさや、通訳精度に起因しているというよりも、被験者がどこまで話題を膨らませたかに依存するところが大きい。例えば、お土産を買う対象を良く想像して細かく条件を伝えたり、対話の目的のない事項まで話すなどの様子も見られた。本評価では、被験者がタスク完了を判定する方式を採ったため、通訳時間もこれに倣って、全体の所要時間で評価している。

表 4 対話に要した平均時間

		英日	中日
観光情報センター	全対話	3:11	2:35
	達成成分のみ	2:56	2:30
国際交流プラザ	全対話	2:22	1:36
	達成成分のみ	2:20	1:32

	元発声	認識結果	翻訳結果
1ej	I'd like to buy some souvenirs.	i'd like to buy some souvenirs	お土産を買いたいのですが。
2ej	Can you recommend anything special?	can you recommend anything special	お勧めの特別の何かはありますか？
3je	最近東京スカイツリーができたので	最近東京スカイツリーができたので	The Tokyo sky tree was made recently.
4je	スカイツリーのグッズがおすすめです	スカイツリーのグッズがおすすめです。	The goods of a sky tree are recommended.
5ej	I plan to go see Tokyo Sky Tree.	i plan to go see this guy tree	この人木を見に行くことを計画します。
6ej	tommorow. That sounds like a great idea.	tommorow that sounds like a great idea.	明日です それは素晴らしい考えのように聞こえます
7je	夜のライトアップがとてもきれいなのでぜひ見て下さい	夜のライトアップがとてもきれいなのでぜひ見て下さい	Lighting of night is very beautiful. Please see.

図 9 英日対話ログ (家族へのお土産を選ぶ)

	元発声	認識結果	翻訳結果
1cj	在哪里可以把外汇兑换成	在那里可以把外汇兑换成	あそこで 外貨にすることができます
2cj	应该怎么走	应该怎么走	どう行くべきである
3jc	(雑音)	えーえーえーえーえーので	是。
4cj	在哪里可以把外汇兑换成日元应该怎么走	在那里可以把外汇兑换成日元应该怎么走	あそこで 外貨にすることができます どう行くべきである
5jc	外貨両替ですね。	外貨両替ですね。	兌換外币。
6cj	二	二	二
7cj	是的	是的	はい。
8jc	ワールドカレンシーショップか	ワールドカレンシーショップか	是世界货币店舗吗？
9jc	ヨドバシカメラのビルの三階にあります。	ヨドバシカメラのビルの三階にあります。	友都八喜照相机の大楼的三楼有。

図 10 中日対話ログ (外貨両替所を探す)

5. おわりに

本稿では、対面業務を支援を目的として開発した同時通訳システムについて述べた。本システムでは、ユーザの連続発声を受け付け、これを適切な単位に区切って順次訳出する。これにより、ユーザは発話の長さや文を意識することなく話すことが可能であり、より短時間でのコミュニケーションが可能となる。

また、開発したシステムを用いた評価実験について報告した。評価実験から、対面して互いの意思を確認しながら対話する状況においては、約 81%の達成率で、対話の目的を達成できることを確認した。

今後の課題としては、通訳精度自体の改善に加え、通訳誤りに起因する誤解の解消と、伝達したい内容を入力する支援機構の開発が挙げられる。

謝辞 実験にご協力いただいた千葉市、公益社団法人千葉市観光協会、公益財団法人千葉市国際交流協会、および、共同で実験に当たっていただいた千葉大学黒岩研究室、東京ベイ通信株式会社に感謝いたします。

参考文献

- [1] 中村 匡伸, 藤村 浩司, 篠原 雄介, 益子 貴史, 河村 聡典, “群遅延に基づく音声特徴量の雑音音響下での評価”, 日本音響学会 2012 年春季研究発表会講演論文集, pp.135-136, (2012).
- [2] Hongfei Ding, Koichi Yamamoto and Masami Akamine, “Comparative evaluation of different methods for voice activity detection”, In *Proc. Interspeech 2008*, pp.107-110, (2008).
- [3] Hiroshi Fujimura, Masanobu Nakamura, Yusuke Shinohara, Takashi Masuko, “N-Best rescoring by adaboost phoneme classifiers for isolated word recognition”, In *Proc. Automatic Speech Recognition and Understanding Workshop (ASRU Workshop 2011)*, pp.83-88, (2011).
- [4] Yang Liu, Andreas Stolcke, Elizabeth Shriberg and Mary Harper, “Using Conditional Random Fields For Sentence Boundary Detection In Speech”, In *Proc. of the 43rd Annual Meeting of the ACL*, pp.451-458, (2005).
- [5] Taku Kudo 2005. “CRF++: Yet Another CRF toolkit”, Available at <https://code.google.com/p/crfpp/>
- [6] Kikuo Maekawa, Hanae Koiso, Sadaoki Furui and Hitoshi Isahara, “Spontaneous Speech Corpus of Japanese”, In *Proc. of LREC2000*, pp.947-952, (2000).
- [7] M. Cettolo, C. Girardi, and M. Federico, “WIT³: Web inventory of transcribed and translated talks”, In *Proc. of EAMT*, pp.261-268, (2012).
- [8] Takanashi Katsuya, Takehiko Maruyama, Kiyotaka

- Uchimoto, and Hitoshi Isahara, “Identification of “Sentence” in Spontaneous Japanese – Detection and modification of clause boundaries –”, In *Proc. of ISCA & IEEE Workshop on Spontaneous Speech Processing and Recognition*, pp.183-186, (2003).
- [9] 東芝ソリューション株式会社, “The 翻訳シリーズ”, http://pf.toshiba-sol.co.jp/prod/hon_yaku/index_j.htm
- [10] Tetsuro, Chino and Satoshi Kamatani, “Partial forest transfer for spoken language translation”, In *Proc. of Recent Advances in Natural Language Processing (RANLP 2005)*,(2005).
- [11] Satoshi Kamatani, Kentaro Furihata and Tetsuro Chino, “Forest Driven Dependency Analysis Enhanced by Japanese Clause Structure Estimation”, In *Proc. of The 20th Pacific Asia Conference on Language, Information and Computation*, pp.265-273, (2006).
- [12] Satoshi Kamatani, Tetsuro Chino and Kazuo Sumita, “Hybrid Spoken Language Translation Using Sentence Splitting Based on Syntax Structure”, In *Proc. of Machine Translation Summit XII*, (2009).
- [13] 知野哲朗, 釜谷聡史, 降幡建太郎, 住田一男, “日中英 3 言語 6 方向音声翻訳システム”, 情報処理学会研究報告. SLP, 音声言語情報処理, pp.15-22, (2008).
- [14] Haifeng Wang, Hua Wu, Xiaoguang Hu, Zhanyi Liu, Jianfeng Li, Dengjun Ren and Zhengyu Niu, “The TCH Machine Translation System for IWSLT 2008”, In *Proc. of International Workshop on Spoken Language Translation 2008 (IWSLT 2008)*, pp.124-131, (2008).