

音声翻訳技術概観

Brief Overview of Speech Translation Technology

中村 哲

Abstract

2020年の東京オリンピック・パラリンピックが決まり、増加する訪日外国人とのコミュニケーションツールとして音声翻訳技術に対する期待が高まっている。本稿ではこれまでの音声翻訳の研究開発、現状、世界の研究動向、音声翻訳システムの構成、音声翻訳に関する国際評価型ワークショップ IWSLT、音声翻訳国際コンソーシアム、実用化動向、そして、現在の技術的課題、及び今後の音声翻訳研究開発の展望について述べる。

キーワード：音声翻訳，機械翻訳，音声認識，音声合成

1. はじめに

2020年の東京オリンピック・パラリンピックの決定等により、多くの外国人の訪日が予想されている。観光立国推進閣僚会議の報告書⁽¹⁾によれば、「2014年の訪日外国人旅行者数は約1,036万人と、2012年の約836万人から大きく飛躍し、2003年のビジット・ジャパン事業開始以来の政府目標であった年間訪日外国人旅行者数1,000万人を史上初めて達成することができた。これにとどまることなく、観光立国推進閣僚会議は、2015年1月、総理大臣の指示を受け、「2020年オリンピック・パラリンピック東京大会の開催という絶好の機会を捉え、これを追い風として、更なる観光立国の推進を図るべく、2020年に向けて、訪日外国人旅行者数2,000万人の高みを目指すこととした。」とし、具体的には、

- ① 「2020年オリンピック・パラリンピック」を見据えた観光振興
- ② インバウンドの飛躍的拡大に向けた取組み
- ③ ビザ要件の緩和など訪日旅行の容易化
- ④ 世界に通用する魅力ある観光地域づくり
- ⑤ 外国人旅行者の受入環境整備

- ⑥ MICEの誘致・開催促進と外国人ビジネス客の取込み (M: Meeting, I: Incentive, C: Convention, E: Exhibition)

に取り組むとしている。また、2020年頃の外国人居住者は300万人に増加すると予想されており、オリンピック・パラリンピックにより居住者についても更なる増加が見込まれる。

近年、日本はグローバル化が一層進み、英語の案内が増え、英語教育の強化、メディアでの英語が増えたこともあり、英語への対応が進みつつある。しかし、これらは一部の都市、場所に限られ、地方では依然大きな問題として残っている。例えば、2008年の日本の宿泊施設に対する調査では、外国人の受入れを望まない理由のトップは、「外国語対応ができないため」(75.5%)となっていることが報告されている⁽²⁾。ここに来て訪日外国人、在住外国人にとっての日本語という言語の壁の問題が大きくクローズアップされてきた。

一方で、日本が世界に先駆けて研究をスタートした音声翻訳技術が近年大きな進歩を遂げている。音声翻訳は、話した言葉をその場で認識し、相手言語に翻訳する技術である。1983年に日本電気がITU-テレコムで音声翻訳のコンセプト展示を行ったこともあり、我が国でも世界に先駆けて、1986年に政府の基盤技術促進法に基づき、国際電気通信基礎技術研究所(ATR)に音声翻訳に特化した研究開発会社であるATR自動翻訳電話研究所が発足した。1980年代後半は、バブル期であり日

中村 哲 正員 奈良先端科学技術大学院大学情報科学研究科情報科学専攻
E-mail s-nakamura@is.naist.jp
Satoshi NAKAMURA, Member (Graduate School of Information Science, Nara Institute of Science and Technology, Ikoma-shi, 630-0192 Japan).
電子情報通信学会誌 Vol.98 No.8 pp.702-709 2015年8月
©電子情報通信学会 2015

本初の技術を創出するための多くの研究開発の取組みが行われた。音声翻訳は、音声認識、機械翻訳、音声合成の三つの要素から構成されているが、1980年代後半は、それぞれ、音声認識は1,000語の不特定話者の音声認識が何とか動作し始めたレベル、機械翻訳は書き言葉で短くて曖昧性のない文章を手でルールに基づき翻訳するレベル、音声合成はテキスト音声合成では、やっと素片接続が動き始めたレベルで、明瞭度が低く、不自然で平板なイントネーションの音声しか生成できなかった。約30年の期間を経て、旅行など分野を限定すれば異言語間コミュニケーションが可能な技術レベルに達し、更には実サービスとしても一般の利用が可能となるまで到達しつつある⁽³⁾。本稿では、これまでの経緯の簡単な紹介と現在の課題、今後の展開について述べる。

2. 音声翻訳の歴史

音声翻訳に関する紹介記事については、2013年にも本会誌において寄稿の機会を頂いているので、まずはそちらを一読されたい⁽³⁾。最近の動向を追記したものを図1に示す。図の縦軸には、音声から音声へのコミュニケーションを目的とした音声翻訳と、多言語の音声情報を収集し、それらを認識・翻訳、情報抽出する音声翻訳プロジェクトを区分して掲載している。国内では、ATR及びそれを継承した情報通信研究機構（NICT）

で継続的な研究開発が精力的に実施されてきた。また、C-Star⁽⁴⁾に引き続き、A-Star⁽⁵⁾におけるアジア国際音声翻訳実験のシステムをベースにITU-Tにおいて音声翻訳用のプロトコルが標準化された^{(6),(7)}。一方、米国では、国防総省DARPAプロジェクトを中心に、戦場でのコミュニケーション用途⁽⁸⁾と情報分析用途⁽⁹⁾の両面から研究開発が進められた。欧州では、会議を対象にし、最近では欧州議会の公用23言語の同時通訳を支援することを目的としたEU-Bridgeプロジェクト⁽¹⁰⁾が実施されている。

実用化については、我が国において、世界初のネットワーク商用サービスがATRの技術をベースにATR-Trek社からリリースされ、その後、スマートフォン上のアプリVoiceTraがNICTからリリースされている^{(11),(12)}。最近では、Google Translateでの画像入力での翻訳機能の追加や、Microsoftが買収したSkypeに音声翻訳機能を付与したトライアル版がリリースされるなど、いろいろなレベルのトライアル、実用化、サービスアプリが拡大している。特に、我が国では2020年の東京オリンピック・パラリンピックにおける訪日外国人への言語サービス強化を目的に、総務省においてグローバルコミュニケーション計画が立案され、今まさに様々な施策が実施されようとしている⁽¹³⁾。

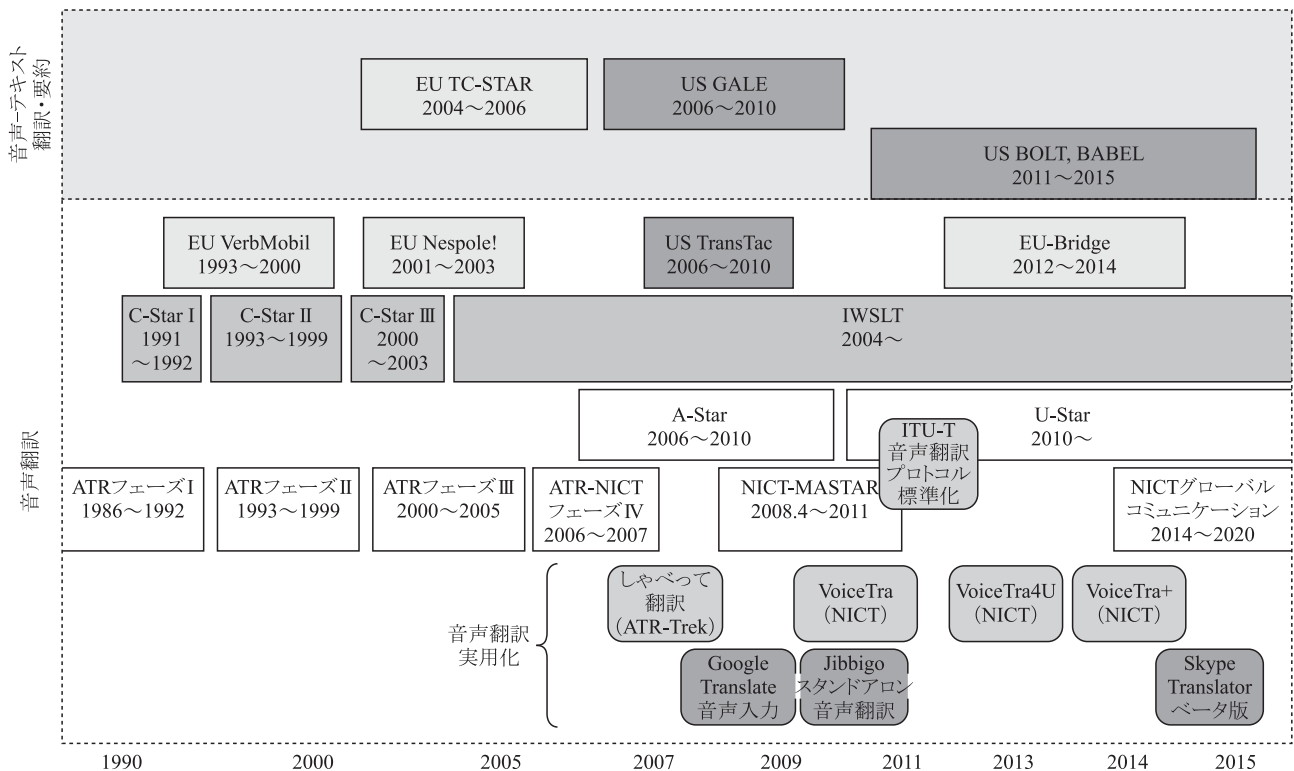


図1 音声翻訳プロジェクトの歴史

3. 音声翻訳システム

3.1 音声翻訳システムの構成と最近のトピックス

音声翻訳技術は大規模な音声言語コーパス（コーパスとは音声や言語データベースの総称）の集積と機械学習技術により大きな進化を遂げた⁽¹⁴⁾。音声翻訳システムは、これら音声言語技術の統合システムとして構成される。図2に音声・マルチモーダル翻訳システムの構成を示す。

音声翻訳は、標準的には、言語情報に特化して、音声認識・機械翻訳・音声合成の三つのコンポーネントとそれらを統合する技術から成る。音声翻訳は、テキストの翻訳と異なり、話し言葉を対象とする。話し言葉の文には非文法的な口語表現が含まれること、音声認識出力には、句読点がない、疑問符や感嘆符、引用符などの記号も含まれないことから、テキスト翻訳とは異なる問題を有する。特に、音声認識の誤りは、時に重大な翻訳誤りを起こす。これまでは、話題を対象を絞り込むことにより、音声認識の精度を利用可能なレベルまで向上させ、音声翻訳全体の誤りを低減するという手法がとられた。このため、比較的容易な翻訳から順次高度な翻訳へと研究開発が進められ、対象とする会話は、会議予約、ホテル予約、旅行会話、多様な日常会話へと簡単なタスクから実用的なタスクへ順を追って進められてきた。現在は、講演の翻訳、更に高度なビジネス会話をもスコープに入れた研究が行われている。

各モジュールに関するより詳細な解説についてはそれぞれの章に譲るが、最近の研究動向としては、

- ・ 音声認識：ディープニューラルネットワーク（DNN）による性能改善

- ・ 機械翻訳：構文情報を考慮した Tree-to-string 翻訳⁽¹⁵⁾、また、リカレントニューラルネットワーク（RNN）を用いた翻訳⁽¹⁶⁾
- ・ 音声合成：隠れマルコフモデル（HMM）に基づく音声合成、調音モデルを考慮した統計的音声合成
- ・ 統合システム：音声認識結果の出力単語ラティスに機械翻訳を適用することで最適な候補列を生成する方法、更に翻訳性能に基づくモジュール最適化

などが、新たな動向として見られる。1990年代に隆盛を見たニューラルネットワークが再び着目され、特に音声認識で大きな性能改善に貢献している点が興味深い。

詳細は後述するが、図2の点線部分のモジュールの研究開発により、より精度が高く、ドメイン知識、対話状態、非言語情報を考慮した音声翻訳などが今後可能になると考えられる。

3.2 IWSLT 音声翻訳評価型ワークショップ

IWSLT (International Workshop on Spoken Language Translation) は、音声翻訳研究の国際的なコンソーシアムである C-Star の主要メンバーが始めた評価型ワークショップであり、2004年から開催されている。毎年、参加機関数も増え、現在では世界の音声翻訳研究の中核的イベントとなっている。対象が旅行会話や講演という一般向きな話題で、学習データ、開発データ、評価データが入手可能であり、共通の土俵で音声翻訳の研究開発ができることが IWSLT の特徴である⁽¹⁷⁾。図3にこれまでの IWSLT のタスクを示す。縦軸の MT は機械翻訳のみのトラック、SLT は音声翻訳のトラック、ASR は音声認識のトラックである。最近になるにつれて、次第にそれぞれ対象言語数が増え、対象も TED 講

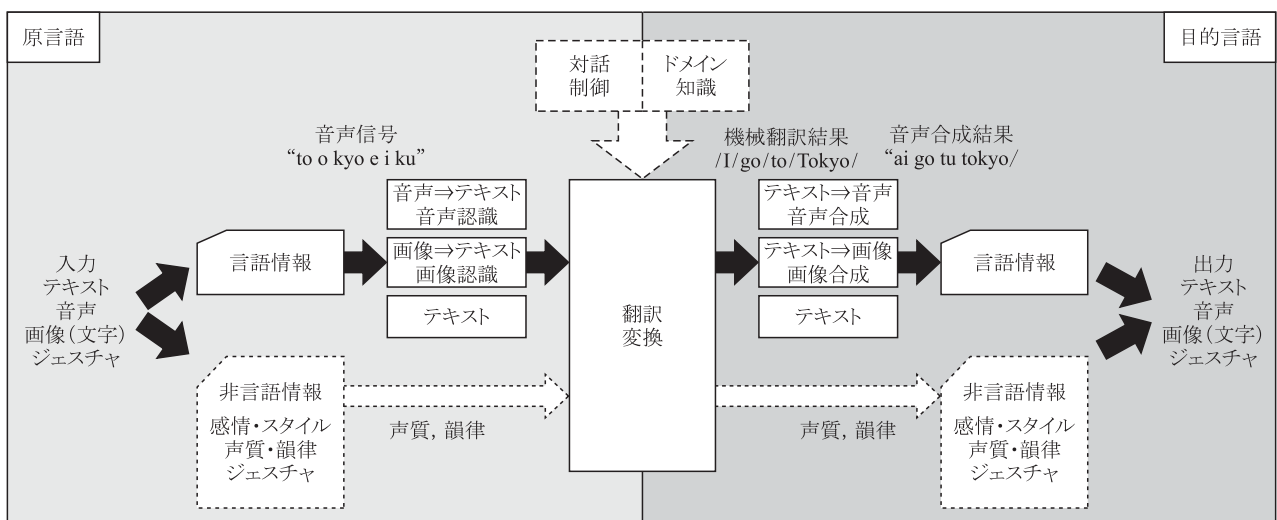


図2 音声翻訳システムの構成

凡例: Ar: アラビア語 En: 英語 De: ドイツ語 Fr: フランス語 Ko: 韓国語
It: イタリア語 Ja: 日本語 SpE: スペイン語 Tu: トルコ語 Zh: 中国語

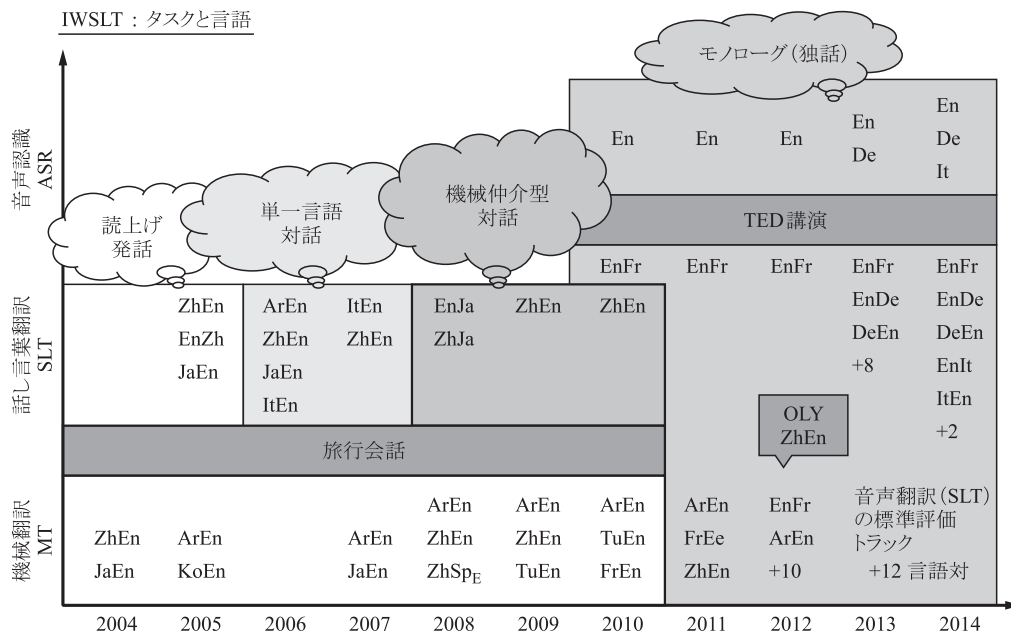


図3 IWSLTのタスクの変遷 (S. Stuker氏の発表資料⁽¹⁷⁾より引用)

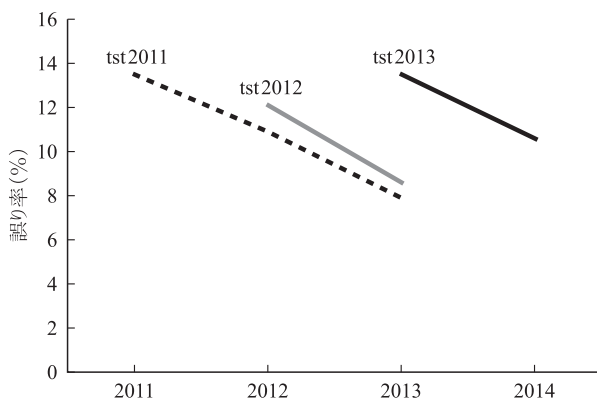


図4 IWSLTの音声認識単語誤り率の変遷⁽¹⁷⁾

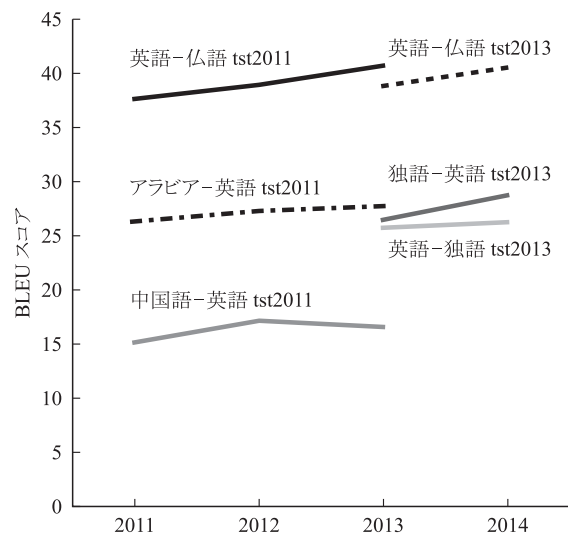


図5 IWSLTの翻訳性能の変遷⁽¹⁷⁾

演と難易度が高まっていることが分かる。2014年の評価トラックの構成は、音声認識では、英語、ドイツ語、イタリア語 TED 講演の音声認識タスク、音声翻訳では、英→仏独伊、独伊→英に2言語を加えた音声翻訳タスク、テキスト翻訳では、この音声翻訳の言語に12言語を加えた言語と英語間での双方向テキスト翻訳タスクとなっている。12言語には、アラビア語、スペイン語、ブラジルポルトガル語、中国語、ヘブライ語、ポーランド語、ベルシャ語、スロベニア語、トルコ語、オランダ語、ルーマニア語、ロシア語が含まれる。TED 英語講演の音声認識の最高性能は単語誤り率8.4%であり、図4に音声認識で最高精度を達成したシステムの過去のテストセットに対する年度ごとの単語誤り率を縦軸に示し

ている。1年ごとに約4%程度の誤り削減を達成していることが分かる。また、TED 英語講演の英仏テキスト翻訳の最高性能は自動評価尺度 (BLEU スコア⁽¹⁸⁾) で37.85であった。図5は機械翻訳で最高性能を達成したシステムの年度ごとの BLEU スコアを縦軸に示している。同様に持続的な改善が見られる。TED 講演を対象にした音声翻訳の結果では、英仏で最高性能が BLEU スコアで28.16となっており、機械翻訳と比較して音声翻訳では BLEU スコアが9.7低下している。

4. 音声翻訳システムの現状と課題

近年では、スタンドアロン型に加えて、スマートフォンを端末としたクラウド型の音声翻訳サービスが広がり、データを蓄積しながら機械学習で性能改善を漸次的に行うシステムが一般的になってきている。

4.1 国際共同研究コンソーシアムと標準化

アジアにおけるネットワーク型音声翻訳の先端研究コンソーシアムである A-Star⁽⁸⁾に引き続き、全世界に向けた音声翻訳コンソーシアムである U-Star⁽¹⁹⁾が結成された。現在 23 か国・26 機関が加盟し、23 言語の音声翻訳技術の研究が行われている。2010 年 10 月 14 日に ITU-T で承認された音声翻訳の接続プロトコル勧告 F.745 及び勧告 H.625 に従い、U-Star として「多言語音声翻訳サービスを提供するシステム」を共同で開発している。このシステムは、加盟機関の音声翻訳サーバを、ネットワーク型音声翻訳通信プロトコルで相互接続し、音声翻訳を提供する。

4.2 その他の動向

現在、一般に利用可能な音声翻訳サービスとしては、図 1 中に示したように NICT の開発した VoiceTra に加えて、Google の Google Translate、Microsoft の Skype Translator、そして、Jibbiggo のスマートフォンスタンドアロン型の音声翻訳ソフトがある。ほとんどのシステムは、同様なシステム構成になっており、ニューラルネットワークをフロントエンドに用いた HMM に基づく音声認識、フレーズベースの統計的機械翻訳、そして、素片ベースあるいは HMM をベースにした音声合成が使われている。Google Translate は、2015 年 7 月現在で 90 言語間のテキスト翻訳が可能で、うち約 40 言語で音声認識が可能であり、更に最近ではカメラのキャプチャ画像の翻訳、手書き文字入力からの翻訳も可能になっている。一方、Skype Translator は、現在ベータ版で英語とスペイン語間の音声翻訳を可能としている。

5. 技術的課題

これまでの研究は主として旅行会話等の音声翻訳と TED 講演の音声翻訳を対象に進められてきた。旅行会話を対象にした音声翻訳では比較的多言語への対応が可能であった。これは、文が短く構文の曖昧性が低いことにより、高い音声認識性能及び機械翻訳性能が確保できることが大きい。一方、TED 講演の場合には、英仏のような類似の文構造を有する言語を対象に試みられてきた。ところが、日英の場合は語順が大きく異なるため翻訳性能が大きく劣化してしまう。また、日本語が原言語の場合には、主語などの省略が頻発し更なる性能劣化を

引き起こす。また、音声翻訳の場合には、話し言葉と書き言葉との違いに対応する必要がある。このために、話し言葉の収集、話し言葉の対訳の収集、話し言葉と書き言葉間の言い換えパラフレーズなどを考慮する必要がある。

下記に技術的な課題を示す。

- 性能上の課題
 - ・ 全体：高速化，高精度化，大語彙化，多言語化，未知言語対応
 - ・ 音声認識：話者，方言，言語，スタイル，雑音
 - ・ 機械翻訳：文脈利用，主語省略・照応への対応，構文・依存構造の利用，対話構造・談話構造の考慮，曖昧性解消，意味解析，クロスカルチャ対応
 - ・ 音声合成：音質，自然性，個人性，スタイル，文生成
 - ・ 全体システム：統合システムとしての最適化
- 実時間性の課題
 - ・ 同時性，同時通訳（音声から音声，音声から字幕）
- 知識とのリンク
 - ・ Web，Wikipedia，オントロジー，意味の利用
- ほかの情報とのリンク
 - ・ 画像，位置，利用者，履歴，状況
 - ・ 韻律利用，パラ言語・非言語制御
 - ・ 実システム運用と自動学習
 - ・ 持続的なデータ自動収集と教師なし学習

筆者らのグループを含め、文末を待たずに処理を進める同時通訳アルゴリズム^{(20), (21)}と評価手法⁽²²⁾，音声翻訳性能から音声認識や機械翻訳のモデルの重みを最適化する手法^{(23), (24)}，強調などの非言語意図情報を翻訳後の音声に付与する音声翻訳⁽²⁵⁾，利用者の声で翻訳出力する声質変換を備えた音声翻訳^{(26), (27)}，発話画像の変換も行う音声・発話顔音声翻訳⁽²⁸⁾などの研究についても積極的に進められている。

6. 音声翻訳研究開発の将来展望

6.1 研究開発の展望

図 6 に筆者が 2008 年に作成した予想ロードマップを示す⁽²⁹⁾。2008 年は旅行会話を対象にした音声翻訳の実用化が開始された翌年であり、言語数も日英中に限定されていた。図中の下の矢印で示された旅行会話の多言語化のラインとその上の実用旅行会話とそのサービスのラインについては、インターネットに接続したクラウドサービスとしたことと、国際コンソーシアムでの協力体制の構築もあり、ほぼ予測どおりに進展している。しかし、一番上の矢印のラインの五月雨型の同時音声翻訳の

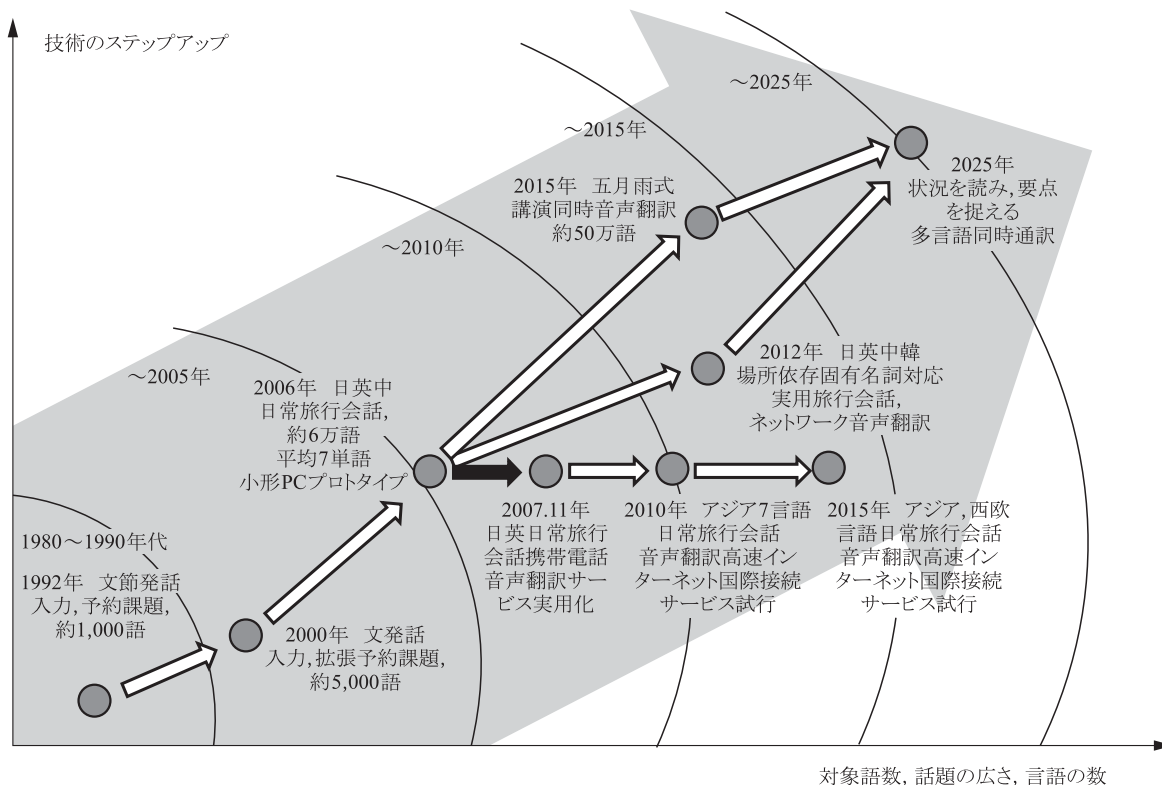


図6 音声翻訳の研究開発ロードマップ⁽²⁹⁾ (2008年)

表1 研究開発の展開イメージ

現在	2020 実験システムレベル @ 2015	2020 以降 基礎研究レベル @ 2015
利用シーン 旅行会話 音声翻訳システム	利用シーン オリンピック, 日常会話音声翻訳, 画像, テキスト翻訳	利用シーン 講演等のオフライン字幕 講演・会議等の同時通訳, 字幕 多様な情報, モダリティの言語変換
システム スマートフォン アプリ	システム ウェアラブル型	システム ウェアラブル アンピエント
多言語化 4言語の音声翻訳	多言語化 10言語対の音声翻訳 音声翻訳クラウドサービス	多言語化 20×20言語対音声翻訳 未知言語即時適応能力
基本技術 フレーズベース翻訳 GMM-HMM 音声認識 DNN-HMM 音声認識基本方式	基本技術 DNNによる高精度版 横文構造利用翻訳 GPU 高速処理 文字認識などとの連携	基本技術 文脈, 意味解析 主語省略, 照応への対応 混合言語対応 クロスカルチャー対応 文字, 画像, 動作翻訳

研究開発については、技術的難易度がとても高く、また、利用可能なコーパスがほとんど存在していないため、本格的な研究開発はまだ緒についたばかりである。筆者らのグループでは、TED 英語講演に対して、上級、中級、初級のプロの同時通訳者による同時通訳コーパス

を現在収集、分析、モデル化の基礎研究を進めている⁽³⁰⁾。

表1に2020年及び、2020年以降の予測イメージを示す。現時点で実験システムレベルの技術が2020年頃に実用レベルになると思われ、一方で、現状において基礎

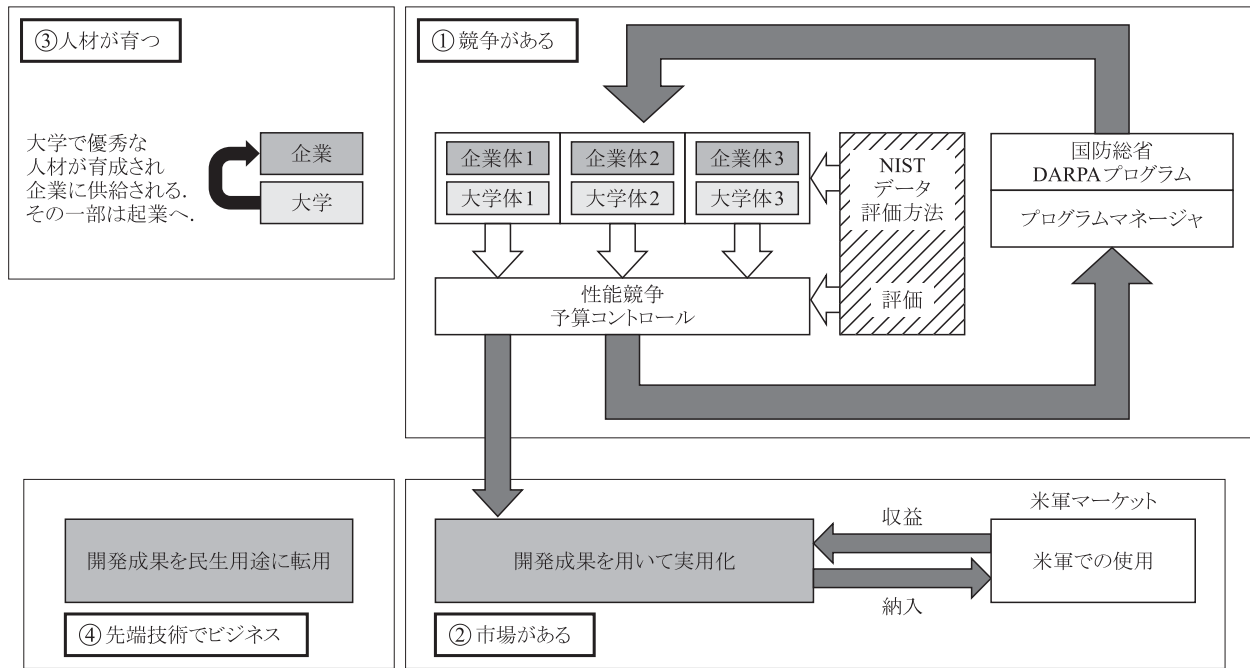


図7 DARPAの研究開発モデル

研究レベルの技術が実用に達するまでには5~10年以上の基盤研究と実証の期間を要すると考えられる。例えば、同時通訳、非言語情報の利用、意味解析、対話・談話構造の利用、日本語の照応などについては、更なるコーパス収集、研究開発の時間を要すると考えられる。もちろん、コーパス構築の枠組みを工夫して加速すれば所要の期間は短縮され得る。

6.2 研究開発のモデル

日本では、これまで、基礎研究の推進を目的とした科研費と、そのほかの省庁からの委託研究費、また、省庁から国立研究開発法人への運営交付金として研究費が投下されてきた。しかし、人材育成、成果の展開速度が米国などに比べて遅いという課題があった。米国の音声、言語分野の研究開発は、DARPAにより、同一テーマについて、企業と大学が複数のチームを作って、競わせる形式で進められてきた(図7)。コーパス構築や配布はLDC(Linguistic Data Consortium)や、評価はNIST(National Institute of Standards and Technology)などの標準化機関が行う。また、成果はプロダクトとして、プロジェクト参加企業から米軍が買い取ることで一定のマーケットと利益を提供するモデルになっている。これにより、性能競争が起こる、一定のビジネスが回る、大学で人材の育成ができ企業に供給あるいは起業を行う、軍用に開発された技術が民間に転用できる、といったサイクルが構築できる。日本でも競争的な仕組みを導入し、成果を地方自治体や国が買い取る仕組みがあれば(例えば、訪日、在住外国人用言語サービス)、米国

DARPAモデルと同様なスキームが構築可能で、研究開発に競争が起こり、実用化による初期マーケットの開拓、人材の育成と流動性の確保、成果展開への道が開ける可能性がある。

7. 終わりに

2020年の東京オリンピック・パラリンピックが決まり、増加する訪日外国人とのコミュニケーションツールとして音声翻訳技術に対する期待がますます高まっている。音声翻訳は、我が国が世界に先駆けて、国レベルのプロジェクトとして着手した技術であり、2020年は本技術の真価を世界に広める絶好の機会である。実用化には、技術開発だけでなく、利用者の観点からのインタフェースデザイン、場所や場面による適応化、他モダリティとの併用等による性能保証、データを蓄積して性能改善を持続的に行う枠組みの構築が不可欠である。更に、2020年以降の講演・映像コンテンツや会議の同時通訳技術を見据え、次世代の技術開発で常に世界をリードしていくロードマップ作りや研究開発の枠組み作りが重要である。

謝辞 奈良先端科学技術大学院大学情報科学研究科知能コミュニケーション研究室の皆様、日頃御協力頂く国立研究開発法人情報通信研究機構の皆様へ感謝する。

文献

- (1) 観光立国推進閣僚会議, “観光立国実現に向けたアクション・プログラム 2014,” 2014.

- (2) 総務省, 訪日外国人旅行者の受入れに関する意識調査結果, 2008, http://www.soumu.go.jp/menu_news/s-news/2008/pdf/081009_1_11.pdf
- (3) 中村 哲, “話し言葉の音声翻訳技術,” 信学誌, vol. 96, no. 11, pp. 865-873, Nov. 2013.
- (4) A. Waibel, “Translation systems under the C-STAR framework,” MT Summit VII, p. 121, 1999.
- (5) 中村 哲, 隅田英一郎, 清水 徹, S. Sakti, 坂井信輔, J. Zhang, A. Finch, 木村法幸, 葦荊 豊, “アジア言語音声翻訳コンソーシアム: A-STAR について,” 音響論集, no. 1-3-14, pp. 45-46, 2007.
- (6) ITU-T F. 745, “Functional requirements for network-based speech-to-speech translation,” 2010.
- (7) ITU-T H. 625 “Architecture for network-based speech-to-speech translation services,” 2010.
- (8) N. Bach, M. Eck, P. Charoenpornasawat, T. Köhler, S. Stüker, T. Nguyen, R. Hsiao, A. Waibel, S. Vogel, T. Schultz, and A.W. Black, “The CMU TransTac 2007 eyes-free and hands-free two-way speech-to-speech translation system,” Proceedings of International Workshop on Spoken Language Translation, IWSLT, 2007.
- (9) H. Soltan, G. Saon, B. Kingsbury, H. Kwang, J. Kuo, and L. Mangu, “Advances in arabic speech transcription at IBM under the DARPA GALE program,” IEEE Transactions on Audio, Speech, and Language Processing, vol. 17, no. 5, pp. 884-894, July 2009.
- (10) <http://www.eu-bridge.eu/index.php>, 及び, M. Freitag, S. Peitz, J. Wuebker, H. Ney, M. Huck, R. Sennrich, N. Durrani, M. Nadejde, P. Williams, P. Koehn, T. Herrmann, E. Cho, and A. Waibel, “EU-BRIDGE Mt: Combined machine translation,” Proceedings of the Ninth Workshop on Statistical Machine Translation, Association for Computational Linguistics, Baltimore, Maryland, USA, 2014.
- (11) 松田繁樹, 林 輝昭, 葦荊 豊, 志賀芳則, 柏岡秀紀, 安田圭志, 大熊英男, 内山将夫, 隅田英一郎, 河井 恒, 中村 哲, “多言語音声翻訳システム “VoiceTra” の構築と実運用による大規模実証実験,” 信学論(D), vol. J96-D, no. 10, pp. 2549-2561, Oct. 2013.
- (12) <http://voicetra-plus.jp/voicetra.html>
- (13) 総務省, “グローバルコミュニケーション計画,” April 2014, http://www.soumu.go.jp/main_content/000285578.pdf
- (14) 中村 哲, 磯谷亮輔, 乾 健太郎, 柏岡秀紀, 河井 恒, 河原達也, 木俣 豊, 黒橋禎夫, 隅田英一郎, 関根 聡, 鳥澤健太郎, 堀 智織, 松田茂樹, “Web 時代の音声・言語技術,” 信学誌, vol. 94, no. 6, pp. 502-517, June 2011.
- (15) J. Graehl and K. Knight, “Training tree transducers,” Proceedings of NAACL-HLT 2004, pp. 105-112, 2004.
- (16) N. Kalchbrenner and P. Blunsom, “Recurrent continuous translation models,” Proc. EMNLP, pp. 1700-1709, 2013, <http://workshop2014.iwslt.org/>
- (17) M. Cettolo, J. Niehues, S. Stüker, L. Bentivogli, and M. Federico, “Report on the 11th IWSLT evaluation campaign, IWSLT 2014,” Proceedings of International Workshop on Spoken Language Translation, IWSLT, pp. 2-17, 2014.
- (18) P. Kishore, S. Roukos, T. Ward, and W.-J. Zhu, “BLEU: A method for automatic evaluation of machine translation,” 40th Annual meeting of the Association for Computational Linguistics, pp. 311-318, 2002.
- (19) <http://www.ustar-consortium.com/index.html>
- (20) K. Ryu, S. Matsubara, and Y. Inagaki, “Simultaneous English-Japanese spoken language translation based on incremental dependency parsing and transfer,” Proceedings of the COLING Association for Computational Linguistics, pp. 683-690, 2006.
- (21) Y. Oda, G. Neubig, S. Sakti, T. Toda, and S. Nakamura, “Optimizing segmentation strategies for simultaneous speech translation,” The 52nd Annual Meeting of the Association for Computational Linguistics (ACL), pp. 551-556, Baltimore, USA, June 2014.
- (22) 三重野隆史, G. Neubig, S. Sakti, 戸田智基, 中村 哲, “同時音声翻訳における翻訳精度と遅延時間を同時に考慮した評価尺度,” 情処学 NL 研報, vol. 2014-NL-219, no. 3, pp. 1-5, Dec. 2014.
- (23) X. He, L. Deng, and A. Acero, “Why word error rate is not a good metric for speech recognizer training for the speech translation task,” Proceedings of ICASSP, pp. 5632-5635, 2011.
- (24) 大串正矢, G. Neubig, S. Sakti, 戸田智基, 中村 哲, “音声認識と機械翻訳のランク学習による同時最適化,” 言語処理学会第19回年次大会, pp. 564-567, 2013.
- (25) T. Kano, S. Sakti, S. Takamichi, G. Neubig, T. Toda, and S. Nakamura, “A method for translation of paralinguistic information,” Proceedings of IWSLT 2012, pp. 158-163, 2012.
- (26) M. Abe, S. Nakamura, K. Shikano, and H. Kuwabara, “Voice conversion through vector quantization,” Proceedings of ICASSP, pp. 655-658, 1988.
- (27) T. Toda, A.W. Black, and K. Tokuda, “Voice conversion based on maximum likelihood estimation of spectral parameter trajectory,” IEEE Transactions on Audio, Speech and Language Processing, vol. 15, no. 8, pp. 2222-2235, Nov. 2007.
- (28) S. Ogata, K. Murai, S. Nakamura, and S. Morishima, “Model-based lip synchronization with automatically translated synthetic voice toward a multi-modal translation system,” Proceedings of ICME 2001, pp. 28-31, 2001.
- (29) 中村 哲, “音声翻訳技術の現状と今後の展開,” 文部科学省科学技術政策研究所科学技術動向研究センター科学技術動向, no. 89, pp. 8-19, Aug. 2008.
- (30) H. Shimizu, G. Neubig, S. Sakti, T. Toda, and S. Nakamura, “Collection of a simultaneous translation corpus for comparative analysis,” Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14), pp. 670-673, May 2014.

(平成 27 年 3 月 16 日受付 平成 27 年 4 月 6 日最終受付)



なかむら たかし
中村 哲 (正員)

1981 京大工織大・工芸・電子卒。1992 京大博士(工学)。1981 シャープ株式会社中央研究所, 情報技術研究所。1986~1989(株)ATR 自動翻訳電話研究所所長。1994~2000 奈良先端大・情報科学研究科・助教授。2000(株)国際電気通信基礎技術研究所音声言語コミュニケーション研究所第一研究室長, 2005 所長, 取締役。2006 独立行政法人情報通信研究機構兼務。けいはんな研究所音声言語グループリーダー, 上席研究員, MASTAR プロジェクトリーダー, 知識創成コミュニケーション研究センター長, けいはんな研究所長などを経て, 現在, 奈良先端大・情報科学研究科・教授。ATR フェロー。ドイツカールスルーエ大客員教授。音声翻訳, 音声認識, 自然言語処理, マルチモーダル情報処理の研究に従事。日本音響学会粟屋奨励賞, 技術開発賞, 本会論文賞, 情報処理学会インタラクティブ 2001 ベストペーパー賞, 山下記念研究賞, 喜安記念業績賞, 人工知能学会研究会優秀賞, AAMT 長尾賞, 日本 ITU 協会国際協力賞, 電気通信普及財団テレコム技術賞, ドコモモバイルサイエンス賞, 総務大臣表彰, 文部科学大臣表彰, ELRA Antonio Zampoli 賞など各受賞。人工知能学会, 日本音響学会, 情報処理学会, IEEE Senior Member, IEEE SLTC 委員(2001~2003, 2013~), Signal Processing Magazine 編集委員(2011~2014), International Speech Communication Association (ISCA) 理事。