

グローバルコミュニケーション計画の推進

-多言語音声翻訳技術の研究開発及び社会実証-

(I. 多言語音声翻訳技術の研究開発)

基本計画書

1. 目的

2013年の訪日外客数は、日本政府観光局が1964年に統計を開始して以来、初めて1,000万人を突破した。2014年はさらに多くの外国人が日本を訪れ、1,300万人超えを記録し、今後も日本を訪れる外国人は増加していくものと予想される。

一方、訪日外国人と日本人の間には依然として「言葉の壁」が立ちはだかつており、コミュニケーション不足による様々な問題や機会の損失が多数存在している。

そこで本研究開発及び社会実証では、「言葉の壁」をなくし、自由でグローバルなコミュニケーションを実現するため、多言語音声翻訳技術を高度化し、かつ翻訳サービスを提供する企業等が共通して利用可能な多言語音声翻訳プラットフォームを構築し、実社会に実装することを目指す。

2. 政策的位置付け

「グローバルコミュニケーション計画 ～多言語音声翻訳システムの社会実装～」(平成26年4月11日 総務大臣発表)においては、世界の「言葉の壁」をなくしグローバルで自由な交流を実現するため、情報通信研究機構(以下、NICTという。)が開発した多言語音声翻訳技術を高度化し、社会実装を推進していくこととしている。

「観光立国実現に向けたアクション・プログラム2014」(平成26年6月17日 観光立国推進閣僚会議)においては、2020年までに訪日外国人旅行者数2000万人にすることを目標に掲げるとともに、「外国人旅行者の受入環境整備」として「多言語対応の改善・強化<多言語アプリの活用>」、として、「豊富な観光情報や地図情報等を備えた多言語対応観光アプリの活用により、外国人旅行者のスムーズな情報取得を促進するとともに、総務省「グローバルコミュニケーション計画」に基づいて多言語通訳・翻訳アプリ技術の研究開発の強化等を行い、精度向上を図ることにより、様々な地域・場面での多言語対応への活用を促進する」とされている。

「日本再興戦略 改訂2014-未来への挑戦-」(平成26年6月24日 閣議決定)においては、「世界に通用する魅力ある観光地域づくり、外国人旅行者の受入環境整備及び国際会議等(MICE)の誘致・開催の促進と外国人ビジネス客の取り込み」として、「全

国各地で多言語対応を改善・強化する」と掲げている。

また、「世界最先端 IT 国家創造宣言」（平成 26 年 6 月 24 日 改定 閣議決定）では、「東京オリンピック・パラリンピック等の機会を捉えた最先端の IT 利活用による「おもてなし」の発信」として、「言葉の壁をなくす多言語音声翻訳システムの高度化（中略）など、安全・安心の確保を図りつつ、最先端の IT 利活用による「おもてなし」を提供し、広く世界に発信することにより、IT 利活用の裾野を拡大するとともに、産業競争力の強化を図る」としている。

総合科学技術・イノベーション会議が取りまとめた「科学技術イノベーション総合戦略 2014 ～未来創造に向けたイノベーションの懸け橋～」(平成 26 年 6 月 24 日 改定 閣議決定)において、「科学技術イノベーションが取り組むべき課題」として「個人が言語や文化の壁を超えるための多言語音声認識や翻訳技術」が明記されている。

3. 目 標

(1) 政策目標（アウトカム目標）

前述のとおり、政府は訪日外国人の増加とそれによる対応に向けて多言語音声翻訳システムの研究開発及び社会実装に取り組むこととしており、都市のインフラとして多言語サポート整備の必要性が急速に高まっている。特に災害が発生した時や病気になった時などでも、日本語を話すことができない外国人の安心・安全を確保するためには意思疎通が不可欠であるが、多国籍の訪日外国人へ通訳・翻訳サービスを人手で提供するのには限界がある。

このため、多言語音声翻訳技術を用いた翻訳サービスが病院、ショッピングセンター、観光地、公共交通機関等の生活拠点に導入され、日本語を理解できない外国人でも日本国内で「言葉の壁」を感じることなく、我が国の生活に必要なサービスを利用できるようにすることを政策目標とする。

(2) 研究開発目標（アウトプット目標）

NICTが開発した多言語音声翻訳システムは、日英中韓の4か国語間の短い旅行会話の翻訳を比較的精度よく実現しており、音声認識技術は世界トップクラスの評価を得ている。他方で、このシステムを実際の社会において実用レベルで使うためには、雑音抑圧技術、位置情報を活用した翻訳精度向上技術、翻訳自動学習技術及び特殊文字認識技術のそれぞれについて、実際の社会の複数の場面における十分な内容の実証実験を通じて得られた知見をフィードバックしながら研究開発に取り組む必要がある。

この研究開発の取り組みによって、実際の社会で利用可能であり、かつ翻訳サービスを提供する企業が共通して利用可能な多言語音声翻訳プラットフォームを構築することを研究開発目標とする。

4. 研究開発内容

以下の技術に係る研究開発を実施するとともに、それらの技術が実装され、翻訳サービスを提供する企業等が共通して利用することを想定した、多言語音声翻訳プラットフォームを構築すること。その際、(a)～(d)の各技術及び多言語音声翻訳プラットフォームを実現するために必要な技術実証を実際の社会の複数の場面で実施し、得られた知見を研究開発にフィードバックすることにより、実際の社会に広く利用できる多言語音声翻訳プラットフォームとなるように取り組むこと。

(a) 雑音抑圧技術

① 概要

多言語音声翻訳技術が、医療やショッピング、タクシー等様々なアプリケーションに適用され、観光地、公共交通機関等、屋内外においても様々な利用がなされることを鑑みると、発話者が発する音声に雑音がまぎれた状態でマイクに入力され、その影響で音声認識精度が劣化し、更には翻訳精度の劣化が引き起こされるケースが発生することが考えられる。

その悪影響を軽減するため、発話者がマイクに入力した音声情報から雑音成分を抑圧し、発話者の音声をより明瞭化した上で、当該音声情報を多言語音声翻訳機能部に入力する雑音抑圧技術の研究開発を実施する。多言語音声翻訳システムの利用が想定される実際の社会環境において技術実証を実施し、得られた知見を研究開発にフィードバックする。

② 技術課題

雑音抑圧技術は環境（例：雑音の種類、大きさ、音声入力デバイス、認識エンジンの音響モデル）に大きく依存する。たとえば高雑音環境下では、雑音抑圧を行わなければ精度高く認識できないが、一方で雑音抑圧量を上げ過ぎると利用者の音声自体も抑圧してしまい、認識結果の誤りを増加させる場合もある。このため、利用環境によって適切な動作パラメータを設定する必要がある。この問題に対して、以下の2つの研究開発に取り組むことで解決を行う。

ア) 音声入力デバイスにおける雑音抑圧のための集音技術

音声入力デバイスに内蔵した複数のマイクにより発話者の音声を指向性集音することで雑音下での認識精度の改善が期待されると考えられるが、これを実現するためには指向性の向きと範囲を最適化する必要がある。

利用者の音声自体を抑圧せずに、十分な雑音抑圧を達成するためには、発話・騒音を識別して分離する音源分離技術、及び発話者の位置と周囲の騒音の位置・音量に基づき、指向性の向きと範囲を制御する適応ビームフォーム技術を、利用者の状況に応じてリアルタイム処理する必要があると考えられるが、現状は音声入力デバイス毎にチューニングした固定方向・固定範囲の指向性となっている。発話者の位置と周囲騒音の位置・音量の変化に追従して雑音抑圧を実現可能とす

る適応集音技術を確立する。

イ) 多様な環境における雑音抑圧最適化技術

音声認識に対する雑音抑圧の効果は利用環境（周囲雑音の種類や大きさ、音声入力デバイス、音声認識エンジンの音響モデル、等）に大きく依存するため、雑音抑圧を実行する各機能部の処理方式の選択や動作パラメータの設定を利用環境に応じて適切に行う必要がある。

従来、このような最適化を行うに当たっては、専門家がそれぞれの利用環境に合わせて経験的なチューニングを行うことで対応をしてきたが、その作業に大きなコストがかかるため、多言語音声翻訳技術を広く社会実装する上での阻害要因となり得る。

このような課題を解決するため、多様な環境における雑音抑圧技術の適応を低コストで実施可能とする雑音抑圧最適化技術を確立する。

③ 到達目標

ア) 音声入力デバイスにおける雑音抑圧のための集音技術

利用シーンとして多様な騒音が存在する場面を想定し、音声認識システムに対して S/N 比=0dB の雑音環境下でも S/N 比=20dB の場合と同等の音声認識精度を得て、例として以下の内容が実現可能な雑音抑圧技術を開発する。

- ・ 交差点で道案内ができる。チケット売り場で販売ができる。
- ・ 鉄道、バスの中で会話ができる。
- ・ タクシーの車内で走行中の会話ができる。
- ・ 道路に面した販売店で会話ができる。
- ・ 百貨店の売り場で会話ができる。

イ) 多様な環境における雑音抑圧最適化技術

従来、専門家が数名で取組み、一つの機器に対して延べ 150~300 人時以上必要としていた雑音抑圧技術を適応するためのコストを、複数の利用シーン（例：病院、ショッピングセンター、観光地、公共交通機関等）に対して、それぞれ 75 人時以内で完了する雑音抑圧最適化技術を実現する。

(b) 翻訳自動学習技術（自動翻訳チューニング技術）

① 概要

対訳コーパスに基づく自動翻訳技術が進展したが、専門分野向け高精度システムにおいても、モデルの不完全性等に起因して一定量の翻訳の誤りは避けられない現状である。自動翻訳技術は仮名漢字変換システムのように自動的にチューニングされエラーが削減されることはなく、現状ではログのオフラインでの分析とそれに基づく改良が必要であり迅速には成し得ないという点に限界がある。

こうした状況において、本研究開発においては、誤訳と正訳の識別技術、及び、誤訳を正訳に変える技術の研究開発を、多言語音声翻訳システムの利用が想定される実際の社会環境において技術実証を実施し、得られた知見を研究開発にフィードバック

する。

② 技術課題

ア) 自動翻訳結果に信頼度を付与する技術

複数の機械翻訳結果の出力や翻訳の信頼度に関する自動学習機能を活用することで、自動翻訳結果に信頼度を付与する技術を確立する。また翻訳結果に対してユーザが評価をつけるなどの外部入力を用いて信頼度を付与する技術を創出する。

イ) 自動翻訳結果の信頼度を利用してシステムを改良する技術

入力された翻訳対象の文章と、それに対する信頼度の高い翻訳結果の集合からモデルを生成し、最初に作られた翻訳モデルと混合し、自動的に改良する技術を確立する。

翻訳結果の信頼度の低い文を選択し、人手翻訳することで、翻訳量を削減する。

ウ) フォールバック機能を利用したシステムの改良に関する技術

翻訳の信頼性が低い時に、ユーザが人手翻訳を呼び出す機能（フォールバック）を実現することで、フォールバックによって得られる翻訳と原文の対を機械翻訳の改良に自動的に利用できるような技術を開発する。

エ) 複数フィードバックを利用したシステムの改良に関する技術

同じ文章について複数個の自動翻訳結果を生成し、どの翻訳結果のどの部分を集めれば正しい意味の翻訳となるかに関して、自動改良する技術を開発する。また、複数の評価者の評価結果の利活用も検討する。

③ 到達目標

初期の翻訳率が70%の場合で、10%改善して80%とすることを旨とする。

(c) 特殊文字認識技術

① 概要

訪日外国人が本邦で買い物、飲食、観光、交通機関による移動等を行う場合、日本語で記載されているメニューや説明等に接することになるが、必ずしも全ての店舗、施設等で多言語対応が出来ているわけではないことや、英語、中国語及び韓国語以外の言語については事前の準備も困難であることを鑑み、メニューや説明等の画像をカメラ、スマートフォン等により取得し、画像を認識して文字に置き換え、その上で多言語音声翻訳を行う技術の開発を、多言語音声翻訳システムの利用が想定される実際の社会環境において技術実証を実施し、得られた知見を研究開発にフィードバックする。

② 技術課題

ア) 文字認識技術（画像処理、文字位置検出、特殊フォント文字認識）

文字傾き補正、台形補正、湾曲及びカメラ映像品質補正機能等を用いて、商品

パッケージや瓶等に印刷された湾曲等された文字を読み取るための画像処理技術
を確立する。

また、画像の中でどの位置に文字があるかを検出する技術およびメニューで使
われるような特殊なフォント文字に対応する文字認識技術を確立する。

イ) 言語処理による補正技術

ユーパス（対訳）を利用した言語処理による補正で、文字認識精度を向上する
技術を確立する。

③ 到達目標

ア) 文字認識技術（画像処理、文字位置検出、特殊フォント文字認識）

レストランにおいてメニューの文字が認識できる、買い物では店先看板の文字
を認識できる、フロアガイドの文字を認識できる及び食品成分表の文字を認識で
きる、観光においては、案内板の文字を認識できる、交通等の場面においては案
内板の文字を認識できる、行き先案内の電光掲示板の文字を認識できる。

イ) 言語処理による補正技術

ア) において想定される各場面において、言語処理による補正を行うことによ
り、文字認識率（対象文字に対する正解文字の割合）90%以上を目標とする。

(d) 位置情報を活用した翻訳精度向上技術

① 概要

訪日外国人が日本国内で観光等のために交通機関を利用して移動する場合、タクシ
ー・電車・バスの利用が想定される。利用者が言語の障壁なく移動・観光・ショッピ
ング・支払が行える環境を提供するため、多言語音声・文字入出力環境の位置情報の
活用技術の研究開発を、多言語音声翻訳システムの利用が想定される実際の社会環境
において技術実証を実施し、得られた知見を研究開発にフィードバックする。

② 技術課題

ア) ユーザ利用機器と交通機関に搭載された機器との連携技術

ユーザ保有の通信機器と交通機関等に搭載された機器（例：タクシーの料金端
末やカーナビ等）との連携処理をすることで、翻訳結果や翻訳に係るやりとりの
利便性向上を図る。例えば、音声の入力でユーザのスマートフォンを利用した際、
結果の表示にはドライバーも分かるようにタクシーの画面に表示することや、電
車やバスであっても音声による入出力等は手元のデバイスを利用するが、結果の
表示は交通機関側の画面であること等の連携技術を確立する。

イ) ユーザの移動・位置情報を活用した多言語翻訳精度向上技術

翻訳が必要とされる単語については、利用者の位置や旅行計画内容等に大きく
依存するため、翻訳が必要とされる対訳のデータベースは、利用者の位置情報や
旅行計画情報を参照することで翻訳精度を向上させることが可能と考えられるた
め、ユーザの移動・位置情報を活用した多言語翻訳精度向上技術を開発する。

③ 到達目標

タクシー等において日本語を理解できない乗客と運転手の会話のうち、行き先の聞き取り、料金收受など輸送サービスに必須な会話について、言い直しを2回まで許容した場合において、話者の用務達成率80%を目標とし、経路周辺の観光スポットの案内等付随的な会話については、同等の条件下で用務達成率60%を目標とする。

なお、用務達成率の評価にあたっては、現状の多言語音声翻訳技術のみを使用した場合と、今回研究開発を実施する「位置情報を活用した翻訳精度向上技術」を組み合わせて使用した場合について、用務達成率、ユーザエクスペリエンス、運用コストなども含めて多面的に比較評価を行うこと。

5. 研究開発期間

平成27年度から平成31年度までの 5年間

6. その他 特記事項

(1) 特記事項

- ① 提案者は、下記課題 (a)、(b)、(c)、及び (d) のいずれか又は複数の課題に提案することができる。
 - (a) 雑音抑圧技術
 - ア) 音声入力デバイスにおける雑音抑圧のための集音技術
 - イ) 多様な環境における雑音抑圧最適化技術
 - (b) 翻訳自動学習技術 (自動翻訳チューニング技術)
 - ア) 自動翻訳結果に信頼度を付与する技術
 - イ) 自動翻訳結果の信頼度を利用してシステムを改良する技術
 - ウ) フォールバック機能を利用したシステムの改良に関する技術
 - エ) 複数フィードバックを利用したシステムの改良に関する技術
 - (c) 特殊文字認識技術
 - ア) 文字認識技術 (画像処理、文字位置検出、特殊フォント文字認識)
 - イ) 言語処理による補正技術
 - (d) 位置情報を活用した翻訳精度向上技術
 - ア) ユーザ利用機器と交通機関等に搭載された機器との連携技術
 - イ) ユーザの移動・位置情報を活用した多言語翻訳精度向上技術
- ② 課題 (a) の代表研究者は課題 (a) ~ (d) の取りまとめを行うこととし、提案にあたっては3. (2) に記載された「多言語音声翻訳プラットフォーム」の構築が全体の研究開発目標となっていることに留意すること。

(2) 提案および研究開発に当たっての留意点

- ① 提案に当たっては、基本計画書に記されているアウトプット目標に対する達成度を評価することが可能な評価項目を設定し、各評価項目に対して可能な限り数値目標を定めるとともに、目標を達成するための研究方法、実用的な成果を導出するための共同研究体制又は研究協力体制、及び達成度を客観的に評価するための実験方法について、具体的に提案書に記載すること。
- ② アウトカム目標の達成に向けた適切な研究成果の取扱方策（研究開発課題の分野の特性を踏まえたオープン・クローズ戦略を含む）について提案書に記載すること。
- ③ 本研究開発成果を確実に展開し、アウトカム目標を達成するため、事業化目標年度、事業化に至るまでの実効的な取組計画（標準化活動、体制、資金等）についても具体的に提案書に記載すること。
- ④ 複数機関による共同研究を提案する際には、研究開発全体を整合的かつ一体的に行えるよう参加機関の役割分担を明確にし、研究開発期間を通じて継続的に連携するための方法について具体的に提案書に記載すること。
- ⑤ 研究開発の実施に当たっては、関連する要素技術間の調整、成果の取りまとめ方等、研究開発全体の方針について幅広い観点から助言を頂くと共に、実際の研究開発の進め方について適宜指導を頂くため、学識経験者、有識者等を含んだ研究開発運営委員会等を開催する等、外部の学識経験者、有識者等を参画させること。
- ⑥ 本研究開発は総務省施策の一環として取り組むものであることから、総務省が受託者に対して指示する、研究開発に関する情報及び研究開発成果の開示、関係研究開発プロジェクトとのミーティングへの出席、シンポジウム等での研究発表、共同実証実験への参加等に可能な限り応じること。
- ⑦ 2. に記載のとおり、「NICTが開発した多言語音声翻訳技術を高度化し、社会実装を推進していく」ために、提案書の内容作成、研究開発及び社会実証の実施にあたっては、NICTが開発した多言語音声翻訳技術との互換性が確保されていること。

(3) 人材の確保・育成への配慮

- ① 研究開発によって十分な成果が創出されるためには、優れた人材の確保が必要である。このため、本研究開発の実施に際し、人事、施設、予算等のあらゆる面で、優れた人材が確保される環境整備に関して具体的に提案書に記載すること。
- ② 若手の人材育成の観点から行う部外研究員受け入れや招へい制度、インターンシップ制度等による人員の活用を推奨する。これらの取組予定の有無や計画について提案書に記載すること。

(4) 研究開発成果の情報発信

- ① 本研究開発で確立した技術の普及啓発活動を実施すると共に、その活動計画・方策については具体的に提案書に記載すること。
- ② 研究開発成果については、原則として、総務省としてインターネット等により発

信を行うとともに、マスコミを通じた研究開発成果の発表、講演会での発表等により、広く一般国民へ研究開発成果を分かりやすく伝える予定であることから、当該提案書には、研究成果に関する分かりやすい説明資料や図表等の素材、英訳文書等を作成し、研究成果報告書の一部として報告する旨の活動が含まれていること。さらに、総務省が別途指定する成果発表会等の場において研究開発の進捗状況や成果について説明等を行う旨を提案書に記載すること。

- ③ 本研究開発終了後に成果を論文発表、プレス発表、製品化、ウェブサイト掲載等を行う際には「本技術は、総務省の『グローバルコミュニケーション計画の推進－多言語音声翻訳技術の研究開発及び社会実証-』による委託を受けて実施した研究開発による成果です。」という内容の注記を発表資料等に付すこと。