

# 多言語同時通訳と音声マルチスポット再生技術



国立研究開発法人情報通信研究機構 ユニバーサルコミュニケーション研究所  
先進的音声翻訳研究開発推進センター 先進的音声技術研究室 研究マネージャー

おかもと たくま  
岡本 拓磨

## 1. 多言語同時通訳技術

NICT（国立研究開発法人情報通信研究機構）では、言語の壁を越えた音声コミュニケーション実現のため、音声翻訳アプリVoiceTra<sup>[1]</sup>をはじめとする多言語翻訳技術の研究開発を行っている。VoiceTraでは1発話ごとに音声認識（入力音声をテキストへと変換）・機械翻訳（入力テキストを別の言語へ自動翻訳）・音声合成（入力テキストを音声波形へと変換）を行う逐次型音声翻訳であるが、近年では人間の同時通訳者の振る舞いと同様に、入力音声の終端を待つことなく連続的に音声認識・機械翻訳・音声合成が可能な多言語音声同時通訳技術の研究開発を行っている<sup>[2]</sup>。これらの成果を社会実装し、2025年の大阪・関西万博をはじめ、数多くの場所で使われている。NICTにおける多言語同時通訳技術については文献<sup>[3]</sup>において詳しく解説しているため、本稿では音声マルチスポット再生技術に焦点を当てて解説する。

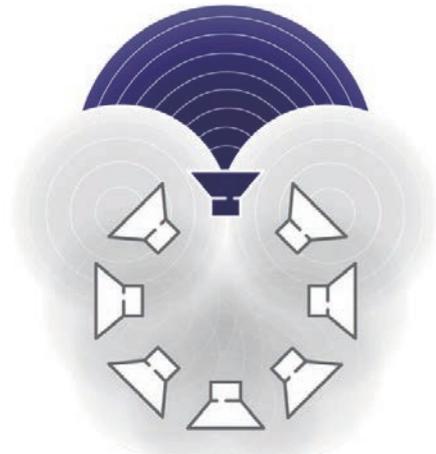
## 2. 音声マルチスポット再生技術

通常、音は全方向に広がるため、これらの合成音声を通常のスピーカで同時再生（単純再生）すると、図1（a）のように音が混ざり合ってしまう、聴取者は目的音を聞き取ることが困難になる。また、駅や空港で多言語対応を行う場合と同様に、日本語→英語→中国語のように順番にアナウンスすると、それらの音が混ざることはいないが、言語の数だけ再生時間が必要となり、多言語同時通訳を実現するメリットが薄れてしまう。音声マルチスポット再生技術は、これらの問題を解決し、多言語で同時通訳された合成音声を、その言語を必要とする人々に同時に提供することができる技術である。



■図1. 単純再生とマルチスポット再生の比較

音声マルチスポット再生技術は、多数のスピーカを用いて、音が聞こえる空間と、その音が聞こえない（または別の音が聞こえる）空間を創出することを可能にする<sup>[4]</sup>。音声マルチスポット再生技術の原理（ここでは端的な紹介にとどめており、正確なものではない。）としては、ノイズキャンセルの原理と同様に、目的方向以外の方向に広がった音を、別のスピーカから再生された音で打ち消すことにより、目的方向にのみ目的音を届ける「局所再生」が可能となり（図2）、この局所再生を方向ごとに重ね合わせることで、異なる方向に異なる音を届ける「マルチスポット再生」を実現している（図1（b））。



■図2. 多数のスピーカを用いた局所再生の原理

音声マルチスポット再生技術と多言語同時通訳技術を融合させることで、日本語話者の発話が即座に翻訳され、英語、中国語、韓国語等の合成音声を別々の方向に同時に届けることが可能となる。言い換えれば「音を時間ではなく空間で分ける」ことが可能であり、これにより、複数の音声が混ざり合って聞き取りづらくなる問題やそれらを順番にアナウンスした際に時間的なロスが生じる問題を解決することができる。また、本技術は多言語対応の用途のみならず、観光施設やエンターテインメント施設、美術館等での同時解説音提示（図3）、車内における活用等も期待できる。さらに、2025年3月には世界防災フォーラムに出展し、緊急時における多言語での避難誘導への応用も期待されている。

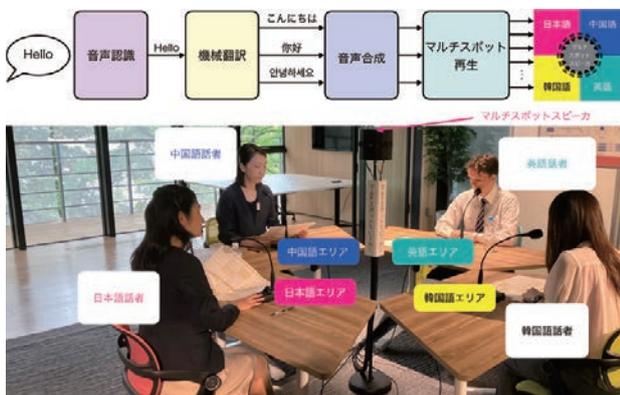


■ 図3. 美術館での応用例

### 3. 社会実装に向けた取組み

音声マルチスポット再生技術の技術紹介を行うホームページ<sup>[5]</sup>や動画<sup>[6]</sup><sup>[7]</sup>を公開し、社会実装を目指すに当たってパートナーとなる企業の探索を行っていることが挙げられる。特に、YouTubeでの技術紹介動画『NICTステーション～音声マルチスポット再生技術～』（NA上白石萌音）<sup>[6]</sup>は、公開から1年1か月の2026年1月に100万回再生を実現している。また、スーツケース1つで持ち運び可能な16チャンネル円形スピーカを用いて、多言語同時通訳と音声マルチスポット再生をノートPC1つで動作可能なデモシステムを実装した<sup>[8]</sup>。このデモシステムを用いて、大阪関西万博、CEATEC及びMobile World Congress (MWC) @バルセロナ等の大規模展示会におけるデモ展示や水族館や博物館等における実証実験を実施しており、本技術の有効性の確認及びフィードバックのヒアリングを行っていることが挙げられる。以下では、その代表例について紹介する。

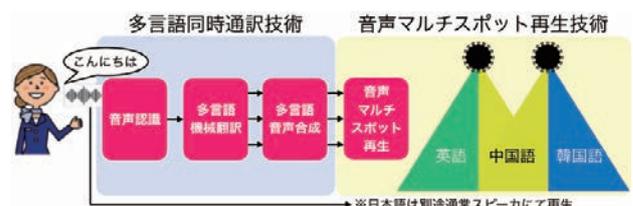
2023年6月のNICTオープンハウス2023では、本技術と多言語同時通訳技術を組み合わせた日英中韓4言語会議システムのデモ展示を行った（図4）。これは、日本語話者の



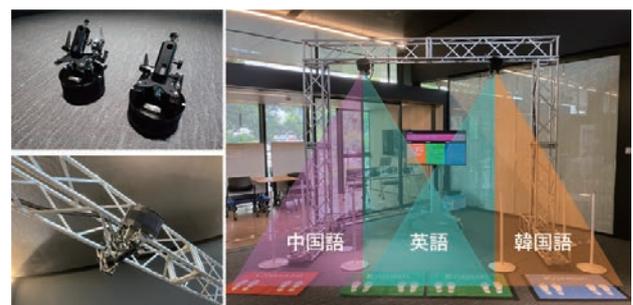
■ 図4. 音声マルチスポット再生技術と多言語同時通訳技術を組み合わせた日英中韓4言語会議システム（2023年6月 NICTオープンハウス2023）

発話が英語、中国語、韓国語に同時通訳され、それらの合成音声が多言語マルチスポット再生により各言語の話者に届くというデモ展示である。これにより、各言語の話者は自身の母語で話し、通訳された母語を聞くことにより異なる母語の話者と会議を行うことが可能となることを提示した。

2025年1月には、NTTデータ カスタマサービス株式会社により、本技術を用いた実証実験（総務省委託）が大阪市の水族館「海遊館」で実施された<sup>[9]</sup>。本実証実験では、解説員が水生動物の様子を日本語で解説し、同時通訳された英語、中国語、韓国語の合成音声を、音声マルチスポット再生技術によって異なる方向に提供された。体験者からは、「ほかの言語の音声は混ざることなくはっきりと聞こえる」といった声が多数あり、本技術の有効性を確認することができた（図5）。なお、実験現場は水生動物がいる水槽の上部（バックヤード）であるため、指向性超音波スピーカや多言語対応のための解説用ヘッドセット等が使えないという課題があったが、本技術によりそれらが解決できることが示された。また、従来、円形スピーカを用いたマルチスポット再生においては、図1 (b) に示されるように聴取者がスピーカの周りを歩き回って音を聞く「円形分割再生方式」としていたが、本実証実験では、図6のとおり円形スピーカを壁面に縦方向に取り付けることにより、聴取者は頭上から降り注ぐ音を聞き、左右に移動することによりマルチスポット再生を体験できる「直線エリア分割型再生方式」を採用した。「円形分割再生方式」は、考えられる本技術の



■ 図5. 多言語同時通訳と音声マルチスポット再生技術を用いた実証実験（2025年1月 海遊館での実証実験）

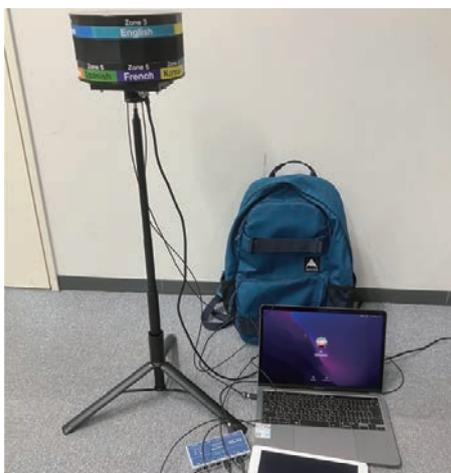


■ 図6. 円形スピーカを用いた直線移動型マルチスポット再生デモシステム（2025年6月 NICTオープンハウス2025）

応用先候補が限定的であったが、「直線エリア分割型再生方式」は、本技術の応用先候補を広げることにつながっており、画期的な再生方式であると言える。

2025年4月には、多言語同時通訳も含めてノートPC1台で動作可能であり、かつリュックサック1つで持ち運び可能なコンパクトデモシステムの試作に成功し<sup>[10]</sup>、これまで以上にどこにでもデモシステムを持ち運べるようになった(図7)。2025年8月には、このコンパクトデモシステムを音声分野のフラッグシップ国際会議Interspeech 2025におけるSpeech Science Festival<sup>[11]</sup>へ出展(招待)し、国際的にもアピールすることに成功した。

2025年9月16日～22日には、大阪・関西万博におけるFuture Life Experienceにて直線エリア分割型再生方式の音声マルチスポット再生技術を出展し、一万人を超える数多くの来場者に音声マルチスポット再生技術を体験いただくことができた(図8)。そして、これらの成果が認められ、CEATEC



■図7. リュックサック1つで持ち運び可能なコンパクトデモシステム



■図8. 大阪関西万博2025 Future Life Experienceにおけるデモ展示

AWARD 2025イノベーション部門賞を受賞した<sup>[12]</sup>。CEATECでは2023年から3年連続で多言語同時通訳と音声マルチスポット再生技術を用いたデモシステムを展示した。また、2025年12月には多言語同時通訳と音声マルチスポット再生技術を用いた実証実験を日本科学未来館にて実施した<sup>[13]</sup>。さらに、2026年3月2日～5日にはMWC26のNICTブースにおいて、言語同時通訳と音声マルチスポット再生技術を用いたデモ展示を実施する。

現在はこのコンパクトデモシステムを用いて様々な場所での実証実験を行い、多言語同時通訳と音声マルチスポット再生技術の応用先候補を検討するとともに、更なる高精度な再生方式の研究開発にも取り組んでいる。

## 謝辞

多言語同時通訳・音声マルチスポット再生技術の研究開発と社会展開はNICTユニバーサルコミュニケーション研究所及びイノベーションデザインイニシアティブ共創デザインプロジェクトとして行われた。ここに感謝の意を表する。

## 参考文献

- [1] <https://voicetra.nict.go.jp>
- [2] <https://youtu.be/uyTRd5Hu6hw>
- [3] PAUL Michael, 今村賢治, 王曉林, 東山翔平, 内山将夫, 藤本雅清, 岡本拓磨, 菊池武文, 塩飽裕彦, 香山健太郎, 「言葉の壁」から解放された万博体験の実現, 情報通信研究機構研究報告, vol. 71, no. 1, pp. 53-70, Dec. 2025.
- [4] 岡本拓磨, 「スピーカアレイを用いたマルチスポット再生技術の理論と実装」, 音響誌, vol. 81, no. 10, pp. 711-718, Oct. 2025.
- [5] <https://ast-astrec.nict.go.jp/MultipleSoundSpotSynthesis/>
- [6] <https://youtu.be/fTyYs6AqtNM>
- [7] <https://youtu.be/G1qkR-B40PM>
- [8] [https://www.jst.go.jp/pr/jst-news/backnumber/2024/202407/pdf/2024\\_07\\_p12-13.pdf](https://www.jst.go.jp/pr/jst-news/backnumber/2024/202407/pdf/2024_07_p12-13.pdf)
- [9] <https://www.nict.go.jp/publicity/topics/2025/01/15-1.html>
- [10] 76. T. Okamoto and M. Kono, "Simultaneous speech translation integrated compact multiple sound spot synthesis system on a laptop carried out with a backpack", in Proc. Interspeech, Aug. 2025, pp. 3539-3540.
- [11] <https://www.interspeech2025.org/science-fest>
- [12] <https://www.nict.go.jp/publicity/topics/2025/10/10-1.html>
- [13] <https://www.nict.go.jp/publicity/topics/2025/12/08-1.html>