

ウェアラブル型音声翻訳端末向け雑音抑圧技術

Noise Reduction Technology for Wearable Speech Translation Device

石川 智一*
Tomokazu Ishikawa

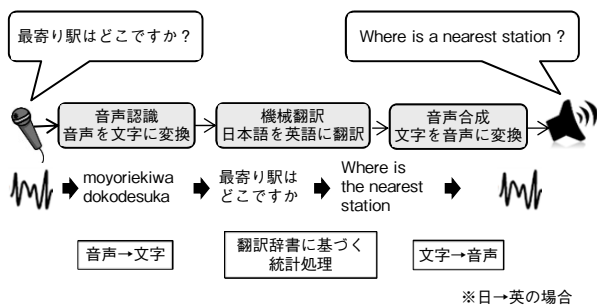
星見 昌克*
Masakatsu Hoshimi

近年、訪日外国人が急激に増加するなかで、言語の壁が社会課題となってきた。本稿では、言語の壁を解決する技術の1つとして期待されている音声翻訳技術の概要と、ウェアラブル型音声翻訳端末の実現に向けた雑音抑圧の技術課題について解説する。

In recent years the number of foreign tourists has been increasing rapidly in Japan, and hence language barriers are becoming a social problem. Regarded as one of the techniques to overcome such language barriers, in this paper we will briefly introduce speech translation technology, and explain in detail the problem of the noise reduction technology to be solved for our wearable speech translation device.

1. 音声翻訳技術の概要

音声翻訳技術は、人の話した言葉を別の言語に翻訳し、その言語の音声で再生する技術であり、第1図に示すように、音声認識技術、機械翻訳技術、音声合成技術から構成されている。



第1図 音声翻訳技術の構成図

Fig. 1 Overview of speech translation technology

当社は、第2図に示すように、小売業界や鉄道業界、ホテル業界など、外国語対応が必要になるさまざまなビジネスシーンで、音声翻訳技術の実用化を目指している。

例えば、百貨店やコンビニエンスストアのような小売業界では、急増している訪日外国人観光客による大量の商品購入に伴い、外国語対応ができる店員の不足が課題となっている。その解決手段の1つとして、音声翻訳技術が期待されている。

スマートフォンやタブレットなどの情報端末でも音声翻訳アプリが出てきており、これを使うことで音声翻訳



第2図 外国語対応の代表的シーン

Fig. 2 Typical examples of situation where foreign language is required

を行うことは可能である。しかし、音声翻訳アプリを情報端末で利用する際には情報端末を手にとって操作することや、会話のたびにマイクを話者に向けなおすことなどが必要であり、小売店などでの対面接客時に手がふさがるため、業務効率が低下するという問題がある。そこで、筆者らは、接客時にハンズフリーで使えるウェアラブル型音声翻訳端末を開発することを目指している。

2. ウェアラブル型音声翻訳端末向け雑音抑圧技術

2.1 プロトタイプによる課題抽出

ウェアラブル型音声翻訳端末を開発するに当たり、技術課題を明確にするため、第3図のプロトタイプのウェアラブル型音声翻訳端末を試作し、筆者らの想定顧客の1つである小売業界に対し、期待する端末サイズなどに関してヒアリングを実施した。

その結果、

1) 4インチ程度の小型スマートフォンよりも小型であ

* AVC ネットワークス社 技術本部
R&D Div., AVC Networks Company

ること

2) 店員が首からぶら下げられるほど軽量であること
3) 接客側と来客側の間で双方向に翻訳ができること
4) 店舗内の雑音が多い状況下で確実に使えること
などのコメントが得られた。サイズ、重量、双方向での翻訳に関してはおおむね肯定的であった。一方、雑音が多い状況下では、音声認識誤りが多くなり、十分な評価が得られなかった。

店舗は、さまざまな部屋の音響特性があり、またその内部は、音量の大小や時間変化が異なる雑音が存在するさまざまな環境のため、確実に使えるためには、周囲の雑音を抑圧し、接客側と来客側の音声のみを收音する技術の開発が必要である。



スピーカー

マイクアレイ

サイズ [mm]	W: 94 × H: 56 × D: 14
重量 [g]	64
通信方式	Bluetooth (注1) (HFP1.6+SPP)

HFP : Hands Free Profile 1.6
SPP : Serial Port Profile

第3図 ウェアラブル型音声翻訳端末の外観・仕様

Fig. 3 Appearance and specifications of wearable speech translation device

2.2 ウェアラブル型音声翻訳端末向け雑音抑圧の技術課題

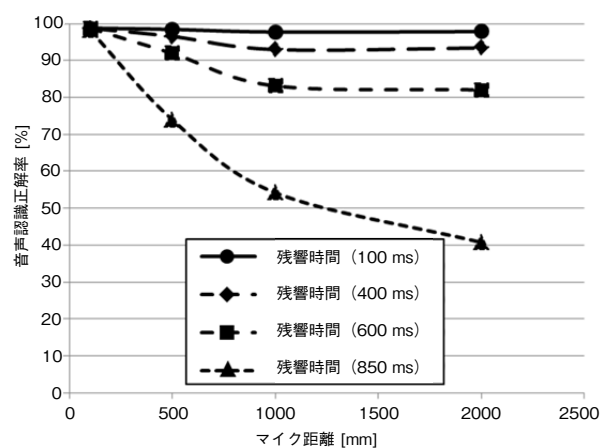
雑音が多い状況下で正しく音声翻訳を行うためには、音声認識の前処理として、雑音を抑圧する技術が必要である。

本技術の課題として、1) マイクとの距離（以下、マイク距離と記す）と、2) 雑音の音量・種類という2つが考えられる。

1つ目が、発話者とマイク距離が長い場合に、周囲の雑音が混入しやすくなる点である。最も効果的な雑音対策はヘッドセットなどのデバイスを用いて口からマイクまでの距離を短くすることであるが、お互いに向き合って使用するユースケースでは、これらのデバイスの利用は

困難である。

そこで、マイクをどの程度口から離して利用できるのかを把握するために、マイク距離が音声認識に与える影響を測定した。第4図に、マイク距離と音声認識正解率の関係を示す。4本のプロット線は、測定環境である部屋の残響時間の違いを示す。



第4図 マイク距離ごとの音声認識正解率の関係

Fig. 4 Recognition ratio related to distance between microphone and sound source

第4図によれば、部屋の残響時間が長くなると音声認識正解率が低下することがわかった。これは、残響時間が長くなることにより、発話者以外の方向から到達する雑音が増加したためと考えられる。この結果から、小売店舗などの残響時間が長い環境で首からぶら下げて使う場合には、残響を拾わずに“発話者の方向の音のみ”を收音するマイクの指向特性の実現が課題である。

2つ目が、雑音の音量や雑音の種類が、音声認識正解率に影響を及ぼす点である。そこで、発話者の声と雑音の比 (SN比) および雑音の種類が音声認識正解率に与える影響を調べた。第5図にその結果を示す。

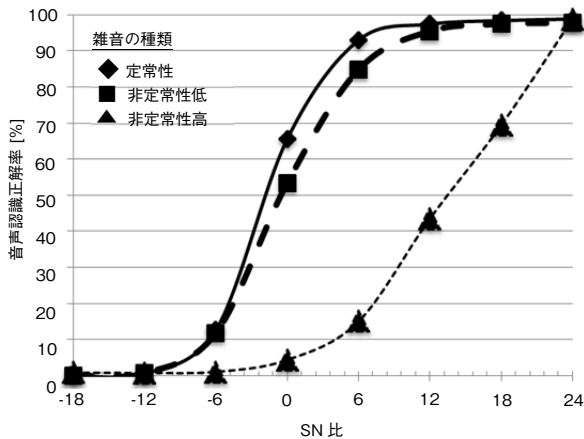
第5図の3本のプロット線は雑音の種類による結果を示している。雑音の種類は、1) 定常性 (エアコンの吹き出し音のような変動の小さい雑音)、2) 非定常性低 (周囲の人の声やBGMの歌声など音声認識に強く影響すると思われる雑音)、3) 非定常性高 (ドアの開閉音や食器が触れ合う音などのある程度変動が大きい雑音) の3種類である。定常性の尺度として、(1) 式で示す尖度 (せんど) β を用い、3種類はそれぞれ次のように定義した。

- 1) 定常性: $\beta < 4$
- 2) 非定常性低: $4 \leq \beta \leq 10$
- 3) 非定常性高: $\beta > 10$

(注1) Bluetooth SIG, Inc. の登録商標

$$\beta = \frac{\sum_{t=0}^{T-1} x(t)^4}{\left(\sum_{t=0}^{T-1} x(t)^2\right)^2} - 3 \dots\dots\dots (1)$$

$x(t)$ は時刻 t における音声信号, T は5 s.



第5図 雑音種別ごとのSN比と音声認識正解率の関係

Fig. 5 Recognition ratio related to SN ratio and noise category

第5図から, SN比の大小だけではなく, 定常性, 非定常性(高低)が音声認識正解率に大きく影響するため, “発話者の方向の音のみ”を取音するだけでなく“発話者の声だけ”を取音する狭い指向特性が必要である. その一方, 狭い指向特性では発話者の声自身も抑圧されて音質が劣化して音声認識正解率を低下させるため, 音質劣化の少ない狭指向特性の実現が課題である.

3. 今後の取り組みと展望

2章で示した2つの課題に対して, 当社の監視用の全方位マイクや音声操作可能なテレビなどに搭載された雑音抑圧技術をベースに, 音源の方向に基づき指向性の範囲と向きを制御する適応型ビームフォーマー技術の研究開発を進め, さまざまな環境で雑音による音声認識精度の劣化を防止し, 音声翻訳技術を実用レベルで利用可能にする技術開発を行う予定である.

今後, さまざまなビジネス用途に音声翻訳技術が展開されていくと予想される. 音声翻訳技術という要素技術をサイネージや構内放送システムなどのさまざまな当社商材と組み合わせて, 鉄道業界, ホテル業界向けのソリューションとして提供していきたい.