

# 音声認識を目的とした音響レンズによる距離選択收音

## Distance-Selectivity Recording using Acoustic Lens for Speech Recognition

A-5

小林 真萌, 松丸 愛実, 高橋 義典

Maho KOBAYASHI, Manami MATSUMARU and Yoshinori TAKAHASHI

東京都立産業技術高等専門学校

Tokyo Metropolitan Collage of Industrial Technology

### 1. 前書き

音声認識では雑音に強健な收音手法が重要である。指向性マイクロホンは、音源と同方向に雑音がある場合は困難である。一方、マイクロホンアレーと信号処理を用いた手法[1]では、実時間処理は厳しい状況である。本稿は、音響レンズを用いて特定距離での感度を高める距離選択收音[2]を提案し、音声認識率の評価について報告する。

### 2. 音響レンズによる距離選択收音

本報告では、特定距離からマイクロホンまでの音波の到達時間差を、管長の異なる伝声管によって取り除く音響レンズを用いた距離選択收音を提案する。Fig. 1 は試作した音響レンズである。伝声管はマイクロホンに集結、收音に最適な音源距離がレンズ面から 0.4m となる構造である。

Fig. 2 に指向特性(a)と距離選択特性(b)を示す。距離選択特性は、距離減衰で正規化され”○”は実測値破線は近似曲線である。強い指向性と、0.4m 付近での高い收音効果が確認できる。

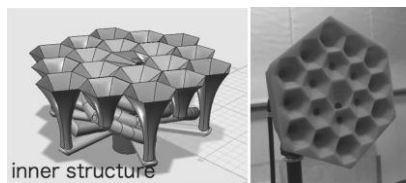
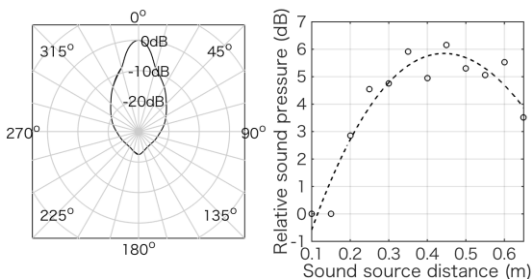


Fig. 1 距離選択收音を目的とした音響レンズ



(a) 指向特性 (b) 距離選択特性

Fig. 2 音響レンズの特性

### 3. 音声認識率による評価手法

本報告では、スピーカ正面に指向性マイクロホンと騒音計を設置、音源距離を 0.2m - 0.6m まで 0.1m 毎に、スピーカより音声を再生・收音した。次に、音源距離 2.5m として雑音を收音した。同様の收音を音響レンズが装着された状態でも行った。音声は、「○○さんお願いします」という文章の名前「○○」を変えた9パターンを、雑音は白色雑音と駅のホームで録音した構内放送等を含む環境雑音を使用した。音声は 0.6m の位置で S/N 比がそれぞれ 15dB, 5dB となるよう正規化、收音した音声と雑音信号を合成した。

用意した合成音声を大語彙連続音声認識エンジン Julius[3]にて音声認識を行い、認識率を評価した。言語モデルは記述文法とし、辞書ファイルは使用した単語の他に 20 単語が登録されている。認識率は正解単語数と正解文の構成単語数の割合で求められる。

### 4. 評価結果

Fig. 3 に音声までの音源距離に対する S/N 比を示す。指向性マイクロホンを用いた場合の S/N 比の方が、総じてわずかに高いことが確認できる。音響レンズでは、音声だけでなく同方向にある雑音も強調され、S/N 比が悪くなったと考えられる。

Fig. 4 に音声認識率の結果を示す。音源距離 0.4m までの結果を見ると、音響レンズを使用した方が音声認識率が高いことが確認できる。

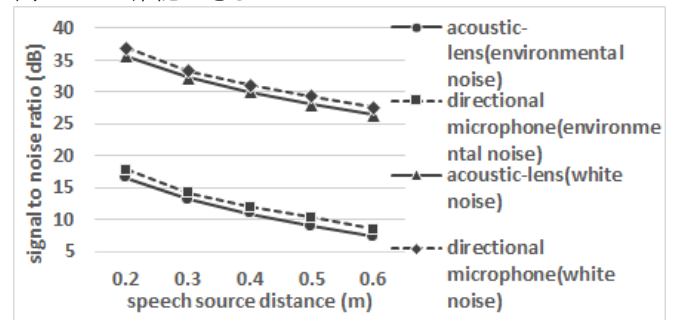


Fig. 3 音源距離と收音された音声の S/N 比

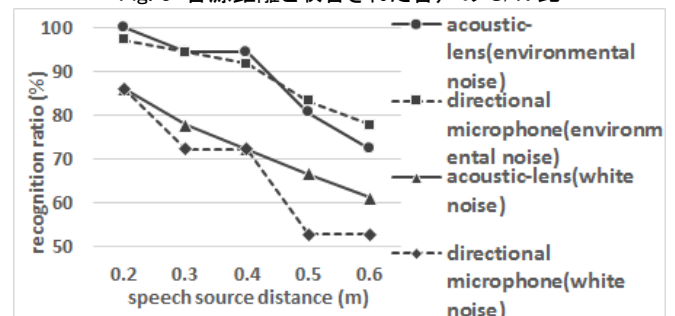


Fig. 4 音源距離と音声認識率

### 5. 後書き

本報告では、音響レンズを用いて特定距離での感度を高める距離選択收音の手法を提案し、認識率での評価を試みた。音響レンズを使うことで、所望する音源距離からの音声に対して、音声認識率が上がることが確認できた。また S/N 比が良くても実際の認識率が悪くなる場合もあることから、認識率を用いて評価することが重要と考えられる。

### 参考文献

- [1] 浅野, コロナ社, ISBN:978-339-0116-6 (2009)
- [2] 高橋, 他, 音講論集春季, 1-P-40 (2016)
- [3] 荒木雅弘, 森北出版, SIBN:978-4627847118 (2007)