# 最短ゴロム定規間隔配置 Delay-and-Sum 型 マイクロホンアレーを用いた雑音環境下の音声認識*

◎鎌本優 (東大・情報理工)　堀内俊治 (ATR-SLT/長岡技科大院・工)
水町光徳　中村哲 (ATR-SLT)　西本卓也　嵯峨山茂樹 (東大・情報理工)

## 1　はじめに

近年, 自動音声認識 (Automatic Speech Recognition; ASR) が, 擬人化エージェントやカーナビゲーションシステムなどへ応用されてきている. 実環境では雑音や残響の影響で認識率が大幅に低下することから, 雑音や残響に頑健な ASR システムを目指す研究がなされてきている [1]. マイクロホンアレーを用いることで, 対象音源と雑音源の空間的位相差を利用し, 雑音や残響を抑圧することにより, 遠隔発話音声の認識性能を向上させることができる.

マイクロホンアレーには様々な技術があるが, Griffith-Jim や AMNOR などの適応型マイクロホンアレーでは, 無音声区間を予め入力し学習させることが必要である [2]. 実際に音声認識を行う場合に, 雑音環境下で学習のための無音声区間を検出することは必ずしも容易ではない. また, 定常雑音に対して頑健な雑音除去を行うことはできるが, 非定常雑音に対しては性能が低下する. このような雑音や残響が時々刻々変化する環境では, 認識性能が低下してしまう.

そこで本研究では, ASR の性能と利便性の両立を目指し, 学習を必要とせず, 雑音および残響の抑制に効果のある Delay-and-Sum(DS) に着目し, そのマイクロホン間隔と配置に関して改良を試みたので報告する.

## 2　Delay-and-Sum 型マイクロホンアレー

### 2.1　予備検討

予備検討として, 様々な条件における DS の性能を比較するため, シミュレーションにより音声認識実験を行った. 特に, マイクロホンの数, マイクロホン間隔, 雑音のマイクロホンアレーに対する角度, SNR に注目した.

音声データには, ATR の BTEC テストセット 01 を用いた. この評価用データは旅行の際に用いられる会話を朗読したもので, 全部で 510 文あり, 16kHz サンプリングで収録されたものである.

雑音はマイクロホンアレー正面から到来すると仮定し, マイクロホンの受音信号として, 適切な時間差を伴う音声に同一の雑音を加えた. 雑音は音声の周波数帯域に合わせて, 125Hz から 6kHz のランダム帯域雑音を用いた. SNR は音声データの無音声区間を除いた区間の平均振幅から信号のエネルギーを求め, 目的の SNR となるように雑音の振幅を変化させた.

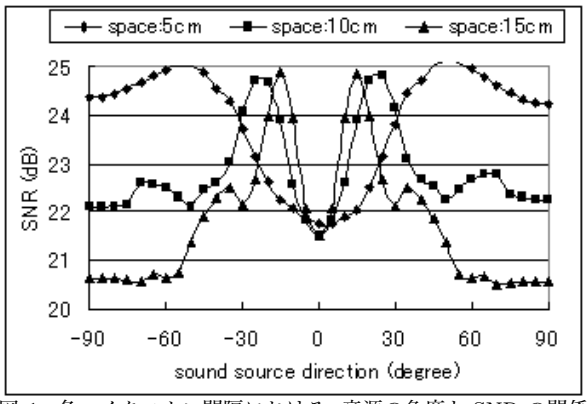


図 1: 各マイクロホン間隔における, 音源の角度と SNR の関係 (マイクロホン 2 個, SNR20dB)

その後, DS により雑音抑圧した音声を認識した.

結果として, マイクロホンの数が多いほど認識率 (単語正解精度) が向上した. さらにマイクロホン間隔に応じて音声のマイクロホンアレーに対する角度と関係して DS 処理後の SNR が変化し, 入力信号の SNR が高いほど認識率も向上することが分かった. 各マイクロホン間隔 (5cm, 10cm, 15cm) での, 音源と雑音源の角度の変化による SNR の変化を図 1 に示す. マイクロホンを 2 個, SNR を 20dB として, 音声を −90 度から +90 度まで 5 度毎に変化させたものを表している.

### 2.2　マイクロホン間隔と音源角度の検討

予備検討から得られた結果より, 音源方向と雑音方向が既知ならば, 図 1 にしたがってマイクロホン間隔を調節することにより, DS 処理後の SNR を向上させることができる.

あらかじめ多数のマイクロホンを用意しておけば, 適切な間隔のマイクロホンの対を選択することにより, 同様の効果が得られる.

できるだけマイクロホン数を増やさずに, 様々な間隔が得らるような配置があれば, 音源方向や雑音方向に合わせて最適な距離を選択することができる.

## 3　最短ゴロム定規間隔の導入

前述の要求を満たすために, 本研究では DS における最短ゴロム定規 (Optimal Golomb Ruler; OGR) 間隔の導入を試みた.

OGR は X 線センサの配置や電波望遠鏡の配置に使われている. この間隔は, センサの数が少なくても, 測ることができる距離の種類が増えるというものである. 例えば, 4 個のマークならば {0-1-4-6}, 10 個のマークならば {0-1-6-10-23-26-34-41-53-55} のようになる. これを用いると, 図 2 に示すように, 等間隔配置よりも多くの間隔を得ることができる.

---

*“Speech recognition under noisy environment using Delay-and-Sum beamformer employing the microphone arrangement based on Optimal Golomb Ruler,”
by Yutaka KAMAMOTO (The University of Tokyo), Toshiharu HORIUCHI (ATR Spoken Language Translation Research Labs./Nagaoka University of Technology), Mitsunori MIZUMACHI, Satoshi NAKAMURA (ATR Spoken Language Translation Research Labs.), Takuya NISHIMOTO and Shigeki SAGAYAMA (The University of Tokyo).

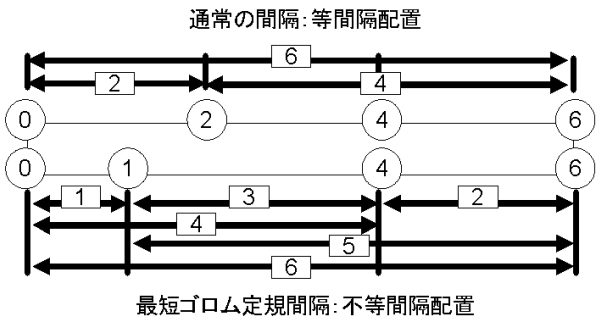

図 2: 通常の等間隔と最短ゴロム定規間隔の比較

ゴロム定規は, 2 組の数字の差が同一ではない正の整数の集合である. M 個のマークがあるときに,

$$\delta_{ij} = a_j - a_i \quad (1 \leq i < j \leq M) \tag{1}$$

が全て異なり, かつ,

$$0 = a_1 < a_2 < \cdots < a_M \tag{2}$$

を満たす数列 $a_k \quad (k = 1, 2, \cdots, M)$ の数値を目盛とした定規を作れば, それがゴロム定規である. この $a_M$ が最も短くなるものを OGR という.

これをマイクロホンアレーに用いることにより, 通常の DS よりも音声認識率を向上させることができると考えた.

また, 最適な間隔を強調するために, 推定された音源と雑音のなす角に応じて, その角度で処理後の SNR が高くなるマイクロホン間隔になる 2 組のマイクロホン対に大きな重みを付け, 低くなるマイクロホン対に小さな重みを付けるため,

$$k_1 + k_2 + \cdots + k_N = 1 \tag{3}$$

という条件の下で重み $k$ を変化させた. この条件であれば音源の振幅は変化しない.

## 4 評価実験

### 4.1 実験条件

提案手法の効果を確かめるために, 音声認識率による性能評価実験を行った.

計算機上のシミュレーションにより, マイクロホンアレーを用いた場合の雑音環境下の音声信号を作成し, そのデータをもとに音声認識実験を行った.

マイクロホンアレーのパラメータとしては, マイクロホンの列に正面から音声を入力し, 30 度傾いた方向から予備検討と同じ雑音を入力した. マイクロホンを 4 個とし, 通常の DS と提案手法である最短ゴロム定規間隔型 Delay-and-Sum(OGR-DS) を比較した. ここで, マイクロホン間隔は 2 つの手法において同規模とするために, DS では, {0cm-6cm-12cm-18cm} にマイクロホンを配置し, OGR-DS では 4 個のマークの OGR をもとにして, {0cm-3cm-12cm-18cm} にマイクロホンを配置した. また, 対照実験としてマイクロホンアレーを用いない場合, つまりマイクロホン 1 個の場合の認識率も求めた.

各発声ごとに, 各マイクロホンの重みを 0.1 ずつ変化させ, 処理後の音声区間と無音声区間を検出し, 比較して得られる SNR が全 84 通りの中で最も高くなるものを音声認識への入力とした.

音声認識エンジンには Julius3.1p2 を用い, IPA-testset の 200 文の新聞朗読音声を評価データとして用いた [3] . 音響特徴量は 12 次の MFCC とその Δ MFCC および Δ Power の計 25 次元とし, フレーム長 25ms・フレームシフト 10ms で分析した.

### 4.2 結果と考察

音声認識実験の結果を表 1 に示す. OGR-DS はマイクロホンの配置を変え重みを付けただけの簡単な方法にも関わらず, 認識率を向上させることができた.

10dB 雑音環境下において, マイクロホン 5 個を {0cm-6cm-12cm-18cm-24cm} に配置した DS の認識率が 46.9% であった. これに対し, OGR-DS はマイクロホン 4 個で認識率 51.1% となり, さらにマイクロホンアレーの規模も 6cm 小さい.

このように提案手法によりマイクロホン数を少なくし, マイクロホンアレーの規模を小さくすることが可能となった.

今回の実験条件において, 各マイクロホンの重みは, 0cm と 3cm に配置されたマイクロホンの重みを 0.3 とし, 12cm と 18cm に配置されたマイクロホンの重みを 0.2 としたものが全 200 文中 193 文にのぼった. このことから, 重みの定式化ができれば処理速度を向上させることができると考えられる.

表 1: 認識率 (単語正解精度) の比較 (DS, OGR-DS:マイクロホン 4 個)

| SNR | 1 microphone | DS | OGR-DS |
|---|---|---|---|
| ∞ dB | 89.3% | — | — |
| 20 dB | 62.4% | 78.2% | **79.3**% |
| 10 dB | 19.4% | 39.4% | **51.1**% |
| 5 dB | 9.0% | 16.6% | **25.3**% |

## 5 まとめ

ASR システムの認識率向上のために, 最短ゴロム定規間隔配置 Delay-and-Sum 型マイクロホンアレーを提案した. 不等間隔配置を積極的に利用することで, 認識率を向上させることができた.

今後は, マイクロホン間隔と音源の角度によって生じる時間差から与えられる各マイクロホンの重みを定式化し, あらゆる個数のマイクロホンアレーに応用できるように細かく分析する必要がある. また, シミュレーション結果を確かめるために, 実環境での実験を行いたい.

## 謝辞

本研究は ATR での実習をもとに発展させたものである. 実習中にご助言を下さった ATR-SLT の皆様に感謝致します.

## 参考文献

[1] 中村哲, “実音響環境に頑健な音声認識を目指して,” 信学技報, SP 2002-12, pp. 31-36, 2002.

[2] 大賀寿郎ら：音響システムとディジタル処理, 電子情報通信学会, 1995.

[3] 鹿野清宏ら：音声認識システム, オーム社, 2001.