アクティブマイクロフォンによる音響伝達特性を用いたシン グルチャネル音源方向推定*

高島遼一, 滝口哲也, 有木康雄 (神戸大)

1 はじめに

音声を用いたインターフェースにおいて,マイクロ フォンは位置や向きが変わらず常に固定されている というのがこれまでの考え方であった.しかしながら 人間は日常,目的の音を注意して聞くときはその方 向に耳を向けるなど,様々な状況に合わせて耳の位置 や方向を変えている.

本研究では,これまで常に固定されていたマイク に,動くという概念を加えることによってどのような 恩恵が得られるかを調べることを目的とし,本稿で は提案するアクティブマイクロフォンを用いることに より,通常複数のマイクが必要であった音源方向推定 をマイクロフォンーつで行う.

2 アクティブマイクロフォン

本研究で用いたアクティブマイクロフォンは, Fig. 1 のように放物面型の反射板の焦点位置に無指向性マ イクが取り付けられている.反射板からの位相のず れた反射波は観測信号のパワースペクトルに影響を 与えるため,マイクロフォンアレーを用いない音源方 向推定法として反射板が用いられることがある[1,2]. 本手法では反射板とマイクの向きを手動で変えなが ら各方向毎に音声を収録し,角度毎の音声を比較す ることで音源方向を検出する.

2.1 パラボラ反射板

平面波を仮定した場合,放物面の軸と平行に入射 した波はすべて放物面の焦点に向かって反射される. 一方別の角度から入射した波が反射によって焦点に 到達することはない.音源が放物面の正面に位置して いる場合の観測信号を Fig. 2 に示す. $x_0(t)$ は焦点 Oへ直接到達する信号, $x_n(t)(n = 1, \dots, N)$ は放物面 を反射して O へ到達する信号である.このとき,Oにおける観測信号は $x_0(t)$ と $x_n(t)$ の和で表される.

$$o(t) = x_0(t) + \sum_{n=1}^{N} x_n(t).$$
(1)

ここで,放物線の性質より $x_0(t) \ge x_n(t)$ の焦点までの時間差はnの値に関わらず $\tau = 2d/c$ (cは音速)であるので, $x_0(t)$, $x_n(t)$ の伝達関数をそれぞれ $h_0(t)$,



Fig. 1 アクティブマイクロフォン



Fig. 2 放物面が音源方向を向いているときの焦点に おける観測信号

$$h_n(t)$$
, クリーン音声を $s(t)$ とすると

$$o(t) \approx s(t) * h_0(t) + \sum_{n=1}^{N} s(t-\tau) * h_n(t)$$
 (2)

周波数領域では

$$O(\omega) \approx S(\omega) \cdot (H_0(\omega) + e^{-j2\pi\omega\tau} \cdot \sum_{n=1}^N H_n(\omega))$$

= $S(\omega) \cdot (H_p(\omega) + H_r(\omega))$ (3)

として,反射板の影響を受けない伝達関数 H_p と反射板により追加される伝達関数 H_r を用いて表すことができる.一方音源が放物面の正面に位置していない場合,Oへ到来する信号は $x_0(t)$ のみであるので,Oにおける観測信号は

$$O(\omega) \approx S(\omega) \cdot H_0(\omega) = S(\omega) \cdot H_p(\omega) \tag{4}$$

となる.

^{*}Monaural sound-source-direction estimation using acoustic transfer function of active microphone. by Ryoichi Takashima, Tetsuya Takiguchi and Yasuo Ariki (Kobe Univ.)

2.2 k-means による音源方向クラスの検出

(3), (4) 式より, アクティブマイクロフォンの向き を θ , 音源方向を $\hat{\theta}$ とすると観測信号 O_{θ} は

$$O_{\theta}(\omega) \approx S_{\theta}(\omega) \cdot H_{\theta}(\omega)$$

$$H_{\theta}(\omega) = \begin{cases} H_{p}(\omega) + H_{r}(\omega) & (\theta = \hat{\theta}) \\ H_{p}(\omega) & (\theta \neq \hat{\theta}) \end{cases} (5)$$

となる.ここでアクティブマイクロフォンの角度が変 化しても H_p の値がほぼ一定であれば, H_θ はマイク が音源方向を向いていない状態ではほぼ同じ空間上 に分布し,音源方向を向いたときのみそれらとは離 れた位置に分布することになる.そこで各角度で得 られる H_θ をk-meansにより2クラスのクラスタリ ングを行うと, H_θ は音源方向での $H_{\hat{\theta}}$ とそれ以外の $H_{\theta}(\theta \neq \hat{\theta})$ とに分類される.その後分類された二つの クラスタの内,属するデータの数が少ない方のクラス タに属する H_θ を音源方向の伝達特性 $H_{\hat{\theta}}$ とみなし, $H_{\hat{\theta}}$ に対応する角度,すなわち

$$H_{\hat{\theta}} = \begin{cases} H_{\theta} \in C_1 & (N_1 \leq N_2) \\ H_{\theta} \in C_2 & (N_1 > N_2) \end{cases}$$
(6)

を満たす $\hat{\theta}$ を出力する.ここで N_k はクラス C_k に属 するデータの数を表す.

2.3 音響伝達特性の推定

*H*の求め方について,(5)式はケプストラム領域では次のように表される.

$$O_{cep}(d) \approx S_{cep}(d) + H_{cep}(d) \tag{7}$$

dはケプストラムの次元を表す.(7)式より, $O \geq S$ を 観測することができればHを求めることができるが, 実際にはSを観測することはできない.そこでSの 代わりにあらかじめクリーン音声のGMM(Gaussian mixture model)を作成しておき,Oに対して,その GMMの尤度が最大となるようにHを求める[3].

$$\hat{H} = \operatorname*{argmax}_{H} \Pr(O|\lambda_S, H) \tag{8}$$

 λ_S はクリーン音声のモデルパラメータを表す.

3 実環境下における実験

まず予備実験として (7) 式に正解の S を与えて求め た各角度における H をプロットした.マイクの角度 は 30°から 150°までの 7 方向,音源距離は 2 m,音 源方向は 90°,反射板は直径 24 cm,焦点距離 9 cm のものを用いている.また,サンプリング周波数は 16 kHz,特徴量は MFCC 16 次元を用いた.MFCC 16 次元のうち,反射板の影響が特に強く現れた1番



Fig. 4 角度毎の H が音源方向として検出された割合

目の次元と2番目の次元をFig.3に示す.図中の各 角度のプロット数はデータ数に対応し,各点は各デー タの平均MFCCをプロットしている.Fig.3による と音源方向である90°の伝達特性が他の角度の伝達 特性より離れた位置に分布していることがわかる.

次に提案手法により角度毎の H が音源方向クラス に分類された割合を Fig. 4 に示す.クラスタリング には MFCC 16 次元のうち1番目の次元と2番目の次 元を用いている.Fig. 4 より 90°の伝達特性は 100% の割合で音源方向クラスに分けられたが,その周辺 の角度である 110°の伝達特性も反射板の影響を受け てしまったため,約 60%の割合で音源方向クラスと して分類されてしまった.

4 おわりに

本稿ではアクティブマイクロフォンを用いて音源方 向推定を行った.反射板の影響により伝達特性は音源 方向とその他の方向に区別されたが,反射板の影響 を若干受ける音源方向周辺の角度も音源方向として 区別されることがあったため,反射板の指向性をさら に鋭くするなどして改善する必要がある.

参考文献

- [1] 市川他, 音講論(春), 687-688, 2003.
- [2] 小野他,信学技報,25-32,2000.
- [3] 住田他, 音講論(春), 771-772, 2008.