

F-12

サンプリング周波数に依存しない話者方向識別

Recognition of Speaker Direction Independent of Sampling Frequency

池葉 雅之†  
Masayuki Ikeba

桑山 敬司†  
Takashi Kuwayama

関口 芳廣†  
Yoshihiro Sekiguchi

鈴木 良弥†  
Yoshimi Suzuki

1. まえがき

音源方向識別の研究に関しては既に多くの研究があり、実用システムも存在する[1]。基本的にはマイクロフォンアレーを利用して、音の位相差を検出し、その時間差から音源方向を識別する方法がよくとられる。精度よく、音源方向を識別するためには、サンプリング速度を上げ、複雑な計算を導入する必要がある。

マンマシンインタフェースとして、ロボットが対話相手の方向を識別するような応用では、高い精度での角度識別は必要なく、話者とロボットの距離がある程度離れているという環境で、音声認識の傍ら話者の方向を識別できればよい。そこで、筆者らは簡潔な方法で話者方向を識別できる方法を検討している。

2. 話者方向識別の概要

(1) マイクロフォンアレーの構成

360度全方向の話者が識別できるように、図1のように円形に8個の無指向性マイクロフォンを配置する。

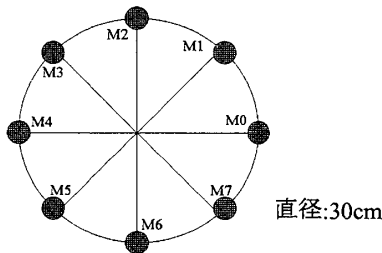


図1. マイクロフォンアレーの構成

(2) 2つのマイクロフォンによる予備実験

図2のように、2つのマイクロフォンM1、M2を配置して、マイクロフォンを結ぶ線分の垂直2等分線と話者とのなす角を $\alpha$ とする。話者に母音/a/を発声してもらい、その時2つのマイクロフォンに入った音声の相関を求めた。データの数は500である。色々な $\alpha$ について実験した結果を図3に示す。図3より、 $\alpha$ の値が大きくない範囲では、求めた相関値は、 $\cos \alpha$ に近似していることがわかる。このことから、2つのマイクロフォンに入ったデータの相関を求めれば、音源方向をある程度推定できると考えられる。

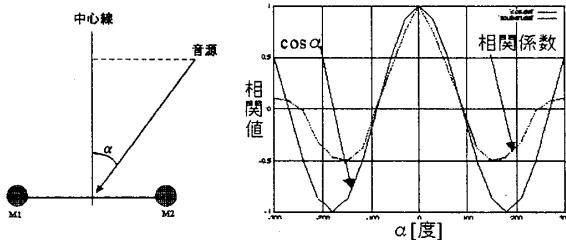


図2. 予備実験の概略

図3. 音源方向 $\alpha$ と相関の関係

(3) 話者方向の識別方法

図1の全てのマイクロフォンの対に対して、データの相関をとる。これはマイクロフォン間の垂直2等分線と話者のなす角度 $\alpha$ に対する $\cos \alpha$ に対応しているので、音声の垂直2等分線方向の成分と考えて、垂直2等分線と同じ向きに $\cos \alpha$ の大きさのベクトルを考える。全てのマイクロフォンの対に対して、この操作を行う。全てのベクトルを加算して、その方向を話者の方向と決める。ただし実用的には、マイクロフォンの間隔が長い場合は誤差が大きくなるため、相関値が0.4以下の場合には使用しない。

3. 音声データの相関係数と音源方向の関係

例えば、図4のマイクM0とM1の間に音源がある場合、マイクM0、M1に入った音声データの相関係数の理論値は1になる。音源が中間よりずれると相関係数は1より小さくなる。

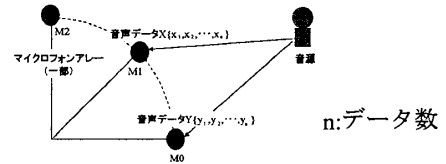


図4. マイクロフォンアレーと音源の関係

図4のマイクロフォンM1に入った音声を正規化したものを $X=(x_1, x_2, \dots, x_n)$ 、マイクロフォンM0に入った音声を正規化したものを $Y=(y_1, y_2, \dots, y_n)$ とすると、音声 $X, Y$ は $n$ 次元のベクトルで表すことができる。ベクトル要素の平均は $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ 、 $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$ である。そうすると、次の式(1)、(2)より $X, Y$ の偏差ベクトル $X', Y'$ が求まる。

$$X' = X - \bar{X} = (x_1 - \bar{x}, x_2 - \bar{x}, \dots, x_n - \bar{x}) \dots (1)$$

$$Y' = Y - \bar{Y} = (y_1 - \bar{y}, y_2 - \bar{y}, \dots, y_n - \bar{y}) \dots (2)$$

この時、2つの偏差ベクトル $X', Y'$ のなす角度 $\theta$ に対する余弦は $X', Y'$ の相関係数になる[2]。

$$\cos \theta = \gamma_{XY'} \quad (X' \text{ と } Y' \text{ の相関係数}) \dots (3)$$

一方、次の式(4)で $\|X'\| = \|Y'\| = 1$ なので、

$$\cos \theta = \frac{(X', Y')}{\|X'\| \cdot \|Y'\|} \dots (4)$$

式(3)、(4)より内積=相関係数という関係が得られる。よって、相関係数が得られれば $\cos \theta$ が分かり、音声データをパラメータとみなしたベクトルの角度 $\theta$ を求めることができる。

また、2つのマイクロフォンの中心線と音源方向のなす角度を $\alpha$ とすると、図3より $\alpha$ が $\pm 100$ 度の範囲では $\cos \alpha \approx \cos \theta$ となる。

† 山梨大学工学部, 甲府市

$\cos \alpha$  は、音源方向ベクトルのマイク間の垂直 2 等分線に平行な成分と考えられる。また、環状マイクロフォンアレーの中心に向かう成分である。

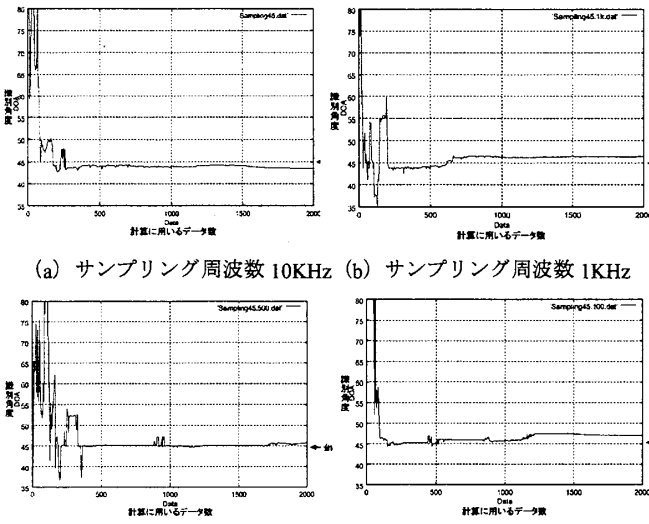
複数のマイクロフォン対に対して、この成分を求めて加算することで、音源方向識別ができる。

#### 4. 評価実験

##### (1) データ数と識別精度の関係

図 1 のマイクロフォン M0 の延長上を話者角度 0 度とし、反時計回りに角度を定義する。マイクロフォンアレーの中心から 1m 離れて、角度 45 度付近で発声した/a/に対してサンプリング周波数、使用するデータ数を変化させて、話者方向の識別実験を行った。その結果を図 5 (a) から (d) に示す。

データ数が少ない場合には、角度推定が安定しないが、500 データ以上になると、どのサンプリング周波数でも、角度推定が安定に行われている。また、(a) から (d) で推定角度が若干異なる。マイクロフォンアレーと話者の位置関係等を精密に測定していないので、正確な誤差については言及できないが、対話相手の方向を識別する目的には十分な精度だと思われる。

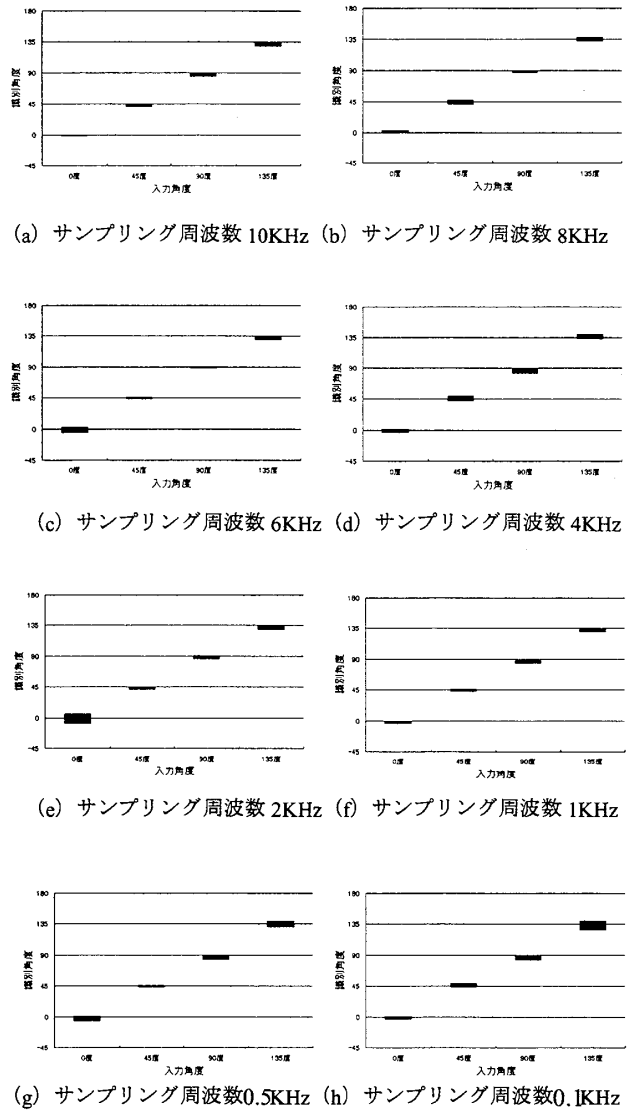


(a) サンプリング周波数 10KHz (b) サンプリング周波数 1KHz  
(c) サンプリング周波数 0.5KHz (d) サンプリング周波 0.1KHz  
図 5. サンプリング周波数とデータ数と話者方向の識別結果の関係 (話者方向約 45 度)

##### (2) サンプリング周波数と識別精度の関係

音源方向識別実験は、マイクロフォンアレーの中心から 1m に音源 (録音した定常波/a/をスピーカーから再生) を固定し行う。サンプリング周波数を 10KHz、8KHz、6KHz、4KHz、2KHz、1KHz、0.5KHz、0.1KHz に設定し、音源方向は約 0 度から 135 度に対して 45 度間隔で音源方向識別を行う。各方向とサンプリング周波数に対して、音源方向の識別実験を 5 回ずつ行う。

実験結果を図 6 (a) から (h) に示す。図 6 の黒いブロックは、縦方向が平均値±標準偏差を表している。実験結果より、低いサンプリング周波数でも、話者方向をほぼ正確に識別できていることが分かる。



(a) サンプリング周波数 10KHz (b) サンプリング周波数 8KHz  
(c) サンプリング周波数 6KHz (d) サンプリング周波数 4KHz  
(e) サンプリング周波数 2KHz (f) サンプリング周波数 1KHz  
(g) サンプリング周波数 0.5KHz (h) サンプリング周波数 0.1KHz  
図 6. サンプリング周波数と話者方向の識別結果の関係 (データ数:500)

#### 5. おわりに

提案した方法は、従来の方法より簡単な方法であるが、サンプリング周波数に影響されず、ある程度小さな誤差で話者方向を識別できる利点がある。

今後は子音を含む音声に対する実験等を進め、実用化を目指したい。

#### 参考文献

[1]大賀・山崎・金田: “音響システムとデジタル処理”、電子情報通信学会 (1995)  
[2]駒澤 勉: “数量化理論”、財団法人放送大学教育振興会 (1992)