

MSC を用いた 2 本のマイクによる複数音源の 2 次元位置推定*

中島弘道 (理研), 河本満 (産総研), 向井利春 (理研)

1 はじめに

ロボットが人と会話する場合, ロボットは話者の方を向いて会話をする方が望ましい. そのためには, 話者の方向を検出する必要がある. また, 実際に次世代ロボットが利用される生活環境においては, マイクで検出される音は複数の音源からの音が混ざり合った混合音になると考えられる. そのような環境においてロボットを使用する場合には, ロボットは混合音を聞き分け, 目的とするものの音源方向を見つける必要がある. 人間や多くの生物は, 2 つの耳だけで複数の音の方向をある程度の正確さで識別する能力を持っているが, ロボットでこれを実現することは未だ困難であり, 多くの音源定位機能を備えたロボットでは, 3 つ以上のマイクや聴覚以外のセンサ情報を用いることにより, 様々な方向からの音源の定位に対応している.

そこで本研究では, 反射板を備えた 2 本のマイクを用いて複数音源の上下左右の音源方向を推定する手法を提案する. マイクから得られた 2 つの音声信号の相関性をコヒーレンス関数から求めることにより単一音源となる区間を推定し, その区間において音源定位処理を行うことにより複数の音源位置の推定を行う.

2 提案する複数音源の定位法

通常, 人の話す音声にはたくさんの無音部分が存在する. よって, 複数の人間が同時に話していたとしても短い時間内であれば話している人間が一人だけとなる区間が存在すると考えられる. そこでなんらかの方法により, 音声の中の 1 音源とみなせる区間を見つけられれば, その区間では 1 音源での音源定位の処理方法を用いる事が可能であると考えられる. 今回はその 1 音源部分の検出に, S.Mohan らにより提案されたコヒーレンス関数を用いた手法 [1] を使用した. 具体的な処理の流れを以下に示す.

1. 2 本のマイクで取得した音声のコヒーレンス関数の値を求める.
2. 処理結果から音源が 1 つのみとなる区間を求め, その区間の音声を抽出する.

3. 抽出した部分の音声信号に対して音源定位処理を行う.

2.1 1 音源区間の検出

2 本のマイクによる入力信号をそれぞれ $y_1(n), y_2(n)$ とする. この入力信号を分割し, おのおののブロックに FFT を施す. $Y_1(m, \omega), Y_2(m, \omega)$ を $\{y_1(n)\}_{n=mK}^{mK+N-1}, \{y_2(n)\}_{n=mK}^{mK+N-1}$ の N 点での離散フーリエ変換とする. これらのスペクトルから以下のように共分散行列を作成する.

$$Y(m, \omega) = [Y_1(m, \omega) \quad Y_2(m, \omega)]^T \quad (1)$$

$$R(m, \omega) = E[Y(m, \omega)Y^H(m, \omega)] \quad (2)$$

この共分散行列から振幅二乗コヒーレンス関数を以下のようにして求める.

$$T_{coh}^{(i,j)}(m, \omega) = \frac{|[R(m, \omega)]_{ij}|^2}{[R(m, \omega)]_{ii}[R(m, \omega)]_{jj}} \quad (3)$$

1 音源の場合この値が 1 に近くなることから 1 音源区間が検出できる.

2.2 音源定位手法

本研究では, 水平方向を推定する手掛かりには, 両耳間到達時間差情報のみを利用し, 左右の到達時間差から相関値を用いて音源方向の推定を行う. また, 垂直方向の定位に関しては, 我々の提案した耳介に相当する反射板を用いた, 上下方向の音源定位学習手法 [2] を用いた. この手法では, 反射板により変化したスペクトルの特徴を, 自己組織特徴マップを用いて分類し, 音源方向を推定する手掛かりとして用いている.

3 複数音源定位実験

3.1 実験環境

実験は無響室で行った. 反射板に無指向性マイクを取り付けたものを左右の耳として搭載したロボット (図 1) を使用し, マイクの地面からの高さは 1.1m とした. マイク間距離は 30cm である. 音源の数は 3 つとして高さの違いは 20cm 刻みの 3 段階とし, 3 つの高さの判別を行うこととした. また, 使用するスピーカ 3 台は図 2 のように配置した.

* A two-dimensional position estimate of multiple sound sources with two microphones by using MSC

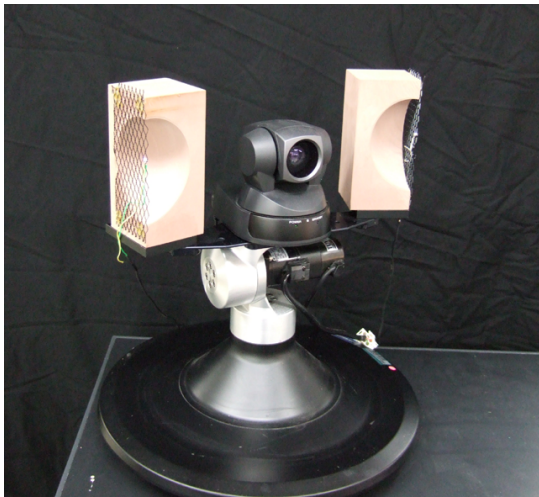


Fig. 1 Head robot consists of two microphones corresponding to ears with reflectors as the pinna

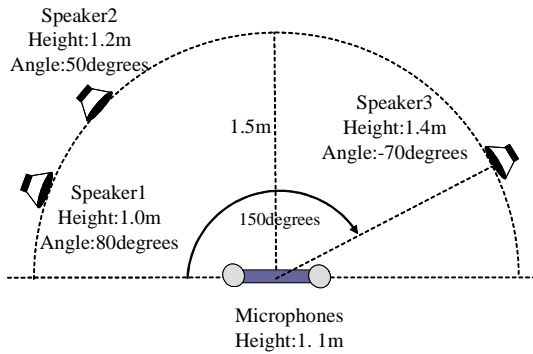


Fig. 2 Arrangement of microphones and loudspeakers

3.2 実験結果

実験に使用した3つの音源信号を図3に示す。図中の影の部分が主な1音源部分と考えられる。MSCを求める際に用いたパラメータの値は、 $N=1024, K=820$ とした。また、(2)式の期待値を求める区間は15ブロックとした。マイクから得られた2つの音声信号の振幅二乗コヒーレンス関数の値を図4に示す。このグラフは関数値の30ポイントの離散周波数で和をとったものである。グラフから1音源の部分において、関数値が大きくなっていることがわかる。検出された1音源区間において水平方向の定位を行った結果を図5に示す。図からほぼ正しい音源方向の推定が出来ていることがわかる。また、上下方向についても3つの高さの判別が正しく行われた。

4 おわりに

本研究では3つの異なる位置から発せられる音声の音源方向を推定する手法を提案し、実験によりその有効性を示した。本手法を用いることにより、2本のマ

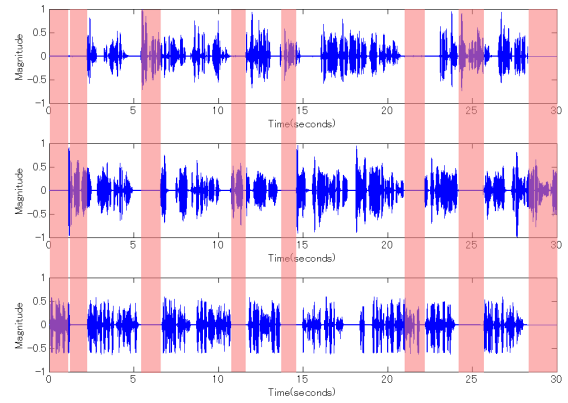


Fig. 3 Sound sources

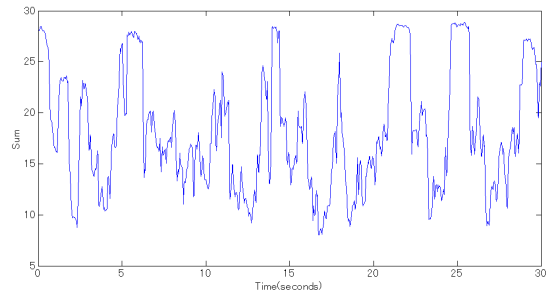


Fig. 4 MSC

イクのみで3つの音源の水平方向及び垂直方向の位置を推定することが可能となった。

参考文献

- [1] S. Mohan et al, "Localization of nonstationary sources using a coherence test", 2003 IEEE Workshop on Statistical Signal Processing, pp.470-473
- [2] H.Nakashima, T.Mukai, "3D Sound Source Localization System Based on Learning of Binaural Hearing", IEEE SMC 2005, 2005.

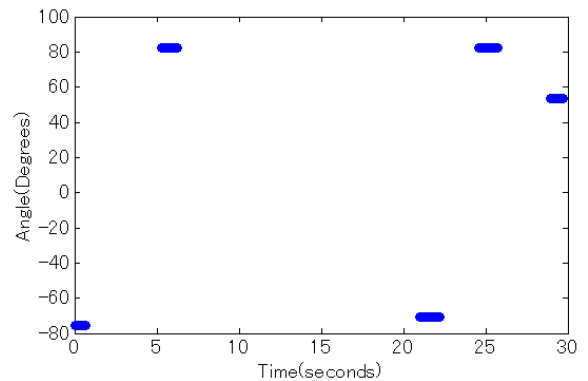


Fig. 5 Result