

多人数参加型会議のためのマイクロホンアレーによる話者位置推定法

山田 貴弘 富野 剛 井上 亮文 市村 哲 松下 温

東京工科大学

1.背景

近年、出張・人件費の削減等が可能になる事からテレビ会議システムを用いて会議や打ち合わせを行う企業が増加している。しかし、現在のテレビ会議システムでは参加者全員が小さく表示されてしまう為に現在誰が発言を行っているか、よくわからないという問題点がある。

本研究では少数のマイクロホンで構成したマイクロホンアレーでの発言者位置を推定するシステムを構築した。そして、発言者の推定を行い、発言者の映像を拡大するテレビ会議システムの開発を目指した。

2.音源方向推定方法

マイクロホンアレーを用いた音源方向推定法を説明する。

2.1 遅延時間による音源推定

発言者位置方向を推定する方法として、遅延時間推定法を使い、2つのマイクロホンに到達する音の時間差から推定を行った。図1で示す様に角度  $\theta_s$  の方向から距離  $d$  だけ離れた二つのマイクロホンを設置した。そして、先にマイク1 (M1)が受信した信号  $x_1(t)$  に対して、遅れてマイク2 (M2)が受信した信号  $x_2(t)$  の遅延時間を  $\tau$  とすると式 (1)により、  $\tau$  [sec]の間に音波が進んだ距離を求める事ができる。

$$\tau = d / c \quad c : \text{音速} \quad (1)$$

これにより、音波の到来方向  $\theta_s$  は式(2)から求める事が可能となる。音速  $c$  とマイク間距離  $d$  は既知数であるので、未知数である遅延時間  $\tau$  を求める事ができる。

$$\theta_s = \arcsin(\tau / d) = \arcsin(c \tau / d) \quad (2)$$

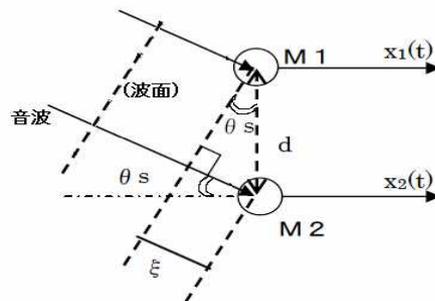


図1 遅延時間による音の推定図

2.2 遅延時間算出方法

遅延時間を算出する方法として、2つのマイクロホンに到達した各音を PCM データとして PC 上で収録を行い、片方の音声波形の時間軸をずらして2つの音声波形の相関を調べて最も波形に近い形で重なった際に移動したデータ数から遅延時間を求めた。相関を調べる時の問題点として、PCM データとして収録したデータには人の声だけでなく、空調やパソコン等の環境音や電源ノイズ等が録音の際に混ざってしまう事があげられる。そこで、録音データに実離散フーリエ変換[1]を行い、人の声の周波数が存在する約 300 ~ 2500Hz 以外の周波数を遮断し、ノイズを除去したデータで相関処理を調べた。

A speaker position detection with Microphone Array for TV conference

Takahiro Yamada, Tomino Takeshi, Akifumi Inoue, Satoshi Ichimura, Yutaka Matsusita

Tokyo University of Technology

### 3. システム構成

#### 3.1 マイクロホンの設置

2つのマイクロホンでは180°までの方向でしか発言者位置の推定を行えない。これは、2つのマイクロホンM1, M2を結んだ直線に対して、上下どちらの方向に音源があったとしても、遅延時間は同じになるからである。そこで、図2のようにマイクロホンM3, M4を正方形の4頂点となるように2本設置した。M3とM4に対してどちらのマイクロホンに大きい音が存在するか判断し、そこからどちらに先に音が到達したかの推定を行う事で360°の範囲で発言者位置の推定を可能にした。

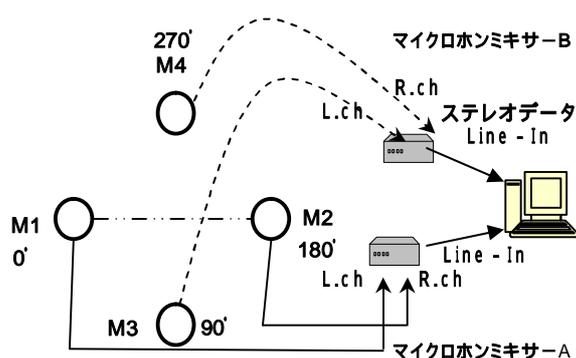


図2 マイクロホン構成図

#### 3.2 ミキサーによるステレオ録音

図2で示すように無指向性モノラルマイクロホンM1~M4を配置する。マイクロホンミキサー-AにM1とM2、BにM3とM4をそれぞれのL,R入力として接続する。サウンドカードを2枚装着したPCの各ライン入力により、ステレオPCMデータとして各マイクロホンミキサーから収録を行った。これにより、特殊な専用装置を使用せずに発言者位置推定処理を1台のPCで実現した。

#### 3.3 発言者位置推定結果の出力方法

PCMデータに対して0.05秒毎に一定値以上の大きさの音があるかの判断をし、存在する場合にM1とM2のステレオPCMデータから式(2)を用いて音源推定を行った。M3とM4では、M3とM4のどちらのマイクロホンに大きい音が存在するかで、音源位置が0~180°側と180~360°側の判断をした。そして、0.05秒毎に出た結果を10°きざみで分別し、1秒間の最頻値データを測定結果として出力した。

### 4. 評価および考察

PCの作動音や、空調等の雑音が存在する室内において、M1とM2、M3とM4のマイク間隔を40cmに設定した。そして、発言者3名を用意して90°、180°、225°の位置にマイクから1m離れた場所から発言した。音源位置推定実験を行った結果、発言者位置が90°の場合90%の割合で誤差±10°の平均した音源位置方向結果が得られた。225°では82%、180°では66%との結果が得られた。180°付近で極端に精度が落ちてしまう原因としては、遅延時間が最大になり先に到達した音波に対して遅れて到達した波形の減衰が大きくなり、相関を求める際に影響を与えてしまう為と考えられる。

### 5. まとめ

本稿では、2本2組のマイクロホンアレイでの遅延時間推定による話者位置推定システムについて述べた。今後は雑音除去の精度の向上していく予定である。

### 参考文献

[1]<http://momonga.t.u-tokyo.ac.jp/~ooura/fftman/index.html>