

マイクロフォンアレイを用いた実環境音の認識による3次元空間の音源定位

Estimation of sound source direction in three-dimensional space with microphone array by the recognition of real environmental sounds

坂 文貴*1
Fumitaka Ban

速水 悟*2
Satoru Hayamizu

*1 岐阜大学大学院工学研究科
Graduate School of Engineering, Gifu University

*2 岐阜大学工学部
Faculty of Engineering, Gifu University

We are aiming at sound localization in the three-dimensional space in real time with the microphone array in the real environment. In this study, we constructed the system that estimates the direction of the sound source in real time in three dimension space and recognizes environment sounds. For improving the detection rate of the direction of sound source, we made the system which recognizes arrival sound. And we made which changes frequency region according to arrival sound with the database which made from frequency spectrum and duration of real sound. Additionally, we change frequency region by using data of direction of sound source in case of two sounds etc. In future, we aim to construct the system which estimates the direction of sound sources with sound information which have not only frequency information but also duration and pattern of sound.

1. はじめに

近年、移動ロボット等の環境認識システムなどで、音や音声の到来方向を推定するためのマイクロフォンアレイが数多く報告されている[持木 04]。我々も3次元空間・リアルタイムにおける音源方向推定の研究を報告した[坂 05]。また、実環境における音認識の研究も報告され[比屋根 98]、音認識を使った製品もある[1]。

本研究では、音源方向推定と音の認識を合わせたシステムを構築した。音源方向の検出率向上のために到来音の音認識を行い、実環境音を元にした周波数・音継続時間のデータベースを参照して、音源方向推定に使う周波数帯域を、到来した音に合わせて変更できるようにした。また、このシステムを評価するために、3次元空間における実環境において実験を行った。また、単音だけでなく、角度・方向の情報を利用して、複数音においても、システムの周波数帯域を、到来した音に合わせて変更できるようにした。将来的には、周波数帯域だけでなく、音の継続時間・音のパターンなどの音情報を利用した音源方向推定を行うシステムを目指す。

2. 3次元空間における音源方向推定の概要

本研究では、正20面体の頂点において任意に選ばれた6個のマイク(図1)と音響信号処理装置 rasp[浅野 03]によって3次元空間において音源推定するシステムを構築した。このシステムの特徴は以下のようになっている。また、システムの流れとして、図2のようになっている。水平方向は0度から360度の5度毎に、垂直方向は上・水平・下方向の3方向の音源の推定ができる。

また、本システムは MUSIC 法を使うことにより、複数の音源方向推定が可能である。



図1: 正20面体マイクロフォンアレイ

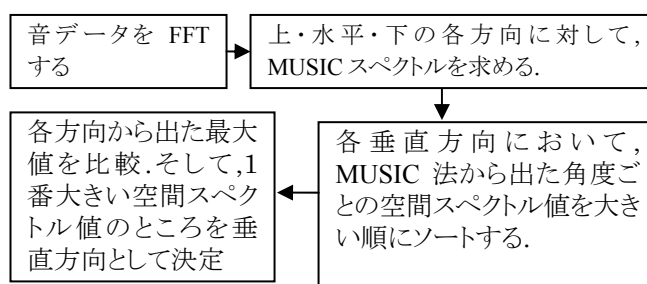


図2: 3次元空間の音源方向推定の流れ

3. 周波数帯域を決定するアルゴリズム

3.1 概要

このシステムでは、音源方向を推定するのに、音の周波数成分を利用している。従来の方法では、認識する周波数帯域を固定して、音源方向の推定を行ってきた。しかし、その方法では、ある一定の周波数の音が流れているときに、空間スペクトルにおいて主となる周波数成分の重みは大きい、それ以外の周波数成分の重みは小さい。また、固定された周波数帯域の内から外へ周波数が遷移する音の場合には、うまく音源方向を推定できない可能性がある。そこで、提案する手法は、音の周波

数スペクトル, 周波数帯域, 音の継続時間長, 音の名前のデータをあらかじめ学習しておく. そして, 音の到来とともに, そのデータからマイクロフォンアレイが認識する周波数帯域をその音に合ったものにチューニングする手法である(図 3). そうすることによって, 限られた周波数帯域で音源方向を推定するので, 音源方向推定率の向上が期待できる.

音のチューニング方法として, 辞書内の周波数スペクトルと, 観測された周波数スペクトルを比べて, 1 番誤差が少ないものをその音として認識させた.

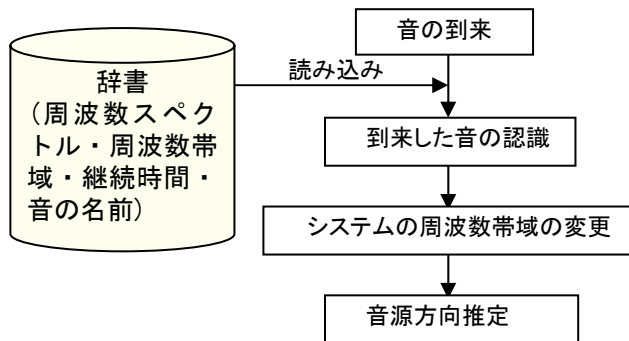


図 3:周波数帯域を決定する流れ

3.2 評価実験

提案手法の有効性を示すために, 実環境において音源方向推定の評価を行った.

(1) 実験手法

音源・マイクロフォンアレイの位置の関係を図 4 に, 目的音は表 1 に掲載する(カッコ内は, 継続時間長). ピアノの音は, RWC 研究用音楽データベース:楽器音[後藤 02]のピアノ(RWC-MDB-2001-W01 No.01)の 3359Hz (ノートナンバー:80)の音を使用して, また, ピッコロの音は, RWC 研究用音楽データベース:楽器音[後藤 02] (RWC-MDB-I-2001-W09 No.32)の 3633Hz (ノートナンバー:82)の音を使用する. 各音源とマイクロフォンアレイとの距離は, 1.5m である. また, 3 次元空間で音源を推定させるために, 各目的音を上方向(higher:1.5m)・水平方向(horizontal:0.9m)・下方向(lower:0.3m)の 3 方向に配置した.

実験で使用する音は, 2 つの発音区間の間に 1 秒ないし 2 秒の非発音区間が含まれている. そして, 発音と非発音の繰り返しを 5 回行う. それを 1 セットとし, 1 つの音に対して 10 セット行う.

なお, これらの目的音を選んだ理由として, 家庭内実環境で使用されそうなものを使用した. 辞書には, それぞれの音を 1 回鳴らしたときの音データと, 1000Hz から 5000Hz に遷移する FM 音のデータの 5 種類が学習されている.

表 1:使用した音

実験名	使用した音
Source A	ガラスの割れる音(1 秒)
Source B	携帯電話の着信音(1 秒)
Source C	ピアノの音(0.5 秒)
Source D	ピッコロの音(2.5 秒)

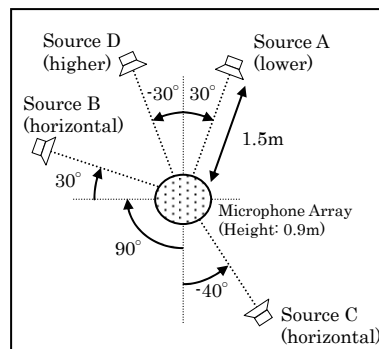


図 4:各音源とマイクロフォンアレイの位置

実験は音源方向推定率向上のため, 次の 2 つの観点から評価する.

評価 1. 音源方向の推定結果が正しいか. (角度と垂直方向の両方が正しければ正解とする.)

評価 2. 空間スペクトルの最大値と平均値との差が従来法と比べて向上するか.

評価 2 は空間スペクトルの最大値と平均値との差が大きいほど空間スペクトルの山の頂点が見つけやすい. 評価式はそれぞれ,

$$\text{評価 1. 検出率}[\%] = \frac{\text{発音区間の推定方向の正解区間}}{\text{全発音区間}} \times 100$$

$$\text{評価 2. } \frac{A(\text{提案法})}{A(\text{従来法})} [\text{倍}]$$

A:空間スペクトルの最大値と平均値の差

で行う. 従来法と提案法の実験は, 同時に同音で行う. また, 空間相関行列を計算する時間間隔は, 従来法, 提案法とも 0.5 秒で行う. さらに, 従来法の周波数帯域は, 1000Hz-5000Hz の間で固定して行う.

(2) 実験結果・考察

実験結果は以下の図 5・表 2 のとおりになった.

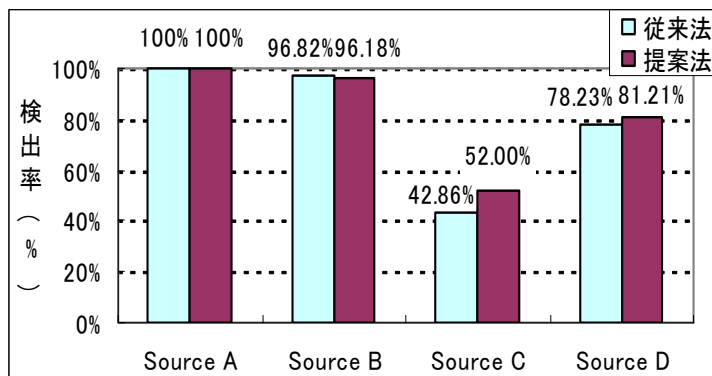


図 5:評価 1 の結果

表 2:評価 2 の結果

実験名	空間スペクトルの差
Source A	1.29 倍
Source B	1.05 倍
Source C	1.01 倍
Source D	1.03 倍

提案法は従来法と比べて、ピアノ・ピッコロの音において、検出率がそれぞれ約 10%・3%向上した。それに加えて、ガラスの音では、空間スペクトルの最大値と平均値との差が従来法に比べて、約 29%向上した。その他も従来法と比べて、その差が 100%を超えることができた。また、ガラスの割れる音、携帯電話の着信音の検出率は従来法・提案法ともによかった。ピアノの音は検出率が悪かったが、5 度の誤差を含めた検出率は、提案法が 100%、従来法が 98%になった。これは、音の継続時間が短いことが起因していると思われる。

評価 2 において、ガラスの音の割合が向上した理由として、従来法の周波数帯域が 1000Hz-5000Hz であるのに対し、4 つの音とも周波数帯域が狭くなったことが理由に挙げられる。提案法による周波数を自動的にチューニングする音源方向の推定が有効であることが分かる。

4. 方向情報を利用して周波数帯域を決定するアルゴリズム

4.1 概要

上で述べた音の認識を行い、周波数帯域を変更して音源方向の推定を行ったが、上記の方法では単音のみで有効なシステムだった。複数の音で音の認識を行い、音源の方向推定を行うと、音同士が干渉してうまく音の認識ができず、音の認識による音源方向の推定率の向上がみられない可能性がある。そこで、その音の干渉を和らげるために、1 ターン前に方向推定した音の位置情報を利用して、各方向の周波数スペクトルと到来音とのマッチングを行なって音の認識を行い、周波数帯域を変更するようにした。(図 6)

4.2 評価実験

(1) 実験手法

音源・マイクロフォンアレイの位置関係を図 7 に、目的音は表 3 に掲載する。(カッコ内は各音の最大周波数成分である。) 各音源とマイクロフォンアレイとの距離は、1.5m である。

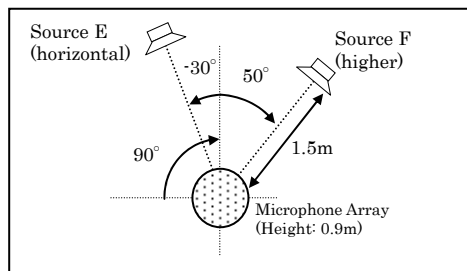


図 7: 各音源とマイクロフォンアレイの位置

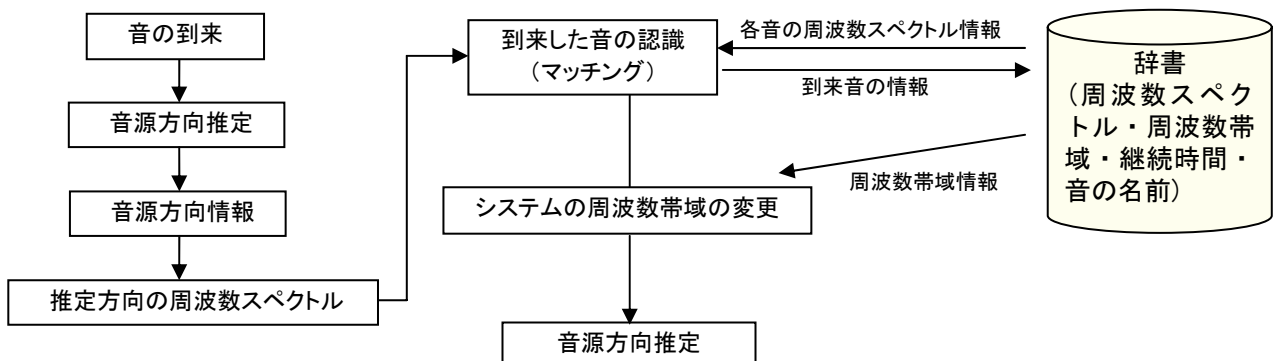


図 6: 周波数帯域を決定する流れ

表 3: 使用した音

実験名	音源位置	使用した音
実験 1	Source E	携帯電話の着信音 (1200Hz)
	Source F	携帯電話の着信音 (3300Hz)
実験 2	Source E	携帯電話の着信音 (2110Hz)
	Source F	携帯電話の着信音 (2630Hz)

実験で使用する音は、発音区間は全て 1 秒で、2 つの発音区間の間に 1 秒の非発音区間が含まれている。そして、発音と非発音の繰り返しを 5 回行う。それを 1 セットとし、1 つの音に対して、10 セット行う。

辞書は、携帯電話の着信音で、変調した音を作成した。各実験の辞書は次の表 4 の通りである。表 4 の値は各音の最大周波数成分である。

表 4: 辞書に収録した音

辞書名	実験 1	実験 2
01	1200Hz	2110Hz
02	1800Hz	2410Hz
03	2110Hz	2630Hz
04	2630Hz	2885Hz
05	3300Hz	3300Hz

実験は前章同様、次の 2 つの観点から評価する。

評価 1. 音源方向の推定結果が正しいか。(角度と垂直方向の両方が正しければ正解とする。)

評価 2. 空間スペクトルの最大値と平均値との差が従来法と比べて向上するか。

評価方法はそれぞれ、

評価 1.
$$\text{検出率}[\%] = \frac{\text{発音区間の推定方向の正解区間}}{\text{全発音区間}} \times 100$$

評価 2.
$$\frac{A(\text{提案法})}{A(\text{従来法})} [\text{倍}]$$

A: 空間スペクトルの最大値と平均値の差

で行う。従来法と提案法の実験は、同時に同音で行う。また、空間相関行列を計算する時間間隔は、従来法、提案法とも 0.5 秒で行う。さらに、従来法の周波数帯域は、1000Hz-5000Hz の間で固定して行う。

(2) 実験結果・考察

実験結果は以下のとおりになった。

評価 1 では、実験 1・2 の両方とも提案法・従来法ともに 100%を示した。

評価方法 2 においては、以下の表 5 の通りになった。

表 5: 評価方法 2 の結果(1)

実験名	空間スペクトルの差
実験 1	1.22 倍
実験 2	1.01 倍

また、2 番目に大きい空間スペクトルの値についても評価方法 2 で評価した。評価結果を表 6 に示す。

表 6: 評価方法 2 の結果(2)

実験名	空間スペクトルの差
実験 1	1.03 倍
実験 2	1.01 倍

実験 1 では、空間スペクトルの最大値と平均値の差が従来法と比べて最大約 22% 向上した。また、その他の評価でもわずかながら 100% を越えることができた。

5. まとめ

本報告では、3 次元空間・実時間において、あらかじめ入力された音データを元に、音源推定に使う周波数帯域を変えることができるシステムの評価実験を行い、その結果について述べた。この結果、複数の音を推定する時においては、従来の周波数帯域を固定する方法より、音源方向の検出と音の認識が 100% を維持することができ、かつ空間スペクトルにおける最大値と平均値との差を大きくすることができ、音源方向を推定する性能は向上した。

今後の課題としては、雑音環境下での提案法の有効性を示し、人間の発話や同じ物体の異なった音や複数の音源があったときの音源分離を用いた音源方向の推定の検討、環境音のデータベースを増やしての提案手法の評価などが挙げられる。

謝辞

音源推定の方法・キャリブレーションなどをご指導いただいた、産総研の浅野太氏、筑波大の山本潔氏、マイクロフォンアレイのハードウェア rasp を開発された日本電子システムテクノロジー株式会社の古河弘光氏、釜島力氏に感謝します。

参考文献

- [持木 04] 持木南生也, 関矢俊之, 小川哲司, 小林哲則: "ロボット頭部に設置した 4 系統指向性マイクロフォンによる音源定位および混合音声認識", 人工知能学会 第 20 回 AI チャレンジ研究会, SIG-Challenge-0420-4, pp.21-26, Dec. 2004.
- [坂 05] 坂文貴, 速水悟: "マイクロフォンアレイを用いた実環境音の認識による音源定位", 日本音響学会講演論文集 I, 2005 春季, pp.615-616
- [比屋根 98] 比屋根一雄, 澤部直太, 飯尾淳: "擬音語表現に基づく衝突音認識システム", 日本音響学会秋季研究発表会, 1998, pp.135-136
- [1] リオン株式会社 "おしらせらんぷ"
<http://www.rion.co.jp>
- [浅野 03] 浅野太, 古河弘光, 釜島力: "マイクロフォンアレイ信号処理用ハードウェアの試作", 日本音響学会講演論文集 I, 2003 秋季, pp.499-500

[後藤 02] 後藤真孝, 橋口博樹, 西村拓一, 岡隆一: "RWC 研究用音楽データベース音楽ジャンルデータベースと楽器音データベース", 情報処理学会 音楽情報科学研究会 研究報告, 2002-MUS-45-4, Vol.2002, No.40, pp.19-26, May 2002.