

# 汎用型マイクロホンを活用した空間情報抽出と音響シーン分析への応用

Acoustic scene analysis using spatial information extracted by general-purpose distributed microphone microphones



**井本 桂右** (Keisuke IMOTO, Ph. D.)  
同志社大学 理工学部 情報システムデザイン学科  
准教授  
(Associate Professor, Faculty of Science and Engineering,  
Doshisha University)  
IEEE、APSIIPA、日本音響学会、電子情報通信学会 他  
受賞：電気通信普及財団賞 (2018 年)、日本音響学会論文賞佐藤賞 (2020  
年)、日本音響学会栗屋潔学術奨励賞 (2013 年) 他  
著書：音響学入門ペディア, コロナ社 (2017 年)  
研究専門分野：音響信号処理

## あらまし

スマートホンやスマートスピーカ、IoT 機器などの普及により、気軽に利用可能な汎用型のマイクロホンが爆発的に増加しており、これらのマイクロホンを利用することで、場所や時間を問わず音を収録可能な環境が整ってきた。本研究では、これら汎用型マイクロホンを複数用いて、音の発生場所や音の広がり方などの空間的なパターン情報を抽出するための新しい信号処理技術の確立を目指す。とりわけ、位置が未知でマイクロホン間の時間同期が保証されていない汎用型マイクロホンの有効活用を目的として、空間ケプストラムとグラフケプストラムと呼ばれる新たな空間音響特徴量を提案し、その有効性を示す。また、音の発生場所や音の広がり方などの空間的なパターンから、周囲の環境（シーン）を理解する音響シーン分析への応用についても議論し、提案する空間ケプストラムやグラフケプストラムの有効性を示す。

## 1. 研究の目的

本研究では、スマートホンやスマートスピーカなどを組み合わせた複数の汎用型マイクロホンから音の発生場所や音の広がりかたといった音の空間的なパターン情報を抽出する手法の確立を目指す。従来の音源定位手法はマイクロホンの位置情報が既知であること

を仮定しており、さらに、複数のマイクロホンが正確に時間同期されている必要があった。そのため、正確な位置情報の取得が困難であったり、マイクロホン間の正確な時刻同期が困難であるスマートホンや IoT 機器などの汎用機器で構成されたマイクロホンアレイには適用困難という問題があった。そこで本研究では、マイクロホンの位置情報を利用しない新たな空間音響特徴量の抽出方法を提案し、その有効性について検討を行う。また、音の発生場所や音の広がり方などの空間的なパターンから、周囲の環境（シーン）を理解する音響シーン分析への応用についても議論する。

## 2. 研究の背景

スマートホンやスマートスピーカ、IoT 機器、情報家電の普及により、気軽に利用可能な汎用型マイクロホンが爆発的に増加しており、場所や時間を問わず音を収録可能な環境が整ってきた。加えて、近年の機械学習理論や計算機環境の発展に伴い、音響分野で扱われる研究対象が、音声や音楽から、あらゆる音（環境音）へと広がってきている。これらを背景として、様々な場所/時間において発生する、あらゆる音事象（音響イベント）を分析する音響イベント検出や、人間のように音から周囲の環境（シーン）を理解する音響シーン分析の研究に関心が高まっている[1][2]。音響イベント検出や音響シーン分析技術が実現されることで、高齢者の見守りやセキュリティシステム、自動運転、ライフログの自動作成、動画の自動タグ付けなど様々な用途に貢献すると期待されている。実際、国内外の学会において音響イベント検出や音響シーン分析の発表件数は飛躍的に増加しており、例えば、信号処理のトップカンファレンスである International Conference on Acoustics, Speech, and Signal Processing (ICASSP)では、2005 年には関連研究の発表件数が 5 件であったが、2020 年には 30 件近くの発表が行われた。これは音響信号処理の発表では最も発表数が多いトピックであり、関心の高さが伺える。

## 3. 研究の方法

### 3.1 従来手法とその問題点

音声の分析や音響イベント検出、音響シーン分析の

# 汎用型マイクロホンを活用した空間情報抽出と音響シーン分析への応用

Acoustic scene analysis using spatial information extracted by general-purpose distributed microphone microphones

研究では、音の周波数情報とパターン認識のアプローチ

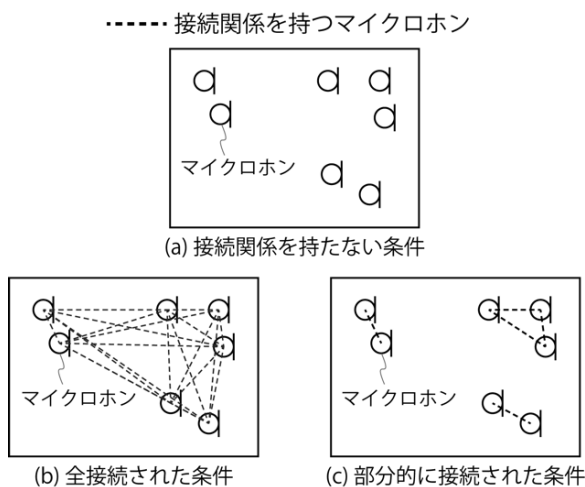


図1 マイクロホンの接続の例

チに基づく手法が盛んに研究開発されている。しかしながら、環境音を分析するためには、周波数情報からどのような音が発生したかを分析するだけでなく、それがどの場所で発生したか、どの程度広がりを持った音かなど、音の空間情報を分析することも重要と言える。例えば、「ガラスが割れる音」が窓から発生する場合とテレビから発生する場合ではその意味は大きく異なり、これらは区別される必要がある。また、公共野スペースでの混雑具合を分析するには、単一の人の足音がある方向から聞こえてくるのか、複数の人の足音がいろいろな方向から聞こえてくるかなど、音の空間的な広がりがある有益な情報となる。音から空間情報を取得するためのマイクロホンアレイ信号処理技術として、これまでも、音の発生位置を推定する音源定位に関して研究が多数行われている。しかしながら、従来の音源定位手法はマイクロホンの位置情報が既知であることを仮定しており、さらに、複数のマイクロホンが正確に時間同期されている必要があった。そのため、正確な位置情報の取得が困難であるスマートホンやIoT機器などの汎用機器の組合せで構成されたマイクロホンアレイには適用困難という問題があった。そこで本研究では、正確に同期され、位置情報も既知である専用のマイクロホンを用いず、スマートホンやIoT機器に備えられた複数の汎用型マイクロホンを利用して音の空間情報を取得する技術の実現を目指す。

## 3. 2 空間ケプストラム

音声の分析や音響イベント検出、音響シーン分析研究において、周波数情報として頻繁に利用され、観測誤差や背景雑音に対しても頑健な特徴量であるケプストラム\*1を空間情報の抽出に応用する手法を検討した。ケプストラムの考え方を空間特徴量抽出に拡張することで、図1(a)のように正確な時間同期が行われておらず、位置情報も不明なマイクロホンから空間特徴量を頑健に取得できる。なお、本稿では時間同期が行われていたり、近接して配置されていることが明らかになっているマイクロホンを「接続された」マイクロホンと称する。また、ケプストラムを空間情報の抽出に応用することで、これまでに多くの研究が行われてきた、(周波数領域における)ケプストラムに関する知見を空間情報の抽出に利用できるという大きな利点もある。具体的には、空間ケプストラムは以下の特徴を持つ。

- ・マイクロホン間の同期誤差に敏感な位相情報は利用せず、振幅情報のみを利用
- ・空間情報の抽出に際し、マイクロホンの位置情報が不要
- ・雑音成分を直交変換により分離し、雑音の影響を低減可能
- ・ケプストラム特徴量に適用可能なケプストラム平均正規化/分散正規化などの手法が適用可能

空間ケプストラムを抽出するためには、マイクロホンアレイで得られた信号から、対数振幅ベクトルを算出し、主成分分析 (PCA) による直行変換を適用すれば良い。図2に空間ケプストラムの変換行列の例を、図3に空間ケプストラムの変換行列の各重み係数をマイクロホン位置に重ねた結果を示す。なお、本実験では、13個のマイクロホンを使用して空間ケプストラムを算出した。各空間ケプストラム係数はそれぞれのマイクロホンでの観測と重み係数の線形和で表現されることを踏まえると、1次の空間ケプストラム係数は、収録音全体の平均的な音圧レベルを表現していることが分かる。2-5次の重み係数は、近接して配置されているマイクロホンの重みが同じ符号となっており、大まかな空間のパターンが取得できていることが見て取れ

# 汎用型マイクロホンを活用した空間情報抽出と音響シーン分析への応用

Acoustic scene analysis using spatial information extracted by general-purpose distributed microphone microphones

る。また、高次の空間ケプストラム係数では、近接し

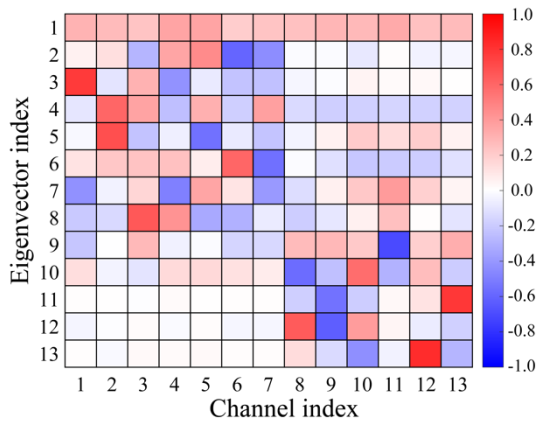


図2 空間ケプストラムの変換行列の例

図1(c)に示すように、部分的に位置情報が既知である

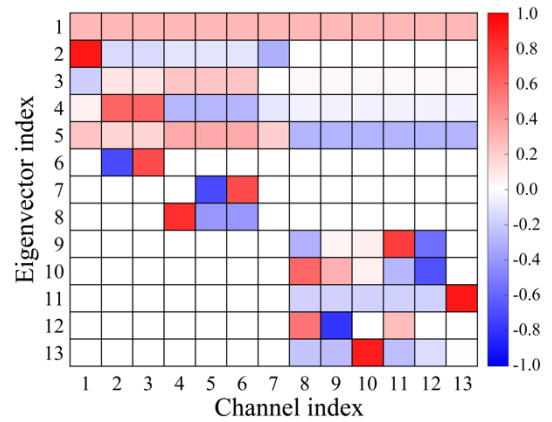


図4 グラフケプストラムの変換行列の例

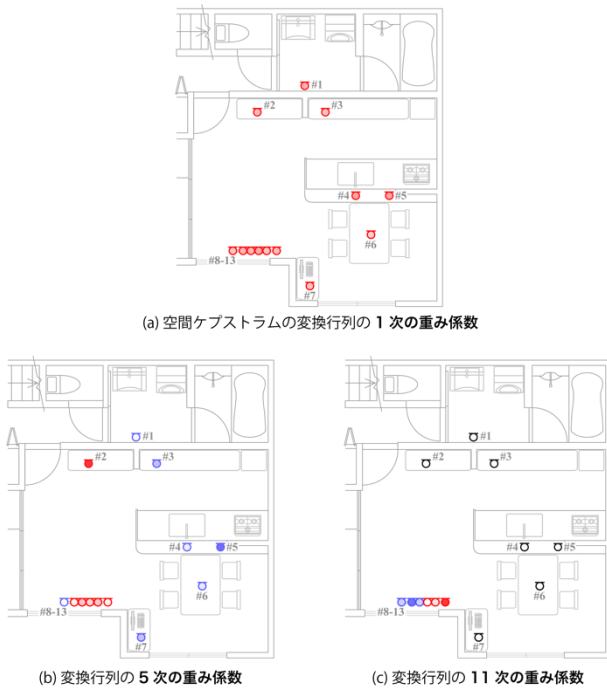


図3 空間ケプストラムの変換重みの例

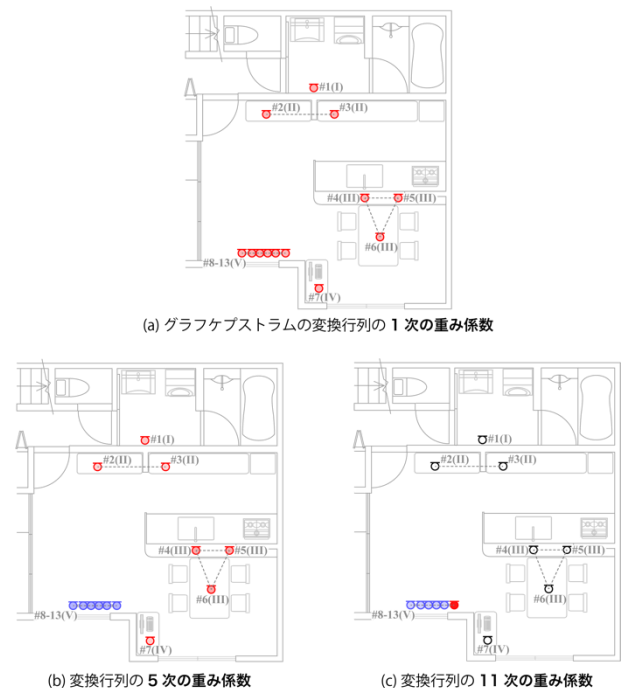


図5 グラフケプストラムの変換重みの例

て配置されているマイクロホンの重みの符号が逆転していることから、微細な空間パターンを取得していることが分かる。このように、空間ケプストラムを用いることで、マイクロホンの位置情報を利用していないにも関わらず、空間特徴量が取得できるという利点がある。

### 3.3 グラフケプストラム

今後の測位技術やネットワーク環境の向上により、

マイクロホンや部分的に同期の取れたマイクロホンを組合せて空間情報を取得する場合も想定される。その場合、空間ケプストラムでは空間情報抽出に有益な位置情報や、マイクロホン間の同期に関する情報を破棄することになる。そこで、マイクロホン間の任意の共通部分（位置の近さや同期関係）を陽に指定し、その構造を考慮した信号処理を行うことで、空間情報抽出の性能向上が可能になると考えられる。本研究では、グラフ信号処理<sup>\*2</sup>[3]とケプストラムを組合せた新

# 汎用型マイクロホンを活用した空間情報抽出と音響シーン分析への応用

Acoustic scene analysis using spatial information extracted by general-purpose distributed microphone microphones

たな特微量として、グラフケプストラムという手法を

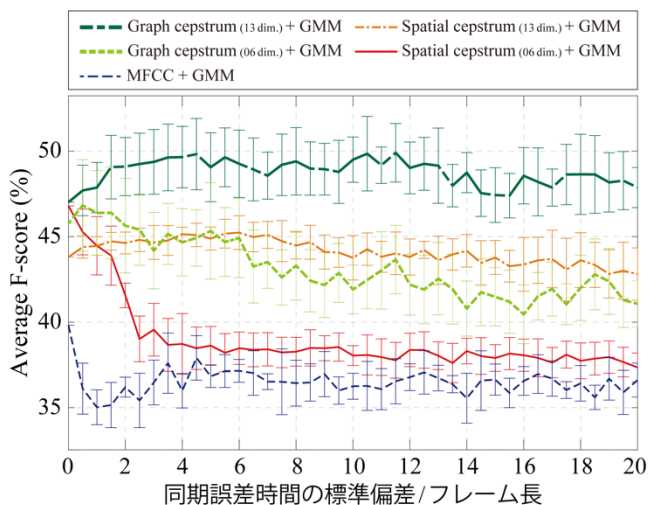


図 6 空間ケプストラムおよびグラフケプストラムを用いた音響シーン分類の性能

提案する[4][5]。具体的には、グラフケプストラムは以下の特徴を持つ。

- ・マイクロホン間の同期誤差に敏感な位相情報は用いず、振幅情報のみを利用
- ・マイクロホン間の共通する構造を考慮可能なグラフフーリエ変換をケプストラムと組合せる
- ・マイクロホン位置情報は必ずしも必要ではない (部分的に既知の場合は考慮することが可能)
- ・事前に大量の学習データを用意不要

グラフケプストラムを抽出するためには、マイクロホンアレイで得られた信号から、主成分分析 (PCA) の代わりにグラフフーリエ変換による直行変換を適用する。なお、グラフケプストラムの詳細については文献 [4]-[5] を参照されたし。図 4 にグラフケプストラムの変換行列の例を、図 5 にグラフケプストラムの変換行列の各重み係数をマイクロホン位置に重ねた結果を示す。グラフケプストラムにおいても、空間ケプストラム同様の傾向を示し、空間特微量が取得できている様子が見て取れる。

### 3. 4 空間情報を用いた音響シーン分析

本研究で提案した空間ケプストラム、および、グラフケプストラムを用いて、音から屋内で発生する 9 つのシーン (料理、洗濯、皿洗い、掃除、読書、会話、

食事、TV 視聴、PC の操作) を分類する実験を行った。詳細な実験条件は文献 [4]-[5] に記載されている。接続されていないマイクロホン間の時間同期ずれを様々に変更させながらシーン分類性能を評価した結果を図 6 に示す。比較手法の音響特徴としてメル周波数ケプストラム係数 (MFCC) を用い、音響シーン分類器として混合ガウスモデル (GMM) を用いた。図 6 より、空間ケプストラムやグラフケプストラムなどを用いることで、周波数情報を用いた場合と同等以上の性能でシーンを分類できていることが確認できる。さらに、グラフケプストラムでは、マイクロホン間の同期誤差時間が大きくなった場合でも大幅な性能劣化を引き起こすことなくシーンを分類できていることが確認できる。この結果より、マイクロホン間の部分的な接続関係をグラフケプストラム効果的に利用できていることが確認される。

### 4. 将来展望

音メディアは情報通信技術の重要な応用先の一つであり、音声通話や映像コミュニケーション、音声認識を用いた検索サービス等に利用されている。これらのサービスを実現するためには音声符号化やエコーキャンセラ、音声認識などの音響処理技術が用いられているが、次世代の技術として注目されているのが音響イベント検出や音響シーン分析である。これら技術と各種情報通信技術を組み合わせることにより、人と人、人と機械のコミュニケーションはより多様な形での発展が期待でき、例えば、遠隔見守りや遠隔監視、自動運転、動画コンテンツの検索/推薦、自動ライフログ生成など様々なサービスが実現可能となる。また、音響イベント検出や音響シーン分析が実用化されることで、人の耳のようにあらゆる音を聞き分けられる機械が実現可能となり、実世界における人工知能が飛躍的に人に近づくと期待される。

### おわりに

位置情報が未知であり、同期されていない複数の汎用マイクロホンを利用して空間音響特微量を取得する方法、並びに、音響シーン分析における応用について研究開発を進めた。これらを可能とする技術として、本

# 汎用型マイクロホンを活用した空間情報抽出と音響シーン分析への応用

Acoustic scene analysis using spatial information extracted by general-purpose distributed microphone microphones

研究では空間ケプストラムおよびグラフケプストラムと呼ばれる特徴抽出法について詳細な検討を行った。また、音響シーン分析にこれらの特徴量を用いた際の性能評価を行い、その有効性を確認した。

## 用語解説

- \*1 ケプストラム：音信号の対数パワースペクトルを時間波形とみなし、離散フーリエ変換を施すことでえられる特徴量
- \*2 グラフ信号処理：交通網やソーシャルネットワークなどのグラフで表現された離散信号をグラフの構造を利用して解析する技術

## 参考文献

- [1] 井本 桂右, "音響イベントと音響シーンの分析," 日本音響学会誌, Vol. 74, No. 4, pp. 198-207, 2018.
- [2] 井本 桂右, "音響イベントおよび音響シーン分析研究の基礎と研究動向," 日本音響学会 音声研究会, 2018.
- [3] A. Ortega, P. Frossard, J. Kovačević, J. M. Moura, and P. Vandergheynst, "Graph signal processing: Overview, challenges, and applications, " Proceedings of the IEEE, 106(5), 808-828, 2018.
- [4] Keisuke Imoto, "Acoustic Scene Analysis Using Partially Connected Microphones Based on Graph Cepstrum," Proceedings of European Signal Processing Conference (EUSIPCO), pp. 2453-2457, 2018.
- [5] Keisuke Imoto, "Graph Cepstrum: Spatial Feature Extracted from Partially Connected Microphones," IEICE Transactions on Information and Systems, Vol. E103-D, No. 03, pp.631-638, 2020.

## 関連文献

- [6] Keisuke Imoto and Nobutaka Ono, "RU Multichannel Domestic Acoustic Scenes 2019: A Multichannel Dataset Recorded by Distributed Microphones with Various Devices," Proc. DCASE Workshop, pp. 104-108, 2019.

- [7] 井本 桂右, 小野 順貴, "RU Multichannel Acoustic Scenes 2019: 複数デバイスで構成された分散マイクアレイによる音響シーンのための環境音データセット" 日本音響学会 2020 年春季研究発表会, 2020.

この研究は、平成29年度SCAT研究助成の対象として採用され、平成30～令和元年度に実施されたものです。