

# 雑音下での音源定位・音源分離に与える 伝達関数測定法の影響の評価

赤堀 渉<sup>†</sup> 増田 太郎<sup>††</sup> 奥乃 博<sup>††</sup> 森島 繁生<sup>†††</sup><sup>†</sup>早稲田大学創造理工学部 <sup>††</sup>早稲田大学創造理工学研究科 <sup>†††</sup>早稲田大学理工学術院総合研究所/JST

## 1. 音源定位・音源分離での伝達関数の役割

実環境下で 1m~3m 程度離れた話者、特に、同時発話が想定される状況での音声認識を行うためには、マイクロフォンアレイ (MA と略記) で取得された音から音源定位や音源分離を行う必要がある。MA は通常ロボットの頭部や筐体に組み込まれているので、各マイクロフォンへの伝達関数を活用して部分空間を構築し、処理性能の向上を狙うことができる。

例えば、音源定位法の一つ MUSIC (Multiple Signal Classification) 法では、MA の幾何学的な配置から部分空間を構築する。また、MA の伝達関数を利用すれば、さらに定位精度が向上する[1]。音源分離法の一つ GHSS (Geometric High-order Decorrelation-based Source Separation) でも、同様に伝達関数を利用し、MA の指向性の形成を行い、分離精度の向上を図っている[2]。

MA の伝達関数は、方向を変えてスピーカから流した測定信号を各マイクで収録し、収録音からインパルス応答を求めた上で算出される。インパルス応答はインパルスを演奏し、収録すれば求まるが、理想的なインパルスは演奏できないので、インパルス応答用の測定信号を選択する必要がある。最近では収録音からインパルス応答が計算しやすい TSP (Time Stretched Pulse) 信号が M 系列に代わって使用される。低音域雑音がある場合には TSP では正確なインパルス応答が求まらないので、低音域の性能を上げた Log-SS 信号の使用が提案され、バイノーラル再生に有効なことが示されている[3]。

一方、伝達関数測定用信号が、MA による音源定位・音源分離に与える影響はこれまで未検証である。この検証は、音源定位・音源分離の処理方法とも密接に関連しており、性能の優れたシステムで評価する必要がある。本稿では、MA を用いてロボット聴覚を実現できるオープンソースソフトウェア HARK[4]を使用し、様々な伝達関数測定信号が音源定位・音源分離の処理性能に与える影響について、その調査方法と結果を報告する。

## 2.

### 2.1 伝達関数の測定

伝達関数は理論的にはインパルス応答と同値である。測定信号の測定系より得られるインパルス応答の周波数表現は以下の式で表される。

$$H(k) + \frac{D(k)}{S(k)} + \frac{N(k)}{S(k)} \quad (1)$$

$H(k)$ は被測定系を表す伝達関数は、 $D(k)$ は非線形系により発生した信号、 $N(k)$ は雑音、 $S(k)$ は測定信号を表している。また、式(1)の第2項は非線形誤差であり、第3項は雑音性誤差を表す。ここで、非線形誤差の現れ方は測定信号の位相特性に依存し、雑音性誤差の現れ方は測定信号の振幅特性に依存する。したがって、応用で許容される測定誤差の性質や大きさに基づいて測定信号を設計し、選択する必要がある。

### 2.2 インパルス応答を求める測定信号の種類

インパルス応答を測定するためにスピーカから発生させる測定信号は次のように分離できる[5]:

種類	対雑音	測定信号	特徴
掃引 正弦波	固定型	upTSP	負の時間方向に高調波歪み が出現
		downTSP	正の時間方向に高調波歪み が出現
		Log-SS	高調波歪みを除去
疑似 雑音		M 系列	波高率が低い
掃引 正弦波	適応型	MNSS	雑音性誤差のエネルギーを 最小化
		NWSS	雑音のパワースペクトルに 比例したパワースペクトル

測定信号は位相特性により、掃引正弦波 (時間と共に周波数が上昇又は下降する正弦波信号) と疑似雑音 (ランダム雑音のような波形を持った周期信号) に大別できる。前者には測定信号は振幅特性を変化させて、周波数パワーを特定の周波数に集中できる。これにより、SN 比が改善し、その結果、雑音性誤差を小さくすることができる。LogSS 信号は、低周波数域の掃引時間が長い信号であるから、低周波数の SN 比を改善する効果がある。また、周囲環境の雑音に応じて振幅特性を変化させる測定信号を適応型測定信号と呼び、代表的な例として、MNSS 信号や NWSS 信号などが挙げられる。これら



図1. 測定信号の収録とMAの配置

の信号では、SN比の改善がさらに期待できる。

図1に測定信号の収録の様子とMAの配置を示す。ロボットHIROの頭部にデジタルMEMSマイクを8本円周上にほぼ等間隔に配置した。MAの中心から1.5m離れた高さ1.5mの位置に置いたスピーカ(Generec 1029A)を使い、測定は5度間隔で行い、収録音はRASP-ZXでWaveファイルに変換し、PCにUSB経由で転送した。収録データからMATLABを用いて、インパルス応答を求めた。インパルス応答から伝達関数への変換はHARKToolを使用した。

### 2.3 音源定位による伝達関数の評価

まず、音源定位の性能比較実験を行うために、1)音源の作成方法、2)音源定位手法、について述べる。

音源として、周波数帯域ごとの定位精度を評価するために、0~8000Hzをmelスケールで20分割した狭帯域信号を使用した。その狭帯域信号に対し、2.1節で得られた特定の方向のインパルス応答を畳み込むことによって、方向性のある音源を仮想的に作成することができる。したがって、音源定位結果の正解方向は、畳み込んだインパルス応答の方向となる。

音源定位手法については、MUSIC法を用いる。MUSIC法とは、以下の式からスペクトルを算出することで、音源の方向を推定する手法である。まず、MAの観測信号の相関行列を固有値分解する。

$$R = XX^H \quad (2)$$

$$R = E\Lambda E^{-1} \quad (3)$$

$$E = [e_1, e_2, \dots, e_M] \quad (4)$$

$$\text{diag}(\Lambda) = [\lambda_1, \lambda_2, \dots, \lambda_M] \quad (5)$$

$R$ は入力信号の相関行列、 $E$ は固有ベクトル、 $\lambda$ は固有値を表している。また、各音源は固有ベクトルと固有値( $e_i, \lambda_i$ )として表される。このとき、固有値が大きい場合は音源を意味し、固有値が小さい場合は雑音を意味すると仮定される。以降、音源数は既知であるという前提を置く。

$$P(\theta) = \frac{|H^H(\theta)H(\theta)|}{\sum_{i \in \text{雑音成分}} |H^H(\theta)e_i|^2} \quad (6)$$

$P(\theta)$ はMUSICスペクトル、 $H(\theta)$ は計測した伝達関数、 $e_i$ は入力音源信号の固有ベクトル、そして添字の $H$ はエルミート転置を表している。

## 3. 音源定位による性能比較の解析と考察

### 3.1 音源定位による性能比較の解析結果

音源定位の精度は、高橋ら[6]の定位誤り分類に従う。リファレンスから $\pm 5^\circ$ 以内に定位される音源を正解、 $\pm 10^\circ$ に定位される音源を置換誤り、それ以上離れて定位された音源を挿入誤りとした。正解音源数を $N$ 、挿入誤り音源数を $I$ 、置換誤り音源数を $S$ とすると、音源正解率 $C$ と音源正解精度 $A$ は以下の式で定義される。

$$C = (N - S)/N \times 100 \quad (6)$$

$$A = (N - S - I)/N \times 100 \quad (7)$$

表1に、帯域190Hz~431Hzの信号を音源とし、MUSICスペクトルの閾値を35.5とし、正解方向角度が $10^\circ$ と $-10^\circ$ であるような2音源の音源定位結果から得られる測定信号による性能差を示す。

表1. 測定信号による性能差

測定信号	N	I	S	C (%)	A (%)
upTSP	7184	187	1496	79.2	76.6
downTSP	2960	5	289	90.2	90.1
M系列	3854	378	228	94.1	84.3
LogSS	5274	443	1487	71.8	63.4
MNSS	2931	635	1146	60.9	39.2
NWSS	3728	287	1826	51.0	43.3

### 3.2 伝達関数測定信号についての考察

upTSP信号、downTSP信号、M系列信号の定位精度が高いことが分かる。これらに共通する性質として、全周波数帯のパワーが一定であることである。一方で、LogSS信号、MNSS信号、NWSS信号のように低域にパワーが強くなる信号では、定位性能が低くなった。

## 4. まとめと今後の課題

本研究では、様々な種類の測定信号を用いることで生じる伝達関数測定法の違いが、音源定位に与える影響を明らかにした。今後、調波構造や音声の音源定位性能を評価するとともに、音源分離精度の評価を行う。本研究の一部は、科研費基盤研究No.24200062とImPACT「タフ・ロボティクス・チャレンジ」の支援を受けた。

### 参考文献

- [1] K. Nakamura et al., "Intelligent Sound Source Localization for Dynamic Environments," *ICASSP-2009*, 664-669.
- [2] H. Nakajima et al., "Blind Source Separation with Parameter-Free Adaptive Step-Size Method for Robot Audition," *IEEE TASP*, **18**(6):1476-1485, 2010.
- [3] 藤本ら, "低域バンドでのSN比改善を目的としたTSP信号に関する検討-高調波歪の除去," 音講論集, 555-556, 2000.
- [4] 奥乃ら, "ロボット聴覚オープンソフトウェアHARK", 日本ロボット学会誌, **28**(1):6-9, 2010.
- [5] 金田, "インパルス応答測定信号と測定誤差", 日本音響学会誌, **69**(10) 549-554 (2013).
- [6] 高橋ら, "実環境下での音源定位・音源検出の検討", *RSJ2011ACIF3-3*.