

時変拡張ビームフォーミングによる移動音源の抽出

Moving Sound Source Extraction by Time-Variant Beamforming

中島 弘史 中臺 一博 長谷川 雄二 辻野 広司
Hirofumi Nakajima Kazuhiro Nakadai Yuuji Hasegawa Hiroshi Tsujino

(株) ホンダ・リサーチ・インスティテュート・ジャパン
Honda Research Institute Japan Co., Ltd.

In this paper, we propose a sound signal extraction method for a moving source. For natural human-robot interaction, a robot should recognize various types of sounds, which are speech, noise, music and so on. In a real environment, there are many moving sound sources that have important information. However, it is difficult to recognize the moving sound sources due to their amplitude and frequency distortions by the movement. For this problem, we propose time-variant beamforming that compensate the amplitude and frequency distortions precisely. Numerical experiments showed that our proposed method improves moving sound source extraction.

1. はじめに

実環境で人とロボットがインタラクションを行うためには、ロボットは音声だけではなく、様々な音を含む周囲環境を認識できる必要がある。周囲環境中の音を大きく移動音源と静止音源とに分けた場合、移動音源は時間的変化を予測する手がかりとなり、静止音源は空間的予測の手がかりとなる。この2つを統合的に認識し、周囲環境の時空間的予測ができることで、環境変化にロバストで、人に有益なヒューマン・ロボット・インタラクションが実現できると考える。このなかで、静止音源認識に関しては従来研究がある[橋爪 2006]が、移動音源に関してはその抽出方法に課題があり、報告例が少ない。本稿ではこの移動音源の抽出に焦点を当てる。移動音源の抽出が困難である理由は、移動により音源信号の振幅や周波数が変化するためである。そこで、この移動による変化を補正し、正確な音源信号を抽出可能な新たなビームフォーミング法(BF)を提案する。提案法は、音源位置を離散化し、離散化した各位置で BF 係数を切り替えて行う従来法と比べると、①係数切替えによる不連続がない、②ドップラー効果による周波数変化および③振幅変動を別途補正する必要がなく、理論上誤差なく推定できるという特長がある。

2. 時変システムと時変畳み込み演算

移動音源の位置を $\mathbf{p}(t)$ 、その信号(体積速度)を $s(t)$ とすれば、位置 \mathbf{q} における観測信号(音圧) $x(t)$ は、次式で与えられることが移動音源の波動方程式[Morse 1968]から導出できる。

$$x(t) = \int s(t_s) h(t - t_s, \mathbf{p}(t_s)) dt_s \quad (1)$$

ここで $h(t, \mathbf{p})$ は、位置 \mathbf{p} にある静止音源から観測位置 \mathbf{q} までのインパルス応答である。この式は移動音源であっても、音源が取りうる各位置からの静止のインパルス応答が既知であれば、その応答と音源信号から出力を求められる事を示す。本稿では式(1)を時変畳み込み演算と定義する。この演算は離散システムにおいても、式(1)と同様に近似的に次式で計算できることが実験的に示されている[奥山 2006]。

$$x(k) = \sum_{k_s=0}^{\infty} s(k_s) h(k - k_s, \mathbf{p}(k_s)) \quad (2)$$

ただしサンプリング周波数は、移動によるドップラー効果を加味した音源の上限周波数の2倍より大きく定める必要がある。式(2)はベクトルと行列を用いて次式で表現できる。

$$\mathbf{x} = \mathbf{H} \mathbf{s} \quad (3)$$

$$\mathbf{s} = [s(1) \quad s(2) \quad \cdots \quad s(L_s)]^T$$

$$\mathbf{x} = [x(1) \quad x(2) \quad \cdots \quad x(L_s + L_h - 1)]^T$$

$$\mathbf{H} = \begin{bmatrix} h(1, \mathbf{p}(1)) & 0 & \cdots & 0 \\ h(2, \mathbf{p}(1)) & h(1, \mathbf{p}(2)) & \ddots & \vdots \\ \vdots & h(2, \mathbf{p}(2)) & \ddots & 0 \\ h(L_h, \mathbf{p}(1)) & \vdots & & h(1, \mathbf{p}(L_s)) \\ 0 & h(L_h, \mathbf{p}(2)) & & h(2, \mathbf{p}(L_s)) \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & h(L_h, \mathbf{p}(L_s)) \end{bmatrix}$$

ここで \mathbf{s} は音源信号ベクトル、 \mathbf{x} は受信信号ベクトル、 \mathbf{H} は時変畳み込み行列[Mastumoto 2003]である。音源の移動パターンを離散時間ベクトル \mathbf{k} の位置ベクトル関数 $\mathbf{p}(\mathbf{k})$ として定義すれば、 \mathbf{H} は、移動パターン $\mathbf{p}(\mathbf{k})$ と受信点 \mathbf{q} により定まる。なお離散時間の原点は 1、音源信号長は L_s 、インパルス応答長は L_h とした。

3. 移動音源に対する高精度 BF

提案する BF は、任意の伝達系で利用できる BF 設計法[中島 2006]を時変システムへ拡張したものである。図 1 に伝播および BF 処理のシステム図を表す。図 1 で N はマイクロホン数、 \mathbf{s} は L_s 行の音源信号ベクトル、 $\mathbf{x}(n)$ は $L_s + L_h - 1$ 行のマイクロホン n に対する受信信号ベクトル、 $\mathbf{H}(\mathbf{p}(\mathbf{k}), n)$ は $L_s + L_h - 1$ 行 L_s 列の移動パターン $\mathbf{p}(\mathbf{k})$ に対するマイクロホン n までの時変畳み込み行列、 $\mathbf{G}(n)$ は $L_s + L_h - 1$ 行 L_s 列のマイクロホン n に接続される BF の係数行列、 \mathbf{y} は L_s 行の出力ベクトルである。なお係数行列 $\mathbf{G}(n)$ の各列は推定する音源信号の時刻に対応し、各行は観測信号の時刻に対応する。出力 \mathbf{y} は、

$$\mathbf{y} = \mathbf{G}^T \mathbf{x} \quad (4)$$

$$\mathbf{G} = [\mathbf{G}(1) \quad \mathbf{G}(2) \quad \cdots \quad \mathbf{G}(N)]^T$$

$$\mathbf{x} = [\mathbf{x}(1) \ \mathbf{x}(2) \ \cdots \ \mathbf{x}(N)]^T$$

であり、入力信号群 \mathbf{x} は、

$$\mathbf{x} = \mathbf{H}^T(\mathbf{p}(\mathbf{k}))\mathbf{s} \quad (5)$$

$$\mathbf{H}(\mathbf{p}(\mathbf{k})) = [\mathbf{H}(\mathbf{p}(\mathbf{k}), 1) \ \mathbf{H}(\mathbf{p}(\mathbf{k}), 2) \ \cdots \ \mathbf{H}(\mathbf{p}(\mathbf{k}), N)]$$

と書ける。BF 係数は、 $\mathbf{y} = \mathbf{s}$ である必要から、

$$\mathbf{H}(\mathbf{p}(\mathbf{k}))\mathbf{G} = \mathbf{I} \quad (6)$$

の解であり、式(6)を擬似逆行列として求めれば、移動音源に対応した最小ノルム重み付き遅延和 BF の係数が得られる。また非目的音源の移動パターン $\mathbf{p}_U(\mathbf{k})$ とその畳み込み行列 $\mathbf{H}(\mathbf{p}_U(\mathbf{k}))$ が既知であれば、

$$\mathbf{H}(\mathbf{p}_U(\mathbf{k}))\mathbf{G} = \mathbf{0} \quad (7)$$

を順次追加した解を求めることで、死角型やサイドローブ最小型の BF 係数が得られる[中島 2006]。

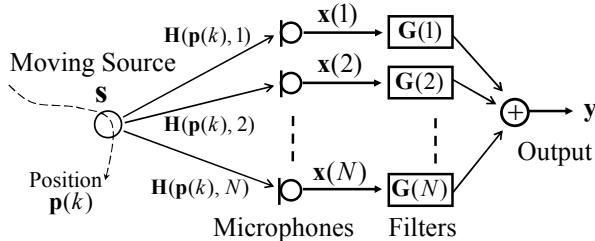


図1 時変伝播系とBF処理系

4. 数値実験

図2に示す設定で数値実験を行った。音源は、移動音源 S1 と静止音源 S2 の2つとした。S1 は 125Hz(図3)の正弦波、S2 は 400Hz の正弦波、サンプリング周波数は 1kHz、S1 の速度は 20m/s、信号長は 0.5s とした。3素子のアレイ(M1~M3)を用いた死角型の BF により、S1 の信号を抽出することを目的とした。従来法は、音源方向を 10 度おきに離散化した各方向で BF を設計し、最大パワーを出力する方向の BF を最終出力とする方法を用いた。提案法では、式(6)と(7)の連立式の解から求めた。図4に式(3)により合成した M1 での観測信号を示す。S2 の信号(高周波)と、振幅と周波数が変化する S1 の信号が見られる。図5に BF の出力信号を示す。(a)の従来法は S2 の高周波成分は除去されるものの、S1 の信号は移動による振幅変化や切り替えによる不連続が残るのに対し、(b)の提案法ではこれらがなく正確に S1 の信号が再現されている。なお従来法の 0.3s 付近の振幅が低い原因は焦点と死角の方向が近いためである。表1は、従来法と提案法による音源波形を、次式の推定誤差 E (dB) で評価した結果である。

$$E = 10 * \log_{10} \left(\frac{\|\mathbf{s} - \hat{\mathbf{s}}\|^2}{\|\mathbf{s}\|^2} \right)$$

ここで、 \mathbf{s} は真の音源信号、 $\hat{\mathbf{s}}$ は推定した音源信号である。表1より、提案法は雑音が無い理想的な環境では完全な再現ができること、入力信号に雑音を含む場合でも従来法より高い精度で再現できることが確認できる。

5. おわりに

本稿では移動による振幅や周波数のひずみを補正し、音源信号を正確に抽出する手法を提案し、数値実験によりその有効性を示した。

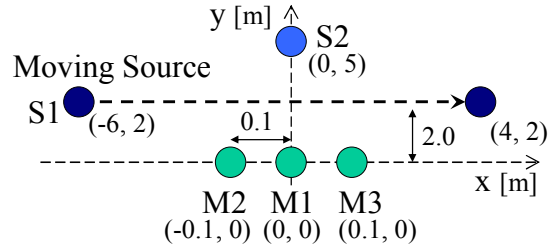


図2 数値実験でのレイアウト

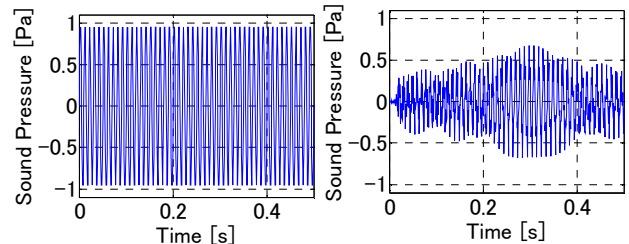
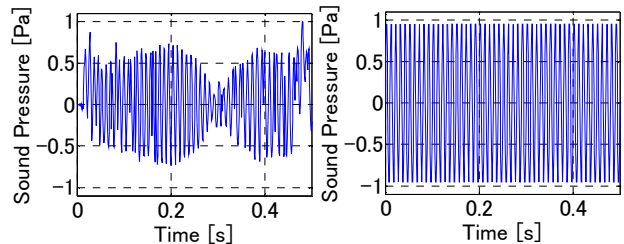


図3 音源信号

図4 観測信号(M1)



(a) 従来法

(b) 提案法

図5 BF出力結果

表1 推定誤差 E (dB)

	雑音なし	雑音あり (-40dB)
従来法	0.6 dB	1.0 dB
提案法	-26.0 dB	-14.3 dB

参考文献

- [橋爪 2006] 橋爪 亜希, 西村 竜一, 入野 俊夫, 河原 英紀, "ロボット対話のための雑音認識手法に関する検討", 日本音響学会秋季大会, 2-P-21, 2006
- [Morse 1968] P. M. Morse and K. U. Ingard, "Theoretical Acoustics", Princeton, USA, pp. 717-732, 1968.
- [奥山 2006] 奥山 智尚, 松久 寛, 宇津野 秀夫, "伝達関数補間を用いた仮想空間移動時の音圧計算", 騒音振動研究会報告 N-2006-46, 2006.
- [Mastumoto 2003] M. Matsumoto, M. Tohyama and H. Yanagawa, "A method of interpolating binaural impulse responses for moving sound images," Acoust. Sci. & Tech. 24, 5, pp. 284-292, 2003.
- [中島 2006] 中島 弘史, "不定項を利用した平均サイドローブ最小ビームフォーミングの実現", 日本音響学会誌 62 巻 10 号, pp. 726-737, 2006.