社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

2マイクロホンの指向性制御による音源分離の一提案

青木　繁治†　　陶山　健仁†

† 東京電機大学 工学部 電気電子工学科
〒 120–8551 東京都足立区千住旭町 5
E-mail: †13kme01@ms.dendai.ac.jp

あらまし　本研究では，2 マイクロホンによる 2 音源分離について検討する．検討手法では，指向性形成に注目する．2 マイクロホンの場合，一方の音源方向にピーク，他方の音源方向にヌルを形成すれば，互いの音源の分離抽出が可能である．しかしながら，ヌル方向以外の方向に対しては感度が高くなるため，現実に存在する音源のような空間的な広がりを有する音源に対しては効果的ではない．そこで，音源定位結果付近の方向にヌルを有する重み付け加算回路を複数用意して抑圧区間を広げ，音源分離する手法について検討する．実環境実験結果により，検討手法の有効性を示す．

キーワード　マイクロホンアレー，指向性，音源分離

A Proposal of Sound Source Separation by Directivity Pattern Control of Two Microphones

Shigeharu AOKI† and Kenji SUYAMA†

† School of Engineering, Tokyo Denki University,
5 Senju-Asahi-cho, Adachi-ku, Tokyo, 120–8551, Japan
E-mail: †13kme01@ms.dendai.ac.jp

Abstract　In this paper, two sound source separation using two microphones is studied. In a directivity pattern of two microphones, forming an appropriate null toward a direction of eliminated source while keeping a constant sensitivity toward a direction of extracted source is effective for the source separation. In the studied method, two directions are estimated by the histogram based method and its results are applied to form the directivity pattern. In addition, multiple directivity patterns are multiplied and thus a wide range of attenuation is realized. Several experimental results in a real environment experiments are shown the effectiveness of the proposed method.

Key words　microphone array, directivity pattern, sound source separation

1. はじめに

音源分離は自動議事録作成やロボット聴覚などにおいて重要な音響信号処理技術である．音源分離は，複数の音声信号が混合した受音信号から，源信号を分離・抽出する問題である．

音源分離法として，マスキングによる手法，独立成分分析 (ICA:Independent Component Analysis) による手法，指向性に基づく手法が挙げられる．マスキングによる手法では，音声信号が「時間—周波数」領域で疎らに分布するスパース性に基づき,「時間—周波数」マスキングにより音源分離を行う．DUET (Degenerate Unmixing Estimation Technique) [1] では，瞬時位相差と振幅比をパラメータとして 2 次元ヒストグラムを作成し，そのヒストグラムのピークから音源方向推定を行う．この推定結果に基づき,「時間—周波数」マスクを推定している．ICA では，音声信号同士の統計的独立性に基づき音源分離を行う．周波数領域 ICA ではパーミュテーション問題がしばしば取り上げられるが，文献 [2] では音源方向を利用してこの問題を解決している．

音源方向が既知である場合，マイクロホンアレーの指向性を利用することが可能である．マイクロホンアレーは同期加減算処理により目的音源方向の信号の強調や，妨害音源方向の信号の抑圧が可能である [3] [4]．一般に減算型アレーではマイクロホン数 −1 個のヌルを形成できる．これにより妨害音源方向にヌルを形成することで妨害音源信号の抑圧が可能である．文献 [5] では，4 マイクロホンで 3 つの妨害音源に対しヌルを形成し，SNR の改善に成功している．近年，小規模なシステムの実

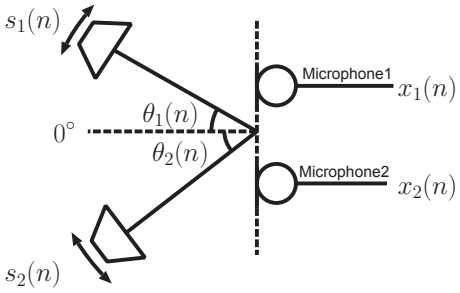

図 1　問 題 設 定

Fig. 1　Problem description

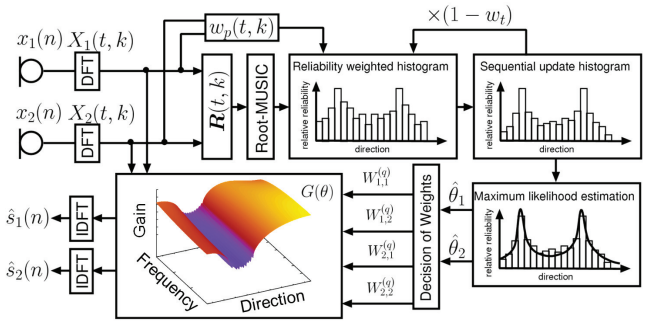

図 2　提案法の流れ

Fig. 2　A procedure of the proposed method

現やコスト削減の面から特に 2 マイクロホンによる信号処理が注目されている．そのため形成可能なヌルは 1 個に限定される．

指向性制御による音源分離手法では，音源方向推定が重要である．固定音源の場合，DUET のように十分な時間の信号を用いて音源方向推定をすることができる．しかしながら，ロボット聴覚などへの応用では，音源が複数存在することや音源が移動することが考えられる．

複数の移動音源分離には，複数移動音源追尾が必須である．逐次更新ヒストグラムによる複数音源追尾手法 [6] では，2 マイクロホンで 2 音源の追尾に成功している．そこで本研究では，逐次更新ヒストグラムによる複数音源追尾手法によって得られた 2 つの音源方向のうち一方を妨害音源方向とみなす．そのとき，他方に対しピークを，妨害音源方向にヌルを形成するように指向性を調整する．この 2 つの方向の関係を入れ替えた分離回路をもう 1 つ用意し，音源分離を行う．前述の通り，2 マイクロホンでは 1 つのヌル形成のみが可能である．しかしながら，実際の音源は点音源ではなく空間的な広がりをもつ．よって，1 つのヌルのみでは抑圧区間が狭く，妨害音源信号を十分に抑圧できない．また，残響環境下ではあらゆる方向から妨害音声信号が到来するため，抑圧区間は広い方が好ましいと考えられる．そこで本研究では，指向性制御として複数の重み付け加算回路を利用し，その出力信号同士を乗じる．これにより妨害音源方向に複数のヌルを形成し抑圧区間の拡大を図る．実環境実験より，提案法の有効性を示す．

2. 問 題 設 定

図 1 に示すようなマイクロホン間隔 d で配置した 2 つのマイクロホンで，2 つの音声信号 $s_1(n)$, $s_2(n)$ を受音する．$s_1(n)$, $s_2(n)$ は時刻とともに移動しているとする．このとき，m 番目のマイクロホンの受音信号は周波数領域で次式で表される．

$$X_m(t,k) = \sum_{i=1}^{2} S_i(t,k) e^{-j\omega_k (m-1)\tau_i(t)} + \Gamma_m(t,k) \qquad (1)$$

ここで $\Gamma_m(n)$ はマイクロホン m での観測雑音，t はフレーム番号，k は周波数帯域番号，ω_k は周波数帯域番号 k のときの角周波数である．また，$S_i(t,k)$ は，$s_i(n)$ の DFT (Disccrete Fourier Transform) である．$\tau_i(n)$ はマイクロホン間の到達時間差であり，次式で表される．

$$\tau_i(n) = \frac{d \sin\theta_i(n)}{c} \qquad (2)$$

ここで c は音速である．音源分離問題はマイクロホン受音信号 $x_m(n)$, $m = 1, 2$ から音声信号 $s_i(n)$, $i = 1, 2$ を分離抽出する問題である．

3. 提　案　法

図 2 に提案法の流れを示す．提案法は以下の流れで行う．

(I) マイクロホンの受音信号 $x_1(n), x_2(n)$ を DFT し，周波数領域に変換し $X_1(t,k), X_2(t,k)$ とする．

(II) 各「時間—周波数」点で Root-MUSIC (Multiple Signal Classification) [7] で用いる相関行列 $\boldsymbol{R}(t,k)$ を次式で算出する．

$$\boldsymbol{R}(t,k) = \boldsymbol{X}(t,k)\boldsymbol{X}^H(t,k) + \beta \boldsymbol{R}(t-1,k) \qquad (3)$$

ここで $\boldsymbol{X}(t,k) = [X_1(t,k), X_2(t,k)]^T$ は受音信号ベクトルであり，T は転置を表す．また，β は忘却係数，H はエルミート転置である．

(III) $\boldsymbol{R}(t,k)$ を用いて各「時間—周波数」点で Root-MUSIC により音源定位を行う．

(IV) 信号パワー比指標 [8] $w_p(t,k)$ を各「時間—周波数」点の推定結果の信頼性として，信頼性重みづけヒストグラムを作成する．信号パワー比指標 $w_p(t,k)$ は次式で算出する．

$$w_p(t,k) = \frac{P(t,k)}{\sum_k P(t,k)} \qquad (4)$$

ここで $P(t,k) = (|X_1(t,k)|^2 + |X_2(t,k)|^2)/2$ である．

(V) 信頼性重み付けヒストグラムに対し更新重み w_t を乗じ，ヒストグラムの逐次更新を行う．

(VI) 逐次更新ヒストグラムに対し混合コーシー分布を当てはめ EM アルゴリズムにより最尤推定を行い，音源方向を推定する．混合コーシー分布は次式で表される．

$$F(\theta) = \sum_{i=1}^{2} \rho_i \left[\frac{1}{\pi} \left\{ \frac{\lambda_i}{\left(\theta_i - \hat{\theta}_{i,t}\right)^2 + \lambda_i^2} \right\} \right] \qquad (5)$$

ここで ρ_i は混合比，$\hat{\theta}_{i,t}$ は最頻値，λ_i は半値半幅を表す．混合コーシー分布の最頻値をフレーム t での音源方向推定結果とする．

(VII) 推定音源方向 $\hat{\theta}_1, \hat{\theta}_2$ を用いて，複素重み係数 $W_{1,1}^{(q)}(t,k)$, $W_{1,2}^{(q)}(t,k)$, $W_{2,1}^{(q)}(t,k)$, $W_{2,2}^{(q)}(t,k)$ を算出する．

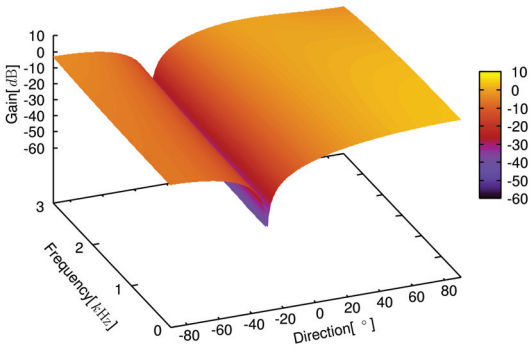

図 3　指　向　性

Fig. 3　Directivity pattern

(VIII) 受音信号に複素重み係数を重み付け加算する．

(IX) 出力結果を IDFT し，音源信号 $\hat{s}_1(t,k), \hat{s}_2(t,k)$ を得る．

3.1　マイクロホンアレーの指向性

マイクロホンアレーの指向性は受音信号に対する重み付け加算によって形成できる．複素重み係数は次の連立方程式で算出する．

$$\begin{cases} W_{1,1}(t,k) + W_{1,2}(t,k)e^{-j\omega_k \tau_s(t)} = 1 \\ W_{1,1}(t,k) + W_{1,2}(t,k)e^{-j\omega_k \tau_n(t)} = 0 \end{cases} \tag{6}$$

式 (6) の第 1 式は，目的音源方向のゲインを 1 に維持すること，第 2 式は妨害音源方向のゲインを 0 に維持することを意味する．複素重み係数 $W_{1,1}(t,k), W_{1,2}(t,k)$ は，目的音源方向 $\theta_s(t)$, 妨害音源方向 $\theta_n(t)$ についてそれぞれ到達時間差を $\tau_s(t)$,$\tau_n(t)$ とするとき次式で決定される．

$$W_{1,1}(t,k) = \frac{-e^{-j\omega_k \tau_n(t)}}{e^{-j\omega_k \tau_s(t)} - e^{-j\omega_k \tau_n(t)}} \tag{7}$$

$$W_{1,2}(t,k) = \frac{1}{e^{-j\omega_k \tau_s(t)} - e^{-j\omega_k \tau_n(t)}} \tag{8}$$

例として，図 3 にマイクロホン間隔 $d = 5.56$[cm]，目的音源方向を 30°, 妨害音源方向を −30° とした場合の「方向ー周波数」に対するマイクロホンアレーの指向性を示す．マイクロホン間隔 $d = 5.56$[cm] のとき，およそ 3000[Hz] 以上の周波数帯域では空間エイリアシングが発生するが，図 3 より空間エイリアシングが生じない周波数帯域において，どの周波数でも目的音源方向と妨害音源方向のゲインが保持されていることが確認できる．

3.2　提案法による指向性制御

2 マイクロホンによる指向性は単一方向のみにヌルを形成可能である．しかし，空間中に存在する音源は点音源ではなく広がりをもつ．空間的広がりをもつ音源に対して，単一方向にヌルを形成しても，抑圧区間が狭いことから，その抑圧が不十分となる．そこで提案法では，ひとつの重み付け加算回路によって単一方向にヌルが形成できることを利用して，複数の重み付け加算を用いて妨害音源方向に複数のヌルを形成し抑圧区間の拡大する．図 4 に提案法の回路構成を示す．提案法の指向性制御は，周波数領域の受音信号に対し，複数の重み付け加算回

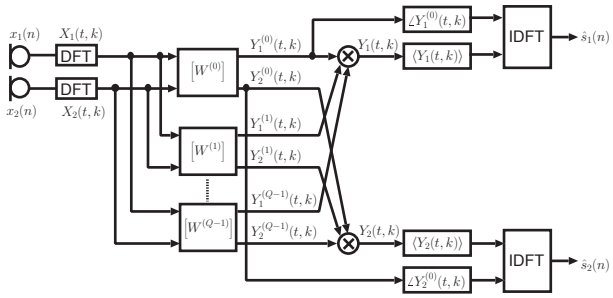

図 4　提案法の回路構成

Fig. 4　A circuit structure of the proposed method

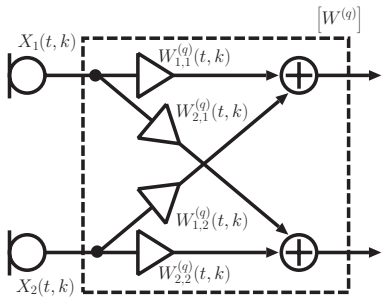

図 5　重み付け加算回路

Fig. 5　A weighted summing circuit

路 $\left[W^{(q)}\right], (q = 0, 1, \cdots Q-1)$ を用いる．図 5 に q 番目の重み付け加算回路を示す．図 5 における $W_{1,1}^{(q)}(t,k)$, $W_{1,2}^{(q)}(t,k)$, $W_{2,1}^{(q)}(t,k)$, $W_{2,2}^{(q)}(t,k)$ は重み係数である．$Q = 1$ のとき，提案法は単一方向のみにヌルを形成する．重み付け加算回路を Q 個使用しその出力結果を乗じることで，Q 個の方向にヌルを形成可能である．提案法では，妨害音源方向に対して微少角度だけ異なる方向にヌルを形成することで抑圧区間を拡大する．ヌルを近接して配置することで，ゲインが減少し抑圧区間を拡大することが可能である．妨害音源方向を $\theta_n(t)$ としたとき，q 番目の重み付け加算回路で形成するヌルの方向 $\theta_q(t)$ は次式で決定する．

$$\theta_q(t) = \begin{cases} \theta_n(t), & q = 0 \\ \theta_n(t) - \dfrac{q+1}{2}\dfrac{w}{Q-1}, & q = 1, 3, 5, \cdots, Q-2 \\ \theta_n(t) + \dfrac{q}{2}\dfrac{w}{Q-1}, & q = 2, 4, 6, \cdots, Q-1 \end{cases} \tag{9}$$

ここで w は抑圧区間の幅であり，複数形成したヌルの両端の幅である．q が奇数の時，妨害音源方向に対して負の方向に，q が偶数のとき妨害音源方向に対して正の方向に妨害音源方向を挟むようにヌルを形成する．Q が奇数の時，妨害音源方向に対して対称にヌルが形成されるため，後の実験では Q が．奇数のときのみを扱う．例として，図 6 に目的音方向 30°, 妨害音方向 −30° とし $Q = 3, w = 10°$ とするときの指向性を示す．同様に，図 7 に $Q = 11, w = 30°$ としたときの指向性を示す．これより，ヌルの数を増加させ，抑圧区間の拡大を行っても，全周波数帯域で目的音源方向のゲインは 0[dB] を維持していることが確認できる．指向性制御によって目的音源方向には影響を与えず，妨害音源方向に対して抑圧区間が拡大され，減衰量が

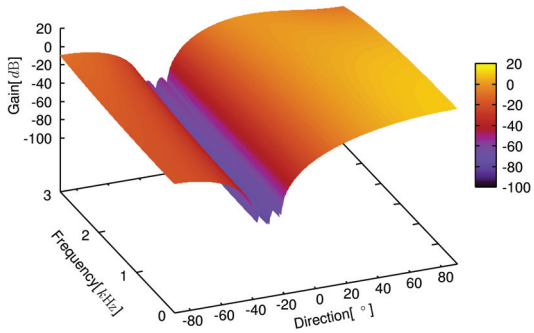

図 6　提案法による指向性 ($Q = 3, w = 10°$)

Fig. 6　A directivity pattern by the proposed method

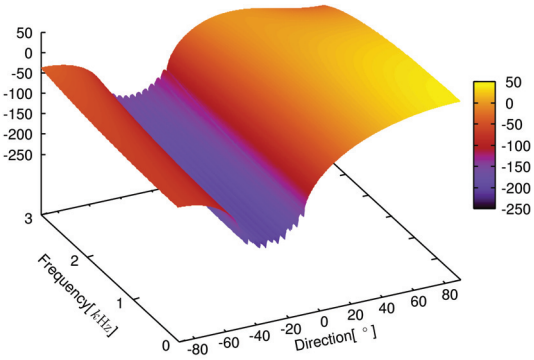

図 7　提案法による指向性 ($Q = 11, w = 30°$)

Fig. 7　A directivity pattern by the proposed method

増加しており，より妨害音源信号が抑圧されるように指向性が形成されているといえる．

提案法では，Q 個の重み付け加算回路の出力を乗じていることから，振幅特性および位相特性の補正が必要である．振幅特性の補正は，音源分離結果のゲインを調整するために必要である．振幅特性は，各重み付け加算回路の出力信号に対し，その平均パワーの $Q-1$ 乗で除する．位相特性は，位相特性の平均を用いた場合，音声信号が重畳する問題が生じる．提案法では，妨害音源方向にヌルを形成した $q = 0$ のときの出力の位相特性を用いた．

4.　実環境実験

最初に実環境実験によりマスキングによる手法と提案法の性能比較を行う．次に実環境実験により移動音源に対する性能検証を行い，提案法の有効性を示す．音源は RWCP 実環境音声・音響データベースの固定音源および移動音源のデータを利用した。性能評価指標として，信号対干渉比 (SIR:Signal-to-Interference Ratio) を用い，その改善値を用いた．演算処理には，Intel (R) Core (TM) i3-2130 2.83[GHz]，メモリ 4[GByte] の PC を用いた．

4.1　分離性能の比較

音源分離性能の検証を行うため，実環境実験を行った．表 1 に実験条件を示す．マイクロホン間隔が 5.66[cm] のとき，およそ 3000[Hz] から空間エイリアシングの影響があるため，音源分離では 3000[Hz] 未満の周波数帯域を用いた．

表 1　実 験 条 件

Table 1　Experimental conditions

サンプリング周波数	8000[Hz]
フレーム長	512
マイクロホン間隔	5.66[cm]
信号長	4[s]
音源分離使用周波数帯域	85-3000[Hz]
音源方向	$-30.0°$, $30.0°$

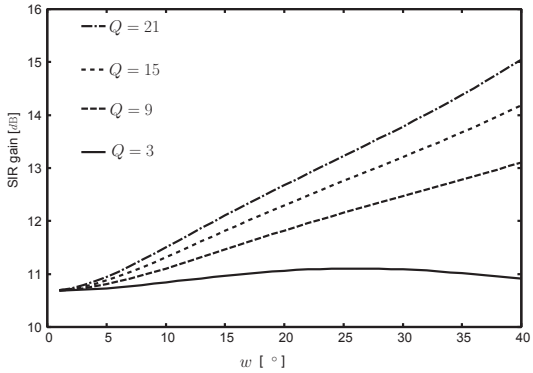

図 8　提案法の SIR 改善値 (残響時間:0.3[s])

Fig. 8　SIR improvement of the proposed method (reverberation time:0.3[s])

表 2　比較手法の SIR 改善値 (残響時間:0.3[s])

Table 2　SIR improvement of the compared method (reverberation time:0.3[s])

手法	SIR 改善値
バイナリマスク	11.32[dB]
連続値マスク	10.44[dB]
指向性 (Q=1)	10.68[dB]

音源分離性能の比較が目的であることから，この実験では音源方向の真値を与え実験を行った．残響時間が 0.3[s]，0.6[s] の部屋で録音されたデータを使用した．残響時間 0.3[s] のときの騒音レベルは 18.9[dB]，残響時間 0.6[s] のときの騒音レベルは 44.7[dB] であった．音源は RWCP 実環境音声・音響データベースにおける mmysda01 から mmysda10 までの固定音源のデータを用いた．提案法は $Q = 3, 9, 15, 21$ についてそれぞれ抑圧区間 w を 1° から 40° まで 1° 間隔で検証した．DUET を用いたバイナリマスクによる分離，DUET の尤度関数比を用いた連続値マスクによる分離，$Q = 1$ のときの指向性による分離と分離性能を比較した．残響時間が 0.3[s], 0.6[s] の両条件において 2 音源の組合わせを 10 パターン試行し，全パターンの SIR 改善値の平均を評価した．

残響時間が 0.3[s] のときの提案法の実験結果を図 8 に示す．バイナリマスクおよび連続値マスクの SIR 改善値を表 2 に示す．残響時間が 0.6[s] のときの提案法の実験結果を図 9 に示す．バイナリマスクおよび連続値マスクの SIR 改善値を表 3 に示す．図 8, 図 9 より，抑圧区間が広く，ヌルの数が多いほど SIR が向上していることが分かる．また，この場合にバイナリマスクおよび連続値マスクよりも提案法の分離性能が高いことが確

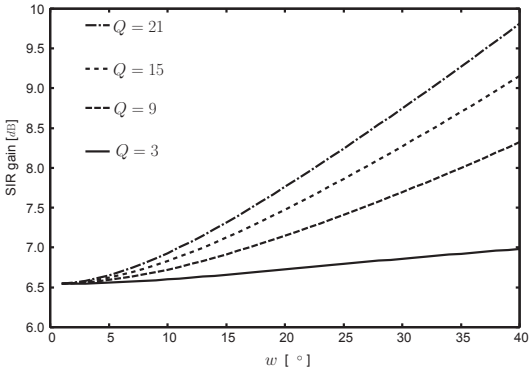

図 9　提案法の SIR 改善値 (残響時間:0.6[s])

Fig. 9　SIR improvement of the proposed method (reverberation time:0.6[s])

表 3　比較手法の SIR 改善値 (残響時間:0.6[s])

Table 3　SIR improvement of the compared method (reverberation time:0.6[s])

手法	SIR 改善値
バイナリマスク	8.71[dB]
連続値マスク	8.33[dB]
指向性 (Q=1)	6.54[dB]

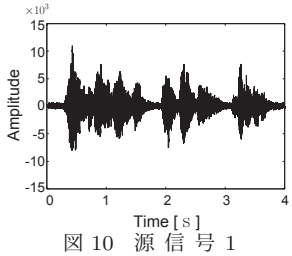

図 10　源 信 号 1

Fig. 10　Source signal 1

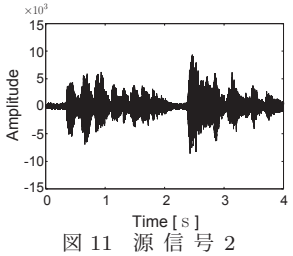

図 11　源 信 号 2

Fig. 11　Source signal 2

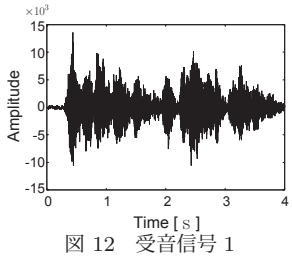

図 12　受音信号 1

Fig. 12　Received sound signal 1

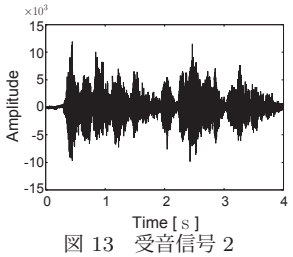

図 13　受音信号 2

Fig. 13　Received sound signal 2

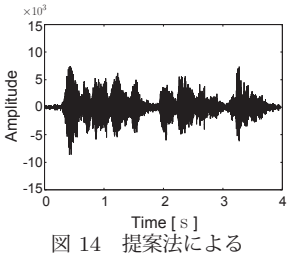

図 14　提案法による分離信号 1

Fig. 14　Separated signal 1 by the proposed method

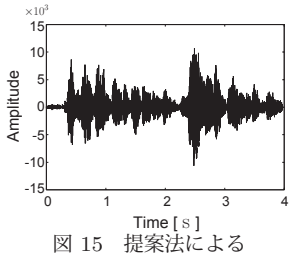

図 15　提案法による分離信号 2

Fig. 15　Separated signal 2 by the proposed method

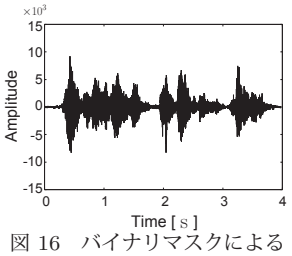

図 16　バイナリマスクによる分離信号 1

Fig. 16　Separated signal 1 by binary masking

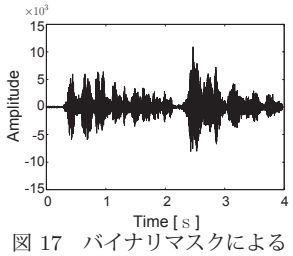

図 17　バイナリマスクによる分離信号 2

Fig. 17　Separated signal 2 by binary masking

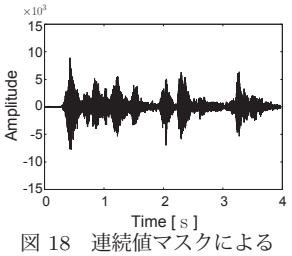

図 18　連続値マスクによる分離信号 1

Fig. 18　Separated signal 1 by soft masking

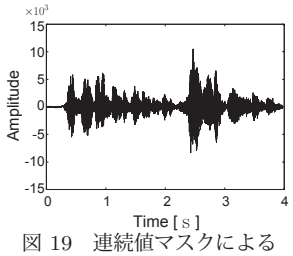

図 19　連続値マスクによる分離信号 2

Fig. 19　Separated signal 2 by soft masking

表 4　実 験 条 件

Table 4　Experimental conditions

サンプリング周波数	8000[Hz]
フレーム長	512
マイクロホン間隔	5.66[cm]
信号長	4[s]
更新重み	0.11
忘却係数	0.02
ヒストグラム分割幅	3°
音源追尾使用周波数帯域	1000-4000[Hz]
音源分離使用周波数帯域	85-3000[Hz]

認できる．図 10 ～ 図 19 に音声信号の例を示す．

4.2　移動音源に対する音源分離性能検証

移動音源に対する音源分離性能を検証した．表 4 に実験条件を示す．音源は RWCP 実環境音声・音響データベースより，mmysda01 から mmysda10 までの移動音源を用い，音源間隔はおよそ 30° とした．固定音源の場合と同様に $Q = 3, 9, 15, 21, 27$ についてそれぞれ抑圧区間 w を 1° から 40° まで 1° 間隔で検証し，全 45 パターンの SIR 改善値の平均値を評価した．図 20 に提案法による SIR の改善値を示す．

図 20 より，SIR 改善値が正であることから，移動音源に対して音源分離可能であることが分かる．また，固定音源の場合と同様に抑圧区間が広く，ヌルの数が多いほど SIR 改善値が向上することが確認できる．図 21 に音源追尾結果の例を示す．図 22 ～ 図 27 に音声信号の例を示す．

提案法の実行時間はヌルの最大数 Q に依存する．これは，複素重み係数の算出とその重み付け加算がヌルの数に比例するためである．提案法の実行時間は $Q = 27$ の場合，信号長 4[s] に対しおよそ 3.755[s] であった．これより，提案法が実時間処理可能であるといえる．信号長 4[s] に対し，$Q = 29$ のとき

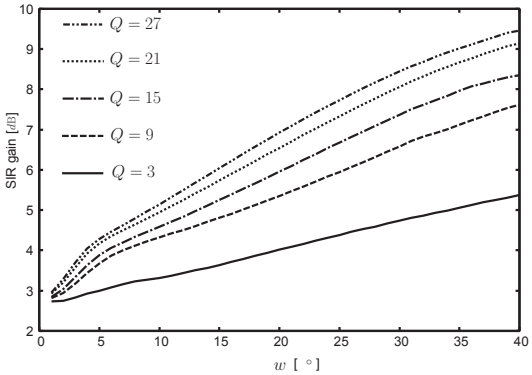

図 20　SIR 改善値 (移動音源，残響時間:0.3[s])

Fig. 20　SIR improvement (moving sources, reverberation time:0.3[s])

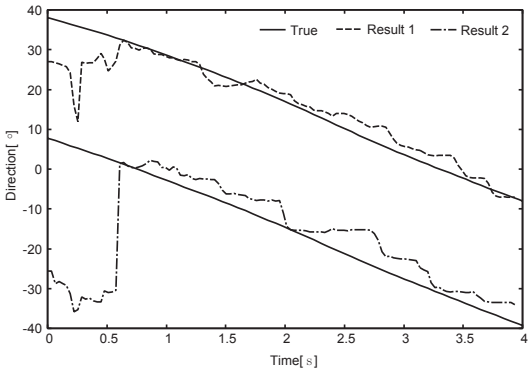

図 21　追 尾 結 果

Fig. 21　Tracking results

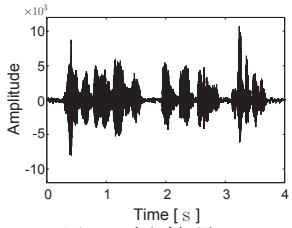

図 22　源 信 号 1

Fig. 22　Source signal 1

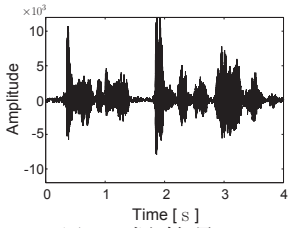

図 23　源 信 号 2

Fig. 23　Source signal 2

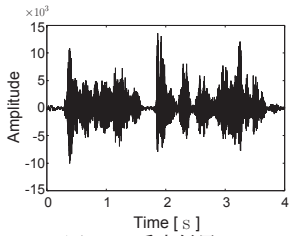

図 24　受音信号 1

Fig. 24　Received sound signal 1

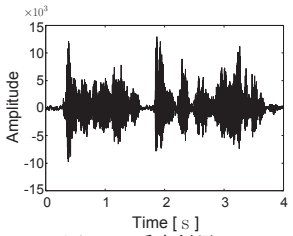

図 25　受音信号 2

Fig. 25　Received sound signal 2

4.001[s] であることから，使用 PC では最大 27 個のヌルを形成することが可能といえる．

5. ま と め

本研究では，2 マイクロホンの指向性制御による音源分離について検討した．提案法では逐次更新ヒストグラムによる音源追尾結果に基づき，複数のヌルを形成することで音源分離を行った．実環境実験より，提案法が移動音源分離可能であること，実時間処理が可能であることを示した．

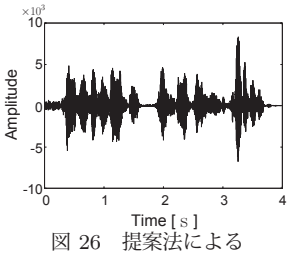

図 26　提案法による分離信号 1

Fig. 26　Separated signal 1 by the proposed method

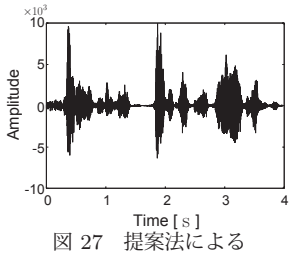

図 27　提案法による分離信号 2

Fig. 27　Separated signal 2 by the proposed method

文　献

[1] Özgür Yılmaz and Scott Rickard, “Blind Separation of Speech Mixtures via Time-Frequency Masking,”IEEE Trans. Signal Processing, vol.52, no.7, pp.1830–1847, July 2004.

[2] H. Sawada, R. Mukai, S. Araki, S. Makino, “Convolutive Blind Source Separation for more than Two Sources in the Frequency Domain,” Proc.of IEEE ICASSP 2004, vol.Ⅲ, pp. 885–888, 2004.

[3] 大賀寿郎, 山崎芳男, 金田豊, 音響システムとディジタル処理, コロナ社, 東京, 1995.

[4] 金田豊, “マイクロホンアレーによる指向性制御,” 日本音響学会誌, vol.51, no.5, pp.390–394, 1995.

[5] 中山雅人, 西浦敬信, 山下洋一, 中迫昇, “話者と音源の位置推定に基づく複数死角制御型ビームフォーマの基礎的検討,” 信学技報 vol.111, no.26, pp.107–112, 2011.

[6] 鈴木毅, 陶山健仁, “逐次更新ヒストグラムに基づく音源追尾の複数音源への拡張,” 信学技報, vol.112, no.486, pp81–86, 2013.

[7] 菊間信良, アレーアンテナによる適応信号処理, pp.191–209, 科学技術出版, 東京, 1998.

[8] 前田繰一朗, 陶山健仁, “音声のスパース性に基づく信頼性重み付け分布による複数音源定位,” 信学論 (A),vol.J95-A, no.3, pp.247–260, 2012.