

■2群 (画像・音・言語) - 6編 (音響信号処理)

3章 音源定位

(執筆者: 浅野 太) [2011年11月受領]

■概要■

複数のマイクロホンを空間的に配置したものをマイクロホンアレーと呼ぶ。音源からマイクロホンまでの距離が異なれば、音波が各マイクロホンに到達する時間が異なる。更に、各マイクロホンまでの到達時間差は、音源の位置の関数となっている。この原理を利用して、音源の位置を推定することができる。これを音源定位と呼ぶ。音源定位に用いられる手法は、地震源の探査や、レーダ、ソナーなどに用いられている技術と類似点が多い。本章では、音源定位に用いられる信号処理技術について述べる。

【本章の構成】

本章では、まず、二つのマイクロホンを用いた単純なマイクロホンアレーにより、マイクロホン間に生じる到達時間差を推定する手法について述べる (3-1 節)。続いて、多数のマイクロホンによるマイクロホンアレーを用いて、音響ビームを構成し、音源定位を行う、ビームフォーミング法について述べる (3-2 節)。更に、空間分解能 (音源が近接している場合に、これを別々の音源として認識する能力) を向上させた、サブスペース法について述べる (3-3 節)。最後に、音源が移動する場合の音源位置推定 (音源追跡) について述べる (3-4 節)。

■2群 - 6編 - 3章

3-1 遅延時間推定法

(執筆著者：西浦敬信) [2011年11月受領]

遅延時間推定法では、二つのマイクロホンで受信した信号の到来時間差 (Time Delay Of Arrival : TDOA) を算出することにより、音源方位 (Direction Of Arrival : DOA) を推定する。到来時間差 (TDOA) の算出手法としては、受信信号間の相互相関関数¹⁾を用いた相互相関法²⁾や受信信号の振幅により白色化 (正規化) した相互相関関数を用いた白色化相互相関 (Cross-power Spectrum Phase Analysis : CSP) 法^{2),3),4)}などがある。

3-1-1 観測信号のモデル

まず観測信号のモデルについて説明する。図3・1は、間隔 d で配置した二つのマイクロホン M_1, M_2 を用いて、 θ 方向から音速 c で平面波として到来する音声 $S(t)$ を受信する流れを表している。ここで、各マイクロホン M_m ($m=1, 2$) の受信信号を $x_m(t)$ 、各マイクロホン M_m に対する音声の到達時間を φ_m とすると、マイクロホン M_1, M_2 の受信信号は式(3・1)、式(3・2)の通り表せる。

$$x_1(t) = s(t - \varphi_1), \quad (3 \cdot 1)$$

$$x_2(t) = s(t - \varphi_2), \quad (3 \cdot 2)$$

このとき、マイクロホン M_1, M_2 間の到来時間差 (TDOA) は式(3・3)となる。

$$\begin{aligned} TDOA_{1,2} &= \varphi_1 - \varphi_2 \\ &= \frac{d \cos(\theta)}{c}, \end{aligned} \quad (3 \cdot 3)$$

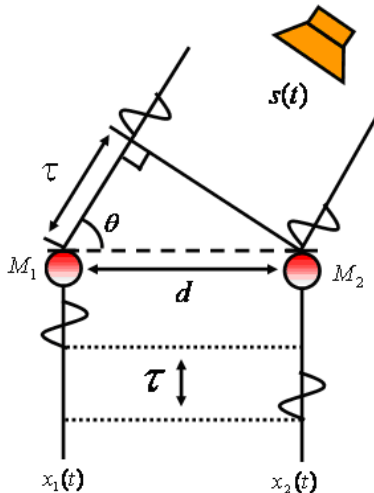


図3・1 2マイクロホンを用いた受信信号

ここで、 $TDOA_{1,2}$ は到来時間差 (TDOA)、 θ は音源方位 (DOA) を示す。式(3・3)より到来時間差 (TDOA) と音源方位 (DOA) の関係^{4),5)}は、式(3・4)で表せる。

$$\theta = \cos^{-1}\left(\frac{TDOA_{1,2} \cdot c}{d}\right) \quad (3 \cdot 4)$$

よって、到来時間差 (TDOA) を算出すれば、式(3・4)により音源方位 (DOA) を推定できることが分かる。

3-1-2 到来時間差 (TDOA) の推定

最初に、受信信号間の相互相関係数を用いた相互相関法による到来時間差 (TDOA) の推定について説明する。相互相関法では、式(3・5)を用いて相互相関関数を算出する。

$$\phi_{1,2}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T x_1(t) \cdot x_2(t + \tau) dt, \quad (3 \cdot 5)$$

ここで、 $\phi_{1,2}(\tau)$ は相互相関関数、 τ は時間指標である。

図 3・2 は、1 音源環境における相互相関関数の例である。この相互相関関数に対して、式(3・6)を用いて相互相関関数が最大となる時間 τ (すなわち受信信号間の相互相関が最大となる時間差) を算出することで、到来時間差 (TDOA) を推定する。

$$TDOA_{1,2} = \arg \max_{\tau} (\phi_{1,2}(\tau)), \quad (3 \cdot 6)$$

その後、推定した到来時間差 (TDOA) に対して、式(3・4)を用いて音源方位 (DOA) を求める。

次に、白色化相互相関 (CSP) 法による到来時間差 (TDOA) の推定について説明する。白色化相互相関 (CSP) 法は、式(3・5)にて求めた相互相関関数に対して、式(3・7)にてクロススペクトルに変換後、式(3・8)にてクロススペクトルの振幅により白色化 (正規化) することで、白色化相互相関関数 (CSP 係数) を求める。

$$\Phi_{1,2}(\omega) = \int_{-\infty}^{\infty} \phi_{1,2}(\tau) \exp(-j\omega\tau) d\tau, \quad (3 \cdot 7)$$

$$CSP_{1,2}(\tau) = \int_{-\infty}^{\infty} \frac{\Phi_{1,2}(\omega)}{|\Phi_{1,2}(\omega)|} \exp(-j\omega\tau) d\omega, \quad (3 \cdot 8)$$

ここで、 $\Phi_{1,2}(\omega)$ はクロススペクトル、 $CSP_{1,2}(\tau)$ は白色化相互相関関数 (CSP 係数) を示す。

図 3・3 は、1 音源環境における白色化相互相関関数 (CSP 係数) の例である。この白色化相互相関関数 (CSP 係数) に対して、式(3・9)を用いて白色化相互相関関数 (CSP 係数) が最大となる時間 τ (すなわち受信信号間の白色化相互相関が最大となる時間差) を算出することで、到来時間差 (TDOA) を推定する。

$$TDOA_{1,2} = \arg \max_{\tau} (CSP_{1,2}(\tau)), \quad (3 \cdot 9)$$

その後、推定した到来時間差 (TDOA) に対して、式(3・4)を用いて音源方位 (DOA) を求める。式(3・8)に示すクロススペクトルの白色化 (正規化) は、受信信号の位相情報 (位相差)

のみを利用して到来時間差 (TDOA) を推定することを意味するため、白色化相互相関 (CSP) 法は受信信号の周波数振幅特性に依存しない到来時間差 (TDOA) の推定が行える。そのため、相互相関関数と比較して白色化相互相関関数 (CSP 係数) は音源方位 (DOA) に対して鋭いピークを得ることができる。特に、白色化相互相関 (CSP) 法は複数の音源が存在する環境下や音源のエネルギーが小さい環境下であっても、高精度に音源方位 (DOA) を推定できるという特徴をもつ。

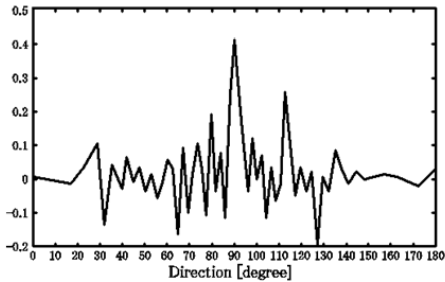


図 3.2 相互相関関数の例

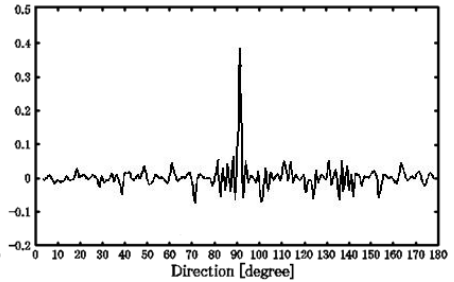


図 3.3 白色化相互相関関数 (CSP 係数) の例

■参考文献

- 1) 磯部 孝, “相関関数およびスペクトル,” 東京大学出版会, 1968.
- 2) Charles H. Knapp and G. Clifford Carter, “The Generalized Correlation Method for Estimation of Time Delay,” IEEE Trans. on Acoustics, Speech, and Signal Processing, vol. ASSP-24, no.4, pp.320-327, 1976.
- 3) M. Omologo and P. Svaizer, “Use of the Crosspower-spectrum Phase in Acoustic Event Location,” IEEE Trans. on Speech and Audio Processing, vol. SAP-5, no.3, pp.288-292, 1997.
- 4) 西浦敬信, 山田武志, 中村 哲, 鹿野清宏, “マイクロホンアレーを用いたCSP法に基づく複数音源位置推定,” 電子情報通信学会論文誌, vol. J83-D-II, no.8, pp. 1713-1721, 2000.
- 5) 大賀寿郎, 山崎芳男, 金田 豊, “音響システムとデジタル処理,” コロナ社, 1995.

■2群 - 6編 - 3章

3-2 ビームフォーミング法

(執筆者：浅野 太) [2011年11月受領]

ビームフォーミング法では、図3・4に示すように、各マイクロホンのあとに遅延子を挿入し、入射波の位相を揃えて、和をとることにより、音響ビームを形成する。このビームを空間的にスキャンさせることにより、音源位置を推定する。この手法は、遅延和法 (Delay-and-Sum Beamformer)^{1), 4), 8)} と呼ばれる。遅延子の部分を適応フィルタに拡張することにより、雑音源を抑制しながら、目的音源の位置を推定することも可能である (例えば最少分散 (Minimum Variance) 法^{1), 4), 8)})。ここでは、最も簡単な遅延和法について述べる。

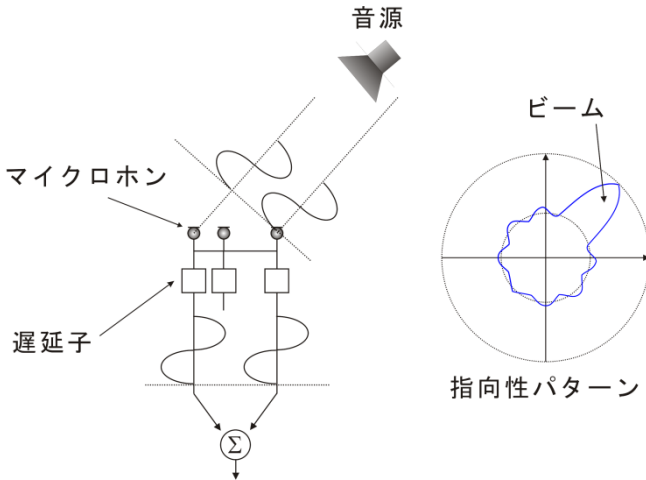


図3・4 遅延和型ビームフォーマの原理

3-2-1 観測信号のモデル

まず、観測信号のモデルについて説明する。本章のこれ以降の部分では、次式のように、観測信号を短区間フーリエ変換²⁾により周波数領域に変換して議論を進める。

$$X_m(\omega, k) = F[x_m(t, k)] \quad (3 \cdot 10)$$

ここで、 $F[\cdot]$ は、短区間フーリエ変換を表す。 $x_m(t, k)$ は、 m 番目のマイクロホンで観測される信号を表し、 t, k は、それぞれ、時間及びフーリエ変換の区間を表すインデックスを表す。 $X_m(\omega, k)$ は $x_m(t, k)$ の短区間フーリエ係数であり、 ω は周波数を表す。全部で M 個のマイクロホンで観測したとして、観測されたデータの短区間フーリエ係数を並べて、次式の観測ベクトルを定義する。

$$\mathbf{x}(\omega, k) = [X_1(\omega, k), \dots, X_M(\omega, k)]^T \quad (3 \cdot 11)$$

ここで、 \cdot^T は行列／ベクトルの転置を表す．図 3・5 のような音環境を観測した場合，この観測ベクトルは，周波数領域では，次式のようにモデル化することができる⁴⁾．

$$\mathbf{x}(\omega, k) = \sum_{n=1}^N \mathbf{a}_n(\omega) S_n(\omega, k) + \mathbf{n}(\omega, k) \quad (3 \cdot 12)$$

ここで， $\mathbf{a}_n(\omega) = [A_{1n}(\omega), \dots, A_{Mn}(\omega)]^T$ は，ステアリングベクトルと呼ばれ，その要素

$$A_{mn}(\omega) = |A_{mn}| \exp(-j\omega\tau_{mn}) \quad (3 \cdot 13)$$

は， n 番目の音源から m 番目のマイクロホンまでの（直接音の）伝達関数を表す．ここで， $|A_{mn}|$ は伝達関数のゲイン（利得）を表し，主に音源からマイクロホンまでの空気減衰や，障害物がある場合は，その遮蔽効果などを反映する．一方， $\exp(-j\omega\tau_{mn})$ は位相であり，このうち， τ_{mn} は， n 番目の音源から m 番目のマイクロホンまでの音の伝達時間を表す． $\mathbf{n}(\omega, k) = [N_1(\omega, k), \dots, N_M(\omega, k)]^T$ は，各マイクロホンで観測される背景雑音を要素にもつベクトルである．

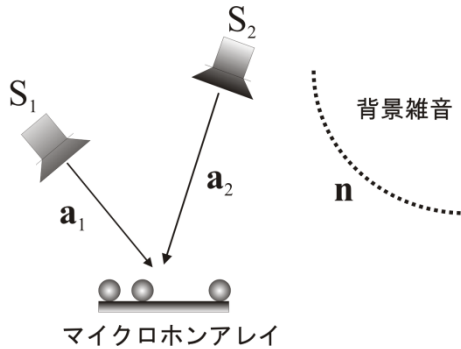


図 3・5 観測信号のモデル

3-2-2 空間スペクトルの推定

ビームフォーミングによる空間スペクトルの一般形は，次式のようになる⁴⁾．

$$P_B(\theta) = \mathbf{w}^H \mathbf{R} \mathbf{w} \quad (3 \cdot 14)$$

ここで，

$$\mathbf{R} = E[\mathbf{x}(\omega, k) \mathbf{x}^H(\omega, k)] \quad (3 \cdot 15)$$

は，空間相関行列と呼ばれる．演算子 $E[\cdot]$ は，期待値を表し，実際の計算では，サンプル平均で代用する． \cdot^H は，共役転置を表す． \mathbf{w} は，ビームフォーマ係数ベクトルと呼ばれ，様々なビームフォーマルゴリズムを用いることができる．例えば，一番シンプルな遅延和法では，次式のようになる．

$$\mathbf{w} = \frac{\mathbf{v}(\theta)}{\mathbf{v}^H(\theta) \mathbf{v}(\theta)}$$

ここで， $\mathbf{v}(\theta)$ は，仮想的な音源が θ 方向にある場合のステアリングベクトルである．音源位

置としては、2次元や3次元の場合も考えられるが、本章では、簡単のため、1次元の場合（平面における入射角度 θ ）を考える．式(3・14)は、ビームフォーマ出力

$$y(\omega, k) = \mathbf{w}^H \mathbf{x}(\omega, k) \quad (3 \cdot 16)$$

の平均パワーを求めていることになる．遅延和法では、係数ベクトル \mathbf{w} により、 θ 方向から到来する音波の位相がそろえられ、加算されるため、 θ 方向から到来する音波が強調される．これにより、音響的なビームが形成される．このビームを、レーダのように、空間的にスキャンすることにより、フォーカスの方向 θ が、真の音源方向と一致したときに、空間スペクトル $P_B(\theta)$ にピークが現れる．このピーク位置により、音源の位置（到来方向）を推定する．

図3・5(a)は、音源が 0° 方向にある場合の空間スペクトルの例である．音源の周波数は1000 Hzである．音源方向にピークが見られる．一方、音源方向以外にも低いピークが観測される．これらは、サイドローブ（Sidelobe）と呼ばれ、その形状は、マイクロホン配置や周波数に依存する．

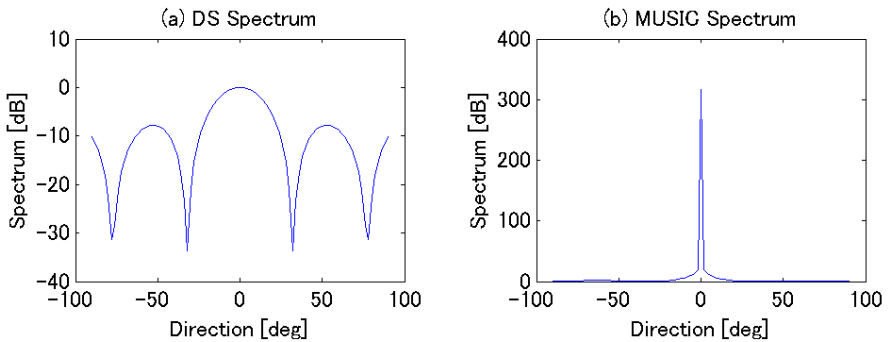


図3・6 空間スペクトルの例

■2群 - 6編 - 3章

3-3 サブスペース法

(執筆者：浅野 太) [2011年11月受領]

前節のビームフォーミング法は、比較的簡単な処理により音源位置を推定する方法として、広く用いられているが、空間分解能が低いという欠点がある。これを克服する手法として、サブスペース法^{1),4),8)}が提案されている。

3-3-1 固有値分布と部分空間

サブスペース法は、空間相関行列の固有ベクトルがもつ性質を用いて、高い分解能で音源位置を推定する方法である⁹⁾。まず、空間相関行列を、次式のように固有値分解する⁹⁾。

$$\mathbf{R}e_i = \lambda_i e_i \tag{3 \cdot 17}$$

ここで、 $\{\lambda_1, \dots, \lambda_M\}$ 及び $\{e_1, \dots, e_M\}$ は、それぞれ固有値及び固有ベクトルである。いま、空間に N 個の音源があり、これを M 個のマイクロホンで観測する場合を考える。背景雑音 $N_m(\omega, k)$ が空間的に白色(各マイクロホンで受音した雑音が互いに無相関)、かつ、音源信号 $S_n(\omega, k)$ が互いに無相関であると仮定すると、固有値は、**図3・7**のような分布となる。ここで、重要なのは、固有値分布は、音源のエネルギー分布を反映しており、例えば音源数が2($N=2$)の場合、大きな固有値が2個存在し、ほかの固有値は小さく、かつ平坦な分布となる。大きな固有値と小さな固有値との比は、信号 $S_n(\omega, k)$ と雑音 $N_m(\omega, k)$ のパワー比(Signal-to-Noise ratio : SNR)により決定される。このことから、大きな固有値の数をカウントすることにより、音源数を推定することができる⁸⁾(実際は、背景雑音が空間的に白色ではないなどのため、固有値分布から直ちに音源数を推定することは困難な場合が多い)。

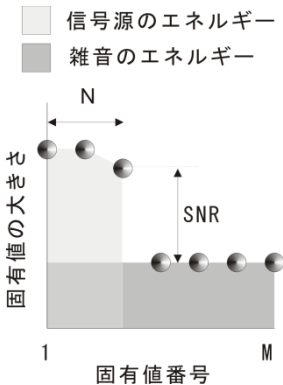


図3・7 固有値分布の例

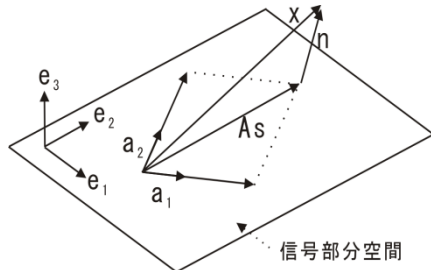


図3・8 観測ベクトルと固有ベクトル

一方、固有ベクトルについては、大きな固有値に対応する固有ベクトルは、音源のステアリングベクトルの張る部分空間(信号部分空間)の基底ベクトルとなる性質が知られて

いる^{6),7)}。図 3・8 は、この関係を幾何学的に表したものである。この例では、2 個の音源に対するステアリングベクトル $\mathbf{a}_1, \mathbf{a}_2$ が張る平面が信号部分空間である。大きい固有値に対応する 2 個の固有ベクトル $\mathbf{e}_1, \mathbf{e}_2$ もこの平面上にあり、かつ、空間相関行列がエルミート (Hermitian) 行列であることから、固有ベクトルは互いに直交する。以上から、固有ベクトル $\mathbf{e}_1, \mathbf{e}_2$ は、信号部分空間の直交基底ベクトルとなる。また、同様に、上述の固有ベクトルが直交する性質から、残りの固有ベクトル \mathbf{e}_3 は、ステアリングベクトル $\mathbf{a}_1, \mathbf{a}_2$ とも直交する。

以上の性質をまとめると、

- ・大きい方から N 個の固有値 $\lambda_1, \dots, \lambda_N$ は、残りの固有値 $\lambda_{N+1}, \dots, \lambda_M$ に比べ、大きな値をもち、その数は音源数に一致する。
- ・残りの $M-N$ 個の固有値 $\lambda_{N+1}, \dots, \lambda_M$ は、 $\lambda_1, \dots, \lambda_N$ に比べ小さな値をもち、平坦な分布を示す。
- ・大きい N 個の固有値に対応する固有ベクトル $\mathbf{e}_1, \dots, \mathbf{e}_N$ は、ステアリングベクトルが張る信号部分空間の基底ベクトルとなる。
- ・小さな $M-N$ 個の固有値に対応する固有ベクトル $\mathbf{e}_{N+1}, \dots, \mathbf{e}_M$ は、信号部分空間に直交する。

3-3-2 空間スペクトルの推定

サブスペース法における代表的な手法として、MUSIC (Mutiple Signal Classification) 法⁹⁾ があげられる。MUSIC 法では、以下のように空間スペクトルを推定する。

$$P_M(\theta) = \frac{\mathbf{v}^H(\theta)\mathbf{v}(\theta)}{|\mathbf{v}^H(\theta)\mathbf{E}_n|^2} \quad (3 \cdot 18)$$

ここで、 $\mathbf{E}_n = [\mathbf{e}_{N+1}, \dots, \mathbf{e}_M]$ であり、前節で述べた直交性から、仮想的なステアリングベクトルが音源のステアリングベクトルと一致 ($\mathbf{v}(\theta) = \mathbf{a}_i$) するとき、

$$\mathbf{v}^H(\theta)\mathbf{e}_{N+1} = \dots = \mathbf{v}^H(\theta)\mathbf{e}_M = 0 \quad (3 \cdot 19)$$

となるため、 $P_M(\theta)$ は $\mathbf{V}(\theta) = \mathbf{a}_i$ において、ピークをもつようになる。図 3・6(b) は、MUSIC 法による空間スペクトルである。音源方向 (0°) に鋭いピークをもつ。

■2群 - 6編 - 3章

3-4 音源追跡

(執筆者：浅野 太) [2011年11月受領]

前節までは、主に、音源が静止している場合について述べた。本節では、音源が移動している場合について述べる。音源が移動している場合は、上述の空間相関行列などの統計量を計算するための十分なサンプルが得られない場合が多く、推定結果の分散が大きくなる。そこで、音源追跡における代表的な枠組みであるカルマンフィルタ (Kalman Filter)¹⁰⁾ では、現在時刻 k におけるデータから得られた推定値 ϕ_k と、過去の時刻 $[1, \dots, k-1]$ におけるデータから現在時刻の位置を予想した予測値 θ_k^- との重み付き和により最終的な推定値 (事後推定値) θ_k を得る。

$$\theta_k = G^- \theta_k^- + G \phi_k \quad (3 \cdot 20)$$

ここで、 G^- 及び G は重みである。このうち G は、カルマンゲインと呼ばれ、カルマンフィルタにより再帰的に計算される。 G^- は G により決定される。重みの働きは、大雑把にいうと、現在時刻の信号対雑音比 (SNR) が低いときは、現在時刻のデータによる推定値の信頼性が低いと考え、 ϕ_k への重み G を相対的に小さくする。現在の推定値 ϕ_k は、上述のビームフォーミング法やサブスペース法を、短区間のデータに適用して求める。予測値 θ_k^- は、運動のモデルを立てることにより求める。例えば、等速度運動をしている場合は、次式のようなモデルを用いられることが多い。

$$\theta_k^- = \theta_{k-1} + v \Delta_T + w_{k-1} \quad (3 \cdot 12)$$

ここで、 θ_{k-1} は、1 時刻前の事後推定値である。 v は音源の移動速度、 Δ_T はサンプル間の時間間隔である。 w_{k-1} は予測誤差を表す項である。このモデルは、連続したサンプル間での移動をモデル化したものであり、1 次マルコフモデル (Markov Model) と呼ばれる。カルマンフィルタでは、運動のモデルや、予測誤差などに制約があるので、近年では、カルマンフィルタをより一般化したパーティクルフィルタ (Particle Filter)^{9), 11), 12)} などが用いられることも多い。

■参考文献 (3-2~3-4 節)

- 1) 磯部 孝, “相関関数およびスペクトル,” 東京大学出版会, 1968.
- 2) 西浦敬信, 山田武志, 中村 哲, 鹿野清宏, “マイクロホンアレーを用いたCSP法に基づく複数音源位置推定,” 電子情報通信学会論文誌, vol. J83-D-II, no.8, pp. 1713-1721, 2000.
- 3) H. L. Van Trees, “Optimum array processing,” Wiley-Interscience, 2002.
- 4) 菊間信良, “アレーアンテナによる適応信号処理,” 科学技術出版, 1998.
- 5) G. Strang, “Linear algebra and its application”, Harcourt Brace Jovanovich College Publishing, 1986.
- 6) A. Cantoni and L. C. Godara, “Resolving the directions of sources in a correlated field incident on an array,” J. Acoust. Soc. Am. 67(4), pp.1247-1255, 1980.
- 7) 浅野 太, “音を分ける,” 計測と制御, vol.43, no.4, pp.325-330, 2004.
- 8) D. Johnson and D. Dudgeon, “Array processing – concepts and techniques,” Prentice hall, Englewood Cliffs, NJ, 1993.
- 9) R. Schmidt, “Multiple emitter location and signal parameter estimation,” IEEE Trans. Antennas Propag. AP-34, no.3, pp.276-280, 1982.

- 10) S. Haykin (Eds.), "Kalman filtering and neural networks," Wiley-Interscience publication, New York, 2001.
- 11) B. Ristic, S. Arulampalam and N. Gordon, "Beyond the Kalman filter," Artech house, Boston, 2004.
- 12) 浅野 太, "パーティクルフィルタを用いた移動音源の追跡技術," 日本音響学会誌, vol.61, no.12, pp.720-727, 2005.
- 13) 金井 浩, "音・振動のスペクトル解析," コロナ社, 1996.
- 14) 浅野 太, "音のアレイ信号処理," コロナ社, 2009.