

2群(画像・音・言語) - 6編(音響信号処理)

---

4章 雑音・残響抑圧

## 2群 - 5編 - 4章

### 4-1 雑音抑圧

(執筆者: 杉山昭彦) [2014年5月受領]

雑音抑圧とは、通常、音声と混在する音声以外の信号の影響を弱くして、音声を聞きやすくする処理を表す。複数の音声のうち一つをほかの音声から区別する処理は、信号分離として別に扱われる。雑音抑圧を行うシステムは、ノイズキャンセラやノイズサプレッサと称されているが、明確な定義はない。本稿では、入力信号のチャンネル数（マイクロホン数）が単数であるものをノイズサプレッサ（雑音抑圧システム）<sup>1)</sup>、複数であるものをノイズキャンセラ（雑音消去システム）<sup>2)</sup>と呼ぶ。ただし、複数のマイクロホンが規則的に配置されているものは、マイクロホンアレイとして2章 音源分離 で取り扱う。

ノイズサプレッサは、唯一の入力である音声と雑音の混合信号（以下、劣化音声と呼ぶ）から、何らかの方法で雑音を推定し、推定した雑音に基づいて、出力となる強調音声を計算する。音声ではなく雑音を推定するのは、一般的に雑音の方が音声よりも定常性が高く、推定が容易なためである。雑音推定に用いるデータが音声と雑音の混合信号だけなので、比較的SNR（信号対雑音比）が低い、すなわち雑音が多い環境や雑音の非定常性が高い環境では、信頼できる雑音推定値を得ることが困難になる。このため、ノイズサプレッサは比較的SNRの高い環境で、それ以外ではノイズキャンセラが、一般的に利用される。

ノイズサプレッサでは、ウィーナーフィルタ<sup>1)</sup>、短時間スペクトル振幅（short-time spectral amplitude: STSA）分析<sup>1)</sup>が代表的な方式であり、ノイズキャンセラでは適応フィルタを用いたシステム同定が基本技術となっている<sup>2)</sup>。STSA分析に基づく方式は、劣化音声、すなわち雑音が重畳した所望音声の振幅スペクトルから、クリーン音声（雑音を含まない音声）の振幅スペクトルを推定する。位相情報は、劣化音声の位相をそのまま利用する。このようにして得られた振幅スペクトルと位相スペクトルに逆フーリエ変換を適用することで、強調音声の時間領域信号を得る。演算量が少ないことから、STSA分析は最も広く利用されている。代表的なSTSA分析としては、ウィーナーフィルタ法<sup>1)</sup>、スペクトル減算（spectral subtraction: SS）法<sup>1,3)</sup>、最小平均二乗誤差（minimum mean-square error: MMSE）STSA法<sup>4)</sup>などがある。後に説明するように、スペクトル減算はウィーナーフィルタの特別な場合に相当する。

#### 4-1-1 時間領域のウィーナーフィルタによる雑音抑圧

図4-1に、時間領域のウィーナーフィルタによる雑音抑圧を実行するための基本構成を示す。連続する入力信号（劣化音声）サンプル  $x(t)$  ( $t$  は離散時間番号) は、音声信号  $s(t)$ 、雑音  $d(t)$  を含む。 $x(t)$  から雑音の推定値  $\hat{d}(t)$  を求め、推定雑音  $\hat{d}(t)$  を用いてウィーナーゲインを求める。求めたウィーナーゲインを劣化音声に作用させることで、出力信号を得る。

雑音の加算された音声信号（=劣化音声） $x(t)$  は、 $s(t)$  と  $d(t)$  を用いて、

$$x(t) = s(t) + d(t) \quad (4\cdot1)$$

と表すことができる。 $x(t)$ 、 $s(t)$ 、 $d(t)$  の時系列サンプルを、サイズ  $L$  のベクトルを用いて式(4.2)–(4.4)で定義すると、劣化音声ベクトル  $\mathbf{x}(t)$  は式(4.5)で表すことができる。

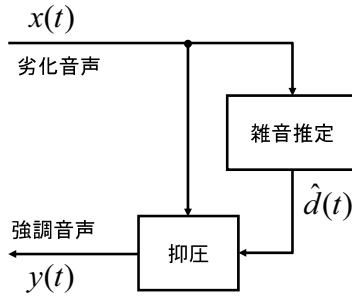


図 4・1 時間領域の雑音抑圧

$$\mathbf{x}(t) = [x(t) \ x(t-1) \ \cdots \ x(t-L+1)]^T \quad (4\cdot2)$$

$$\mathbf{s}(t) = [s(t) \ s(t-1) \ \cdots \ s(t-L+1)]^T \quad (4\cdot3)$$

$$\mathbf{d}(t) = [d(t) \ d(t-1) \ \cdots \ d(t-L+1)]^T \quad (4\cdot4)$$

$$\mathbf{x}(t) = \mathbf{s}(t) + \mathbf{d}(t) \quad (4\cdot5)$$

ここに、 $L$  は自然数、 $[\cdot]^T$  は行列の転置を表す。

係数ベクトル  $\mathbf{h}$  を有する  $L$  タップの有限インパルス応答長 (FIR) フィルタを、劣化音声ベクトル  $\mathbf{x}(t)$  に作用させて、音声  $s(t)$  の推定値  $y(t)$  (強調音声) を得ることによって雑音を抑圧することを考えると、強調音声は式 (4・6) で表される。

$$y(t) = \mathbf{h}^T \mathbf{x}(t) \quad (4\cdot6)$$

$$\mathbf{h} = [h_0 \ h_1 \ \cdots \ h_{L-1}]^T \quad (4\cdot7)$$

ウィーナーフィルタ  $\mathbf{h}_w(t)$  は、 $y(t)$  と  $s(t)$  の二乗誤差の数学的期待値  $E[\{y(t) - s(t)\}^2]$  を最小化するフィルタとして、式 (4・8) で定義される<sup>2)</sup>。

$$\mathbf{h}_w(t) = E[\mathbf{x}(t)\mathbf{x}^T(t)]^{-1} E[\mathbf{x}(t)s(t)] \quad (4\cdot8)$$

ここに、 $E[\mathbf{x}(t)\mathbf{x}^T(t)]$  と  $E[\mathbf{x}(t)s(t)]$  は、劣化音声の自己相関行列及び音声と劣化音声の相互相関行列それぞれの数学的期待値を表す。式 (4・8) のウィーナーフィルタは因果性を保証されず (非因果的)、実際に用いるときにはこれを因果関数で近似する。

音声信号  $s(t)$  は未知なので、このままではウィーナーフィルタを用いて雑音抑圧を行うことはできない。そこで、式 (4・8) に式 (4・1) を適用して、 $s(t)$  を消去する。

$$\begin{aligned} \mathbf{h}_w(t) &= E[\mathbf{x}(t)\mathbf{x}^T(t)]^{-1} \{E[\mathbf{x}(t)x(t)] - E[\mathbf{x}(t)d(t)]\} \\ &= E[\mathbf{x}(t)\mathbf{x}^T(t)]^{-1} \{E[\mathbf{x}(t)x(t)] - E[\mathbf{s}(t)d(t)] - E[\mathbf{d}(t)d(t)]\} \\ &= E[\mathbf{x}(t)\mathbf{x}^T(t)]^{-1} \{E[\mathbf{x}(t)x(t)] - E[\mathbf{d}(t)d(t)]\} \end{aligned} \quad (4\cdot9)$$

ただし、音声と雑音が無相関であると仮定し、 $E[\mathbf{s}(t)d(t)] = \mathbf{0}$  であることを利用した。式 (4・9)

右辺は、劣化音声と雑音に関する情報だけを含み、雑音に関する情報は無音声区間において推定することが可能なので、ウィナーフィルタによる雑音抑圧を実現することができる。

#### 4-1-2 短時間スペクトル振幅 (STSA) 領域の雑音抑圧

各周波数成分に対して異なる、最適な処理を実行するために、雑音抑圧は、通常、短時間スペクトル振幅 (STSA) 領域で実行される。STSA 領域の雑音抑圧は、何らかのかたちで推定して得られた推定雑音を用いて入力信号を処理することで、雑音抑圧を達成する。図 4・2 に、STSA 領域の雑音抑圧を実行するための基本構成を示す。連続する入力信号 (劣化音声) サンプル  $x(t)$  にフレーム分割、窓がけ処理、フーリエ変換などの線形変換を適用することによって、短時間スペクトル振幅  $|X(n, k)|$  を求める。ここに、 $n, k$  はそれぞれ、フレーム番号、離散周波数番号である。通常、一連の窓がけと変換処理においては、直前フレームの一部と現在のフレームを合わせて変換するオーバーラップ処理<sup>5)</sup>を適用する。続いて、雑音パワー推定値  $\lambda^2(n, k)$  を求める。

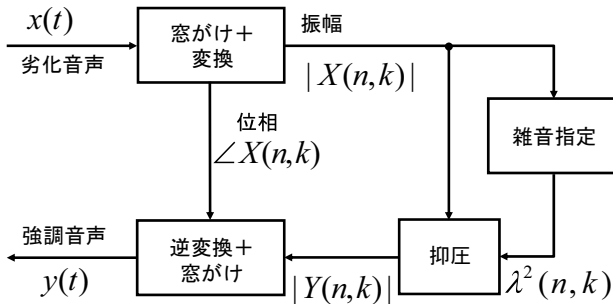


図 4・2 STSA 領域の雑音抑圧

図中の抑圧と書かれた部分で、SS 法は  $\lambda^2(n, k)$  の減算を、ウィナーフィルタと MMSE STSA 法は  $\lambda^2(n, k)$  を用いたウィナー利得または MMSE 利得の計算と入力信号振幅との乗算を実行し、 $|Y(n, k)|$  を求める。抑圧処理は振幅成分  $|X(n, k)|$  に対してだけ適用し、位相成分  $\angle X(n, k)$  はそのままのかたちで逆変換に利用する。逆変換結果は、窓がけ処理した後で直前フレームの一部と加算し、時間領域の出力信号  $y(t)$  とする。逆変換後に窓がけすることで、フレーム境界における信号の連続性を改善し、クリック音などの異音発生を防止することができる<sup>6)</sup>。

#### (1) 雑音推定

最も原始的な雑音推定法は、雑音だけが存在する初期の雑音区間で、入力信号の時間平均を用いて雑音のパワースペクトルを推定する方式である。この推定法は、特性が頻繁に変動する非定常雑音に追従できず、強調信号の品質が低下する。

追従能力を有する継続的な雑音推定法として、最小統計量に基づく最小統計法<sup>7)</sup>が広く知

られている．最小統計法では，周波数成分ごとの劣化音声パワーをフレーム（時間）方向に平均して劣化音声パワー平均値を求め，有限時間長窓内（通常は，連続する複数フレーム）に含まれる複数の平均値のうち，最小のもの（最小統計量）を検出する．この最小値を更に定数倍して，雑音推定値として用いる．これは，STSA 領域の雑音抑圧が比較的高い SNR を対象としていることから，周波数成分ごとの劣化音声最大値が音声，最小値が雑音である確率が高いという観測に基づく．平均値の最小値は雑音の最小値と平均値の間に位置するので，適切な倍率係数で補正することによって，雑音の平均値の推定値とする．統計的な性質上，窓長が長いときに定常雑音に対する推定精度が向上するが，非定常雑音に対する追従性が劣化する．反対に，短い窓長は追従性を向上させるが，過大推定を生じ，高 SNR 時の音質を低下させる．過大推定が発生する理由は，サンプル数が少ない短い窓長では，統計量としての信頼性が低いためである．窓長が短いとき，最小値が分析区間に出現する確率は低い．このように，最小統計法では，窓長の選択にトレードオフが発生する．したがって，過大推定を回避しつつ優れた追従性を達成する適切な窓長を選択することは容易でない．

重み付き雑音推定<sup>8,9)</sup>は，様々な非定常雑音源に対して，過大推定を発生させずに高い追従性を達成するために，推定事後 SNR に基づく重み係数を用いた雑音推定を採用している．重み係数は，事後 SNR が高いときに小さくなるような関数で定められている．最小統計量推定とは異なり，統計的な処理に基づいていないので，窓長選択の問題は存在しない．適切な重み係数を用いた重み付けによって，過大推定をもたらすことなく，音声区間においても継続的に雑音を推定することが可能になる．その結果，重み付き雑音推定は，音声及び非音声区間における雑音の変動に追従できる．文献 8) が最初にノイズ推定の重要性を指摘してから，数々のノイズ推定法が検討されてきた．最近の技術については，文献 10) に詳しい．

## (2) 短時間スペクトル振幅領域のウィーナーフィルタによる雑音抑圧

式 (4・6) で時間領域のウィーナーフィルタを用いたベクトルの内積（畳込み演算）として表現されていた雑音抑圧過程は，周波数領域では式 (4・10) の乗算で表現される．

$$Y(n, k) = H_w(k)X(n, k) \quad (4 \cdot 10)$$

ここに， $Y(n, k)$  は，強調音声  $y(t)$  に窓掛けと短時間離散フーリエ変換を適用して得られた周波数領域表現である．

時間領域ウィーナーフィルタ  $h_w(t)$  に対応する非因果的ウィーナーフィルタ  $H_w(n, k)$  は，次式で与えられる<sup>1)</sup>．

$$H_w(n, k) = \frac{P_s(k)}{P_s(k) + P_d(k)} \quad (4 \cdot 11)$$

ここに， $P_s(k)$  と  $P_d(k)$  はそれぞれ，音声と雑音のパワースペクトル密度である．これらのパワースペクトル密度は未知であるが，音声と雑音が STSA の 1 フレームの間，近似的に定常であると仮定すると，それぞれの短時間パワースペクトルを用いて近似することができる．このとき，非因果的ウィーナーフィルタ  $H_w(n, k)$  は式 (4・12) の時変ウィーナーフィルタ  $\hat{H}_w(n, k)$  となる．

$$\hat{H}_w(n, k) = \frac{E[|S(n, k)|^2]}{E[|S(n, k)|^2] + E[|D(n, k)|^2]} \quad (4 \cdot 12)$$

ここに  $S(n, k)$ ,  $D(n, k)$  は、音声信号  $s(t)$ , 雑音  $d(t)$  に窓掛けと短時間離散フーリエ変換を適用して得られた周波数領域表現である。

式 (4-12) の分母第二項は雑音として、音声の影響を受けない無音声区間において推定することができる。一方、分母第一項及び分子に存在する  $E[|S(n, k)|^2]$  は、

1.  $|X(n, k)|^2$  の時間平均(フレーム平均)で  $E[|X(n, k)|^2]$  を近似し  $E[|S(n, k)|^2] = E[|X(n, k)|^2] - E[|D(n, k)|^2]$  を適用する,
2.  $|X(n, k)|^2$  の部分周波数平均(包絡線)で  $E[|X(n, k)|^2]$  を近似し  $E[|S(n, k)|^2] = E[|X(n, k)|^2] - E[|D(n, k)|^2]$  を適用する,
3. 別の手段で求めた  $|Y(n, k)|^2$  で、 $E[|S(n, k)|^2] \approx |Y(n, k)|^2$  と近似する,
4. 別の手段で求めた  $|Y(n, k)|^2$  の包絡線で、 $E[|S(n, k)|^2]$  を近似する,

などの方法で求めることができる。式 (4-12) の右辺は正の実数であり、式 (4-10) による抑圧は、劣化音声の振幅だけを変化させて雑音抑圧を実行していることになる。

上記「別の手段」として、例えば、 $|Y(n, k)|^2$  が定常であるならば、 $|Y(n, k)|^2 \approx |Y(n-1, k)|^2$  と近似して、 $|Y(n, k)|^2$  を求めることができる。実際、短時間であれば、 $|Y(n, k)|^2$  が定常であるとみなせることは少なくない。また、定常でなくても、非定常性の程度に対応した誤差を受け入れる近似と考えることもできる。また、このようにして求めた  $\hat{H}_w(n, k)$  を用いて  $|Y(n, k)|^2$  を再度求める意味がどこにあるのかという疑問も生ずる。これは、 $\hat{H}_w(n, k)$  を用いて求めた  $|Y(n, k)|^2$  は、より精度が高いと考えれば納得できよう。

非因果的ウィナーフィルタ  $H_w(n, k)$  の一般形に、式 (4-13) で定義されるパラメトリックウィナーフィルタがある。

$$H_w(k) = \left( \frac{P_s(k)}{P_s(k) + \alpha P_d(k)} \right)^\beta \quad (4 \cdot 13)$$

正の実数  $\alpha$  と  $\beta$  の値を変化させることによって、特性の異なるフィルタ  $H_w(k)$  を実現することができる。  $\alpha = \beta = 1$  のときに式 (4-11) に一致する。式 (4-11) と同様に、式 (4-13) のパワースペクトル密度を短時間パワースペクトルで近似すると、式 (4-14) を得る。

$$\hat{H}_w(n, k) = \left( \frac{E[|S(n, k)|^2]}{E[|S(n, k)|^2] + \alpha E[|D(n, k)|^2]} \right)^\beta \quad (4 \cdot 14)$$

### (3) スペクトル減算による雑音抑圧<sup>3)</sup>

劣化音声の現在値を表す式 (4-1) の第一要素を短時間パワースペクトルで表し、両辺の振幅をとってから二乗すると、

$$|X(n, k)|^2 = |S(n, k)|^2 + S_n^*(k)D(n, k) + S(n, k)D_n^*(k) + |D(n, k)|^2 \quad (4 \cdot 15)$$

となる。ここに、 $\{\cdot\}^*$  は、 $\{\cdot\}$  の複素共役である。式 (4・15) 右辺の各信号は利用可能ではないので、数学的期待値をそれらの推定値として利用する。音声と雑音が無相関であるときに、

$$E[|S_n^*(k)D(n, k)|] = E[S(n, k)D_n^*(k)] = 0 \quad (4\cdot16)$$

であることを利用すると、

$$|X(n, k)|^2 = E[|S(n, k)|^2] + E[|D(n, k)|^2] \quad (4\cdot17)$$

となる。  $E[|S(n, k)|^2]$  を強調音声  $Y(n, k)$  として用いると、

$$|Y(n, k)|^2 = E[|S(n, k)|^2] = |X(n, k)|^2 - E[|D(n, k)|^2] \quad (4\cdot18)$$

となる。通常は、雑音の過剰推定による影響を低減するために、 $0 < \alpha < 1$  なる過剰減算係数 (over subtraction factor) を導入して、式 (4・19) のかたちで利用される。

$$|Y(n, k)|^2 = |X(n, k)|^2 - \alpha E[|D(n, k)|^2] \quad (4\cdot19)$$

式 (4・19) は、その式のかたちから、パワースペクトル減算法と呼ばれる。このほかに、式 (4・20) で表される、振幅スペクトル減算法もある<sup>3)</sup>。

$$|Y(n, k)| = |X(n, k) - \alpha E[|D(n, k)|] \quad (4\cdot20)$$

次に、パワースペクトル減算がウィーナーフィルタの特別な場合であることを確認する。式 (4・6) を周波数領域表現に変換してから、両辺の絶対値をとって二乗すると、式 (4・21) となる。

$$|Y(n, k)|^2 = H^2(k)|X(n, k)|^2 \quad (4\cdot21)$$

式 (4・21) を  $H^2(k)$  について解き、式 (4・17) と式 (4・18) を代入すると、

$$H^2(k) = \frac{|Y(n, k)|^2}{|X(n, k)|^2} = \frac{E[|S(n, k)|^2]}{E[|S(n, k)|^2] + E[|D(n, k)|^2]} = \hat{H}_w^2(n, k) \Big|_{\alpha=1, \beta=1/2} \quad (4\cdot22)$$

となる。式 (4・22) のフィルタを劣化音声に作用させて強調音声を求めることが、パワースペクトル減算による雑音抑圧と等価である。このフィルタを表す式 (4・22) は、式 (4・14) において  $\alpha = 1$ 、 $\beta = 1/2$  とすることで得られる。すなわち、パワースペクトル減算は、パラメトリックウィーナーフィルタの特別な場合である。

#### (4) 最小平均二乗誤差 (MMSE) による雑音抑圧<sup>4)</sup>

ウィーナーフィルタにおける短時間スペクトル振幅は最適な信号スペクトルの推定から、スペクトル減算の場合は最適な信号パワー (分散) の推定から、それぞれ導かれる。これらはいずれも、最適なスペクトル振幅の推定ではない。MMSE STSA は、雑音を含む入力から直接導かれる、最適な信号スペクトル振幅の推定である。ほかの推定値ではなく、雑音混じりの入力から直接、振幅推定値を導くことによって、低 SNR 時の残留ノイズをウィーナー

フィルタよりも低減できる。また、音声の存在確率を導入して、音声歪を増すことなく、残留雑音を低減している。

最小平均二乗誤差を抑圧の規範とした MMSE STSA では、劣化音声の振幅スペクトル  $|X(n, k)|$  に MMSE 利得 (スペクトルゲイン)  $G(n, k)$  を乗算することによって、強調信号の振幅スペクトル  $|Y(n, k)|$  を得る。そのために、次の手順でスペクトルゲインを計算する。

1. 雑音のパワースペクトルを推定
2. 事後 (*a posteriori*) SNR を推定
3. 事前 (*a priori*) SNR を推定
4. 事前 SNR の推定値と事後 SNR の推定値からスペクトルゲインを計算

まず、推定雑音パワー  $|D(n, k)|^2$  を用いて、事後 SNR  $\gamma(n, k)$  の推定値  $\hat{\gamma}(n, k)$  を、次式で求める。

$$\hat{\gamma}(n, k) = \frac{|X(n, k)|^2}{|D(n, k)|^2} \quad (4 \cdot 23)$$

式 (4・23) の分子は入力そのものであり、音声と雑音の和となっている。入力の SNR そのものであるために、事前 SNR と呼ぶ方が自然に思えるが、文献 4) にあるように、事後 SNR と定義されている。これは、雑音のない音声の基本状態であり、それに雑音を加算する操作を行った後 (事後) の SNR であると理解することができる。

このようにして求めた  $\hat{\gamma}(n, k)$  と 1 フレーム前 (フレーム  $n - 1$ ) の周波数  $k$  におけるスペクトルゲイン  $G_{n-1}(k)$  から、次に示す判定指向 (decision-directed: 判定結果に基づいて値の変わる) 計算式<sup>4)</sup>を用いて、事前 SNR  $\xi(n, k)$  の推定値  $\hat{\xi}(n, k)$  を求める。

$$\hat{\xi}(n, k) = \alpha \hat{\gamma}_{n-1}(k) G_{n-1}^2(k) + (1 - \alpha) P[\hat{\gamma}(n, k) - 1] \quad (4 \cdot 24)$$

$P[x]$  は半波整流関数であり、 $0 < \alpha < 1$  とする。事前 SNR も、事後 SNR と同じように、音声に雑音を加える前 (事前) の SNR、すなわち音声と雑音の比であると理解することができる。

最後に、事前及び事後 SNR の推定値  $\hat{\xi}(n, k)$  と  $\hat{\gamma}(n, k)$  を用いて、スペクトルゲイン  $G(n, k)$  を求める。

$$G(n, k) = \left[ (1 + v(n, k)) I_0 \left( \frac{v(n, k)}{2} \right) + v(n, k) I_1 \left( \frac{v(n, k)}{2} \right) \right] \cdot \frac{\Lambda(n, k)}{1 + \Lambda(n, k)} \frac{\sqrt{\pi v(n, k)}}{2 \hat{\gamma}(n, k)} \exp \left( -\frac{v(n, k)}{2} \right) \quad (4 \cdot 25)$$

ただし、 $I_0(z)$  と  $I_1(z)$  は、0 次と 1 次の変形ベッセル関数<sup>11)</sup>を表す。ここに、 $v(n, k)$ 、 $\Lambda(n, k)$ 、 $\eta(n, k)$  は、次のように定義される。

$$v(n, k) = \frac{\eta(n, k)}{1 + \eta(n, k)} \hat{\gamma}(n, k) \quad (4 \cdot 26)$$

$$\eta(n, k) = \frac{\hat{\xi}(n, k)}{1 - q} \quad (4 \cdot 27)$$

$$\Lambda(n, k) = \frac{1 - q}{q} \cdot \frac{\exp(v(n, k))}{1 + \eta(n, k)} \quad (4 \cdot 28)$$



音声非存在確率  $q$  は、非音声成分及びパワーが十分小さい音声成分だけが、劣化音声の周波数成分  $k$  に含まれる確率を表す。  $q$  は、すべての  $k$  に対して共通に 0.2 程度の固定値を用いる<sup>4)</sup>が、これを適応的に制御することで音質が向上するという報告もある<sup>12)</sup>。文献 4) では、  $q = 0.2$  とした根拠が示されていないが、実験的に求めた最適値であろうと思われる。

強調信号のフーリエ変換  $Y(n, k)$  は、次のように  $\angle X(n, k)$  と  $|Y(n, k)|$  で構成される。ここに、  $j = \sqrt{-1}$  とする。

$$\begin{aligned} Y(n, k) &= |Y(n, k)| \cdot \exp\{j\angle Y(n, k)\} \\ &= G(n, k)|X(n, k)| \cdot \exp\{j\angle X(n, k)\} \end{aligned} \quad (4.29)$$

MMSE STSA による雑音抑圧を実行するための基本構成を、図 4.3 に示す。

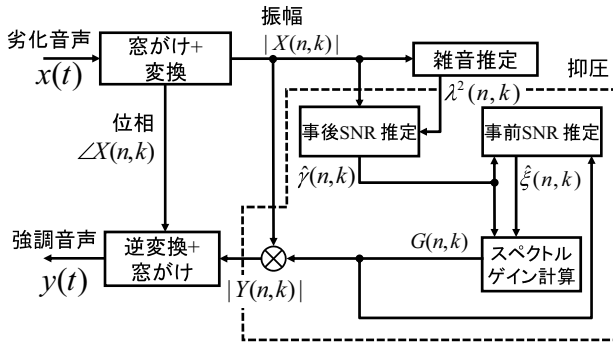


図 4.3 MMSE STSA による雑音抑圧

#### 4-1-3 2 マイクを用いた雑音消去

SNR が低くなると、1 マイク雑音抑圧では、十分に雑音を抑圧できないか、強調信号に歪を生ずる。このような過酷な雑音環境では、Widrow らによって最初に提案された適応ノイズキャンセラ (ANC: adaptive noise canceller)<sup>2)</sup>が効果的である。第一のマイク (主マイク) で、音声と雑音の混合信号である劣化音声を捕捉する。雑音源に十分近い位置に配置された第二のマイク (参照マイク) が、雑音 (参照雑音) を捕捉する。参照雑音は、適応フィルタを駆動して、主マイクにおける信号に含まれる雑音の擬似信号 (擬似雑音) を生成する。擬似雑音を主マイクの信号から減算することで、雑音の消去を行う。出力信号 (強調音声) は、残留雑音を含み、適応フィルタの係数更新に利用される。

図 4.4 に、ANC のブロック図を示す。  $s(t)$ 、  $d_0(t)$ 、  $d(t)$ 、  $g(t)$ 、  $\hat{d}(t)$  はそれぞれ、音声、雑音、主マイク信号に含まれる雑音成分、雑音源から主マイクまでの雑音経路のインパルス応答、適応フィルタの出力である。主マイク信号  $x_P(t)$  と参照信号  $x_R(t)$  は、式 (4.30) と式 (4.31) で表される。

$$x_P(t) = s(t) + d(t), \quad (4.30)$$

$$x_R(t) = d_0(t) \quad (4.31)$$

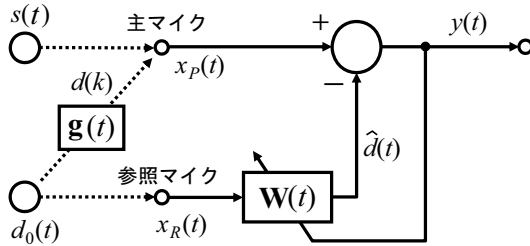


図 4・4 Widrow の適応ノイズキャンセラ.

ANC による強調音声  $y(t)$  は,

$$y(t) = s(t) + d(t) - \hat{d}(t), \quad (4\cdot32)$$

$$d(t) = \mathbf{g}^T(t) \mathbf{d}_0(t), \quad (4\cdot33)$$

$$\hat{d}(t) = \mathbf{w}^T(t) \mathbf{x}_R(t) = \mathbf{w}^T(t) \mathbf{d}_0(t), \quad (4\cdot34)$$

で与えられる。ここに、 $\mathbf{w}(t)$  は適応フィルタの係数ベクトル、 $\mathbf{x}_R(t)$  と  $\mathbf{d}_0(t)$  はサイズ  $L$  の参照信号と雑音のベクトルを表し、式 (4・35)–(4・38) で与えられる。

$$\mathbf{g}(t) = [g_0(t), g_1(t), \dots, g_{L-1}(t)]^T, \quad (4\cdot35)$$

$$\mathbf{w}(t) = [w_0(t), w_1(t), \dots, w_{L-1}(t)]^T, \quad (4\cdot36)$$

$$\mathbf{d}_0(t) = [d_0(t), d_0(t-1), \dots, d_0(t-L+1)]^T, \quad (4\cdot37)$$

$$\mathbf{x}_R(t) = [x_R(t), x_R(t-1), \dots, x_R(t-L+1)]^T = \mathbf{d}_0(t) \quad (4\cdot38)$$

式 (4・32) に式 (4・33) と式 (4・34) を代入すると、式 (4・39) となる。

$$y(t) = s(t) + [\mathbf{g}(t) - \mathbf{w}(t)]^T \mathbf{d}_0(t). \quad (4\cdot39)$$

雑音経路  $\mathbf{g}(t)$  の推定に NLMS アルゴリズム<sup>13)</sup>を用いると、 $\mathbf{w}(t)$  の更新は、

$$\mathbf{w}(t+1) = \mathbf{w}(t) + \frac{\mu y(t) \mathbf{x}_R(t)}{\|\mathbf{x}_R(t)\|^2} = \mathbf{w}(t) + \frac{\mu y(t) \mathbf{d}_0(t)}{\|\mathbf{d}_0(t)\|^2}, \quad (4\cdot40)$$

で実行される。ここに、 $\mu$  はステップサイズである。式 (4・32) と式 (4・39) から、係数更新が十分行われて、 $\hat{d}(t) \approx d(t)$  または等価的に  $\mathbf{g}(t) \approx \mathbf{w}(t)$  のときに、 $y(t)$  が  $s(t)$  になることが分かる。

出力となる強調信号  $y(t)$  は、適応フィルタの係数値と相関がなく、適応フィルタ係数の更新に対して妨害信号となる  $s(t)$  を含む。この問題は、係数更新ステップサイズを、主マイクにおける目的信号と雑音のパワー比の推定値が大きいために小さくなるように制御することで、解決できる<sup>14)</sup>。目的信号の推定値には強調信号そのものを、雑音の推定値には適応フィルタの出力を、近似的に用いることができる。しかし、近似精度を高くするためにこれらの

信号の平均値を用いると、平均操作に伴う遅延を避けることができず、急激な SNR 変化に追従できない。これは、適応フィルタ係数の不適切な更新を通じて、強調信号の歪と残留雑音を増加させることになる。

Ikedra は、新たなフィルタ（パイロットフィルタ）を用いて適応フィルタ部分を並列化し、メインのフィルタ部の経路に遅延を導入することによって、平均操作による SNR 推定遅延の影響を相殺した<sup>14)</sup>。Sugiyama は、SNR 推定の遅延はそのままに、推定遅延による不適切な係数更新が生じないように対策をとった<sup>15)</sup>。すなわち、係数更新ステップサイズを、係数パワーに比例し、その係数に入力される信号サンプルパワーに反比例するように制御する。演算量を削減するために、係数ベクトルを  $M$  個に分割し、各部分ベクトルが同じステップサイズを共用する。この係数更新アルゴリズムは、ブロック分割パワー正規化 PNLMS (PP-PNLMS) と呼ばれている。

これまででは、目的音源から参照マイクに至る経路の存在を無視してきたが、一般的な応用ではそのような経路によって生ずるクロストークの影響が問題となる。Al-Kindi は、クロストーク経路に対応した別の適応フィルタを元の適応フィルタと交差するかたちで有する交差結合構成 (Cross-Coupled Structure : CCS) を導入して、この問題に対応した<sup>16)</sup>。一方、Zinser は、ANC 出力を帰還して別の適応フィルタで処理することによってクロストークを消去し、その出力信号を本来の ANC の参照信号として用いることで、クロストークの影響を低減する耐クロストーク ANC ( Crosstalk resistant ANC : CTRANC ) を提案した<sup>17,18)</sup>。実際、CCS と CTRANC は等価である<sup>19)</sup>。CCS/CTRANC にも、前述のパイロットフィルタまたは PP-PNLMS を導入することができ、係数更新における妨害信号問題に対処することができる<sup>20)</sup>。

#### 参考文献

- 1) J.S. Lim and A. V. Oppenheim, "Enhancement and bandwidth compression of noisy speech," Proc. IEEE, vol.67, no.12, pp.1586–1604, Dec. 1979.
- 2) B. Widrow, J.R. Grover, Jr., J.M. McCool, J. Kaunitz, C.S. Williams, R.H. Hearn, J.R. Zeidler, E. Dong, Jr., and R.C. Goodlin, "Adaptive noise cancelling: Principles and applications," Proc. of the IEEE, vol.63, no.12, pp.1692–1716, Dec. 1975.
- 3) S.F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Trans. Acoust., Speech, Signal Processing, vol.ASSP-27, no.2, pp.113–120, Apr. 1979.
- 4) Y. Ephraim and D. Malah, "Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator," IEEE Trans. Acoust., Speech, Signal Processing, vol.ASSP-32, no.6, pp.1109–1121, Dec. 1984.
- 5) This is to be filled.
- 6) A. Sugiyama, T.P. Hua, M. Kato, M. Serizawa, "Noise suppression with synthesis windowing and pseudo noise injection," Proc. of ICASSP'02, pp.545–548, May 2002.
- 7) R. Martin, "Spectral subtraction based on minimum statistics," EUSIPCO'94, pp.1182–1185, Sep. 1994.
- 8) M. Kato, A. Sugiyama and M. Serizawa, "Noise Suppression with High Speech Quality Based on Weighted Noise Estimation and MMSE STSA," Proc. IWAENC2001, pp.183–186, Sep. 2001.
- 9) M. Kato, A. Sugiyama, M. Serizawa, "Noise suppression with high speech quality based on weighted noise estimation and MMSE STSA," IEICE Trans. Fund. vol.E85-A, no.7, pp.1710–1718, Jul. 2002.
- 10) J. Taghia, J. Taghia, N. Moohammadiha, J. Sang, V. Bouse, and R. Martin, "An evaluation of noise power spectral density estimation algorithms in adverse acoustic environments," Proc. ICASSP2011,

- pp.4640–4643, May 2011.
- 11) 岩波数学辞典, 第 3 版, 日本数学会, 1985 年 12 月 .
  - 12) M.-S. Choi and H.-G. Kang, “An improved estimation of a priori speech absence probability for speech enhancement in perspective of speech perception,” Proc. ICASSP2005, pp.1117–1120, Apr. 2005.
  - 13) G.C. Goodwin and K.S. Sin, “Adaptive filtering, prediction and control,” Englewood Cliffs, NJ: Prentice-Hall Info. Syst. Sci., 1985.
  - 14) S. Ikeda and A. Sugiyama, “An adaptive noise canceller with low signal distortion for speech codecs,” IEEE Trans. Signal Processing, vol.47, no.3, pp.665–674, Mar. 1999.
  - 15) A. Sugiyama, M. Kato, M. Serizawa, “A low-distortion noise canceller with an SNR-modified partitioned power-normalized PNLMS,” Proc. APSIPA2009, pp.222–225, Nov. 2009.
  - 16) M.J. Al-Kindi and J. Dunlop, “A low distortion adaptive noise cancellation structure for real time applications,” Proc. ICASSP’87, pp.2153–2156, Apr. 1987.
  - 17) R.L. Zinser, G. Mirchandani, J.B. Evans, “Some experimental and theoretical results using a new adaptive filter structure for noise cancellation in the presence of crosstalk,” ICASSP’85, pp.1253–1256, Mar. 1985.
  - 18) G. Mirchandani, R.L. Zinser and J.B. Evans, “A New Adaptive Noise Cancellation Scheme in the Presence of Crosstalk,” IEEE Trans. Circuits and Systems, vol.39, no.10, pp.681–694, Oct. 1992.
  - 19) A. Sugiyama, “Low-distortion noise cancellers – Revival of a classical technique,” *Speech and audio processing in adverse environment*, Chap.7, Hänslér and Schmidt, ed. Springer, 2008.
  - 20) S. Ikeda and A. Sugiyama, “An adaptive noise canceller with low signal-distortion in the presence of crosstalk,” IEICE Trans. Fund., vol.E82-A, no.8, pp.1517–1525, Aug. 1999.

## ■2群 - 6編 - 4章

### 4-2 残響抑圧

(執筆著者：古家賢一) [2011年10月受領]

部屋の残響が存在する環境下でマイクロホンから離れた位置にある音源からの音を收音する場合、音源からの直接音に加えて残響も一緒に收音され音が変形する。特に音声收音の場合、この残響が多い場合には音声聞き取りにくくなる問題がある。この残響の影響を取り除き、音源信号を回復する処理を残響抑圧という。残響抑圧と雑音抑圧は、妨害音を抑圧するという点では似ている。しかし、通常、雑音抑圧では音源信号と妨害音(雑音)の無相関性を仮定できるが、残響抑圧の場合にはパワーの大きな残響の初期反射部分は音源信号と相関が強い音であり無相関性は仮定できない。

本節では、まず、室内インパルス応答の逆フィルタ処理によって残響を抑圧する方法について述べる。次に、音源信号波形の厳密な復元を行うのではなく聴感上で残響感が低減される部分のみに非線形処理を行う方法について述べる。

#### 4-2-1 室内インパルス応答の逆フィルタ処理による方法

残響を線形システムである室内インパルス応答による変形と考えた場合、その特性を打ち消す逆フィルタによって残響を除去することができる。ただし、室内インパルス応答は一般には非最小位相系であり、その逆フィルタは不安定になる<sup>1)</sup>。この逆フィルタの不安定性を解決する方法として、不要な零点の除去、遅延の付加、雑音の付加によって逆フィルタを近似的に実現する方法、複数のマイクロホンを用いて多チャンネル逆フィルタにより実現する方法など<sup>2)-6)</sup>がある。

音源信号  $s(t)$  で音源から発音される音を  $M$  本のマイクロホンで收音する場合を考える。音源から  $m$  番目のマイクロホンまでのインパルス応答を  $h_m(t)$  とし、 $x_m(t)$  を  $m$  番目のマイクロホンで收音された信号とすると、

$$x_m(t) = h_m(t) \otimes s(t) \quad (4.41)$$

である。ただし、 $\otimes$  は畳込み演算を表す。

多チャンネル逆フィルタとしては、MINT法<sup>5)</sup>による逆フィルタが広く用いられている。MINT法では、部屋を多チャンネル入出力の線形システムでモデル化することにより、逆フィルタが安定となる条件を導き、その安定な逆フィルタにより原理的には残響を除去できることを示した。MINT法の条件を満たす逆フィルタを用いると各マイクロホンまでのインパルス応答  $h_m(t)$  の間には、次式の関係が成り立つ

$$\sum_{m=1}^M g_m(t) \otimes h_m(t) = \delta(t) \quad (4.42)$$

ここで、 $g_m(t)$  は  $m$  番目のマイクロホン入力に接続される逆フィルタであり、 $\delta(t)$  はデルタ関数を表す。式(4.41)と式(4.42)の関係を用いて、残響によって歪んだ信号  $x_m(t)$  から音源信号は次式で回復される。

$$\hat{s}(t) = \sum_{m=1}^M g_m(t) \otimes x_m(t) \quad (4.43)$$

実際に逆フィルタを計算するには、通常は元となる部屋のインパルス応答を測定しなければならないが、音声の残響抑圧など音源が人間である場合などには、インパルス応答を測定することが難しい。このような場合に対して、インパルス応答を測定することなく、マイクロホンで収録した信号のみから逆フィルタを推定する方法が用いられる。逆フィルタを推定する方法としては、インパルス応答ブラインド推定を用いる方法<sup>7)-16)</sup>、線形予測を用いる方法<sup>17)-19)</sup>、マイクロホン間の到達時間差を用いる方法<sup>20)-22)</sup>、音声の調波構造を用いる方法<sup>23), 24)</sup>などがある。

逆フィルタを用いる方法では、一般的に少ないマイクロホン数で高い残響抑圧性能が得られるが、計算量が大きくリアルタイム処理が難しい。また、到来方向が比較的はっきりしている残響の初期反射成分に対しては効果が大きいですが、統計的に揺らぎのある後部残響成分に対しては効果が小さい。

#### 4-2-2 非線形処理により残響感を低減する方法

逆フィルタ処理による方法は、インパルス応答の変動や揺らぎがある場合にはそれらに追従しなければならない困難が伴う。一方、音源の性質や人間の聴覚特性を利用して、逆フィルタ処理のように厳密な原波形の復元を行うのではなく、環境の変動に影響されな聴感上で残響感が低減される部分のみに非線形な処理を行う方法がある。このような方法として、音源パワースペクトラム推定による方法、MTF (Modulation Transfer Function) 理論に基づく方法、線形予測残差制御による方法を説明する。

##### (1) 音源パワースペクトラム推定による方法

音源パワースペクトラム推定による方法<sup>22), 25)-27)</sup>では、音声の明瞭度を決める要因は主にパワースペクトルであり、位相はあまり問題とならないことを利用し、雑音抑圧で用いられるウイナーフィルタやスペクトル減算法と同様に、パワースペクトル領域で残響パワーを推定し受信信号から差し引くあるいは減衰させることにより残響を抑圧して音源信号を復元する。

短時間フーリエ分析 (STFA : Short-Time Fourier Analysis) された時間周波数領域で、ある時刻  $t$  フレーム、帯域  $\omega$  に含まれる残響パワー  $P^2(\omega, t)$  は、過去の観測信号の複数フレームのパワーに係数を掛けて積和した次式で推定される。

$$P^2(\omega, t) = \sum_{t_0} \alpha(\omega, t_0) X^2(\omega, t - t_0) \quad (4.44)$$

ここで、 $X^2(\omega, t)$  は時刻  $t$  フレーム、帯域  $\omega$  における観測信号  $x(t)$  のパワーを表し、 $\alpha(\omega, t_0)$  は帯域  $\omega$  における時間  $t_0$  前の過去フレームに対する係数である。係数  $\alpha(\omega, t_0)$  の決定方法には部屋のインパルス応答パワーを指数関数でモデル化し算出する方法<sup>25), 26)</sup>、過去フレームとの相互相関から算出する方法<sup>22)</sup>、観測信号の線形予測係数から算出する方法<sup>27)</sup> などがある。推定された残響パワーから観測信号にかける残響抑圧するためのウイナーフィルタは次式で計算される。

$$W(\omega, t) = \frac{X^2(\omega, t) - P^2(\omega, t)}{X^2(\omega, t)} \quad (4.45)$$

このウィナーフィルタを観測信号にかけることによって時間周波数領域で残響を抑圧し音源信号を回復する。

$$\hat{S}(\omega, t) = W(\omega, t)X(\omega, t) \quad (4.46)$$

この方法はパワースペクトルのみに着目し位相にはあまり影響されないで、逆フィルタ法に比較して室内インパルス応答の変動に頑強であり、聴感上も残響感を低減できる。しかし、ミュージカルノイズに代表される非線形処理に伴う人工的な歪が発生し音質が劣化する場合がある。

## (2) MTF 理論に基づく方法

MTF 理論に基づく方法<sup>28), 29)</sup>では、音源信号と部屋の室内インパルス応答を変調雑音信号を用いて次式でモデル化する。

$$s(t) = e_s(t)n_1(t)$$

$$h(t) = e_h(t)n_2(t) \quad (4.47)$$

ここで、 $e_s(t)$ 、 $e_h(t)$ はそれぞれ音源信号と部屋のインパルス応答の包絡線を表し、 $n_1(t)$ 、 $n_2(t)$ は互いに無相関な白色雑音である。このとき観測信号のパワー時間包絡線  $e_x(t)$ は次式で推定される。

$$\begin{aligned} e_x^2(t) &= \langle x^2(t) \rangle \\ &= \langle (s(t) \otimes h(t))^2 \rangle \\ &= e_s^2(t) \otimes e_h^2(t) \end{aligned} \quad (4.48)$$

ここで、 $\langle * \rangle$ は集合平均を表す。更に、室内インパルス応答の時間包絡線を残響時間で決まる指数関数でモデル化すると、式(4.48)から  $e_h^2(t)$ の逆畳み込みを用いて  $e_s^2(t)$ から音源信号の時間包絡線が推定できる。推定された音源信号の時間包絡線  $\hat{e}_s(t)$ から音源信号は次式で復元される。

$$\hat{s}(t) = \hat{e}_s(t) \frac{x(t)}{e_x(t)} \quad (4.49)$$

この方法は、残響時間が長く室内インパルス応答が指数関数でよく近似される場合には、簡単な処理で残響感を抑圧できるが、被変調信号  $n_1(t)$ が回復できないことや実環境において必ずしもモデルに当てはまらない場合には音質の改善が困難であり、実環境への適用にはまだ課題が残されている。

## (3) 線形予測残差制御による方法

残響抑圧の目的を音声強調に限定し、音声生成モデルに基づく線形予測残差に現われる残響による付加的なパルス列の振幅を制御することにより残響を抑圧する方法<sup>30)-35)</sup>がある。この方法においては、線形予測係数で表されるパワースペクトル包絡に対する残響の影響は小さいものとして、振幅制御された残差をそのまま線形予測フィルタを通すことによって音声を復元する。付加的なパルス列の振幅を制御する方法として様々な方法が提案されているが、これらの方法では残差における残響によるパルス列は抑圧されるが、復元された音声の自然性が悪くなる欠点がある。

#### ■参考文献

- 1) S. T. Neely and J. B. Allen, "Invertibility of a room impulse response," *J. Acoust. Soc. Amer.*, vol.66, no.1, pp.165-169, July 1979.
- 2) J. Mourjopoulos, P. Clarkson, and J. Hammond, "A comparative study of least-squares and homomorphic techniques for the inversion of mixed phase signals," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, vol.7, pp.1858-1861, May 1982.
- 3) P. A. Nelson, F. Orduña-Brustamante, and H. Hamada, "Inverse filter design and equalization zones in multichannel sound reproduction," *IEEE Trans. Speech Audio Processing*, vol.3, no.3, pp.185-192, Nov. 1995.
- 4) B. D. Radlović and R. A. Kennedy, "Nonminimum-phase equalization and its subjective importance in room acoustics," *IEEE Trans. Speech Audio Processing*, vol.8, no.6, pp.728-737, Nov. 2000.
- 5) M. Miyoshi and Y. Kaneda, "Inverse filtering of room acoustics," *IEEE Trans. Acoust., Speech, Signal Processing*, vol.36, no.2, pp.145-152, Feb. 1988.
- 6) K. Yamada, J. Wang, and F. Itakura, "Recovering of broad band reverberant speech signal by sub-band MINT method," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp.969-972, 1991.
- 7) G. Xu, H. Liu, L. Tong, and T. Kailath, "A least-squares approach to blind channel identification," *IEEE Trans. Signal Processing*, vol.43, no.12, pp.2982-2993, Dec. 1995.
- 8) Y. Huang and J. Benesty, "A class of frequency-domain adaptive approaches to blind multichannel identification," *IEEE Trans. Signal Processing*, vol.51, no.1, pp.11-24, Jan. 2003.
- 9) L. Tong and S. Perreau, "Multichannel blind identification: from subspace to maximum likelihood methods," *vol.86, no.10, pp.1951-1968, Oct. 1998.*
- 10) S. Gannot and M. Moonen, "Subspace methods for multimicrophone speech dereverberation," *EURASIP Journal on Applied Signal Processing*, vol.2003, no.11, pp.1074-1090, Oct. 2003.
- 11) Y. Huang and J. Benesty, "Adaptive multi-channel least mean square and Newton algorithms for blind channel identification," *Signal Process.*, vol.82, no.8, pp.1127-1138, Aug. 2002.
- 12) N. D. Gaubitch, J. Benesty, and P. A. Naylor, "Adaptive common root estimation and the common zeros problem in blind channel identification," in *Proc. European Signal Processing Conf. (EUSIPCO)*, Antalya, Turkey, Sept. 2005.
- 13) Yiteng Huang, J. Benesty, and Jingdong Chen, "Optimal step size of the adaptive multichannel lms algorithm for blind simo identification," *IEEE Signal Processing Lett.*, vol.12, no.3, pp. 173-176, Mar. 2005.
- 14) Y. Huang, J. Benesty, and J. Chen, "Blind channel identification based two-stage approach to separation and dereverberation of speech signals in a reverberant environment," *IEEE Trans. Speech Audio Process.*, vol.13, no.5, pp.882-896, 2005.
- 15) Md. K. Hasan, J. Benesty, P. A. Naylor, and D. B. Ward, "Improving robustness of blind adaptive multichannel identification algorithms using constraints," in *Proc. European Signal Processing Conf. (EUSIPCO)*, Antalya, Turkey, Sept. 2005.
- 16) Ken'ichi Furuya and Yutaka Kaneda, "Two-channel blind deconvolution for non-minimum phase impulse responses," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, pp.1315-1318, 1997.
- 17) M. Triki and D. T. M. Slock, "Delay and predict equalization for blind speech dereverberation," *Proc. Of the ICASSP 2006*, vol.5, pp.97-100, France, May 2006.
- 18) T. Hikichi, M. Delcroix, and M. Miyoshi, "Blind dereverberation based on estimates of signal transmission



- channels without precise information on channel,” Proc. Of the ICASSP 2005, vol.1, pp.1069-1072, USA, Mar. 2005.
- 19) M. Delcroix, T. Hikichi, and M. Miyoshi, “On the use of LIME dereverberation algorithm in an acoustic environment with a noise source,” Proc. Of the ICASSP 2006, vol.1, pp.825-828, France, May 2006.
  - 20) K. Furuya, “Noise reduction and dereverberation using correlation matrix based on the multiple-input/output inverse-filtering theorem (MINT),” Proc.of International Workshop on Hands-free Speech Communication, pp.59-62, Japan, Apr. 2001.
  - 21) 古家, 片岡, “チャンネル間相関行列と音声の白色化フィルタを用いた Semi-Blind 残響抑圧,” 信学論(A), vol.J88-A, no.10, pp.1089-1099, 2005.
  - 22) K. Furuya and A. Kataoka, “Robust speech dereverberation using Multichannel blind deconvolution with spectral subtraction,” IEEE Trans. Audio Speech Lang. Process., vol.15, no.5, pp.1579-1591, 2007.
  - 23) 中谷, 三好, 木下, “調波構造に基づくモノラル音声信号のブラインド残響除去,” 信学論(D-II), vol.J88-D-II, no.3, pp.509-520, 2005.
  - 24) T. Nakatani, K. Kinoshita, and M. Miyoshi, “Harmonicity based blind dereverberation for single channel speech signals,” IEEE Trans. Audio Speech Lang. Process., vol.15, no.1, pp.80-95, 2007.
  - 25) K. Lebart and J. M. Boucher, “A new method based on spectral subtraction for speech dereverberation,” Acta Acustica, vol.87, pp.359-366, 2001.
  - 26) E. A. P. Habets, “Multi-channel speech dereverberation based statistical model of late reverberation,” Proc. ICASSP 2005, vol.4, pp.173-176, USA, March 2005.
  - 27) K. Kinoshita, T. Nakatani, and M. Miyoshi, “Spectral subtraction steered by multi-step forward linear prediction for single channel speech dereverberation,” Proc. ICASSP2006, pp.817-820, May 2006.
  - 28) 広林, 野村, 小池, 東山, “パワーエンベロープ伝達関数の逆フィルタ処理による残響音声の回復,” 信学論(A), vol.J81-A, no.10, pp.1323-1330, 1998.
  - 29) M. Unoki, M. Furukawa, K. Sakata, and M. Akagi, “A method based on the MTF concept for dereverberating the power envelope from the reverberant signal,” Proc. ICASSP 2003, vol.1, pp.840-843, 2003.
  - 30) M. S. Brandstein and S. M. Griebel, “Nonlinear, model-based microphone array speech enhancement,” in Acoustic Signal Processing for Telecommunication, S. L. Gay and J. Benesty, Eds., pp.261-279. Kluwer Academic Publishers, 2000.
  - 31) B. Yegnanarayana and P. Satyanarayana, “Enhancement of reverberant speech using LP residual signal,” IEEE Trans. Acoust., Speech, Signal Processing, vol.8, no.3, pp.267-281, May 2000.
  - 32) S. M. Griebel and M. S. Brandstein, “Microphone array speech dereverberation using coarse channel estimation,” in Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, vol.1, pp.201-204, 2001.
  - 33) B. Yegnanarayana, S. R. Mahadeva Prasanna, and K. Sreenivasa Rao, “Speech enhancement using excitation source information,” in Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, vol.1, pp.541-544, 2002.
  - 34) B. W. Gillespie, H. S. Malvar, and D. A. F. Florêncio, “Speech dereverberation via maximum-kurtosis subband adaptive filtering,” in Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing, vol.6, pp.3701-3704, 2001.
  - 35) N. D. Gaubitch, P. A. Naylor, and D. B. Ward, “Mult-microphone speech dereverberation using spatio-temporal averaging,” in Proc. European Signal Processing Conf. (EUSIPCO), Vienna, Austria, pp.809-812, Sept. 2004.