

通信会議における遠方音声収録のための残響抑圧方式の研究

古家, 賢一

<https://doi.org/10.15017/459055>

出版情報 : Kyushu University, 2005, 博士 (芸術工学), 論文博士
バージョン :
権利関係 :

第3章

入力相関行列の最小固有ベクトルを用いた二入力残響抑圧

3.1 まえがき

本章では，従来，MINT法では困難であった事前インパルス応答が測定できない応用において室内伝達関数が非最小位相である場合にも，フィルタ次数を決定でき残響除去がおこなえる入力相関行列を用いた二入力残響抑圧法について述べる。

二入力残響抑圧法の特徴は，事前情報として入力チャネル間の相関行列のみを用いて室内インパルス応答を推定し，逆フィルタ計算が可能な点である。事前にインパルス応答の測定が必要でないことは，家庭での使用や日常的に利用する会議などへの応用では極めて重要なことである。また，逆フィルタを用いて残響抑圧を行うことにより，遅延和アレー法に比べ高い効果が期待できる。

この方法では，まず入力相関行列の最小固有ベクトルをいろいろなモデル化次数に対して計算し，原音声が回復できたときに最小となる評価関数を用いて次数を決定する。次にその次数で推定された伝達関数の逆フィルタを計算し波形回復をおこなう。低次の非最小位相伝達系の場合にはほぼ完璧に逆フィルタが計算でき波形回復できる。さらに，部屋の残響などモデル化次数が数千次におよぶ場合には，大規模行列の最小固有ベクトルおよび逆行列を計算する必要

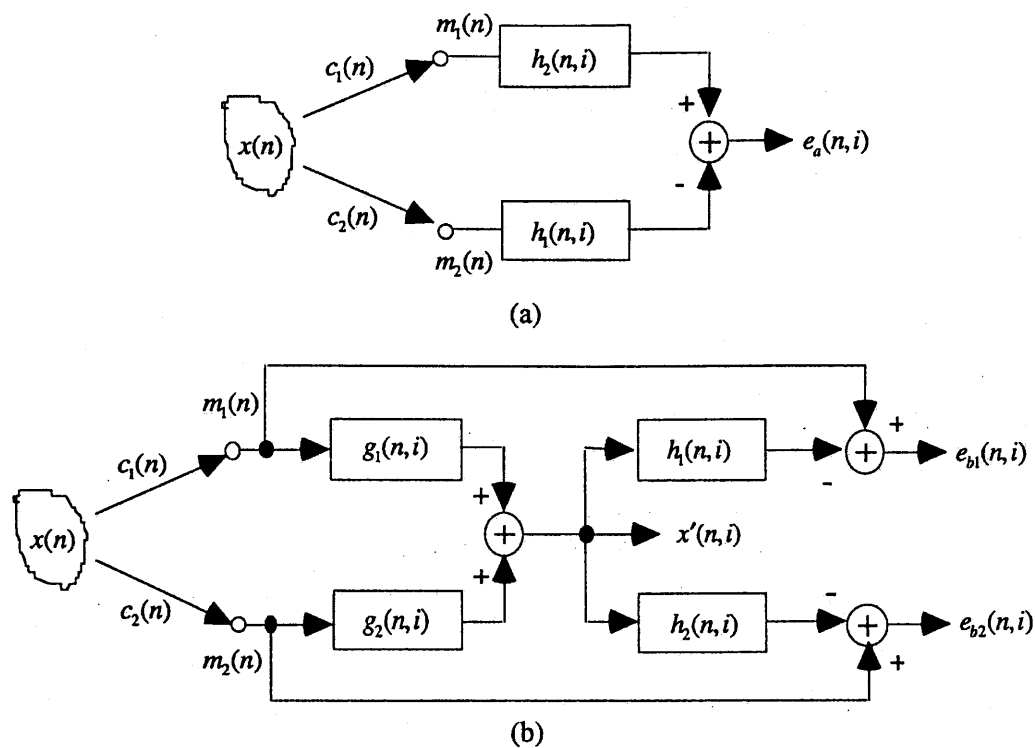


図 3.1: 二入力残響抑圧の構成 (a) 第一ステップ: インパルス応答の推定, (b) 第二ステップ: 最適次数 i の探索

があり, これらを計算することは容易ではなかった点を, 帯域分割処理 [58] を用い, 逆フィルタの計算を容易にして再合成する.

次節では, 二入力残響抑圧法の原理を述べ, 低次の非最小位相伝達系の場合には逆フィルタが計算でき波形回復できることを確認する. 次に, 第 3.3 節において, 残響のような数千次におよぶ伝達関数に対応するため帯域分割処理を導入し, 実験により検討を行い, 第 3.4 節でそれらの結果をまとめる.

3.2 最小固有ベクトルを用いた二入力残響抑圧の原理

二入力残響除去法の処理の基本構成を図3.1に示す。処理は、いろいろなフィルタ次数に対して伝達関数の推定値を計算するステップと最適次数を決定するステップの2段階からなる。以下に、その各処理について述べる。

3.2.1 インパルス応答の推定

図3.1(a)に示すように、音源信号を $x(n)$ 、音源からマイクロホンまでのインパルス応答をそれぞれ $c_1(n), c_2(n)$ 、マイクロホンでの観測信号を $m_1(n), m_2(n)$ とする。ここで、インパルス応答 $c_1(n), c_2(n)$ の z 変換が伝達関数である。 $m_1(n), m_2(n)$ は、それぞれフィルタ $h_2(n, i), h_1(n, i)$ に通された後 (i はフィルタ次数)、差分がとられ誤差信号 $e_a(n, i)$ となる。

$c_1(n), c_2(n)$ は次数 d のFIRフィルタでモデル化できるとし、互いに共通する零点をもたないと仮定する。このとき、 $e_a(n, i)$ は次式であらわせる。

$$\begin{aligned} e_a(n, i) &= m_1(n) \otimes h_2(n, i) - m_2(n) \otimes h_1(n, i) \\ &= x(n) \otimes c_1(n) \otimes h_2(n, i) - x(n) \otimes c_2(n) \otimes h_1(n, i) \\ &= x(n) \otimes \{c_1(n) \otimes h_2(n, i) - c_2(n) \otimes h_1(n, i)\} \end{aligned} \quad (3.1)$$

ここで、記号 \otimes は畳み込みをあらわす。もし、式(3.1)において $i = d$ が成立し、かつ、すべての n に対して $e_a(n, i) = 0$ であるならば、

$$\begin{aligned} h_1(n, i) &= \alpha c_1(n) \\ h_2(n, i) &= \alpha c_2(n) \end{aligned} \quad (3.2)$$

が成り立つ。ここで、 α は任意定数である。逆に式(3.2)がなりたてば、 $e_a(n, i) = 0$ となる。

$e_a(n, i)$ は計算誤差、測定誤差などにより正確には0とならないので、 $e_a(n, i)$ の二乗平均を最小にするような $h_1(n, i), h_2(n, i)$ を計算する。 $e_a(n, i)$ の二乗平均

は次式でかける.

$$\begin{aligned} E\{e_a^2(n, i)\} &= E\{\mathbf{h}^T(i) \mathbf{m}(i) \mathbf{m}^T(i) \mathbf{h}(i)\} \\ &= \mathbf{h}^T(i) \mathbf{R}(i) \mathbf{h}(i) \end{aligned} \quad (3.3)$$

ここで, $\mathbf{h}(i)$ はフィルタ係数ベクトル:

$$\mathbf{h}(i) = \begin{bmatrix} h_2(0, i) \\ h_2(1, i) \\ \vdots \\ h_2(i, i) \\ -h_1(0, i) \\ -h_1(1, i) \\ \vdots \\ -h_1(i, i) \end{bmatrix},$$

$\mathbf{m}(i)$ は入力信号ベクトル:

$$\mathbf{m}(i) = \begin{bmatrix} m_1(n) \\ m_1(n-1) \\ \vdots \\ m_1(n-i) \\ m_2(n) \\ m_2(n-1) \\ \vdots \\ m_2(n-i) \end{bmatrix},$$

$\mathbf{R}(i)$ は入力相関行列:

$$\mathbf{R}(i) = E\{\mathbf{m}(i) \mathbf{m}^T(i)\}$$

であり、 $E\{\cdot\}$ は期待値をあらわす。 $e_a(n, i)$ の二乗平均をフィルタノルム $\|\mathbf{h}(i)\|$ 一定で最小化するフィルタ係数ベクトル $\mathbf{h}(i)$ は、入力相関行列 $\mathbf{R}(i)$ の最小固有値に対応する最小固有ベクトルとして計算される。

3.2.2 最適次数の決定と音源回復

一般的には、伝達関数 $c_1(n), c_2(n)$ の次数は未知であり、フィルタ次数が適当でなければ式 (3.2) は成り立たない。そこで、フィルタ次数を決定するために図 3.1(b) に示すような二層フィルタ構成から導かれる評価関数を導入する。

図 3.1(b) において、 $g_1(n, i), g_2(n, i)$ は $h_1(n, i), h_2(n, i)$ の MINT 逆フィルタ [44] であり、次式のディオファントス方程式を解くことにより計算される。

$$G_1(z, i)H_1(z, i) + G_2(z, i)H_2(z, i) = 1 \quad (3.4)$$

ここで、 $G_1(z, i), G_2(z, i), H_1(z, i), H_2(z, i)$ はそれぞれ $g_1(n, i), g_2(n, i), h_1(n, i), h_2(n, i)$ の z 変換である。次数を i として得られたときの、 $g_1(n, i), g_2(n, i)$ を用いて、音源信号 $x(n)$ の残響回復をした信号を $x'(n, i)$ とあらわすと、

$$x'(n, i) = m_1(n) \otimes g_1(n, i) + m_2(n) \otimes g_2(n, i) \quad (3.5)$$

となる。このとき、次式の評価関数を採用する。

$$PE(i) = \frac{E\{e_{b1}^2(n, i)\}}{E\{m_1^2(n)\}} + \frac{E\{e_{b2}^2(n, i)\}}{E\{m_2^2(n)\}} \quad (3.6)$$

ここで、

$$e_{b1}(n, i) = m_1(n) - x'(n, i) \otimes h_1(n, i)$$

$$e_{b2}(n, i) = m_2(n) - x'(n, i) \otimes h_2(n, i)$$

である。この評価関数は、各次数 i に対して推定された $h_1(n, i), h_2(n, i)$ と $x'(n, i)$ を用いて、実際の残響信号 $m_1(n) = x(n) \otimes c_1(n), m_2(n) = x(n) \otimes c_2(n)$ をどの程度良く表わせるかを評価するものである。このとき、 $PE(i) = 0$ つまり $e_{b1}(n, i) = 0, e_{b2}(n, i) = 0$ ならば、またそのときに限り次式が成り立つ。

$$\begin{aligned} h_1(n, i) &= \alpha c_1(n) \\ h_2(n, i) &= \alpha c_2(n) \\ x'(n, i) &= \frac{1}{\alpha} x(n) \end{aligned} \tag{3.7}$$

式 (3.7) の証明は付録とする。

$c_1(n), c_2(n)$ は共通零点を持たないと仮定しているので、観測信号 $m_1(n), m_2(n)$ の共通零点は音源信号 $x(n)$ を表し、非共通零は $c_1(n), c_2(n)$ の特性を表している。したがって、評価関数 $PE(i)$ は、観測信号から共通零点を取り除くことにより最小化されると考えられる。もし、 $c_1(n), c_2(n)$ が共通零点を持つ場合、その零点は取り除かれ残りの零点が、 $c_1(n), c_2(n)$ として推定される。

推定されたインパルス応答の最適次数 I は、次の手順で決定される。

- (i) 相関行列 $\mathbf{R}(i)$ の最小固有ベクトルを計算することにより、次数 i のときのインパルス応答の推定値 $h_1(n, i), h_2(n, i)$ を計算する。
- (ii) 式 (3.4) を用いて逆フィルタ $g_1(n, i), g_2(n, i)$ を計算する。
- (iii) 式 (3.5) を用いて $x'(n, i)$ を計算する。
- (iv) 式 (3.6) から評価関数 $PE(i)$ を計算する。
- (v) 様々な次数 i に対して (i) から (iv) を繰り返し計算し、評価関数 $PE(i)$ を最小とする次数を最適次数 I として選択する。

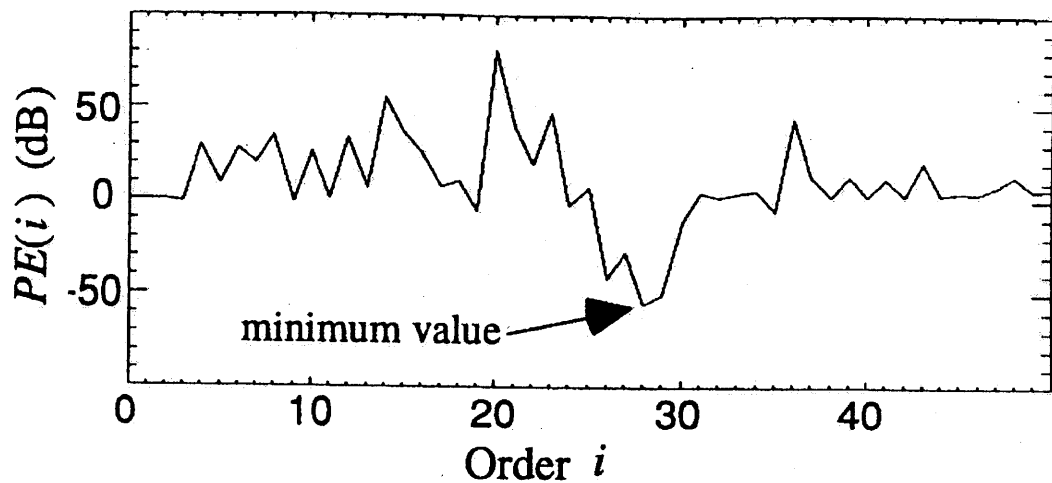


図 3.2: 最適次数 i の探索

そして、最適次数 I に対する推定値 $h_1(n, I), h_2(n, I), x'(n, I)$ を最終的な推定値とする。

3.2.3 シミュレーション実験

本方法の原理確認を行うため、残響除去のシミュレーション実験を行った。伝達系のインパルス応答 $c_1(n), c_2(n)$ の一例として、図 3.3 に示す 30 次のもの（点線で示す真値）を用い、音源信号には白色雑音を用いた。

図 3.2 は、フィルタ $h_1(n, i), h_2(n, i)$ の次数を変えたときの評価関数 $PE(i)$ の値を dB 表示している。図 3.2 より次数が 28 次するとき評価関数が最小となっている。これは $c_1(n), c_2(n)$ の次数と 2 次分の差があるが、各伝達系のインパルス応答の先頭には 2 サンプル分の遅延があり、この遅延が共通零点として除かれて推定されたと考えられる。つまり、この方法は $c_1(n), c_2(n)$ の共通零点を除くという効果もある。

図 3.3 は、フィルタ次数が 28 次するときの、インパルス応答の推定結果を示

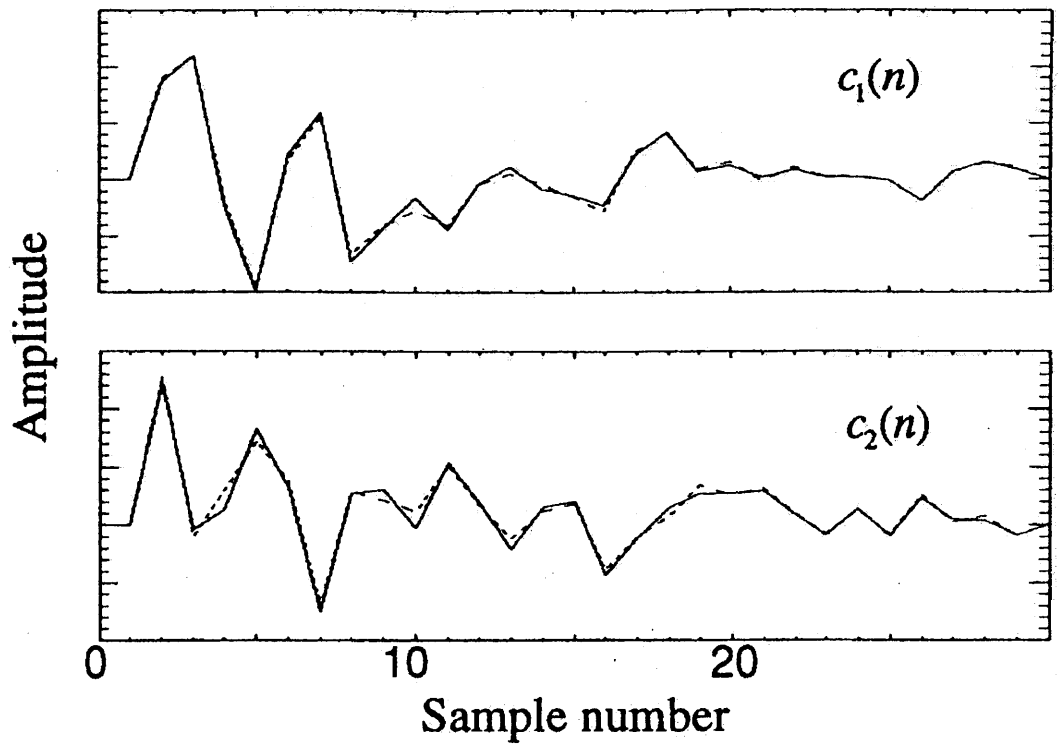


図 3.3: インパルス応答の推定値 (実線) と真値 (点線)

めしている。実線が推定値である $h_1(n, i)$, $h_2(n, i)$, 点線が元のインパルス応答 $c_1(n)$, $c_2(n)$ であり, よく推定できていることがわかる。ただし, フィルタ $h_1(n, i)$, $h_2(n, i)$ は 2 サンプル分の遅延を加えて表示してある。

図 3.4 に, フィルタ次数が 28 次の場合に音源信号としてインパルスを入力したときの残響除去処理前後の信号波形を示す。実線は本手法により処理された残響回復信号, 点線は処理前の残響信号である。図から, インパルスに続く残響成分が処理することにより除去されていることがわかる。

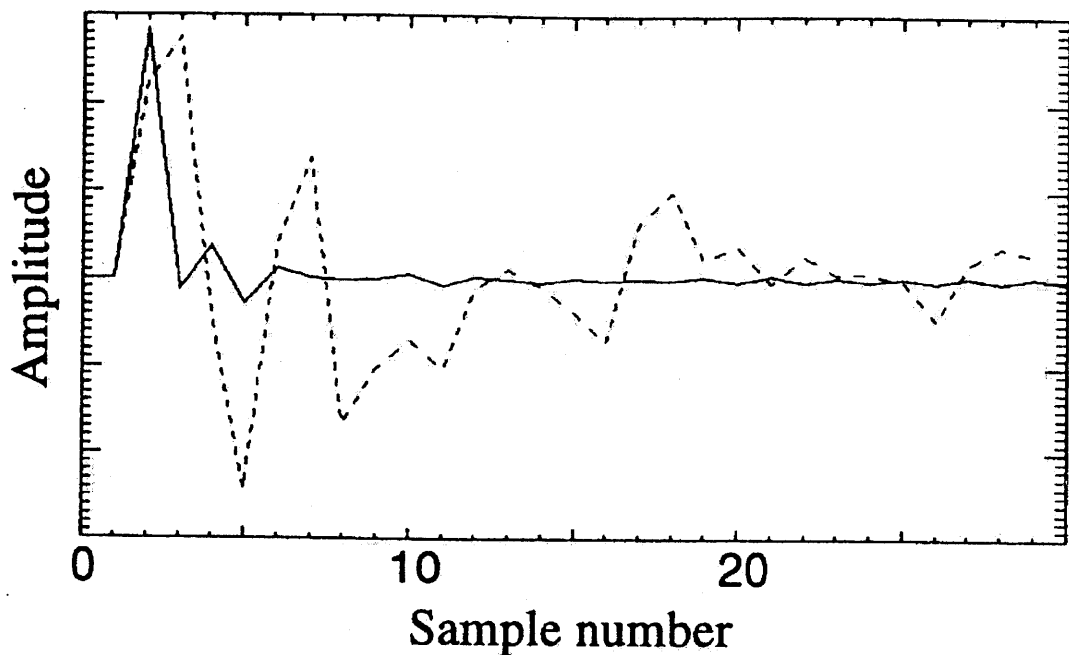


図 3.4: 残響抑圧された音源からのインパルス応答結果 処理後（実線）と処理前（点線）

3.3 帯域分割処理

一般に部屋の伝達関数の次数は数千次にもものぼる。たとえば、残響時間が約 200 msec の部屋のインパルス応答を 12kHz でサンプリングし、-60 dB の誤差でモデル化した場合、その次数は約 2400 次となる。この場合、二入力残響除去法においては、約 5000 次 の入力相関行列の最小固有ベクトル計算および式 (3.4) を解くための逆行列計算が必要となるが、これだけ大規模の行列計算は、計算精度および計算量の観点から実際におこなうことは難しい。

ここでは、帯域分割処理を用いて複数の帯域に信号を分割し、間引き処理をおこなうことにより、伝達関数の各帯域あたりの次数を減らした後、各帯域ごとに二入力残響除去法で残響を取り除き、再び再合成することを検討する。これにより、処理すべき行列の規模が小さくなり、最小固有ベクトル、逆行列の

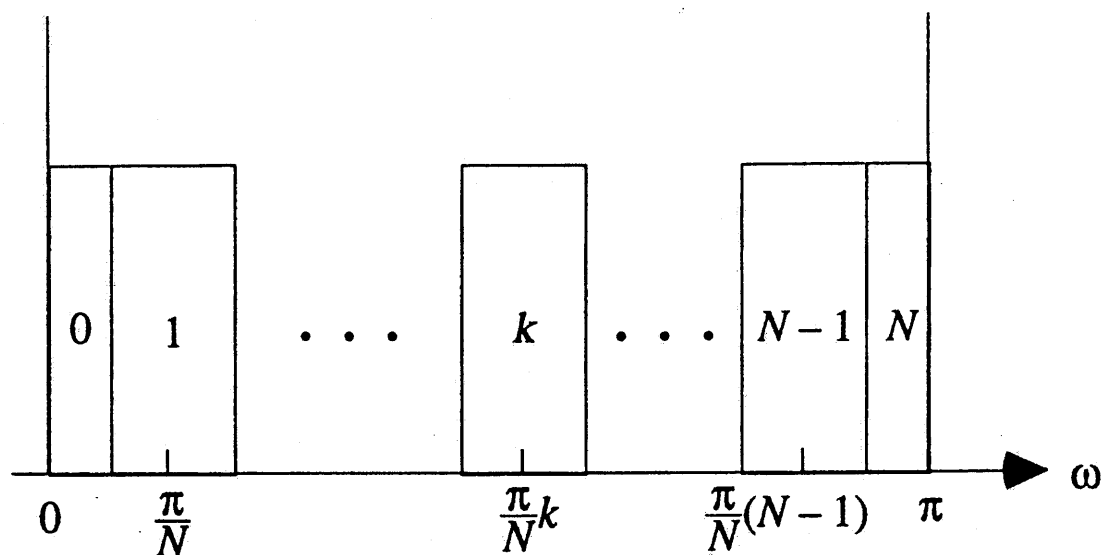


図 3.5: $N + 1$ 帯域分割

計算が容易となる. 図 3.5 に示すような $N + 1$ の帯域に分割を考える. 図 3.6 に帯域分割処理を用いた二入力残響除去処理のブロック図を示す. 入力信号はまず $N + 1$ の帯域に分割され, 各帯域ごとに間引き率 M で間引き処理がおこなわれる. 間引かれた信号に 2 節で述べた残響除去処理を加え残響回復信号を得る. 次に, 各帯域ごとに得られた回復信号を補間率 M で補間して, 全帯域を足し合わせ全帯域の残響回復信号を合成する.

3.3.1 帯域分割および間引き処理

帯域分割処理および間引きは以下の手順でおこなう.

- (i) 入力信号 $m_1(n), m_2(n)$ を $e^{-j\pi kn/N}$ で変調し, 中心周波数 $\pi k/N$ の k 番目の帯域をベースバンドに移動し, カットオフ周波数 $\pi/(2N)$ のローパスフィルタ $w(n)$ を通過させる. これにより k 番目の帯域信号のみ取り出せる. (図 3.7(b) 参照)

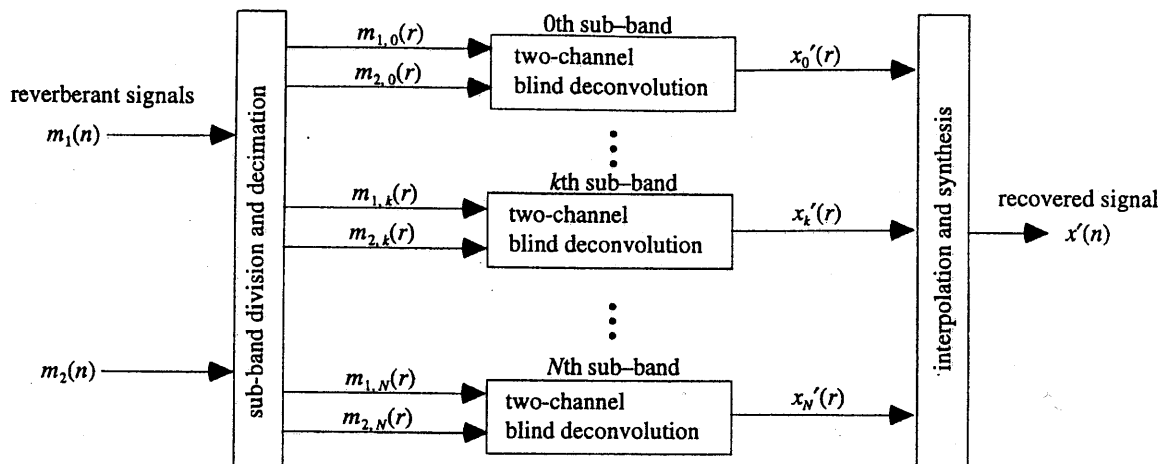


図 3.6: 帯域分割二入力残響抑圧処理ブロック図

(ii) この信号はSSB信号となっているので複素数であり、これを実数にするため $e^{j\pi n/2M}$ で再び変調して中心周波数を $\pi/(2M)$ に移動し、それと複素共役対称な成分を足し込む。(図 3.7(c) 参照)

(iii) 間引き率 M で間引く。(図 3.7(d) 参照)

以上のようにして得られた k 番目の帯域の信号は次式で表わせる。

$$\begin{aligned}
 m_{1,k}(r) &= \sum_{l=-\infty}^{\infty} w(rM-l)m_1(l)\{e^{-j\pi kl/N}e^{j\pi rM/2M} + e^{j\pi kl/N}e^{-j\pi rM/2M}\} \\
 &= 2 \sum_{l=-\infty}^{\infty} w(rM-l)m_1(l) \cos\left(\frac{\pi kl}{N} - \frac{\pi r}{2}\right) \\
 m_{2,k}(r) &= 2 \sum_{l=-\infty}^{\infty} w(rM-l)m_2(l) \cos\left(\frac{\pi kl}{N} - \frac{\pi r}{2}\right)
 \end{aligned} \tag{3.8}$$

3.3.2 補間および合成処理

各帯域ごとに残響除去処理された信号 $x'_k(r)$ は補間率 M で補間され、もとの帯域にアップサンプリングしてから合成される。以下にその手順を示す。

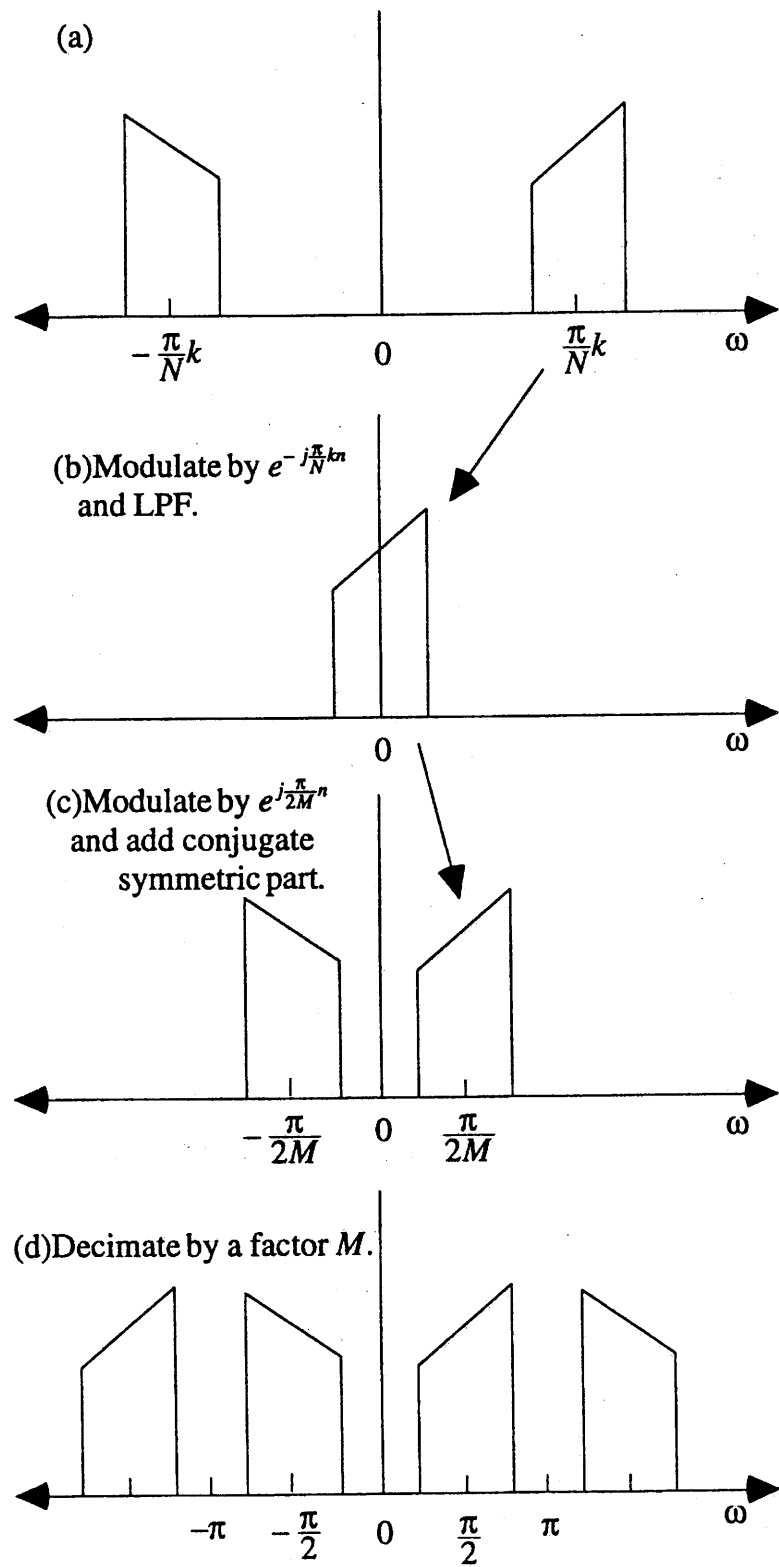


図 3.7: 間引き処理

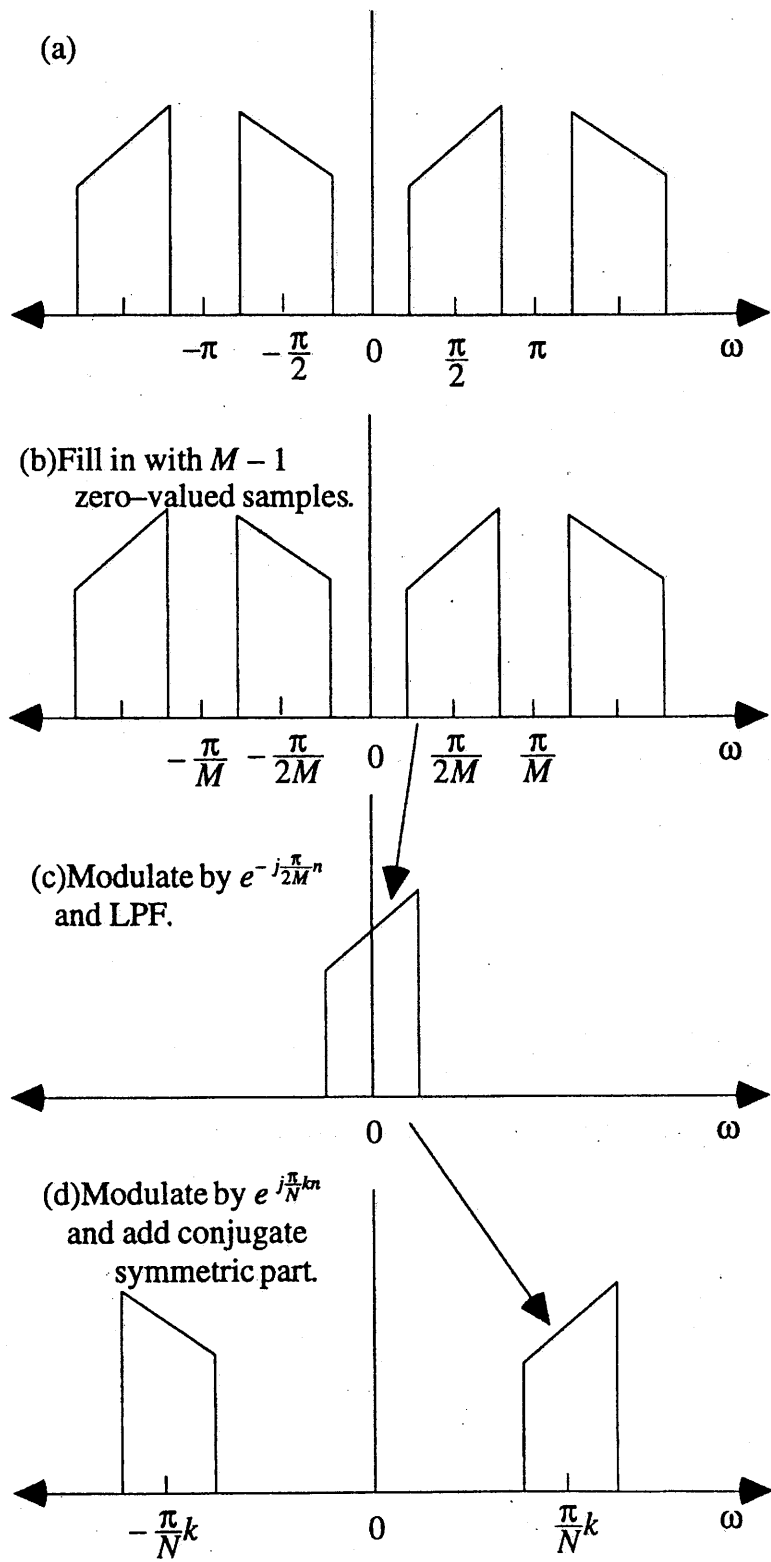


図 3.8: 補間処理

- (i) 各サンプルデータ間に $M - 1$ 個の 0 を詰め、アップサンプリングする。(図 3.8(b) 参照)
- (ii) $e^{-j\pi n/2M}$ で変調し、 k 番目の帯域の中心周波数を原点に移動し、カットオフ周波数 $\pi/(2N)$ のローパスフィルタを通過させる。これにより折り返し成分をカットする。(図 3.8(c) 参照)
- (iii) $e^{j\pi kn/N}$ で変調し、 k 番目の帯域の中心周波数を $\pi k/N$ に移動し、それに複素共役対称な成分を足し込む。(図 3.8(d) 参照)
- (iv) 各帯域信号の和をとり、信号を再合成する。

以上の処理により得られる合成された残響回復信号は次式であらわせる。

$$\begin{aligned}
 x'(n) &= \sum_{k=0}^N \sum_{l=-\infty}^{\infty} w(n-l) \alpha_k x_k''(l) \{ e^{-j\pi l/2M} e^{j\pi kn/N} + e^{j\pi l/2M} e^{-j\pi kn/N} \} \\
 &= 2 \sum_{k=0}^N \sum_{l=-\infty}^{\infty} w(n-l) \alpha_k x_k''(l) \cos\left(\frac{\pi l}{2M} - \frac{\pi kn}{N}\right)
 \end{aligned} \tag{3.9}$$

ここで、 α_k は各帯域の回復信号 $x'_k(r)$ に含まれる任意定数であり、 $x_k''(l)$ は

$$x_k''(l) = \begin{cases} x'_k\left(\frac{l}{M}\right) & l = 0, \pm M, \pm 2M, \dots \\ 0 & \text{otherwise} \end{cases}$$

である。

3.3.3 回復信号に含まれる任意定数の決定

式(3.7)で示されるように、二入力残響除去による回復信号には、任意定数が含まれている。各帯域に信号を式(3.9)で合成するときには任意定数を決定して合成しなければならない。ここでは、完全なブラインド処理とはならないが音

源信号 $x(n)$ を帯域分割したときの各帯域信号の平均パワーが既知と仮定して各帯域の任意定数を決定する。音源を音声だとすると第2.3節で述べたように、音声の平均パワースペクトルを利用して各帯域信号の平均パワーを得ることができる。

音源信号の k 番目の帯域信号を $x_k(r)$ としたときの平均パワーを、

$$\Lambda_{kk} = E\{x_k(r)x_k(r)\} \quad (3.10)$$

とし、同様に k 番目の帯域での回復信号の平均パワーを Λ'_{kk} とする。また、回復信号の p 番目と q 番目の帯域信号を $x'_p(r), x'_q(r)$ としたときの相互相関値を、

$$\Lambda'_{pq} = E\{x'_p(r)x'_q(r)\} \quad (3.11)$$

とする。このとき、音源信号と回復信号との平均パワーが各帯域毎に等しくなり、任意の帯域の位相を基準として符号が一致するように任意定数を決定すると、 p 番目の帯域の位相を基準にした q 番目の帯域の任意定数 α_k は、

$$\alpha_q = \text{sign}(\Lambda'_{pq}) \sqrt{\frac{\Lambda_{qq}}{\Lambda'_{qq}}} \quad (3.12)$$

とかける。

3.3.4 室内インパルス応答の残響抑圧実験

図3.9に示す実験系の残響除去シミュレーションをおこなった。部屋の残響時間は約0.5秒、サンプリング周波数は12kHz、音源とマイクロホンとの距離は1チャンネル目が3m、2チャンネル目が4mである。図3.10に測定された1チャンネル目の部屋のインパルス応答 $c_1(n)$ を示す。シミュレーションでは $c_1(n), c_2(n)$ の次数を3000次に制限した。入力信号 $m_1(n), m_2(n)$ は、計算機上で伝達関数と音源信号を畳み込んで計算された。帯域分割数 N は200、間引き率および補間

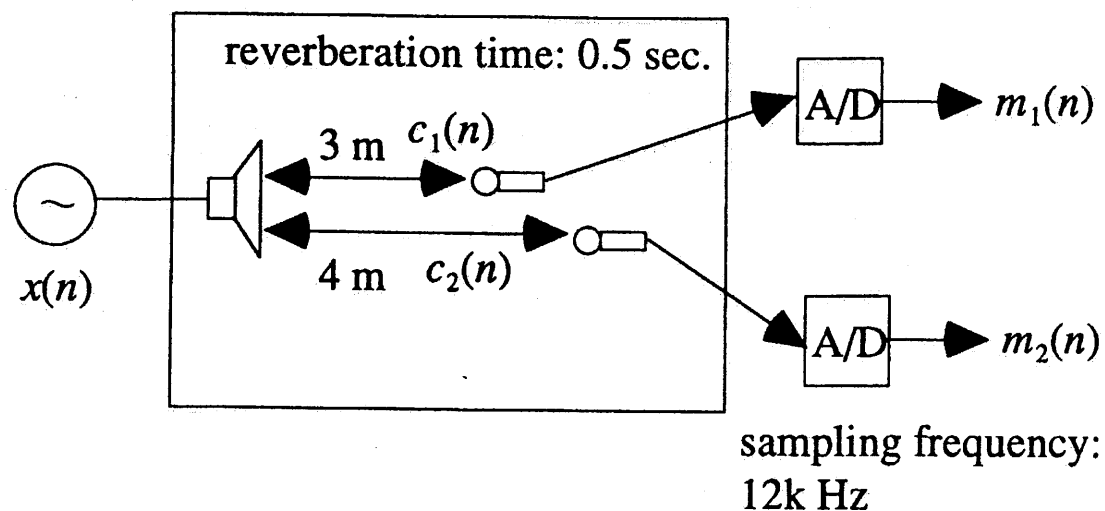


図 3.9: 実験系

率 M は 100 とした. 各帯域におけるフィルタ次数 i は, 間引くことによって 100 分の 1 になっているので 15 次から 30 次の範囲で最適値を探索した.

図 3.11 に音源信号にインパルス信号を入力した場合の残響除去処理結果を示す. 560msec 付近に直接音に対応するすどいピークがあり, 音源のインパルスが復元されていることがわかる. ピーク値に対する残響成分の相対パワーを計算すると, 図 3.10 のインパルス応答に比べ, 図 3.11 の残響成分のパワーが約 5dB 抑えられていることが分かった. しかし, 帯域分割処理が原因で, 560msec 付近に直接音のピークが生じるような大きな遅延が生じてしまっている. また, ピーク以前の時刻にも帯域分割処理により生じた歪のせいではないかと考えられる残響成分があらわれている.

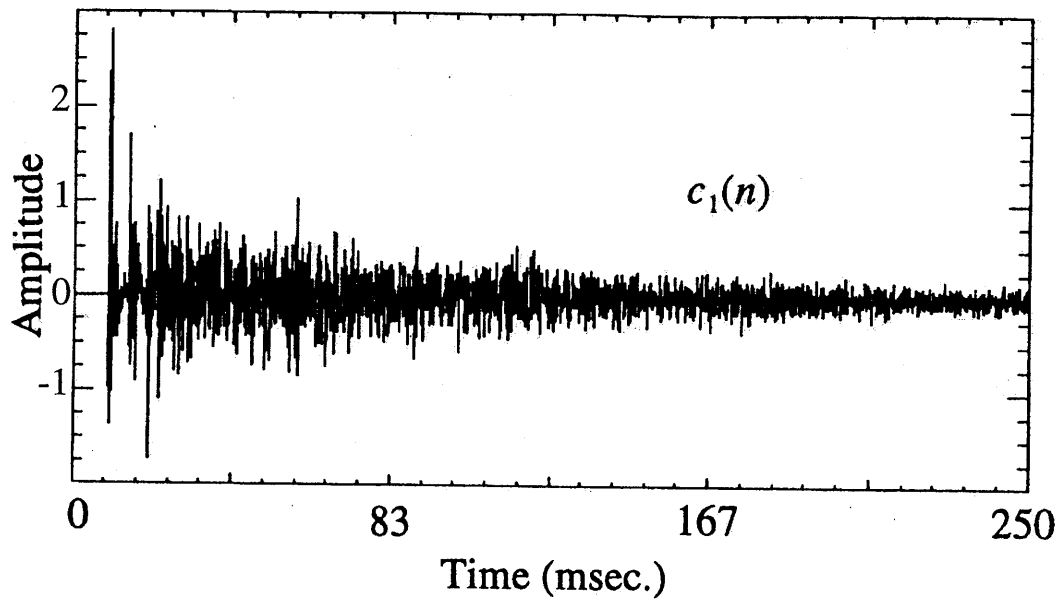


図 3.10: 残響のある部屋のインパルス応答

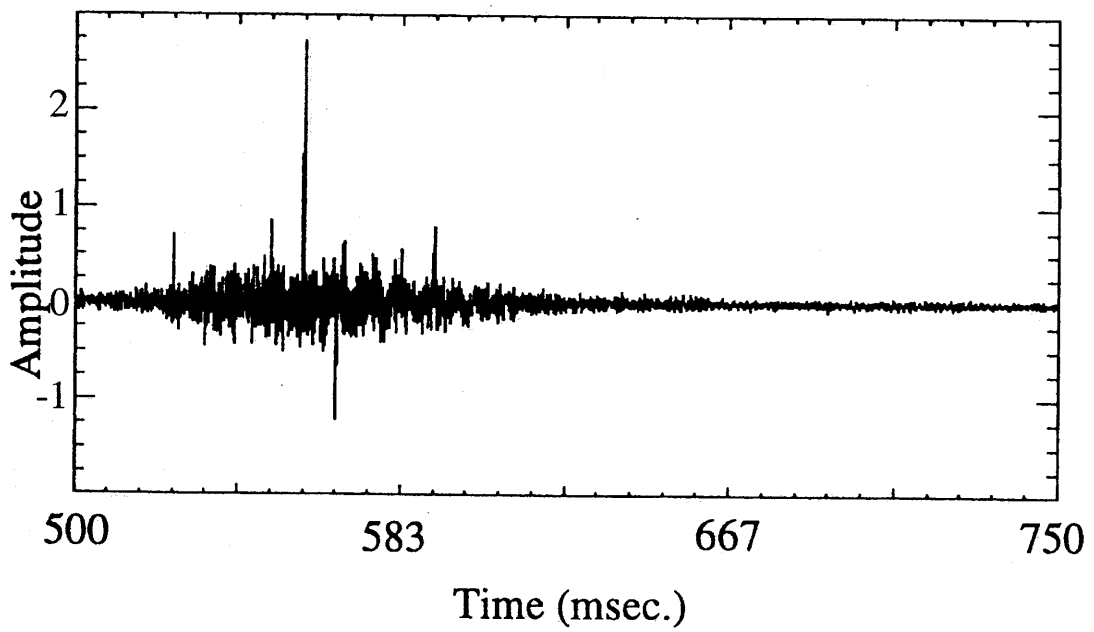


図 3.11: 残響抑圧されたインパルス応答

3.4 むすび

従来、MINT法では困難であった事前インパルス応答が測定できない応用において室内伝達関数が非最小位相である場合にも、フィルタ次数を決定でき残響除去がおこなえる入力相関行列を用いた二入力残響抑圧法を提案した。部屋の伝達関数の大きい次数に対応した大規模行列の計算を避けるため、帯域分割処理を用いることを検討した。信号を帯域分割し、間引き処理をおこなうことにより、伝達関数の各帯域あたりの次数を減らし計算すべき行列のサイズを小さくした。

シミュレーション実験により、低次の非最小位相伝達系の場合にはほぼ完璧に逆フィルタが計算でき残響回復を行い、本手法の原理的な確認ができた。室内インパルス応答を用いた実験ににおいては、伝達関数が数千次の場合においても本手法によって残響成分が抑えられる可能性を示した。

しかし、処理結果には帯域分割処理による遅延があり、また、ピーク以前の時刻にも帯域分割処理により生じた歪のせいではないかと考えられる残響成分があらわれた。この原因のひとつは、今回、各帯域の回復信号に含まれる任意定数をパワー成分のみで決定しているためであると考えられるが、正確に位相まで計算することは困難である。もうひとつの原因は、本手法においては最小固有ベクトルの計算をしており、帯域分割処理などによる僅かな計算誤差や雑音が大きく影響してくる。最小固有ベクトルは最もパワーの小さい固有ベクトルなので、僅かな雑音で埋もれてしまう可能性がある。