

Title	残響時間の自動推定による残響波形パワーエンベロープの回復に関する研究
Author(s)	古川, 正和
Citation	
Issue Date	2002-03
Type	Thesis or Dissertation
Text version	author
URL	<a href="http://hdl.handle.net/10119/1553">http://hdl.handle.net/10119/1553</a>
Rights	
Description	Supervisor:赤木 正人, 情報科学研究科, 修士

# 残響時間の自動推定による残響波形パワーエンベロープの回復に関する研究

古川 正和 (010103)

北陸先端科学技術大学院大学 情報科学研究科

2002年2月15日

キーワード: パワーエンベロープ, 残響時間, パワーエンベロープ逆フィルタ処理, 変調伝達関数 (MTF).

## 1 はじめに

実環境で観測される信号は、音源信号の他に雑音や残響が混在している。そのため実環境下で音声信号処理を行う場合、その精度が著しく低下する。特に残響は、音声信号処理の性能を低下させる主な原因となる。現在、残響信号から音源信号を回復する逆フィルタ処理が数多く検討されており、その有効性が示されている。しかし、室内伝達系は室内環境の変動を伴う時変系であるため、変動するごとに伝達系を測定しなければ回復精度が低下する。

広林らは、MTF (Modulation Transfer Function) の理論 [3] に基づき音源波形のパワーエンベロープ回復を目的としたパワーエンベロープ逆フィルタ処理を提案している [1]。これにより、インパルス応答の測定をすることなく残響の影響を軽減できたが、残響時間が既知であるという制約条件が存在する。

そこで本研究では、残響波形の情報のみから残響時間を自動推定してパワーエンベロープ逆フィルタ処理により残響波形のパワーエンベロープを回復する。パワーエンベロープ逆フィルタ処理の問題点を解決して、室内残響の影響を軽減する可能性を検討する。残響時間を自動推定することにより、室内伝達特性が時変系である空間や音源が移動している場合においても有効に残響を抑圧する手法が構築できる。

## 2 パワーエンベロープ逆フィルタ

残響波形のパワーエンベロープを回復するため、広林らが提案しているパワーエンベロープ逆フィルタ処理を用いる。以下にパワーエンベロープ逆フィルタ処理について説明する。

## 2.1 原理

音源信号を  $x(t)$ , 室内インパルス応答を  $h(t)$  とすると、残響信号 (観測信号)  $y(t)$  は  $x(t)$  と  $h(t)$  の畳込みで表すことができる。MTF の理論に基づき、音源信号と室内インパルス応答を以下のようにモデル化する。

$$x(t) = e_x(t)n_1(t) \quad (1)$$

$$h(t) = e_h(t)n_2(t) \quad (2)$$

$$e_h(t) = ae^{-\frac{6.9t}{T_R}} \quad (3)$$

ここで、 $e_x(t), e_h(t)$  はそれぞれの包絡線であり、 $a$  は定数、 $T_R$  [s] は残響時間 (Reverberation time) である。 $n_1(t), n_2(t)$  は互いに無相関な白色雑音である。

今、ある確率過程から生成される残響信号  $y(t)$  の 2 乗集合平均を考える。残響信号が原信号と室内インパルス応答の畳込みで表すことができ、各々の搬送波は無相関であることから

$$\begin{aligned} \langle y(t)^2 \rangle &= \left\langle \left\{ \int_{-\infty}^{\infty} e_x(\tau)^2 e_h(t-\tau)^2 d\tau \right\}^2 \right\rangle \\ &= e_x(t)^2 * e_h(t)^2 \end{aligned} \quad (4)$$

$$= e_y(t)^2 \quad (5)$$

と表すことができる。これより、音声のパワーエンベロープ特性  $P_x(z)$  は室内インパルス応答パワーエンベロープの逆特性を用いることで

$$P_x(z) = \frac{P_y(z)}{P_h(z)} = \frac{1}{a^2} \left\{ 1 - \exp\left(-\frac{13.8T_s}{T_R}\right) z^{-1} \right\} P_y(z) \quad (6)$$

のように残響音声のパワーエンベロープ特性  $P_y(z)$  とパワーエンベロープ伝達関数  $P_h(z)$  の逆特性の積で表される。 $P_x(z), P_y(z)$  はそれぞれ  $e_x(t)^2, e_y(t)^2$  の  $z$  変換である。 $e_x(t)^2$  と  $e_y(t)^2$  の正の平方根 (すなわち振幅包絡) を用いて、復元音声  $\hat{x}(t)$  は

$$\hat{x}(t) = \hat{e}_x(t) \frac{y(t)}{e_y(t)} \quad (7)$$

で表される。但し、右辺の  $\hat{e}_x(t)$  は式 (6) の逆  $z$  変換で求められた回復エンベロープである。

これにより、残響波形のパワーエンベロープを回復することができる。

## 2.2 問題点

パワーエンベロープ逆フィルタにより残響波形のパワーエンベロープを回復するとき、(1) 回復処理に用いる残響時間の決定に関する問題、(2) MTF の理論を音声に適用するときの原理的な問題がある。そこで本研究では、これらの問題を解決して残響波形パワーエンベロープの回復をする。

### 3 残響時間の推定

パワーエンベロープ逆フィルタ処理は、残響時間が既知でなければならない。そこで、残響時間を残響波形の情報のみから推定する手法を提案する。

残響時間の推定を行うにあたり、変調度に関して考察をする。MTF の理論において、残響信号の変調度は原信号の変調度に対する相対量である。そのため、原信号の変調度が既知でなければならない。また、音声において無音区間ではパワーを持たないことから、原信号の変調度が 1 であると仮定する。

#### 3.1 残響時間の推定法

残響時間の推定法を提案する。残響は、時間的に正方向にパワーエンベロープのピークの山をならし、ならされたものが谷を埋める。パワーエンベロープ逆フィルタは、このならされたものを削り取り時間的に逆方向に谷を強調して山を盛り上げる。もし、推定した残響時間  $\hat{T}_R$  が付加した残響時間  $T_R$  よりも大きいとき、谷が必要以上に強調されることでパワーエンベロープが負の値になることがある。しかし、信号の 2 乗包絡線であるパワーエンベロープにおいて、非負であることは明らかである。本研究では原信号の変調度を 1 と仮定しているため、パワーエンベロープが負の値になる直前の残響時間を求めることで、残響時間の推定ができると考えられる。

#### 3.2 検証結果

提案した残響時間推定法の検証を行う。原信号のパワーエンベロープには以下に示すものを用いる。

1. 最も単純なパワーエンベロープを式 (8) に示す。

$$e_x(t)^2 = 1 + \sin 2\pi Ft \quad (8)$$

ここで  $F$  は変調周波数であり、 $F = 10$  [Hz] とする。以後、このパワーエンベロープを正弦波のパワーエンベロープと呼ぶ。

2. 式 (9) に 1 ~ 20 [Hz] までの周波数成分を含んだパワーエンベロープを示す。

$$e_x(t)^2 = \sum_{k=1}^{20} \sin(2\pi kt + \hat{\theta}_k) \quad (9)$$

但し、 $\hat{\theta}_k$  はランダムな値である。以後、このパワーエンベロープを複合音のパワーエンベロープと呼ぶ。

3. ランダム信号からカットオフ周波数 20 [Hz] の LPF でフィルタリングして求めたパワーエンベロープを用いる。以後、このパワーエンベロープをランダム信号からのパワーエンベロープと呼ぶ。

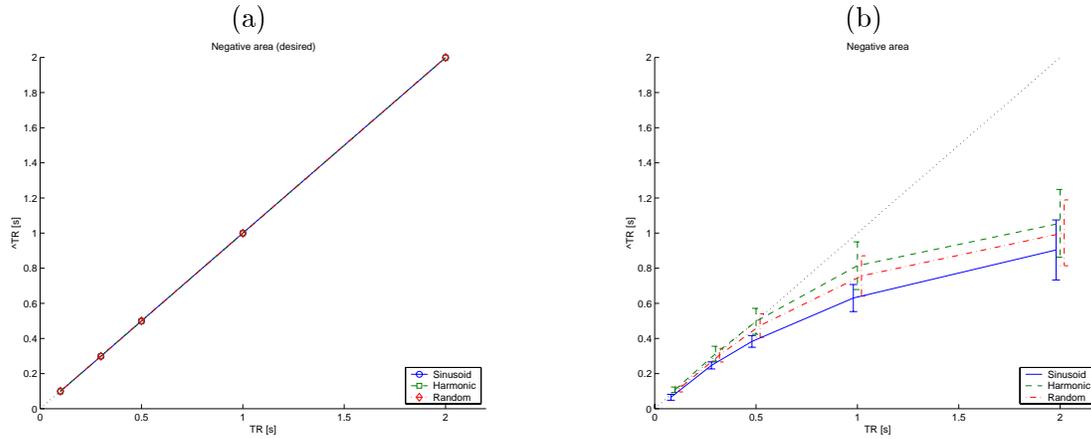


図 1: 負の面積による残響時間の推定.(a) 残響信号パワーエンベロープの理論値から推定した場合,(b) 抽出した残響信号パワーエンベロープから推定した場合

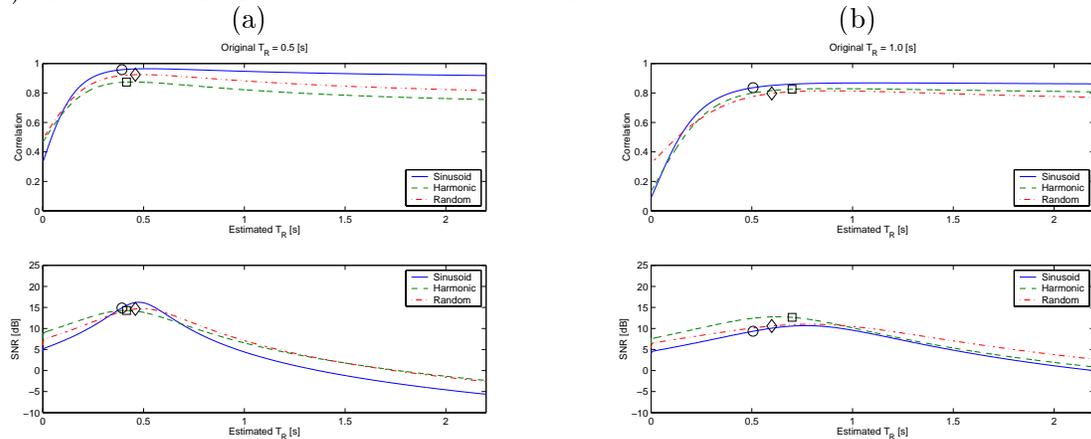


図 2: 推定残響時間と相関,SNR の関係.(a) $T_R = 0.5$  [s], (b) $T_R = 1.0$  [s].(上) 相関値,(下)SNR

搬送波は白色雑音とする。付加する残響の残響時間は  $T_R = 0.1, 0.3, 0.5, 1.0, 2.0$  [s] の 5 種類を用いた。残響を付加した信号からパワーエンベロープを抽出して残響時間の推定シミュレーションを行う。残響信号のパワーエンベロープは、ヒルベルト変換後の絶対値信号を低域フィルタリングすることにより求めた。推定した残響時間の結果を図 1 に示す。(a) は残響信号パワーエンベロープの理論値を用いて推定した残響時間、(b) は残響信号からヒルベルト変換を用いて抽出したパワーエンベロープを用いて推定した残響時間である。横軸が付加した残響の残響時間、縦軸が推定した残響時間である。図 2 は、抽出した残響信号のパワーエンベロープを残響時間を変数として回復処理をしたパワーエンベロープ  $\hat{e}_x(t)^2$  と原信号のパワーエンベロープ  $e_x(t)^2$  の相関値と SNR を示す。(a),(b) は、それぞれ付加した残響の残響時間  $T_R = 0.5, 1.0$  [s] の場合であり、マークしている部分は推定した残響時間である。図 1,2 において、実線、破線、点破線は、原信号のパワーエンベロープがそれぞれ正弦波のパワーエンベロープ、複合音のパワーエンベロープ、ランダム信号からのパワーエンベロープの場合を示している。

図1より、残響信号パワーエンベロープの理論値を用いた場合は正確に残響時間を推定できている。抽出したパワーエンベロープでは、付加した残響の残響時間が長いとき推定誤差が大きくなる。しかし、図2より推定した残響時間を用いてパワーエンベロープを回復するとパワーエンベロープ回復度が向上することがわかる。従って、提案した手法は有効であると言える。

## 4 音声への適用

パワーエンベロープ逆フィルタ処理により、残響音声のパワーエンベロープを回復する。音声のパワーエンベロープは帯域ごとに異なる。しかし、パワーエンベロープ逆フィルタ処理は、全帯域のパワーエンベロープを等しくする。このため、音声のパワーエンベロープを回復するときは帯域分割をする。分割する帯域幅は、残響音声のパワーエンベロープが共変調している周波数領域から決定すれば良いと考えられる。本研究では、帯域幅が400 [Hz]の定帯域フィルタバンクを用いる。

### 4.1 検証結果

原音声にはATR音声データベース mau の“相変わらず”を用いた。残響音声は原音声に残響時間  $T_R = 0.5$  [s]としたインパルス応答を畳み込むことで作成した。パワーエンベロープの回復を評価するため、パワーエンベロープ回復指標 (Improvement index of power envelope distortion: Ip) を式 (10) のように定義して定量的に評価する。

$$I_p = 10 \log_{10} \frac{\int_0^T \{e_x(t)^2 - e_y(t)^2\}^2 dt}{\int_0^T \{e_x(t)^2 - \hat{e}_x(t)^2\}^2 dt} \quad [\text{dB}] \quad (10)$$

ここで、 $e_x(t)^2, e_y(t)^2, \hat{e}_x(t)^2$  はそれぞれ原音声のパワーエンベロープ、残響音声のパワーエンベロープ、回復したパワーエンベロープである。 $T$  は分析時間であり、約 1.6 [s] である。パワーエンベロープが回復すると  $I_p$  は正の値を示す。フィルタバンクの各チャンネルごとに  $I_p$  を求めた。結果を図3に示す。(a) は全帯域での残響時間が 0.5 [s] として回復した場合、(b) は各帯域ごとに残響時間を推定した場合である。

全帯域で残響時間が一定とした (a) では高域を除いてパワーエンベロープが回復されていない。これに対して各帯域ごとに残響時間を推定した (b) では全ての帯域でパワーエンベロープが回復することがわかる。このため、本手法は有効であるといえる。

## 5 まとめ

本研究では、残響波形の情報のみから残響時間を推定して残響波形パワーエンベロープの回復をした。パワーエンベロープ逆フィルタ処理で問題となっている (1) 残響時間の決

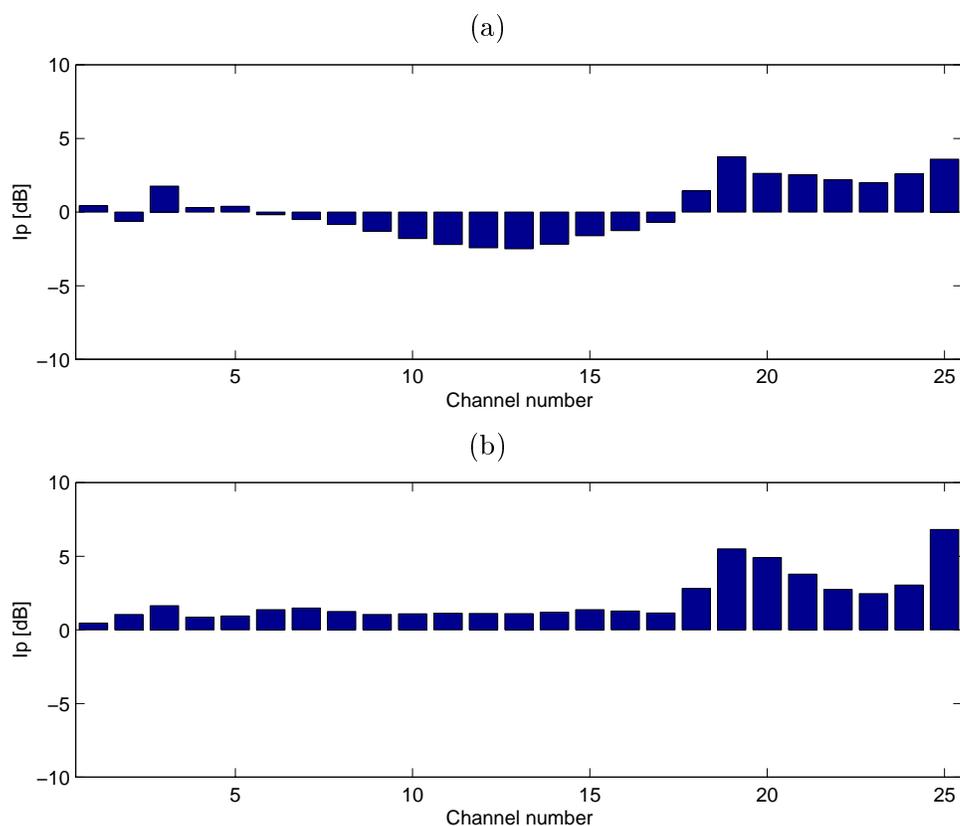


図 3: 各チャンネルにおけるパワーエンベロープ回復指標.(a) 残響時間を一定とした場合,(b) 残響時間を推定した場合

定,(2) 音声への適用について検討を行った。残響時間の推定では、パワーエンベロープが非負であるようにすることで残響時間の推定ができることを示した。音声のパワーエンベロープは、パワーエンベロープが共変調していると思なせる帯域ごとに分割し、分割した帯域ごとに残響時間を推定することで、各帯域におけるパワーエンベロープが回復できることを示した。

## 参考文献

- [1] 広林茂樹, 野村博昭, 小池恒彦, 東山三樹夫, “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, No10, pp.1323-1330,1998.
- [2] 広林茂樹, 山淵龍夫, “帯域分割を用いたパワーエンベロープ逆フィルタ処理の残響抑圧効果” 信学論 A, Vol. J83-A, No8, pp.1029-1033,2000.
- [3] T,Houtgast, H.J.M. Steeneken, and R.Plomp “Predicting speech intelligibility in room acoustics”, Acustica, Vol.46, pp.60-72, 1980.