

Title	瞬時振幅を利用したブラインド残響音声回復の可能性の検討
Author(s)	柴野, 洋平
Citation	
Issue Date	2007-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/3624
Rights	
Description	Supervisor: 鷗木 祐史, 情報科学研究科, 修士

修 士 論 文

瞬時振幅を利用した
ブラインド残響音声回復の可能性の検討

北陸先端科学技術大学院大学
情報科学研究科情報処理学専攻

柴野 洋平

2007 年 3 月

修 士 論 文

瞬時振幅を利用した
ブラインド残響音声回復の可能性の検討

指導教官 鵜木祐史 准教授

審査委員主査 鵜木祐史 准教授

審査委員 赤木正人 教授

審査委員 徳田功 准教授

北陸先端科学技術大学院大学
情報科学研究科情報処理学専攻

510045 柴野 洋平

提出年月: 2007 年 2 月

概 要

実環境の雑音や残響は観測される音声信号に歪みを与える．特に残響環境下では，壁からの反射音が直接音に重なり歪みを与え，音声の明瞭度を著しく低下させる．そのため音声認識システムや，拡声音声通話，補聴システムなどでは性能が低下する．これらの影響を取り除き，残響音声を回復させることが求められる．残響の影響は室内の構造や室温の変化とともに変動する可能性があるため，逐次，室内インパルス応答を測定することなく，適応的に残響音声を回復する必要がある．

先行研究では残響音声を帯域分割し，帯域内で伝達系の測定を必要としないMTFに基づくパワーエンベロープ逆フィルタ処理を用いて音声のパワーエンベロープを回復し，残響環境下から推定された基本周波数 (F_0) と有声無声区間の情報と群遅延を用いることで，自然性の高い音声のキャリアを生成している．まず，古川らは広林らの提案したパワーエンベロープ逆フィルタ法の問題点である室内インパルス応答の振幅項，残響時間の決定方法を示し，残響音声信号を回復させ，音声波形を再合成する手法にモデルを拡張した．パワーエンベロープのみの回復では，明瞭度は回復しないので，酒田らは F_0 と有声無声区間の情報を用いてキャリアを再生成した．また，最適な帯域幅についての検討も行った．戸井らは群遅延を用い自然なキャリアを再生成し，最適な時間-周波数分割処理とカットオフ周波数について検討した．この手法は，残響環境からブラインドで音声を回復できる．しかし，残響環境下では F_0 と有声無声区間の情報の推定精度の低下が懸念されるため，これらの手法では十分な回復効果を得られない可能性がある．

本研究では，戸井らによって提案された伝達系の測定を必要としない残響音声回復法を改良し，残響環境下から F_0 と有声無声区間の情報を推定することなく，ブラインドで音声の明瞭度を回復するために，精度よく推定できる可能性のある瞬時振幅から信号を復元する方法を検討した．帯域分割した元の信号の瞬時振幅を既知としたときに，Hilbert 変換を用いて瞬時位相を推定し，帯域内の信号を復元し，音声を復元した．100 単語に対するシミュレーションの結果，復元音声の対数スペクトル歪み (LSD) は平均 0.1 dB，信号対雑音比 (SNR) は平均 24.0 dB であり，音声を復元させることができた．本報告は，ブラインドで残響音声回復の可能性を検討する．

目 次

第 1 章	序論	1
1.1	研究の背景	1
1.2	目的	2
1.3	本論文の構成	2
第 2 章	MTF に基づく残響音声回復モデル	3
2.1	概要	3
2.2	課題	5
第 3 章	改良モデル	7
3.1	概要	7
3.2	問題設定	7
3.3	瞬時振幅からの瞬時位相の推定と，信号の復元	8
3.3.1	Hilbert 変換の関係	8
3.3.2	最小位相信号	9
3.3.3	瞬時位相の推定	9
3.3.4	周波数領域で片側が 0 となる信号の扱い	10
3.3.5	帯域分割した信号への適用	11
3.4	評価	12
3.4.1	実験条件	12
3.4.2	結果	12
第 4 章	残響音声の回復の可能性の検討	15
4.1	瞬時振幅の回復の可能性	15
4.2	残響音声の回復の可能性	16
第 5 章	結論	22
5.1	本論文で明らかにしたこと	22
5.2	課題	22
付 録 A		28

目 次

2.1	MTF に基づく残響音声回復法の概要	4
2.2	元の音声から抽出した F0	5
2.3	残響音声から抽出した F0	6
3.1	改良モデルの概念	8
3.2	周波数領域において振幅から位相を求められる例	10
3.3	時間領域において振幅から位相を求められる例	11
3.4	信号を正の部分と負の部分に分割する例	12
3.5	チャンネル内の代表的な波形	13
3.6	原音声と復元音声	14
4.1	瞬時振幅の回復	16
4.2	原信号とその正の部分 $x_p(t)$, 負の部分 $x_n(t)$	18
4.3	残響信号とその正の部分 $y_p(t)$, 負の部分 $y_n(t)$	18
4.4	原信号の正の部分のエンベロープ $E_{xp}(t)$, 負の部分のエンベロープ $E_{xn}(t)$	19
4.5	残響信号の正の部分のエンベロープ $E_{yp}(t)$, 負の部分のエンベロープ $E_{yn}(t)$	19
4.6	回復した正の部分のエンベロープ $E_{\hat{x}p}(t)$, 負の部分のエンベロープ $E_{\hat{x}n}(t)$	20
4.7	原信号の正の部分の瞬時振幅 $s_{xp}(t)$, 負の部分の瞬時振幅 $s_{\hat{x}n}(t)$,	20
4.8	残響信号の正の部分の瞬時振幅 $s_{yp}(t)$, 負の部分の瞬時振幅 $s_{yn}(t)$	21
4.9	回復した正の部分の瞬時振幅 $s_{\hat{x}p}(t)$, 負の部分の瞬時振幅 $s_{\hat{x}n}(t)$	21

第1章 序論

1.1 研究の背景

本研究の目的は、室内伝達系を測定することなく、ブラインド的に残響音声を回復するための信号合成法を提案し、残響音声回復の可能性を示すことである。

実環境の雑音や残響は観測される音声信号に歪みを与える。特に残響環境下では、壁からの反射音が直接音に重なり歪みを与え、音声の明瞭度を著しく低下させる。そのため音声認識システムや、拡声音声通話、補聴システムなどでは性能が低下する。これらの影響を取り除き、残響音声を回復させることが求められる。

このため、Neely と Allen は、単一マイクロホンで受音された信号から室内伝達特性の最小位相成分のみを取り除いて信号を回復させる逆フィルタ法を提案している [1]。しかしこの手法は、室内音場が最小位相特性を持つときのみしか有効でない。これに対し、三好と金田は、音源の数に対しマイクロホンを複数個配置することで、音源とマイクロホンの間の伝達特性の零点が重複しない場合、系が最小位相特性を有していなくても音源波形そのものを正確に復元できる音場逆フィルタ理論を提唱している [2]。また、王と板倉は、複数のマイクロホンによる処理に加え、観測信号を帯域分割する処理を行っている。それぞれの帯域ごとに音源信号と回復信号の誤差が最小になる最適なインパルス応答の逆フィルタを求め、音源波形を回復する帯域分割逆フィルタ理論を提唱している [3]。音源と複数のマイクロホン間におけるそれぞれの伝達系に対して、最適な逆フィルタを求め、音源信号を復元している。これらの伝達特性の逆特性を逆フィルタ処理で実現する方法では、逆フィルタを決定する上であらかじめ残響のインパルス応答を測定しておかなければならない。しかし、室内のインパルス応答は室内の構造や室温の変化によって変動するため、その都度伝達系を測定し直さなければいけないという問題がある。逐次室内インパルス応答を測定することなく、適応的に残響音声を回復する必要がある。

一方、室内のインパルス応答の測定を必要としない方法として、変調伝達関数 (MTF) の概念 [6, 7] に基づいて、残響音声のエンベロープを回復させる手法を Mourjopoulos と Hammond [8] や、広林ら [9, 10, 11] によって提案されている。

鵜木らを中心とした先の研究では、有効性が確認されている広林らの手法をベースに改良を行った。まず、古川らは広林らの提案したパワーエンベロープ逆フィルタ法の問題点である室内インパルス応答の振幅項、残響時間の決定方法を示し、残響音声信号を回復させ、音声波形を再合成する手法にモデルを拡張した [15, 17]。パワーエンベロープのみの回復では、明瞭度は回復しないので、酒田らは F_0 と有声無声区間の情報を用いてキャリ

アを再生成した．また，最適な帯域幅についての検討も行った．戸井らは群遅延を用い自然なキャリアを再生成し，最適な時間-周波数分割処理とカットオフ周波数について検討した [12, 13]．これによってエンベロープだけではなく， F_0 を用いることでキャリアを生成し，ブラインドで残響音声を回復できることが示された．また，中谷らは，インパルス応答の測定を必要としない，音声の調波構造に基づくブラインド残響除去法を提案している [4, 5]．これは継続音を抑制した音声から推定した F_0 と有声無声区間の情報を用いて室内インパルス応答を推定することで，伝達系を測定することなく残響音声を回復し，さらに回復音声から F_0 を精度よく推定する同様の処理を多段階で行うことで音声を回復できる手法である．

しかし，残響環境下からの F_0 の推定は非常に難しいとされ， F_0 の推定精度の低下が懸念される．これらの F_0 を用いる方法では，残響音声を十分に回復できない可能性がある．

1.2 目的

本研究も目的は，時変的な性質を持つ室内伝達系のインパルス応答を測定することなく，ブラインド的に残響音声を回復するための手法を改良し，残響音声回復の可能性を示すことである．本研究では，戸井らによって提案された伝達系の測定を必要としない残響音声回復法を改良し，残響環境下から F_0 と有声無声区間の情報を推定することなく，ブラインドで音声を回復するために，精度よく推定できる可能性のある瞬時振幅から信号を復元する方法を検討する．

これにより，音声の回復精度が向上し，音声認識システムや，拡声音声通話，補聴システムにおける性能の向上が期待される．

1.3 本論文の構成

本論文は全 6 章で構成されている．以下にその概要を述べる．第 2 章では，先行研究で提案された MTF に基づく残響音声回復モデルとその問題点について述べる．第 3 章では，本論文で解決する問題を設定する．瞬時振幅から瞬時位相を求めるための条件と方法を示す．第 4 章では，残響音声回復の可能性について検討する．第 5 章では，本研究のまとめと今後の課題について述べる．

第2章 MTFに基づく残響音声回復モデル

2.1 概要

先行研究で提案された MTF に基づく残響音声回復法は，パワーエンベロープ回復部とキャリア再生成部で構成される．N チャンネルのフィルタバンクを用いて残響音声信号を帯域分割し，エンベロープとキャリアそれぞれに対して回復処理，再生成処理を行う（図 2.1）．

パワーエンベロープ回復部では，室内インパルス応答の測定を必要としないパワーエンベロープ逆フィルタ処理を用いてエンベロープを回復している．原信号 $x(t)$ ，室内インパルス応答 $h(t)$ ，残響信号を $y(t)$ をそれぞれ以下のように表す．

$$y(t) = x(t) * h(t), \quad (2.1)$$

$$x(t) = e_x(t)\mathbf{n}_1(t), \quad (2.2)$$

$$h(t) = e_h(t)\mathbf{n}_2(t) = a \exp(-6.9t/T_R)\mathbf{n}_2(t), \quad (2.3)$$

$e_x(t)$ と $e_h(t)$ は $x(t)$ と $h(t)$ のエンベロープ， $\mathbf{n}_1(t)$ と $\mathbf{n}_2(t)$ はお互いに独立なホワイトノイズである． a は振幅， T_R は残響時間を表す．

ここで，残響信号のパワーエンベロープ $e_y(t)^2$ は下式のように表せる．

$$e_y(t)^2 = e_x(t)^2 * e_h(t)^2, \quad (2.4)$$

すなわち， $e_y(t)^2$ に室内インパルス応答のパワーエンベロープ $e_h(t)^2$ の逆特性を畳み込むことで，元の信号のパワーエンベロープ $e_x(t)^2$ を回復することができる． $e_x(t)^2$ ， $e_h(t)^2$ ， $e_y(t)^2$ の z 変換をそれぞれ $E_x(z)$ ， $E_h(z)$ ， $E_y(z)$ とおくと， $E_x(z)$ は

$$E_x(z) = \frac{E_y(z)}{a^2} \left\{ 1 - \exp\left(-\frac{13.8}{T_R \cdot f_s}\right) z^{-1} \right\}, \quad (2.5)$$

となる．このパワーエンベロープ逆フィルタ法をフィルタバンクで分割された各チャンネル信号に適用する． $y_n(t)$ ， $s_{y,n}(t)$ ，と $e_{y,n}(t)$ はそれぞれ 帯域分割した観測信号，その瞬時振幅，エンベロープである．この方法により，回復エンベロープ $\hat{e}_{x,n}(t)$ は $y_n(t)$ から求めることができる．

古川らは残響音声から室内インパルス応答の振幅項 a と残響時間 T_R を求める方法を示している．これにより，ブラインドでパワーエンベロープ逆フィルタ処理を行うことが

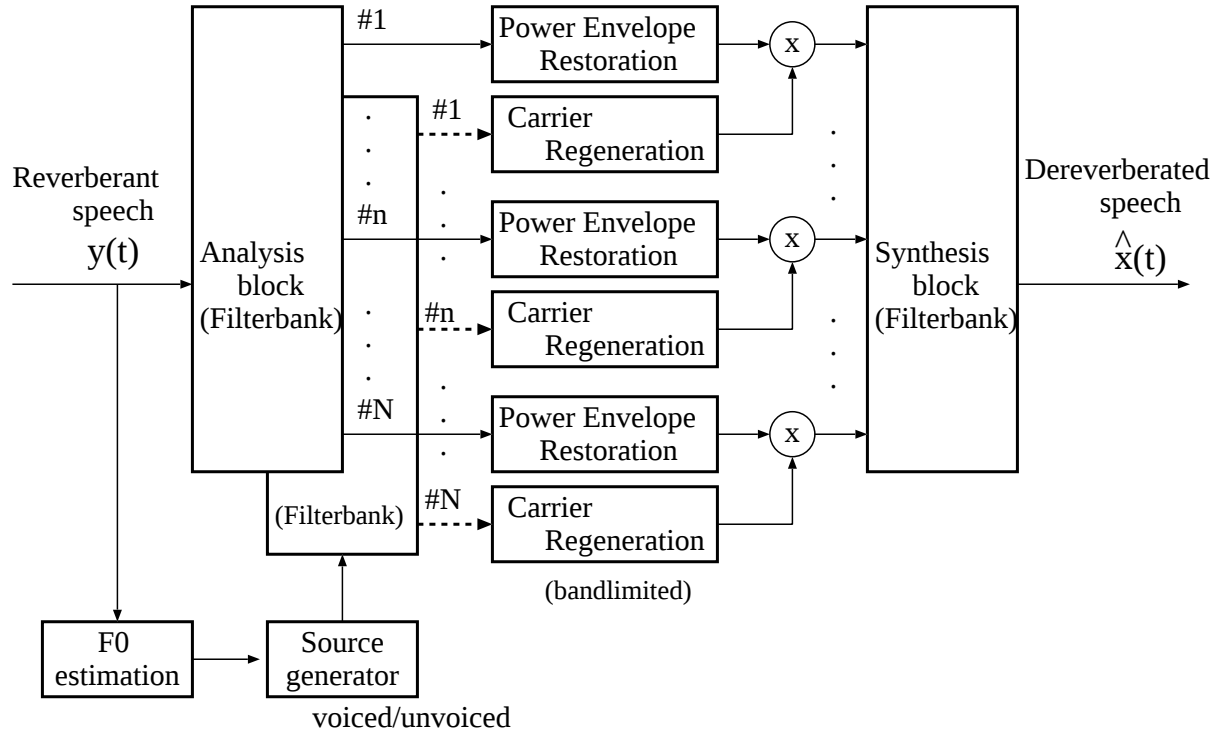


図 2.1: MTF に基づく残響音声回復法の概要

可能になった [15] . しかし , パワーエンベロープのみの回復では音声の明瞭度は回復しなかった .

そこで , 酒田らは最適な帯域幅によるパワーエンベロープの回復に加えて , F_0 , 有聲無声区間の情報を基にキャリアを生成した [16] . キャリア再生成部では , 有聲区間と , 無声区間それぞれで生成処理を行う .

有聲区間では , PIFM 法に基づいたキャリア生成処理を行う [19] . 周期的な情報を持つ有聲音は , F_0 と同値の周波数である基本波と F_0 の倍数からなる複数の調波を足し合わせた波形構造である . 基本波は時間的に緩やかに変化し , 調波もそれに従って変化する . そこで , 残響音声中から推定された F_0 から基本波およびその倍音である調波を作成し , それらの波形を足し合わせてできた調波複号音をキャリアとする . 生成されたキャリア $\hat{c}_v(t)$ は下式のように表される .

$$\hat{c}_v(t) = \frac{1}{K(t)} \sum_{k=1}^{K(t)} \sin \left(k \int_0^t 2\pi F_0(\tau) d\tau + \phi_k(t) \right), \quad (2.6)$$

ここで , $F_0(t)$ は基本周波数 , $\phi_k(t)$ は初期位相 , k は調波の次数 ($1 \leq k \leq K(t)$) , $K(t)$ は調波の最大次数である .

次に , 無声区間のキャリア再生成では , 白色雑音を $\hat{c}_u(t)$ キャリアとする . これより , 音声信号のキャリア $\hat{c}(t)$ は , $\hat{c}(t) = \hat{c}_v(t) + \hat{c}_u(t)$ として再生成される . キャリアはフィルタバンクによって各チャンネルに分解され , 回復されたパワーエンベロープと掛け合わせら

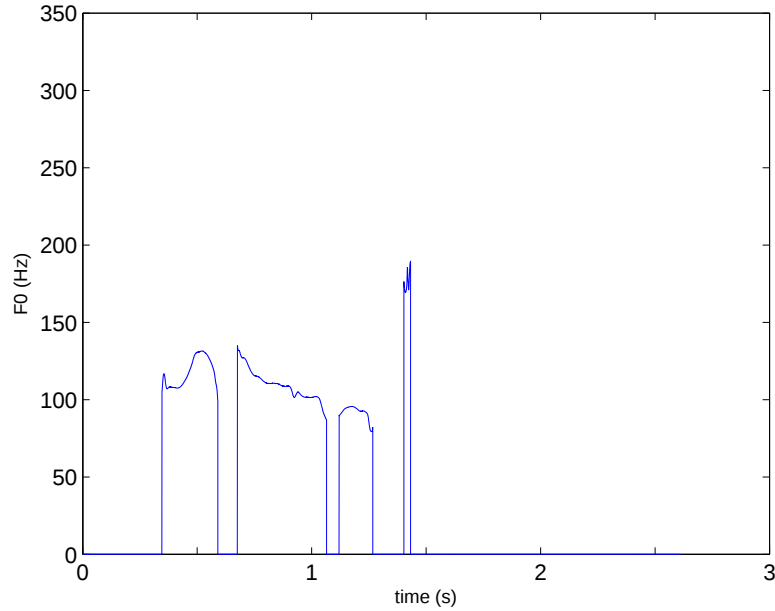


図 2.2: 元の音声から抽出した F0

れ，合成フィルタバンクで回復音声を得られる．

さらに，戸井らはパワーエンベロープ回復処理において最適な時間-周波数分割とカットオフ周波数を設定し，キャリア生成時には STRAIGHT 合成系 [20] で利用された群遅延操作を行うオールパスフィルタを用いた [12, 13]．その結果，残響時間が長い音声に対しても効果的な回復が得られ，残響音声に比べて残響感が解消され，酒田らの方法と比べて音声品質が向上した．

以上の研究により，エンベロープだけではなく，キャリアについても残響環境下から十分に回復できることが示された．

2.2 課題

前述したが，MTF に基づく残響音声回復モデルでは，キャリアの再生に音声の F0 や有声無声区間の情報を用いている．しかし，残響環境下からの F0 の推定は非常に難しいとされ，F0 の推定精度の低下が懸念される．図 2.2，2.3 に元の音声と残響音声から音源情報抽出法である TEMPO を用いて抽出した F0 の例を示す．用いた音声は ATR 音声データベースの男性話者「冗談」．残響時間は 1 秒．残響音声から抽出した F0 は推定誤りが目立つ．F0 が正確に推定できなければ，基本波とその倍音からなる有声音を正確に構成できない．また，有声音と無声音区間も正確に分けられない．したがって，F0 を用いなくてはならない方法では，残響音声を十分に回復できない可能性がある．

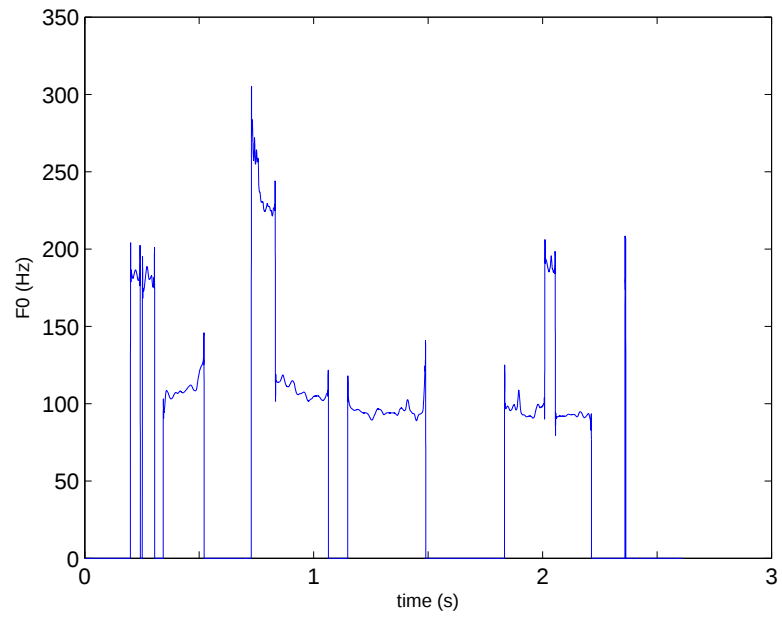


図 2.3: 残響音声から抽出した F0

第3章 改良モデル

3.1 概要

F0 を用いることなく，残響音声を回復するためのモデルの概念を示す（図 3.1）．まず残響音声を帯域分割した信号を正の部分と負の部分に分割する．それぞれの信号の瞬時振幅から残響の影響を取り除き，瞬時振幅を回復する．回復した瞬時振幅から瞬時位相を求め，帯域内の信号を合成する．

3.2 問題設定

本研究では，まず瞬時振幅から残響の影響が取り除かれたものとして，元の信号の瞬時振幅から瞬時位相を推定し，元の信号を合成する方法について検討する．

フィルタバンクによって帯域分割した信号 $x_n(t)$ を AM-PM 信号として考える．

$$x_n(t) = s_{x,n}(t) \cos(\omega_{c,n}t + p_{x,n}(t)), \quad (3.1)$$

$s_{x,n}(t)$ は瞬時振幅， $p_{x,n}(t)$ は瞬時位相である． $\omega_{c,n}$ はフィルタバンクの中心周波数， n はチャンネル番号である．

ここで，瞬時振幅とフィルタバンクの中心周波数は既知であるから，残る未知の項である元の信号の瞬時位相が分かれば，元の信号を復元することができる．

しかし，1 つの式に対して未知の項が 2 つ存在しており，このままでは式を解くことができない．また，既知の項である瞬時振幅や中心周波数と未知の項である瞬時位相は一般に独立で，既知の項から未知の項を推定することも難しい．本研究では，瞬時振幅と瞬時位相を関係付けるための条件を示し，元の瞬時振幅から信号を合成する方法について検討する．

今，瞬時振幅 $s_n(t)$ と瞬時位相 $p_n(t)$ は，以下のように表せる．

$$s_{x,n}(t) = \sqrt{x_n(t)^2 + \{\text{Hilbert}(x_n(t))\}^2}, \quad (3.2)$$

$$\omega_{c,n}t + p_n(t) = \arctan(\text{Hilbert}[x_n(t)]/x_n(t)). \quad (3.3)$$

$x_n(t)$ の Hilbert 変換は

$$\text{Hilbert}(x_n(t)) = \mathcal{F}^{-1}[-j\text{sgn}(\omega)X_n(\omega)], \quad (3.4)$$

であらわされる． $X_n(\omega)$ は $x_n(t)$ の Fourier 変換である．

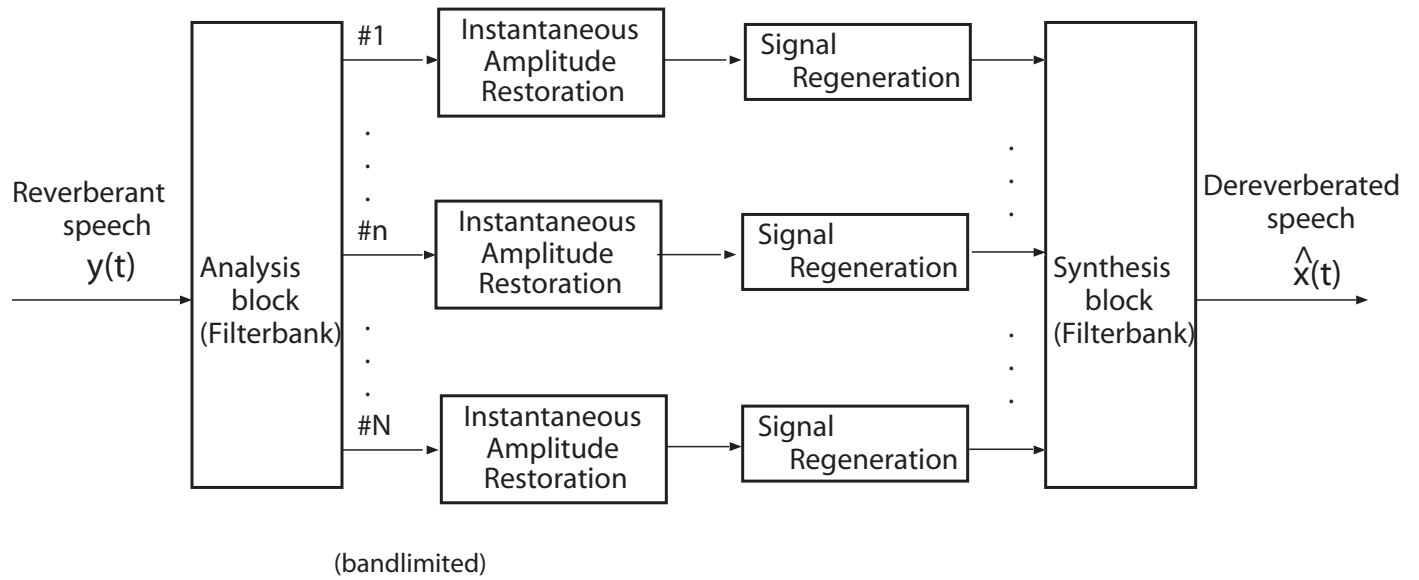


図 3.1: 改良モデルの概念

3.3 瞬時振幅からの瞬時位相の推定と，信号の復元

ここでは，周波数領域と時間領域のアナロジーを利用して，周波数領域で振幅スペクトルから Hilbert 変換の関係を用いて位相スペクトルを推定する方法と条件を参考に，時間領域で瞬時振幅から瞬時位相を推定し，信号を復元する方法について検討する．

3.3.1 Hilbert 変換の関係

ここでは Hilbert 変換の関係について説明する．

ある信号が $t < 0$ で 0 となるならば周波数領域の実部と虚部は Hilbert 変換の関係である．そのような信号 $U(t)$ のフーリエ変換は

$$F[U(t)] = \pi\delta(\omega) + j/\omega \quad (3.5)$$

実部と虚部が Hilbert 変換の関係であるか確かめるために実部を Hilbert 変換する．

$$\text{Hilbert}[\pi\delta(\omega)] = F^{-1}[-j\text{sgn}(\omega')\pi]$$

$$\begin{aligned}
&= 1/(\pi\omega)\pi \\
&= 1/\omega
\end{aligned} \tag{3.6}$$

$t < 0$ で 0 となる信号のフーリエ変換の実部と虚部には Hilbert 変換の関係があることが分かる．

3.3.2 最小位相信号

最小位相信号であれば，位相スペクトルを振幅スペクトルから求めることができる．時間領域では最小位相信号は片側が 0 となる．周波数領域では

$$\begin{aligned}
X(\omega) &= s(\omega) \cos(p(\omega)) + j s(\omega) \sin(p(\omega)) \\
&= s(\omega) \exp(jp(\omega)),
\end{aligned} \tag{3.7}$$

と表せる．時間領域で片側が 0 であるから，周波数領域の実部と虚部には Hilbert 変換の関係がある．さらにこの対数をとると，振幅成分と位相成分が実部と虚部に分離する．

$$\log[X(\omega)] = \log[s(\omega)] + jp(\omega) \tag{3.8}$$

最小位相信号では， $\log[X(\omega)]$ の時間領域でも片側が 0 となる．すなわち，周波数領域の実部と虚部も Hilbert 変換で結ばれており，対数振幅スペクトルの Hilbert 変換から位相スペクトルを一意に求めることができる．

図 3.2 に $X(\omega)$ と $\log[X(\omega)]$ の時間領域と周波数領域の表現の例を示す．時間領域では片側が 0 となっている．また，周波数領域での実部と虚部には Hilbert 変換の関係がある．このようなときに，周波数領域の実部は正になっていることが分かる．したがって，周波数領域の実部が正である信号は最小位相信号である．詳しい証明は付録で行う．

3.3.3 瞬時位相の推定

時間領域と周波数領域のアナロジーから，時間領域でも瞬時振幅から瞬時位相を推定できる．帯域分割した信号 $x_n(t)$ とその Hilbert 変換から，解析信号 $x_{a,n}(t)$ を下式のように定義する．

$$\begin{aligned}
x_{a,n}(t) &= x_n(t) + j\text{Hilbert}[x_n(t)] \\
&= s_{x,n}(t) \cos(\omega_{c,n}t + p_{x,n}(t)) + j s_{x,n}(t) \sin(\omega_{c,n}t + p_{x,n}(t)) \\
&= s_{x,n}(t) \exp(jp_{x,n}(t)) \exp(j\omega_{c,n}t)
\end{aligned} \tag{3.9}$$

今，簡略のため $s_{x,n}(t) \exp(jp_{x,n}(t))$ を中心に検討する．この実部と虚部は Hilbert 変換の関係にある．また， $x_{a,n}(t)$ はフーリエ変換すると片側が 0 となる．この対数をとると

$$\log[x_{a,n}(t)] = \log[s_{x,n}(t)] + j(p_{x,n}(t)). \tag{3.10}$$

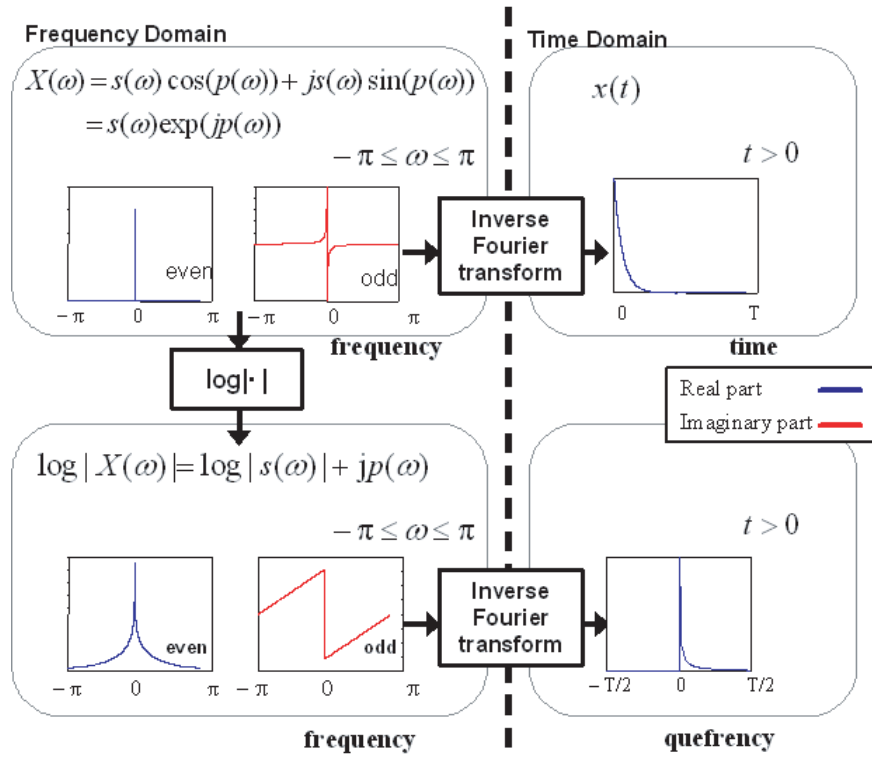


図 3.2: 周波数領域において振幅から位相を求められる例

ここまでの式の展開は，周波数領域と時間領域が入れ替わっただけで同じである．本研究の着想点は，もし対数瞬時振幅と瞬時位相が Hilbert 変換の関係にある場合，瞬時振幅から信号を復元できるということである．その場合は，下式により対数瞬時振幅の Hilbert 変換から瞬時位相を求めることができる．

$$\hat{p}_{x,n}(t) = \text{Hilbert}(\log[s_{x,n}(t)]). \quad (3.11)$$

図 3.3 に $x_{a,n}(t)$ と $\log[x_{a,n}(t)]$ の時間領域と周波数領域の表現の例を示す．周波数領域では片側が 0 となっている．また，時間領域での実部と虚部には Hilbert 変換の関係がある．このようなときに，時間領域の実部，すなわち元の信号は正になっていることが分かる．

元の信号が正であるという条件下であれば，一般的には独立である対数瞬時振幅と瞬時位相が Hilbert 変換の関係で結ばれ，問題設定で未知であった瞬時位相を既知である瞬時振幅から一意に求めることが可能になる．これらが分かれば，問題設定で難しいとした元の信号の復元ができるようになる．

3.3.4 周波数領域で片側が 0 となる信号の扱い

前述したが，周波数領域と時間領域のアナロジーから，瞬時振幅から瞬時位相を求められる信号は周波数領域で片側が 0 になる．周波数領域で片側が 0 となるような信号とは，

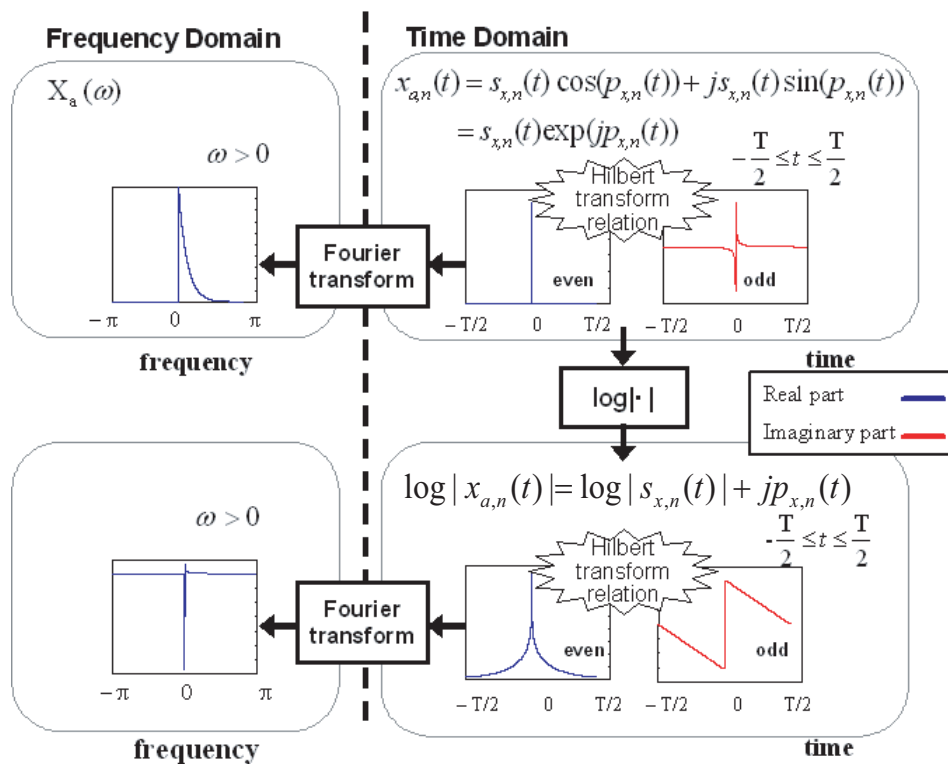


図 3.3: 時間領域において振幅から位相を求められる例

負の時間でも値を持っていなければならない．また，元の信号にあたる時間領域の実部は偶関数でなければならないという条件もある．しかし，現実的に観測できる信号は負の時間で値を持たない．

そこで，本来なら $0 < t < T$ に存在する元の信号を， $-T/2 < t < T/2$ に存在する偶関数とみなして処理を行うこととする．これによって，解析信号は周波数領域で片側が 0 となる．

3.3.5 帯域分割した信号への適用

上記の手法を帯域分割した信号に適用する．一般に帯域分割した信号は正の値のみもつことはない．このままでは，瞬時振幅から瞬時位相を推定できず，元の信号を復元することはできない．そこで，図 3.4 のように信号を正の部分と負の部分に分割して，負の部分についてはマイナスをかけることにする．正の部分と負の部分に分割した信号からは，半波整流と同様にエンベロープを求めることができる．それぞれのエンベロープの残響の影響を取り除き，回復瞬時振幅を求め，正の部分と負の部分信号を回復する．最終的に，正の部分とマイナスをかけた負の部分は足し合わされ，帯域内の信号は回復される．

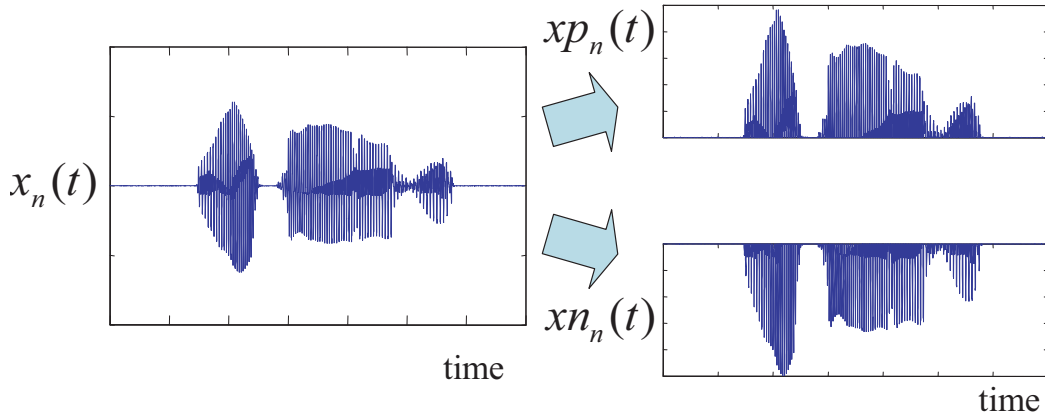


図 3.4: 信号を正の部分と負の部分に分割する例

3.4 評価

ここでは帯域分割した信号の瞬時振幅から帯域内の信号を復元し，音声を復元する方法の評価を行う．

3.4.1 実験条件

ATR 音声データベース [21] の 10 話者 (MAU, MHT, MNM, MTM, MTT, FAF, FFS, FKN, FSU, FYN), 10 単語 (相変わらず, 季節, 新聞, 冗談, 中間, 滑らか, 施す, 間に合う, 楽観, わがまま), 合計 100 単語を対象にシミュレーションを行った．元の音声と復元音声を LSD と SNR を用いて比較した．LSD と SNR 下式のように表される．

$$\text{LSD} = \sqrt{\frac{1}{W} \sum_{\omega} \left(20 \log_{10} \frac{|S(\omega)|}{|\hat{S}(\omega)|} \right)^2}, \quad (\text{dB}) \quad (3.12)$$

$$\text{SNR}(x(t), \hat{x}(t)) = 20 \log_{10} \frac{\int_0^T x(t) dt}{\int_0^T (x(t)^2 - \hat{x}(t)^2) dt} \quad (\text{dB}). \quad (3.13)$$

ここで, W は周波数の上限 (ここでは 10 kHz とした), $|S(\omega)|$ と $|\hat{S}(\omega)|$ は $x(t)$ と $\hat{x}(t)$ の振幅スペクトルである．サンプリング周波数は 20 kHz, 帯域分割に用いたフィルタバンクは帯域幅 100 Hz の定帯域フィルタバンクである．

3.4.2 結果

図 3.5 に, 代表的な例として MAU10003.AD (冗談) を用いた場合の, 2 チャンネル目の信号を示す．2 チャンネル目の帯域分割した音声とその正の信号, その瞬時振幅, その瞬時位相と Hilbert 変換を用いて生成した瞬時位相を示す．図 3.5 より, 実際の瞬時位相

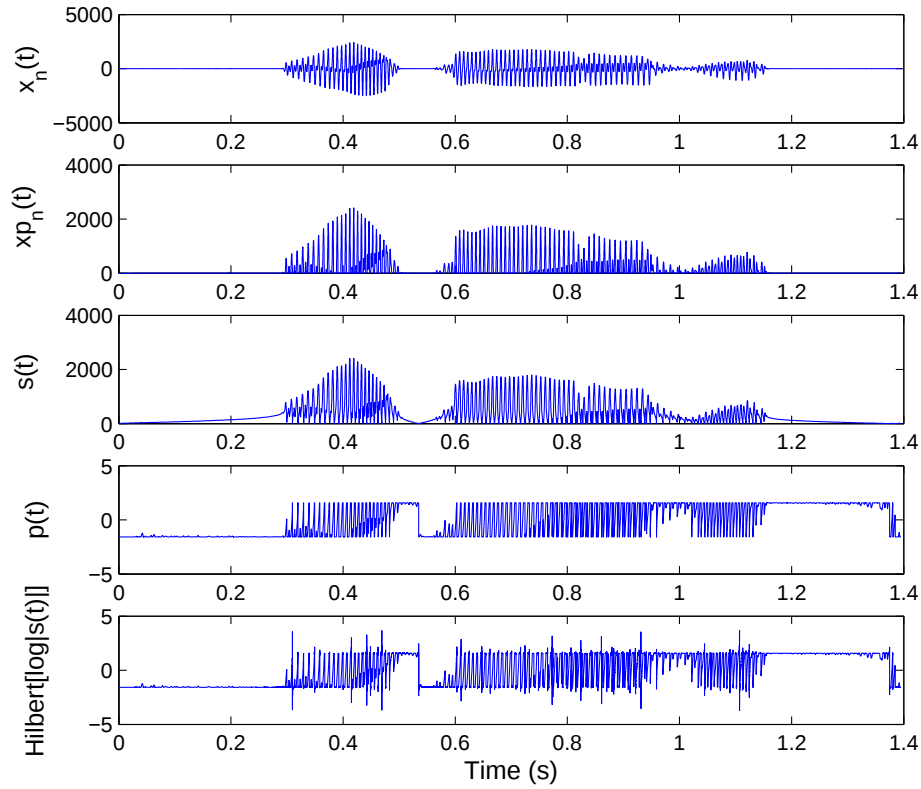


図 3.5: チャンネル内の代表的な波形

と生成した瞬時位相はほぼ同じであることがわかった。

負の部分の信号についても同様で、これらをあわせて帯域内の信号を復元し、合成フィルタバンクで各帯域を合成した。結果を図 3.6 に示す。 $x(t)$ は元の音声 $\hat{x}(t)$ は復元音声である。図 3.6 より、音声が復元できていることが分かる。復元音声の LSD は約 0.1 dB, SNR は約 37.8 dB と元の音声に近いものであった。

次に 100 単語に対するシミュレーションの結果について示す。結果は、平均 LSD 0.1dB, 平均 SNR 24.0 dB と音声を復元できていることが分かった。すなわち、元の帯域分割した信号の瞬時振幅が分かれば、瞬時位相を推定でき、帯域内の元の信号を復元し、音声を復元できることが分かった。

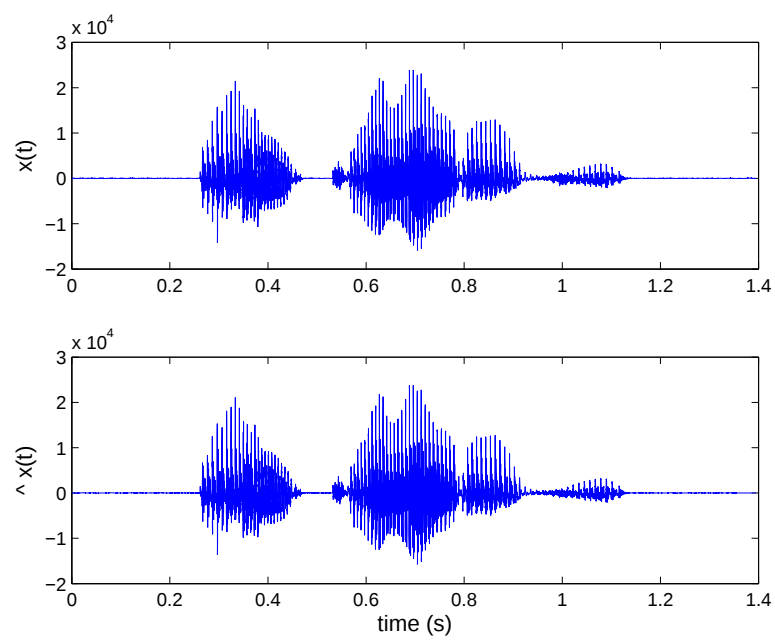


図 3.6: 原音声と復元音声

第4章 残響音声の回復の可能性の検討

4.1 瞬時振幅の回復の可能性

先行研究で回復エンベロープを求める際、まず、残響信号の瞬時振幅をローパスフィルタ (LPF) にとおして、2 乗して残響信号のパワーエンベロープを得ている。その後、パワーエンベロープ逆フィルタ処理により残響の影響を取り除き、このルートをとることで回復エンベロープを得ている。回復瞬時振幅は回復エンベロープに LPF の逆特性をかけることで求まる (図 4.1)。

ここで、もし 2 乗と LPF を通す処理が可換であるならば、LPF と LPF の逆特性をかける処理はお互いに相殺される。結局、残響信号の瞬時振幅にパワーエンベロープ逆フィルタ処理を行うことで、回復瞬時振幅が求められることになる。

ここでは、回復エンベロープを求める処理と、上記の方法で回復瞬時振幅を求める処理を比較しながら、瞬時振幅が回復する可能性について検討する。音声にみたてた原信号、インパルス応答、残響信号を以下のように定義した。

$$x(t) = (1 + \cos(2\pi f_e t)) \times \sum_{k=1}^M \alpha_k \cos(2\pi f_k t + \rho_k) \quad (4.1)$$

$$h(t) = a \exp(-6.9t/T_R) \mathbf{n}(t) \quad (4.2)$$

$$y(t) = h(t) * x(t) \quad (4.3)$$

ここで、 f_e は 20 Hz、 M は 100 とし、 α_k 、 f_k 、 ρ_k はそれぞれ $[0, 1][0, f_s/2][-\pi, \pi]$ の範囲で乱数を用いて定めた。サンプリング周波数 f_s は 20 kHz、残響時間 T_R は 1 秒とした。エンベロープを求める際の LPF のカットオフ周波数は 20 Hz とした。

図 4.2、4.3 に元の信号と残響信号を、正の部分と負の部分に分けた信号を示す。

図 4.4-4.6 に原信号のエンベロープ、残響信号のエンベロープ、回復エンベロープを示す。まず、時間領域で先行研究で提案されている方法で求めた回復エンベロープを残響信号のエンベロープと比較すると、減少していた変調度がより高くなり、原信号のエンベロープに近づいていることが分かる。また、周波数領域で比較する。残響信号のエンベロープでは高い周波数成分が減少していたが、回復エンベロープでは高い周波数成分が強調され、原信号のエンベロープに近づいていることが分かる。

次に、図 4.7-4.9 に原信号の瞬時振幅、残響信号の瞬時振幅、回復瞬時振幅を示す。時間領域をみると回復瞬時振幅は残響信号の瞬時振幅と比べて、振幅が大きくなっている。周波数領域をみると瞬時振幅はエンベロープと比較して広い帯域に周波数が広がっている。

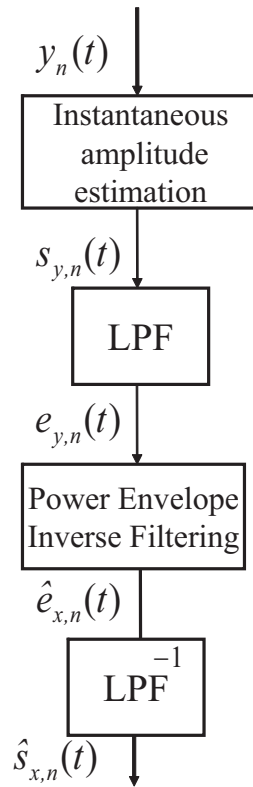


図 4.1: 瞬時振幅の回復

ことが分かる．回復瞬時振幅は原信号の瞬時振幅と比較し高域の値が大きくなっている．また，原信号の瞬時振幅と残響信号の瞬時振幅では大きさがほとんど変わらないことが分かる．

瞬時振幅の回復にパワーエンベロープ逆フィルタをそのまま適用できないことが分かった．しかし，パワーエンベロープ逆フィルタ処理が低周波成分については回復するが，高周波成分については必要以上に強調処理が行われるものであることから，パワーエンベロープ逆フィルタを瞬時振幅の低周波成分にのみ作用させることで，瞬時振幅の低域のみを回復させ，高周波成分をそのままとしたり，瞬時振幅が低周波成分のみをもつように帯域分割幅を決めることで，エンベロープのときと同様に瞬時振幅が回復できる可能性がある．

4.2 残響音声の回復の可能性

前章では，難しい問題である瞬時位相の復元方法と条件を示した．これによって，元の信号の瞬時振幅さえ分かれば，瞬時位相を求められ，元の信号を完全に復元できる．この方法を音声に適用した結果，瞬時振幅から元の音声を完全に復元できることが実験で示された．すなわち，帯域分割した信号の瞬時振幅が回復した場合，残響音声を回復すること

ができる．

あとは，戸井らの方法を用いて，残響音声を帯域分割した信号の瞬時振幅から残響の影響を取り除くことさえできれば，残響音声を回復することができる．

戸井らの方法をそのまま用いて残響音声を帯域分割した信号の瞬時振幅を回復することはできないが，今後はパワーエンベロープ逆フィルタ処理を低域のみに対応させたり，瞬時振幅の周波数成分が低域のみ存在するように残響音声の帯域幅を決めることで，瞬時振幅は回復する可能性があり，これによって残響音声回復の可能性がある．

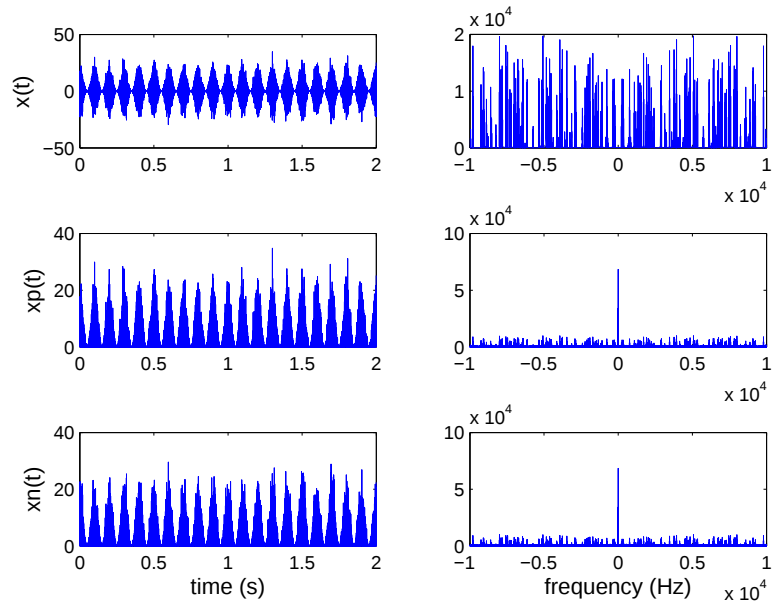


図 4.2: 原信号とその正の部分 $x_p(t)$, 負の部分 $x_n(t)$

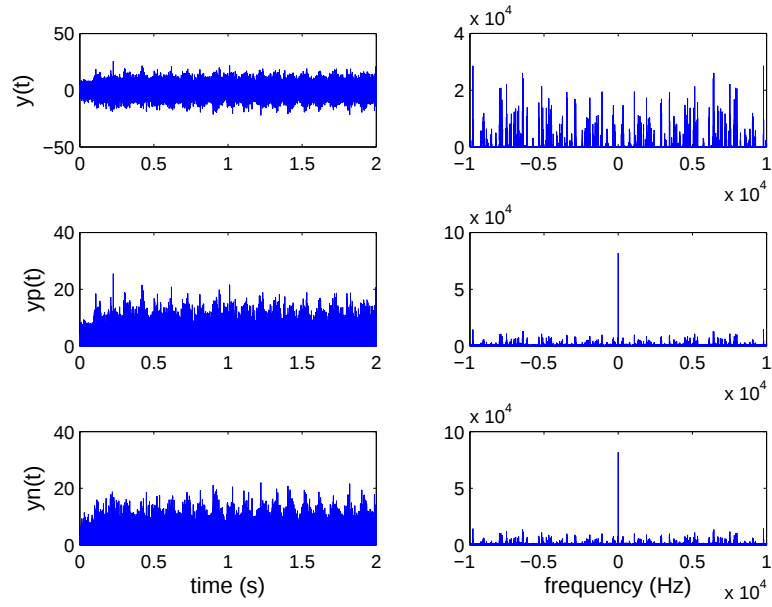


図 4.3: 残響信号とその正の部分 $y_p(t)$, 負の部分 $y_n(t)$

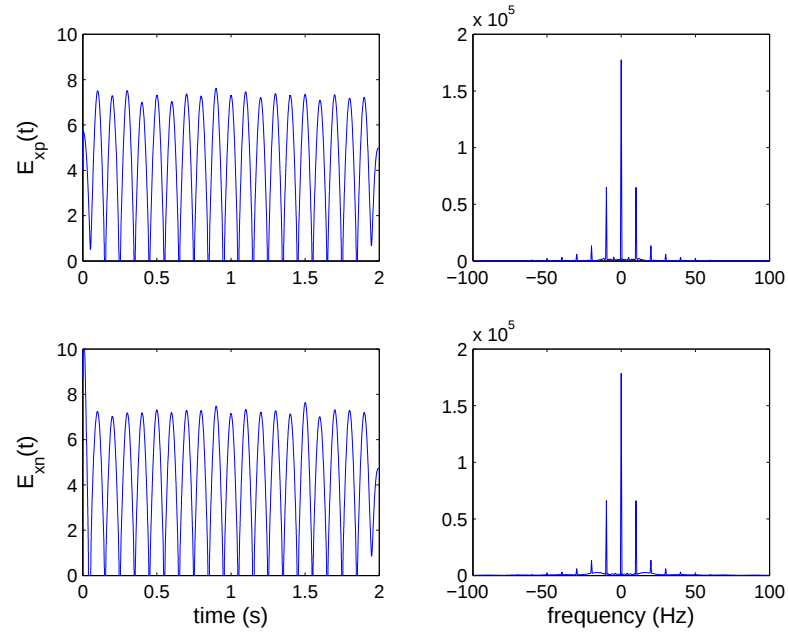


図 4.4: 原信号の正の部分のエンベロープ $E_{xp}(t)$, 負の部分のエンベロープ $E_{xn}(t)$

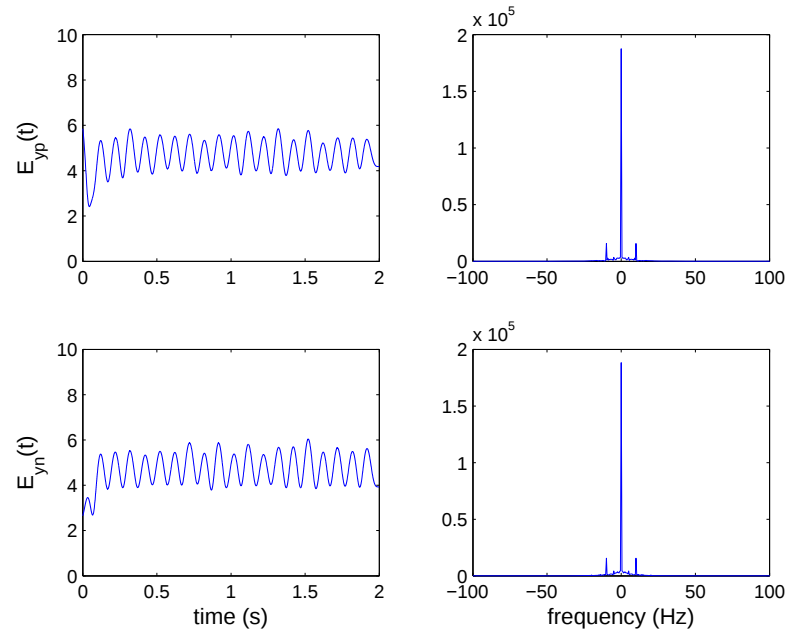


図 4.5: 残響信号の正の部分のエンベロープ $E_{yp}(t)$, 負の部分のエンベロープ $E_{yn}(t)$

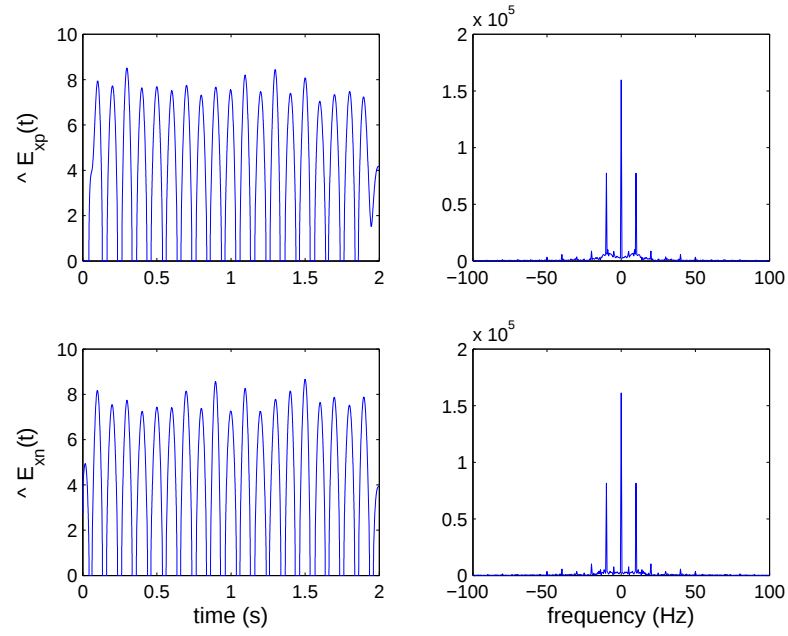


図 4.6: 回復した正の部分のエンベロープ $E_{\hat{x}p}(t)$, 負の部分のエンベロープ $E_{\hat{x}n}(t)$

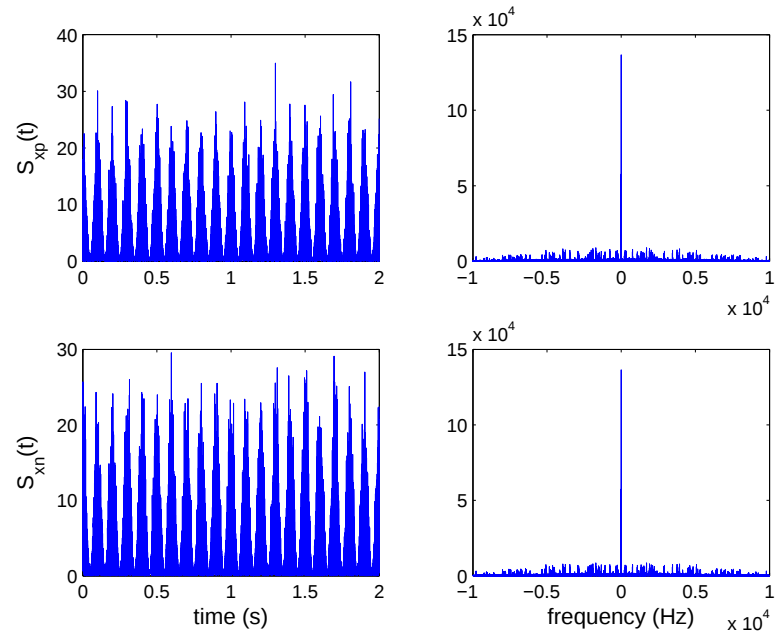


図 4.7: 原信号の正の部分の瞬時振幅 $s_{xp}(t)$, 負の部分の瞬時振幅 $s_{xn}(t)$,

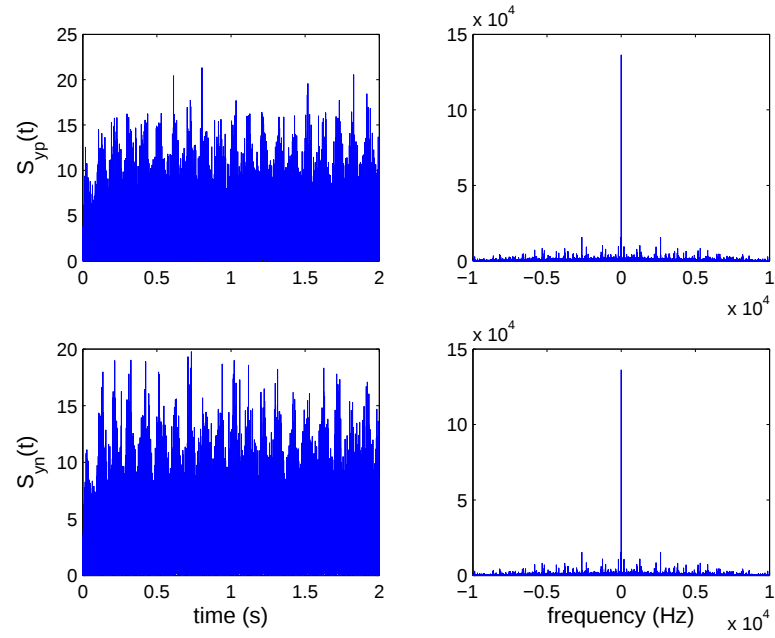


図 4.8: 残響信号の正の部分の瞬時振幅 $s_{yp}(t)$, 負の部分の瞬時振幅 $s_{yn}(t)$

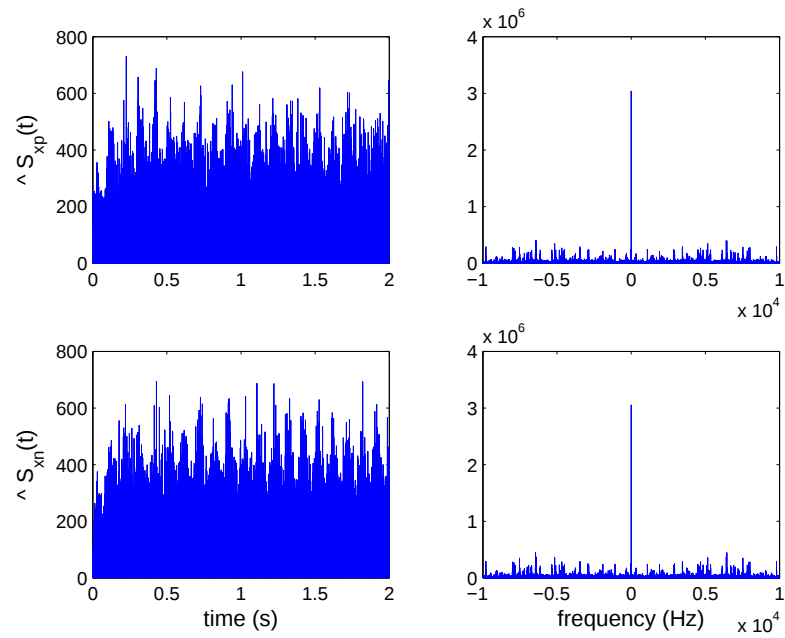


図 4.9: 回復した正の部分の瞬時振幅 $s_{xp}(t)$, 負の部分の瞬時振幅 $s_{xp}(t)$

第5章 結論

5.1 本論文で明らかにしたこと

戸井らは、伝達系の測定を必要としないMTFに基づく残響音声回復法を提案しているが、キャリアの再生成で F_0 を用いているため、残響音声が十分回復しない可能性があった。

本研究の目的は、 F_0 を用いることなく、残響音声回復の可能性を示すことである。まず、先に提案されたMTFに基づく残響音声回復法の改良モデルの概念を示した。改良モデルでは残響音声を帯域分割し、先行研究で用いられたパワーエンベロープ逆フィルタ処理を用いて瞬時振幅を回復する。回復した瞬時振幅から位相を推定し、帯域内の信号を回復し、最終的にそれらを足し合わせ音声を回復する。本研究ではこの、難しい問題である瞬時振幅から位相を推定し、信号を復元する方法について検討した。

この問題は1つの式に未知の項が2つあり、また既知の項である瞬時振幅と未知の項である瞬時振幅は一般に独立であり非常に難しい。しかし、本論文ではこれを、元の信号が正であるという条件下であれば、対数瞬時振幅のHilbert変換から瞬時位相を推定し、信号を完全に復元できるということを示した。この方法を音声に適用するために、音声を帯域分割した信号を正の部分と負の部分に分割し、それぞれの瞬時振幅から帯域内の信号を復元させ、音声を復元するシミュレーションを行った。シミュレーションはATR音声データベースの100単語を対象に行われた。その結果、復元音声は平均LSD 0.1 dB、平均SNR 24.0 dBであり、瞬時振幅から音声を復元することができた。これにより、残響音声を帯域分割した信号の瞬時振幅が回復した場合、残響音声をブラインドで回復することができる。

瞬時振幅の回復については、戸井らの方法をそのまま用いて残響音声を帯域分割した信号の瞬時振幅を回復することはできないと考えられるが、今後はパワーエンベロープ逆フィルタ処理を低域のみに対応させたり、瞬時振幅の周波数成分が低域のみ存在するように残響音声の帯域幅を決めることで、瞬時振幅は回復する可能性がある。瞬時振幅が回復すれば、残響音声は回復することができる。したがって、 F_0 を用いることなく、残響音声回復の可能性がある。

5.2 課題

今後は、残響信号の瞬時振幅を回復させ、回復した瞬時振幅から信号を復元するシミュレーションを行う。パワーエンベロープ逆フィルタ処理は低周波成分については適切な処

理を施すが，高周波成分については必要以上に強調処理が行われるものであることから，パワーエンベロープ逆フィルタを瞬時振幅の低周波成分にのみ作用させることで，瞬時振幅の周波数成分が低域のみ存在するように残響音声の帯域幅を決めることで，瞬時振幅は回復する可能性がある．これにより，本手法を総合的に評価することができる．音声認識システムや，拡声音声通話，補聴システムでの性能の向上が期待できる．

謝辞

本研究を進めるにあたり，多大な御助言を頂き熱心なご指導をして頂きました鵜木祐史准教授に心から感謝致します．本研究を進めるにあたり，多大な御助言を頂き熱心なご指導をして頂きました赤木正人教授に心から感謝致します．本研究に関して，多大な御助言，御討論をして頂いた党建武教授に心より感謝致します．本研究に関して，多大な御助言，御討論をして頂いた徳田功准教授に心より感謝致します．本研究に関して，多大なご助言を頂いた研究生の西本博則氏，博士後期課程の羽二生篤氏，黄純芳氏，Vu Tat Thang 氏，Binh Phu NGUYEN 氏に心より感謝致します．また，本研究を進めるにあたり，有意義な討論並びに，有益なご助言を賜った赤来・鵜木研究室，党・徳田研究室の皆様方に感謝いたします．

参考文献

- [1] Neely S. T. and Allen J. B., “Invertibility of a room impulse response,” J. Acoust. Soc. Am. Vol. 66, No. 1, July 1979.
- [2] Miyoshi, M. and Kaneda, Y., “Inverse filtering of room acoustics,” IEEE Trans. ASSP, Vol. 36, No. 2, pp. 145–152, Feb. 1988.
- [3] Wang, H. and Itakura, F., “Realization of acoustic inverse filtering through multi-microphone sub-band processing,” IECE Trans. Fundamentals, vol.E75-A, no. 11, pp. 1474-1483, Nov. (1992).
- [4] Nakatani, T., and Miyoshi, M., “Blind dereverberation of single channel speech signal based on harmonic structure,” Proc. ICASSP2003, vol. 1, 92-95, 2003.
- [5] 中谷智広, 三好正人, 木下慶介, “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,” 電子通信情報学会論文誌, Vol. J88-D-II, No. 3, pp.509-520, Mar. 2005.
- [6] Houtgast, T., Steenken, H. J. M., and Plomp, R., “Predicting speech intelligibility in room acoustics,” Acoustica, Vol. 46, pp. 60-72, 1980.
- [7] Houtgast, T., and Steeneken, H. J. M., “A review of the MTF-concept in room acoustics,” J. Acoust. Soc. Am. 77, 1069-1077. 1985.
- [8] Mourjopoulos J., and Hammond J. K., “Modelling and Enhancement of Reverberant Speech using an Envelope Convolution Method,” Proc. IEEE ICASSP 1983. pp. 1144-1147. 1983.
- [9] 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,” 電子情報通信学会論文誌 A, Vol. J81-A, No. 10, pp. 1323–1330, 1998.
- [10] 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,” 電子情報通信学会論文誌 A, Vol. J83-A, No. 8, pp. 1029-1033, 2000.
- [11] 広林 茂樹, 寺島 洋行, 山淵 龍夫, “残響音場における音響信号のエンベロープ推定法の評価,” シミュレーション学会誌, Vol. 22, No. 3, pp. 68-75, 2003.

- [12] 戸井真智, “音声明瞭度向上のための残響音声回復法に関する研究,” 北陸先端科学技術大学院大学修士論文, 平成 17 年 3 月.
- [13] Unoki, M., Toi, M., and Akagi, M., “Development of the MTF-based speech dereverberation method using adaptive time-frequency division,” Forum Acusticum 2005 Budapest.
- [14] Unoki, M., Toi, M., and Akagi, M. “Refinement of an MTF-based speech dereverberation method using optimal inverse-MTF filter,” Proc. SPECOM2006, pp. 323-326, St. Petersburg, Russia, June 2006.
- [15] 古川正和, “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,” 北陸先端科学技術大学院大学修士論文, 平成 14 年 3 月.
- [16] 酒田恵吾, “残響音声からの音声特徴量抽出法と音源波形再合成に関する研究,” 北陸先端科学技術大学院大学修士論文, 平成 15 年 3 月.
- [17] Unoki, M., Sakata, K., Furukawa, M., and Akagi, M., “A Method based on the MTF concept for dereverberating the power envelope from the reverberant signal,” Proc. of ICASSP 2003, vol. I, pp. 840-843, 2003.
- [18] Unoki, M., Sakata, K., Akagi M., “A speech dereverberation method based on the MTF concept,” Proc. EuroSpeech2003, pp. 1417-1420, 2003.
- [19] Hamagami, T., “Speech synthesis using source wave shape modification technique by harmonic phase control,” J.Acoust. Soc. Jpn, 54(9), 623-632, 1998.
- [20] 河原 英紀, 津崎 実, Roy, D, Patterson., “オールパスフィルタの位相操作による時間構造制御とその知覚への影響について,” 日本音響学会聴覚研究会資料, H-96-79, 1996.
- [21] Takeda, T., Sagisaka, Y., Katagiri, S., Abe, M., and Kuwabara, H., Speech Database User's Manual, ATR Technical Report, TR-I-0028, 1988.

研究業績

Shibano, Y., Unoki, M., “An improved carrier regeneration process in the MTF-based speech dereverberation method using instantaneous amplitude and phase,” Nonlinear Circuits and Signal Processing 2007, Shanghai, China, 2007 (to be appear).

付 録 A

周波数領域の実部と虚部に Hilbert 変換の関係があり，実部が正となる信号は最小位相信号であることを証明する．最小位相信号は極も零も単位円の内側に入っている．

$$\begin{aligned} X(z) &= R(z) + jI(z) \\ &= \frac{R_P(z)}{R_Q(z)} + j \frac{I_P(z)}{I_Q(z)} \\ &= \frac{R_P(z)I_Q(z) + jR_Q(z)I_P(z)}{R_Q(z)I_Q(z)} \end{aligned} \quad (\text{A.1})$$

$R(z)$ は実部, $I(z)$ は虚部である． $R_P(z)$ と $I_P(z)$ は分子多項式, $R_Q(z)$ と $I_Q(z)$ は分母多項式を表す．

もし $X(z)$ の逆 z 変換が因果的であるならば，

$$R_Q(z)I_Q(z) = 0, |z| < 1 \quad (\text{A.2})$$

$$I_Q(z) = 0, |z| < 1 \quad (\text{A.3})$$

また，もし $R_P(z) = 0$ であるならば，

$$R_P(z)I_Q(z) = 0, |z| < 1 \quad (\text{A.4})$$

$$(R_P(z)I_Q(z) + jR_Q(z)I_P(z)) = 0, |z| < 1 \quad (\text{A.5})$$

となり， $X(z)$ の極も零も単位円内に入っているから，最小位相信号となる．実部が正 $R(z) > 0$ のとき $R_P(z) = 0$ が成り立ち，すなわち周波数領域の実部と虚部に Hilbert 変換の関係があり，実部が正であれば最小位相信号である．よって証明された．最小位相信号は振幅スペクトルから位相スペクトルを求めることができる．