JAIST Repository
https://dspace.jaist.ac.jp/

Title	変調伝達関数に基づく骨導音声のブラインド回復法に関する研究
Author(s)	衣笠，光太
Citation	
Issue Date	2009-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/8094
Rights	
Description	Supervisor:鵜木祐史　准教授，情報科学研究科，修士

JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

修 士 論 文

変調伝達関数に基づく骨導音声のブラインド回復法に関する研究

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

衣笠 光太

2009 年 3 月

修 士 論 文

変調伝達関数に基づく骨導音声のブラインド回復法に関する研究

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　党 建武　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

710021 衣笠 光太

提出年月: 2009 年 2 月

Copyright © 2009 by Kota Kinugasa

概 要

骨導マイクは発話による頭部の振動をピックアップし，音声を収録するものである．頭蓋骨と空気のインピーダンスの差が非常に大きい事から，骨導マイクは外部雑音の影響を受けることなく音声を収録することが可能である．しかし，骨導マイクを使って収録された音声（骨導音声）は，音質が悪く，音声明瞭度が低い．そのため，骨導音声を利用して音声コミュニケーションを行うには，骨導音声の音質や明瞭度を回復する必要がある．

骨導音声は高い周波数帯域ほどパワー減衰することがわかっているため，骨導音声の音質改善の最も簡単な手法として，高域強調が用いられている．しかし，骨導音声の減衰作用は複雑であり，高域強調のみでは補えない．クロススペクトル法や長時間 Fourier 変換などの逆フィルタ法と呼ばれる方法もあるが，これらの手法はエコーなどのアーティファクトを生み出してしまう上，気導音声の情報を必要とするためブラインド処理ではない．高騒音環境下では気導音声を録音することが非常に難しいため，骨導音声回復法は，ブラインド処理であることが必要である．

一方，ブランド処理である骨導音声回復法として，線形予測分析に基づく骨導音声回復法が提案されている．この手法は，骨導音声の周波数成分を非常によく回復することができたが，事前に学習を必要とする手法である．また，骨導音声を利用して音声コミュニケーションを行うことを考えた際，最も重要視すべきは音声明瞭度であり，音声明瞭度を直接回復できる手法が望まれる．

音声明瞭度を回復する手法として，MTF に基づく骨導音声回復法がある．この手法は，音声明瞭度と関係のある MTF に基づいており，音声明瞭度を直接回復することができる．しかし，気導パワーエンベロープと骨導パワーエンベロープ間の MTF をどのようなモデルで表現するのが最適か明らかにされていない．また，この手法は，骨導音声回復に気導音声の情報を必要とし，ブラインド処理になっていない．

本研究では，MTF に基づくブラインド骨導音声回復法を提案する．まず，気導/骨導データベースを用いて気導パワーエンベロープと骨導パワーエンベロープ間の変換関係の解析を行い，気導パワーエンベロープと骨導パワーエンベロープの振幅比を回帰曲線で近似できることを明らかにした．また，この回帰曲線は観測点のみに依存し，話者や発話内容にはほぼ依存しないことを明らかにした．次に，気導/骨導データベース内の音声から求めた MTF と 3 つのモデルをフィッティングすることで，気導エンベロープと骨導エンベロープ間の MTF をモデリングした．このモデルは 2 つのパラメータを持っており，減衰特性を制御するパラメータ b は先行研究により推定法が提案されている．ゲインを制御するパラメータ a は，解析結果より近似曲線を用いることで気導音声の情報なしに決定することが可能である．これらの結果から，MTF に基づく処理体系のブラインド骨導音声回復法を提案した．

最後に，シミュレーションにより提案法の評価を行い，提案法が骨導音声の音質，明瞭度を回復できている事を確認した．

目 次

第 1 章　序論 **1**
1.1 はじめに 1
1.2 研究の背景 1
1.3 MTF に基づく骨導音声回復法 2
1.3.1 変調伝達関数 2
1.3.2 パワーエンベロープ逆フィルタ法 3
1.3.3 MTF に基づく骨導音声回復法 3
1.3.4 残された課題 5
1.4 研究の目的 6
1.5 本論文の構成 6

第 2 章　骨導/気導パワーエンベロープ間の変換特性 **7**
2.1 気導/骨導データベース 7
2.2 骨導/気導エンベロープ間の変換特性の解析 7
2.3 解析結果の考察 9
2.4 気導パワーエンベロープと骨導パワーエンベロープ間の MTF のモデリング 17

第 3 章　MTF に基づくブラインド骨導音声回復法 **28**
3.1 MTF モデルのパラメータ a と b の決定方法 28
3.1.1 パラメータ a の決定方法 28
3.1.2 パラメータ b の決定方法 43
3.2 回復条件の変更 46

第 4 章　提案法の評価 **47**
4.1 評価方法 47
4.2 評価結果 48

第 5 章　結論 **59**
5.1 本研究で明らかにしたこと 59
5.2 今後の課題 59

図 目 次

1.1 気導音声と骨導音声の間の伝達特性の定義．(1) 波形レベル間の伝達特性，(2) パワーエンベロープ間の伝達特性． 4
1.2 変調フィルタバンクを用いた MTF に基づく骨導音声回復法の概要． 5

2.1 気導/骨導データベース構築の際の音声の収録環境 (数字 1〜5 は観測点)． . 9
2.2 全観測点での解析結果 （実線: 平均，破線: 平均 ± 標準偏差）．(a) 相関係数，(b) SNR，(c) MTF の回帰直線の傾き，(d) 伝達関数，(e) パワーエンベロープの平均パワーの比（パラメータ $1/a_n^2$），(f) 各チャンネル毎の $e_y^2(t)$ の平均（点線は相対パワーが-40 dB 下がった位置を表す）． 11
2.3 観測点 1 での解析結果．体裁は図 2.2 と同じ． 12
2.4 観測点 2 での解析結果．体裁は図 2.2 と同じ． 13
2.5 観測点 3 での解析結果．体裁は図 2.2 と同じ． 14
2.6 観測点 4 での解析結果．体裁は図 2.2 と同じ． 15
2.7 観測点 5 での解析結果．体裁は図 2.2 と同じ． 16
2.8 実際の MTF とモデルの比較． MTF without internal noise:内部雑音を取り除いた MTF $e_h(t) = at\exp(-bt)$: 指数関数 $e_h(t) = a\exp(-bt)$: 先行研究で用いられているモデル LPF: ローパスフィルタ MTF:気導/骨導音声データベースのデータから求めたの MTF． 18
2.9 内部雑音を除去したパワーエンベロープ． (a) 骨導音声 (b) 内部雑音 (c) 内部雑音除去後の骨導音声のパワーエンベロープ． 19
2.10 内部雑音を除去した骨導音声の変調スペクトル (実部)． (a) 骨導音声の変調スペクトル (b) 内部雑音の変調スペクトル (c) 内部雑音除去後の骨導音声の変調スペクトル． 20
2.11 内部雑音を除去した骨導音声の変調スペクトル (虚部)． (a) 骨導音声の変調スペクトル (b) 内部雑音の変調スペクトル (c) 内部雑音除去後の骨導音声の変調スペクトル． 21
2.12 MTF を表現するのに最も適切なモデル $a^2\exp(-2bt)$ と気導/骨導データベース内の全音声から求めた MTF との RMS 誤差の平均と標準偏差（実線: 平均 破線: 平均 ± 標準偏差 (std)）． 22
2.13 データベース内の音声から求めた MTF に，MTF を表現するのに最も適切なモデル $a^2\exp(-2bt)$ をフィッティングした際のモデルの回帰直線の傾き（実線: 平均 破線: 平均 ± 標準偏差） (観測点 1)． 23

2.14 データベース内の音声から求めたMTFに，MTFを表現するのに最も適切なモデル $a^2\exp(-2bt)$ をフィッティングした際のモデルの回帰直線の傾き（実線: 平均 破線: 平均 ± 標準偏差） (観測点 2). 24
2.15 データベース内の音声から求めたMTFに，MTFを表現するのに最も適切なモデル $a^2\exp(-2bt)$ をフィッティングした際のモデルの回帰直線の傾き（実線: 平均 破線: 平均 ± 標準偏差） (観測点 3). 25
2.16 データベース内の音声から求めたMTFに，MTFを表現するのに最も適切なモデル $a^2\exp(-2bt)$ をフィッティングした際のモデルの回帰直線の傾き（実線: 平均 破線: 平均 ± 標準偏差） (観測点 4). 26
2.17 データベース内の音声から求めたMTFに，MTFを表現するのに最も適切なモデル $a^2\exp(-2bt)$ をフィッティングした際のモデルの回帰直線の傾き（実線: 平均 破線: 平均 ± 標準偏差） (観測点 5). 27

3.1 パラメータ a の平均の回帰曲線と，発話内容ごとの最適なパラメータ a の値とRMS誤差の平均と標準偏差（観測点1）. 29
3.2 パラメータ a の平均の回帰曲線と，発話内容ごとの最適なパラメータ a の値とRMS誤差の平均と標準偏差（観測点2）. 30
3.3 パラメータ a の平均の回帰曲線と，発話内容ごとの最適なパラメータ a の値とRMS誤差の平均と標準偏差（観測点3）. 31
3.4 パラメータ a の平均の回帰曲線と，発話内容ごとの最適なパラメータ a の値とRMS誤差の平均と標準偏差（観測点4）. 32
3.5 パラメータ a の平均の回帰曲線と，発話内容ごとの最適なパラメータ a の値とRMS誤差の平均と標準偏差（観測点5）. 33
3.6 パラメータ a の平均の回帰曲線と，話者ごとの最適なパラメータ a の値とRMS誤差の平均と標準偏差（観測点1）. 34
3.7 パラメータ a の平均の回帰曲線と，話者ごとの最適なパラメータ a の値とRMS誤差の平均と標準偏差（観測点2）. 35
3.8 パラメータ a の平均の回帰曲線と，話者ごとの最適なパラメータ a の値とRMS誤差の平均と標準偏差（観測点3）. 36
3.9 パラメータ a の平均の回帰曲線と，話者ごとの最適なパラメータ a の値とRMS誤差の平均と標準偏差（観測点4）. 37
3.10 パラメータ a の平均の回帰曲線と，話者ごとの最適なパラメータ a の値とRMS誤差の平均と標準偏差（観測点5）. 38
3.11 実際の a の値のRMS誤差が小さい話者（観測点3，話者2）のパラメータ a の平均. 39
3.12 実際の a の値のRMS誤差が大きい話者（観測点3，話者3）のパラメータ a の平均. 40
3.13 観測点3，話者3のパワーエンベロープのパワーの平均. 上: 気導音声 下: 骨導音声. 41

3.14 観測点 3，話者 2 のパワーエンベロープのパワーの平均．上: 気導音声 下: 骨導音声． 42
3.15 周波数ドメインでのパラメータ B の推定． 44
3.16 パラメータ B の推定法の比較．restored1: 時間ドメインでのパラメータ推定 restored2: 周波数ドメインでのパラメータ推定． 45

4.1 提案法による SNR の改善度． 49
4.2 提案法による相関の改善度． 50
4.3 提案法による MTF の回帰直線の傾きの改善度． 51
4.4 提案法による変調度 1 の MTF と骨導/回復音声の RMS 誤差の改善度． . . 52
4.5 提案法による伝達関数の改善度． 53
4.6 LSD による総合評価．BCspeech: 骨導音声，MTF previous: 従来の MTF に基づく骨導音声回復法，MTF nonblind: 気導音声の情報を用いてパラメータ a を求めた提案法，MTF blind: 提案法． 54
4.7 LP-LSD による総合評価．体裁は，図 4.6 と同じ． 55
4.8 ケプストラム距離による総合評価．体裁は，図 4.6 と同じ． 56
4.9 メルケプストラム距離による総合評価．体裁は，図 4.6 と同じ． 57
4.10 明瞭度を考慮した LSD による総合評価．体裁は，図 4.6 と同じ． 58

表 目 次

2.1 気導/骨導音声の収録条件. 8
2.2 骨導パワーエンベロープと気導パワーエンベロープのパワー比に対する観測点毎の近似曲線のパラメータ. 10

第1章　序論

1.1　はじめに

工場や作業現場といった高騒音環境下では，空気伝播された音声（気導音声）は騒音にの影響で歪んでしまい，音声によるコミュニケーションや，音声認識などの音声アプリケーションを阻害する．作業の安全化や効率化のため，高騒音環境下での音声による円滑なコミュニケーションを可能とする方法や，騒音に頑健な音声アプリケーションの開発が求められている．現在までに，雑音抑圧法や音圧強調，特殊マイクを用いて音声を録音する方法など数多くの方法が提案されている．中でも骨導マイクを用いて音声を収録する方法は，外部雑音の影響を受けることなく音声を収録可能であり，非常に有効な手法である．これは，頭蓋骨と空気のインピーダンスの差が非常に大きい事から，骨導マイクは外部雑音の影響を受けることなく音声を収録できるからである．

しかし，骨導マイクを使って収録された音声（骨導音声）は，空気とは減衰特性の異なる頭部の骨や皮膚を伝って伝達されることから，音質が悪く，音声明瞭度が低い [1]，[2]．そのため，音声了解度の低下や，音声認識装置の認識率の低下を招いてしまう [3]．骨導音声を利用して音声コミュニケーションを行う，あるいは音声アプリケーションを動作させるには，骨導音声の明瞭度を気導音声と同等に回復する必要がある．高騒音環境下での問題は，マンマシンの問題と，マンマンの問題に大きく分かれるが，本研究では，マンマンの方を取り扱う．そのため，音声が人間に明確に伝わっているかの指標である音声明瞭度に着目した．

1.2　研究の背景

骨導音声は，高い周波数帯域ほどパワーが減衰することが分かっている．そのため，骨導音声の音質改善の最も簡単な手法として現在用いられているのが高域強調である．しかし，骨導音声の減衰作用は，骨導マイクを設置する位置や，話者，発話内容などにより複雑に変化するため [4]，[5]，[6]，高域強調ではその変化に対応することができない．

この問題に対処した方法として，クロススペクトル法 [7]，[8] や長時間 Fourier 変換 [9] を用いて骨導音声と気導音声の間の伝達特性を求め，その逆特性を利用して音声回復を行う方法がある．また，伝達特性を適時学習しながら適応フィルタリングにより音声回復を行う手法も提案されている [10]．逆フィルタリング法と呼ばれるこれらの方法 [7]，[8]，[9]，[10] は，骨導音声の周波数成分を回復させるが，同時にエコーといったアーティファ

クトを生み出してしまう問題点がある上，気導音声の情報を必要とするため，ブラインド処理ではない．高騒音環境下では気導音声を録音することが非常に難しいため，骨導音声回復法は，ブラインド処理であることが必要である．

一方，鵜木らは，気導音声と骨導音声の間の変換関係を伝達特性とみなし，気導・骨導音声を同時収録した大規模データベースを用いて変換関係の解析を行い，骨導音声の音質・明瞭度の改善方法を検討してきた [11]〜[15]．彼らの研究のコンセプトは，音源フィルタモデルを仮定し，線形予測分析と変調伝達関数 (MTF) の二つの側面から分析を行ことである．これらの解析の結果，鵜木らは音源信号ではなくフィルタ情報の回復が骨導音声の回復に重要であることを明らかにした．また，これに基づき 2 つの手法が提案されている．一つは，Vu らが提案した周波数領域での線形予測分析に基づく骨導音声回復法である [11]．この手法は，ブラインド処理を実現している．しかし，この手法は学習の過程を必要としており，回復精度は学習に依存してしまうため，多様な環境に対処し難い．また，骨導音声を利用して音声コミュニケーションを行うことを考えた際，最も重要視すべきは音声明瞭度であり，音声明瞭度を直接回復できる手法が望まれる．もう一つは，木村らが提案した時間領域での MTF に基づく骨導音声回復法である [12]．この手法は，音声明瞭度と関係のある MTF に基づいており，音声明瞭度を直接回復することができ，骨導音声を利用して音声コミュニケーションを行うために非常に有効な手法である．よって，本研究では変調伝達関数に基づいた手法に着目した．しかし，この手法は，骨導音声回復に気導音声の情報を必要とするためブラインド処理になっていない．

1.3 MTF に基づく骨導音声回復法

1.3.1 変調伝達関数

円滑に音声コミュニケーションを行うためには音声明瞭度を高く保つ必要がある．音声明瞭度を定量的な物理指標から予測する MTF に基づく音声明瞭度予測理論が，Houtgast と Steeneken により提案されている [16]〜[18]．これは，音声の時間エンベロープを入出力とした時の伝達関数である MTF の変調度から音声明瞭度を予測する方法である．MTF を音声伝達指数 (STI) に変換することで，MTF と音声明瞭度が直接関係づけられる [19]．Houtgast と Steeneken は入力パワーエンベロープ $e_x^2(t)$，出力パワーエンベロープ $e_y^2(t)$ を以下の式で定義した．

$$e_x^2(t) = \overline{I_x^2}(1+\cos(2\pi f_m t)), \tag{1.1}$$

$$e_y^2(t) = \overline{I_y^2}\{1+m(f_m)\cos(2\pi f_m(t-\tau))\} \tag{1.2}$$

ここで，$\overline{I_i^2}$ と $\overline{I_o^2}$ は，入出力の強度，f_m は変調周波数，τ は位相である．また，$m(f_m)$ が MTF である．この MTF は，以下の式で定義されている．

$$m(f_m) = \frac{\left|\int_0^\infty h^2(t)\exp(-j2\pi f_m t)dt\right|}{\int_0^\infty h^2(t)dt} \tag{1.3}$$

1.3.2 パワーエンベロープ逆フィルタ法

Drullman により，音声明瞭度に最も重要な影響を与えるのは時間エンベロープであるということが示されている [20]．このことから，MTF に着目し，変調度を回復させる方向にパワーエンベロープを回復させることで音声の明瞭度を直接回復できる可能性がある．この点に着目し，広林らは残響の影響を受けた音声の明瞭度を回復させる手法としてパワーエンベロープ逆フィルタ法を提案した [21]．MTF の逆フィルタ $E_h^{-1}(z)$ は以下の様に表現される．

$$E_h^{-1}(z) = E_x(z)/E_y(z) \tag{1.4}$$

ここで，$E_h(z)$，$E_x(z)$，$E_y(z)$ はそれぞれ $e_h^2(t)$，$e_x^2(t)$，$e_y^2(t)$ の z 変換，$e_h^2(t)$ は，入出力をパワーエンベロープとしたときのシステムのインパルス応答である．パワーエンベロープ逆フィルタ法では，逆フィルタは Schroeder の確率論的近似インパルス応答 [22] を用いて以下の式で定義されている．[12]

$$e_h(t) = a\exp(-6.9t/T_R) \tag{1.5}$$

ここで，a はゲインを制御するパラメータ，T_R は残響時間である．a，T_R については，以下の式で求められる．

$$a = \sqrt{\frac{1}{T_R}\int_0^T e_x^2(t)/e_y^2(t)dt \Big/ \int_0^T \exp(-13.8t/T_R)dt} \tag{1.6}$$

$$\hat{\mathrm{T}_\mathrm{R}} = \max\left(\underset{0\le \mathrm{T_R} \le \mathrm{T_{R,max}}}{\mathrm{argmin}} \left\{\int_0^{\mathrm{T}} \min(\hat{\mathrm{e_x}}, \mathrm{T_R^2(t)}, 0)\mathrm{dt}\right\}\right) \tag{1.7}$$

ここで，$\hat{e_x}, T_R^2(t)$ は T_R を関数として回復されたパワーエンベロープ、$T_{R,max}$ は T_R の上限，$\hat{T_R}$ は残響時間の推定値である．T_R の推定式については古川らによって提案された方法である [23]．これは，パワーエンベロープが負の値を持たないこと，また無音区間のパワーエンベロープの値が 0 であることに着目した手法である．パワーエンベロープ逆フィルタ法はパワーエンベロープの山谷を強調させるため，T_R の値を大きくとっていくと，パワーエンベロープが負の値を持つようになる．つまり，パワーエンベロープが負の値を持つ直前の T_R が最適な T_R の推定値となる．パワーエンベロープ逆フィルタによる回復は以下の式で表される．

$$\hat{E}_x(z) = \frac{E_y(z)}{a^2}\left\{1 - \exp\left(-\frac{13.8}{T_R \cdot f_s}\right)\right\} z^{-1} \tag{1.8}$$

ここで，$\hat{E}_x(z)$ は逆フィルタにより回復されたパワーエンベロープである．

1.3.3 MTF に基づく骨導音声回復法

木村らは，Drullman の考えに基づき，骨導音声の明瞭度を回復するのには時間エンベロープの回復が重要であると考えた．解析を行った結果から，気導パワーエンベロープ

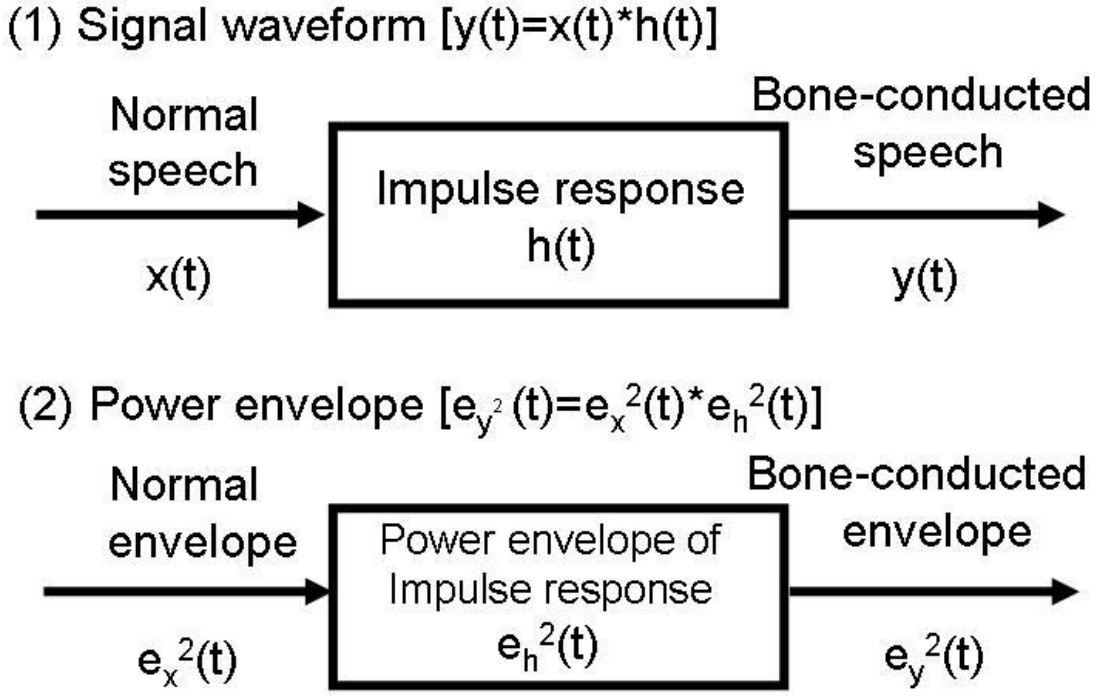

図 1.1: 気導音声と骨導音声の間の伝達特性の定義．(1) 波形レベル間の伝達特性，(2) パワーエンベロープ間の伝達特性.

と骨導パワーエンベロープの間の MTF はローパス特性であることを明らかにし，気導パワーエンベロープと骨導パワーエンベロープの間の伝達特性を図 1.1 のように定義することでパワーエンベロープ逆フィルタ法を骨導音声の回復に適応した．ここで，$e_h(t)$ はパワーエンベロープ逆フィルタ法で用いた $e_h(t)$ と同じものである．骨導音声回復の概要を図 1.2 に示す．音声は N チャンネル等帯域フィルタバンクによって時間エンベロープ $e_x(t)$ と $e_y(t)$，キャリア $c_x(t)$ と $c_y(t)$ に分割される．ここで，信号 $x(t)$ と $y(t)$ は以下のように表現される.

$$x(t) := \sum_{n=1}^{N} x_n(t) = \sum_{n=1}^{N} e_{x_n}(t) \cdot c_{x_n}(t) \tag{1.9}$$

$$y(t) := \sum_{n=1}^{N} y_n(t) = \sum_{n=1}^{N} e_{y_n}(t) \cdot c_{y_n}(t) \tag{1.10}$$

ここで，$x_n(t)$ と $y_n(t)$，$e_x(t)$ と $e_y(t)$，$c_x(t)$ と $c_y(t)$，はそれぞれフィルタバンクにより帯域分割された信号，時間エンベロープ，キャリアである．パワーエンベロープは以下の式

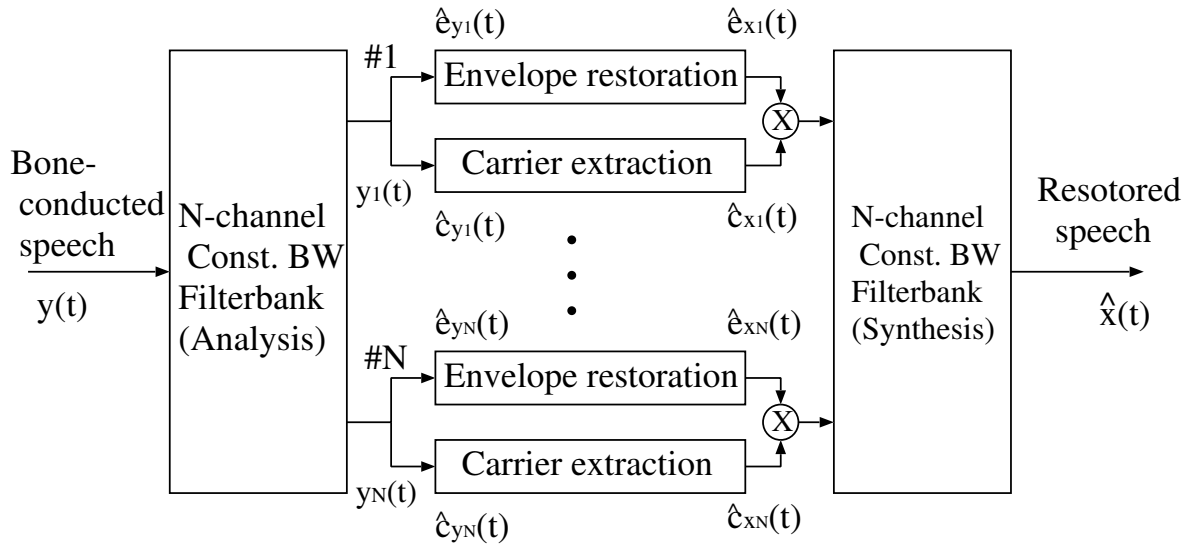

図 1.2: 変調フィルタバンクを用いた MTF に基づく骨導音声回復法の概要.

で算出される.

$$e^2_{y_n}(t) = \mathrm{LPF}[|y_n(t) + j\mathrm{Hilbert}(y_n(t))|^2] \tag{1.11}$$

$$c_{y_n}(t) = y_n(t)/e_{y_n}(t) \tag{1.12}$$

ここで，Hilbert(·) はヒルベルト変換を，LPF[·] はカットオフ周波数 20-Hz[24]〜[26] のローパスフィルタを表す．$e_x(t)$ と $c_x(t)$ も同様の方法で算出される．以降はパワーエンベロープ逆フィルタ法の流れと同様である．ここで，木村らは解析結果より気導パワーエンベロープと骨導パワーエンベロープ間の相関係数が 0.8 以上かつ気導パワーエンベロープの相対パワーが −20 dB 以内の時のみ逆フィルタによる骨導音声回復を行い，それ以外の範囲で気導パワーエンベロープの相対パワーが −40 dB の範囲まではパラメータ a を用いたゲイン補正のみを行うと回復条件を定めた.

1.3.4 残された課題

MTF に基づく骨導音声回復法では，MTF は Schroeder の確率論的近似インパルス応答を用いて表現されているが，この形で表現される MTF が骨導音声の回復にパワーエンベロープ逆フィルタ法を応用する際に適切かどうかの議論はなされていなかった．また，パラメータ a を導出するためには気導音声の情報を必要とするため，この手法はブラインド処理になっていない．気導音声の収録が難しい状況で骨導音声を利用する事を考えた際，これは大きな問題である．この 2 点を解決しない限り，実環境での MTF に基づく骨導音声回復は行なうことができない.

1.4　研究の目的

骨導音声を利用した音声コミュニケーションを実現するには，ブラインド処理で骨導音声の音声明瞭度を回復させなければならない．そのような骨導音声回復法は今までに提案されていない．そこで，本研究は，MTF に基づくブラインド骨導音声回復法の提案を目指す．MTF の概念に基づくことにより，骨導音声の音声明瞭度を直接回復することを可能とする．気道パワーエンベロープと骨導パワーエンベロープの間の変換関係の解析を行うことにより，変換関係を表現する最適な MTF を明らかにし，その逆特性を利用した逆フィルタにより骨導音声の回復を行う．また，逆フィルタは話者や発話内容に依存せず，気導音声の情報を必要としないように設計し，手法をブラインド処理にする．

1.5　本論文の構成

第 2 章では，気導パワーエンベロープと骨導パワーエンベロープ間の変換関係の解析を行い，解析の結果に基づいて MTF を表現できる最適なモデルの提案を行う．第 3 章では第 2 章で述べたモデルのパラメータを，気導音声の情報を必要とせずに決定する方法を述べ，MTF に基づく骨導音声回復法をブラインド処理に改良する．第 4 章では，提案法の評価を行う．最後に，第 5 章では，まとめと今後の展望を記す．

第2章　骨導/気導パワーエンベロープ間の変換特性

MTF に基づき骨導音声の回復を行うには，適切な形の逆フィルタをどのように設計するかという問題がある．本研究では，気導パワーエンベロープと骨導パワーエンベロープ間の変換特性の解析を行い，気導パワーエンベロープと骨導パワーエンベロープ間の関係を MTF モデルで表現し，適切な逆フィルタの設計を行う．

2.1　気導/骨導データベース

本研究では，気導パワーエンベロープと骨導パワーエンベロープ間の変換特性の解析を行うため，気導/骨導データベースを用いる [13]．表 2.1 にデータベースの構築に使用した機材，図 2.1 に収録環境を示す．データベースに収録されている音声は，5 つの観測点（1: 下顎横，2: こめかみ，3: 頬骨，4: 額，5: 頭頂部）で収録された．観測点 1〜4 はマイク C，観測点 5 はマイク B，気導音声はマイク A を用いて収録されている．発話内容は NTT データベース [27] から 4 つの親密度 [28] 毎に 25 単語ずつ選ばれた．話者は男性女性各 5 名である．

2.2　骨導/気導エンベロープ間の変換特性の解析

気導パワーエンベロープと骨導パワーエンベロープ間の変換特性の解析を行う際に，以下の項目に着目した．

- 気導/骨導パワーエンベロープ間の相関係数

$$\mathrm{Corr}(e_x^2, e_y^2) = \frac{\int_0^T (e_x^2(t) - \overline{e_x^2(t)})(e_y^2(t) - \overline{e_y^2(t)})dt}{\sqrt{\left\{\int_0^T (e_x^2(t) - \overline{e_x^2(t)})^2 dt\right\}\left\{\int_0^T (e_y^2(t) - \overline{e_y^2(t)})^2 dt\right\}}} \tag{2.1}$$

表 2.1: 気導/骨導音声の収録条件.

Measurement site	Soundproof room
Number of pick-up points	5
Number of speakers	10
Recorder	MARANZ, PMD671
Coding method	PCM
Sampling frequency	48 kHz
Sample size	16 bits
Number of channels	2 (Left:AC, Right:BC)
Mic. A for AC speech	SONY, C536P
Mic. power supply A	SONY, AC148F
Mic. B for BC speech	TEMCO, HG-17
Mic. C for BC speech	TEMCO, SK-1
Mic. amp. B & C	Handmade
Speakers (4 set)	JBL, CM62

- 気導/骨導パワーエンベロープ間の SNR

$$\mathrm{SNR}(e_x^2, e_y^2) = 10\log_{10}\frac{\int_0^T (e_x^2(t))^2 dt}{\int_0^T \left(e_x^2(t) - e_y^2(t)\right)^2 dt} \tag{2.2}$$

- MTF

$$M(\omega) = \left| \int e_h^2(t)\exp(-j\omega t)dt \Big/ \int e_h^2(t)dt \right| \tag{2.3}$$

- 気導/骨導パワーエンベロープ間のパワー比（従来法のパラメータ a）

$$a = 10\log_{10}\left(\int_0^T e_y^2(t) / \int_0^T e_x^2(t)dt\right) \tag{2.4}$$

- 気導/骨導音声間の伝達関数

$$H(\omega) = \mathcal{F}[y(t)] / \mathcal{F}[x(t)] \tag{2.5}$$

ここで，$\mathcal{F}[\cdot]$ は長時間フーリエ変換である．先行研究で，木村らにより観測点 5 で収録された音声についての解析がすでに行われている．本研究では，全観測点について解析を行う．これにより，観測点毎の骨伝導の影響の差が明らかとなる．

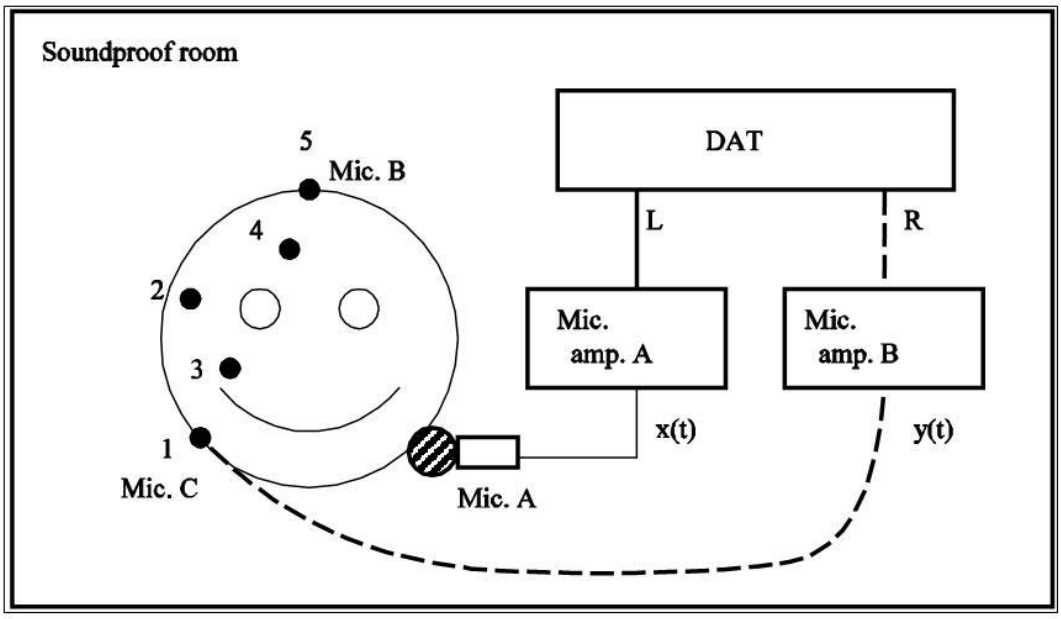

図 2.1: 気導/骨導データベース構築の際の音声の収録環境 (数字 1〜5 は観測点).

2.3 解析結果の考察

全観測点での解析結果を図 2.2 に示す．また，観測点毎に骨導音声の性質が変わることが先行研究により明らかとなっているため，各観測点毎の結果を図 2.3〜2.7 に示し，観測点毎の差について考察する．ここで，(a) 相関係数，(b)SNR，(c)MTF の回帰直線の傾き，(d) 気導音声と骨導音声の伝達関数，(e) 骨導パワーエンベロープと気導パワーエンベロープの平均パワーのパワー比（パラメータ a），(f) 骨導パワーエンベロープの平均パワーを表し，実線は平均，破線は平均 $\pm$ 標準偏差を示す．また，(a)，(c) の図において，1〜10 チャンネルの値が他のチャンネルと比較して大きく異なった値を持っているのは，振幅変調の定義を満たさない範囲であるため，エンベロープが上手く抽出できていないためである．図 2.2 の（a）と（b）及び（f）から，音声のパワーが低い帯域では相関も低いという傾向が見て取れる．各観測点毎に見ていくと，観測点 1 と 5 は高周波数成分があまり上手く録音できておらず，観測点 2，3，4 は高周波数成分がよく録音できているのがわかる．図 2.2 の（c）は 1〜10 Hz までの範囲の MTF に対して回帰直線を引き，その傾きをプロットしたものである．MTF の範囲を 1〜10 Hz と限定したのは，10 Hz 以上の範囲の MTF は血流や伝送系の持つ雑音，あるいはノイズフロア，その他録音時の外乱 [29] といった内部雑音の影響を受けるためである．ここで，MTF の回帰直線の傾きが正であれば MTF はハイパス特性，負であればローパス特性であることを意味している．図 2.2 の（f）と見比べると，骨導パワーエンベロープの相対パワーが −30〜−40 dB 以内の範囲まで MTF はローパス特性を示している．骨導パワーエンベロープの相対パワーが −40 dB 以下になると相対的に内部雑音が大きくなり，パワーエンベロープの形状に大きな影響を

表 2.2: 骨導パワーエンベロープと気導パワーエンベロープのパワー比に対する観測点毎の近似曲線のパラメータ.

	観測点 1	観測点 2	観測点 3	観測点 4	観測点 5
パラメータ　c	-17.5	-17.1	-15.8	-11.8	-13.8
パラメータ　d	8.54	7.98	7.90	6.74	9.48

与えるため，骨導パワーエンベロープの相対パワーが −40 dB 以下になると MTF の回帰直線の傾きが正の値になる傾向がある．さらに骨導パワーエンベロープの相対パワーが減少すると，MTF の 0 Hz の成分（直流成分）が内部雑音の影響で増加するため，再び MTF の回帰直線の傾きが負の値を持ち始めると考えられる．考察の結果，内部雑音の影響がなければ MTF はローパス特性であると示唆される．観測点毎に見ても，この傾向は変わらずに見られる．図 2.2 の（d）は骨伝導の影響がローパス特性であることを示している．観測点毎に見ると，細かい傾向は違うものの骨伝導の影響はローパス特性であることに変わりはなかった．図 2.2 の（e）は骨導パワーエンベロープと気導パワーエンベロープの平均パワーのパワー比であり，逆フィルタ法に用いられているパラメータ a である．これも，伝達関数と同じく骨伝導の影響がローパス特性であることを表している．各観測点毎に見ていくと，音声の高周波数成分を比較的よく録音できる観測点 2 と 3 は平均の値の形が非常に似ていることがわかる．また，観測点 1 と 4 については，低域側で多少の誤差があるものの，観測点 2 と 3 と平均の値が同じような傾向である．観測点 5 は，もっとも音声の高周波数成分を録音できていない箇所であり，また録音に使用したマイクも異なるため，他 4 つの観測点と少し誤差が大きいものの，平均の値の形状は他の 4 点と似ている．この結果から，骨導パワーエンベロープと気導パワーエンベロープの平均パワーのパワー比（パラメータ a）は，回帰曲線で近似することができるのではないかと考えられた．(e) の図に点線で記されている曲線が，$1/a_n^2 = cn^{-1} + d$ というフィルタバンクのチャンネル数を従属変数とする関数でパワー比の平均を近似したものである．各観測点毎のパラメータ c と d の値を表 2.2 に示す．図 (e) から，回帰曲線はパワー比の平均と非常によくフィットしていることがわかる．以上の考察から，骨伝導の影響はローパス特性であること．観測点 2，3，4 は音声の高周波成分を比較的良好に録音することができ，骨導マイクで音声を収録する場合に良い点であること．骨導パワーエンベロープと気導パワーエンベロープ間の MTF の特性は全帯域においてローパス特性であるということ．骨導パワーエンベロープと気導パワーエンベロープのパワー比は $cn^{-1} + d$ という関数で近似可能である事が示唆された.

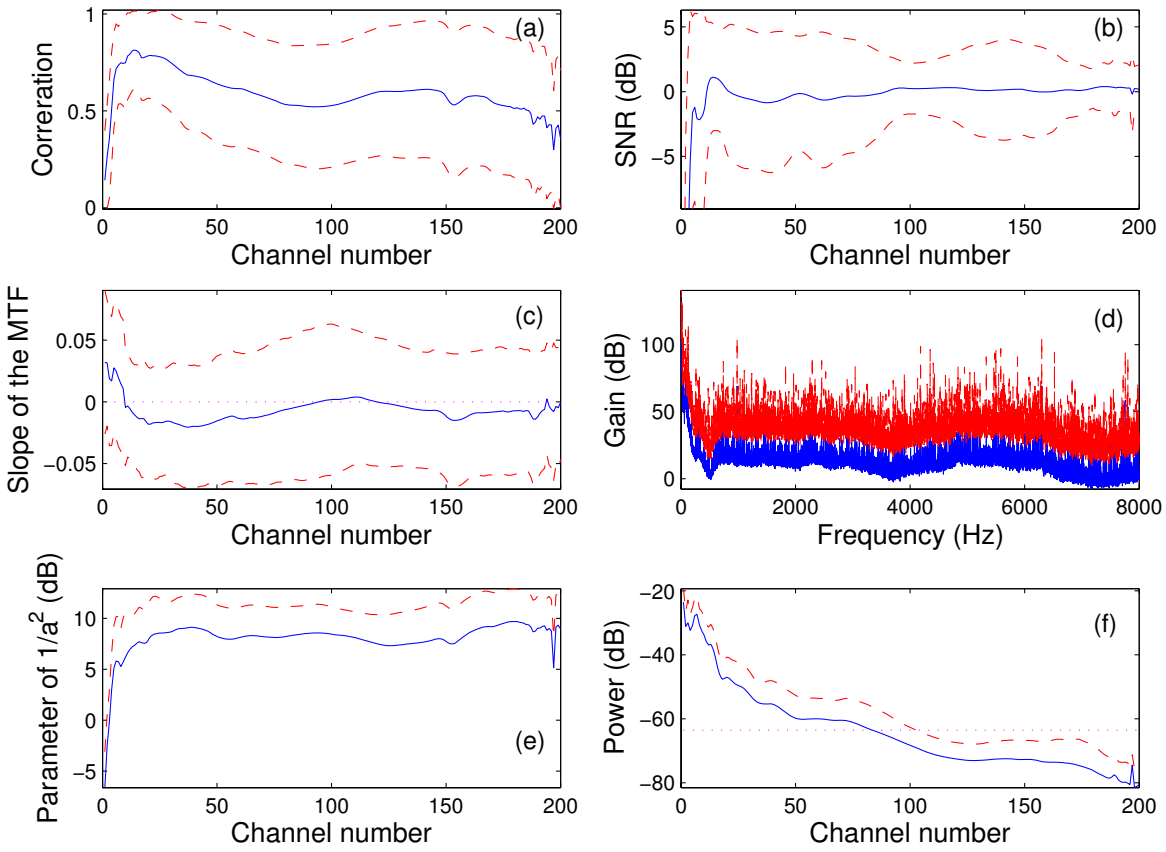

図 2.2: 全観測点での解析結果 （実線: 平均，破線: 平均 ± 標準偏差）．(a) 相関係数，(b) SNR，(c) MTF の回帰直線の傾き，(d) 伝達関数，(e) パワーエンベロープの平均パワーの比（パラメータ $1/a_n^2$），(f) 各チャンネル毎の $e_y^2(t)$ の平均（点線は相対パワーが-40 dB 下がった位置を表す）．

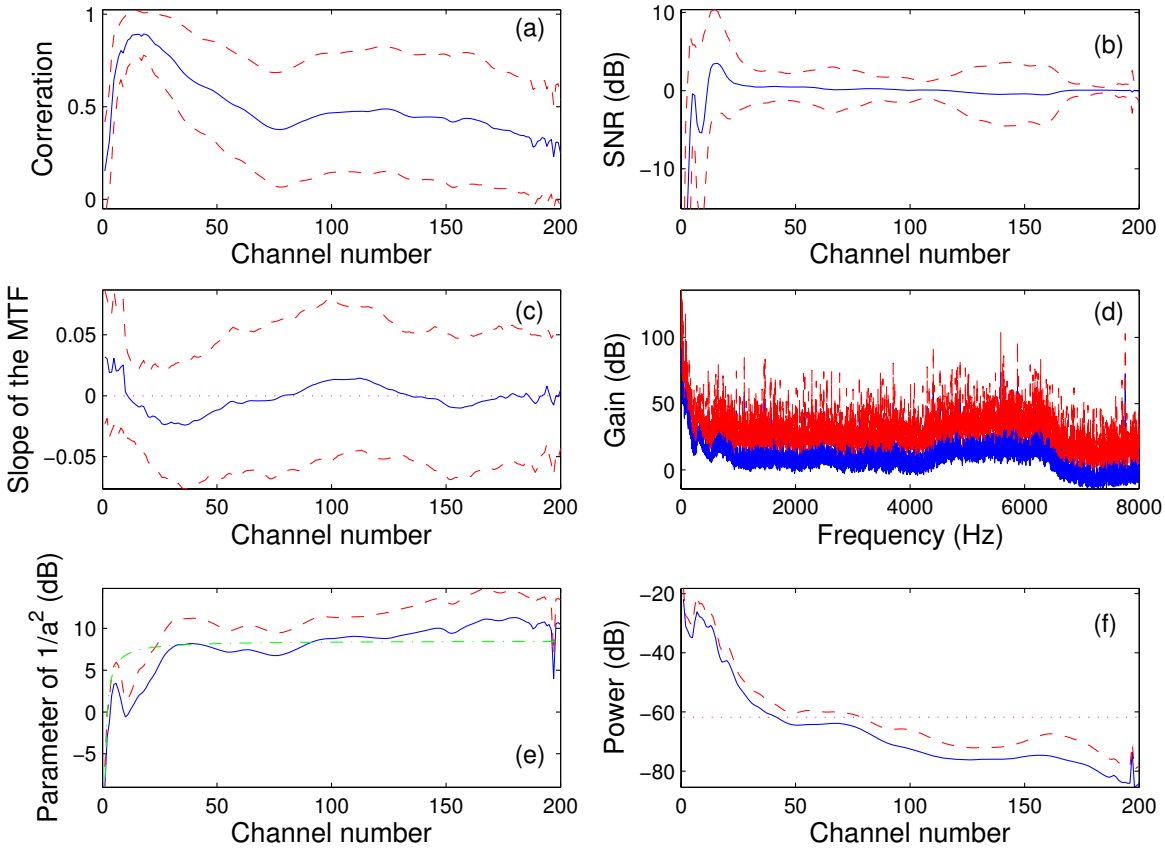

図 2.3: 観測点 1 での解析結果．体裁は図 2.2 と同じ．

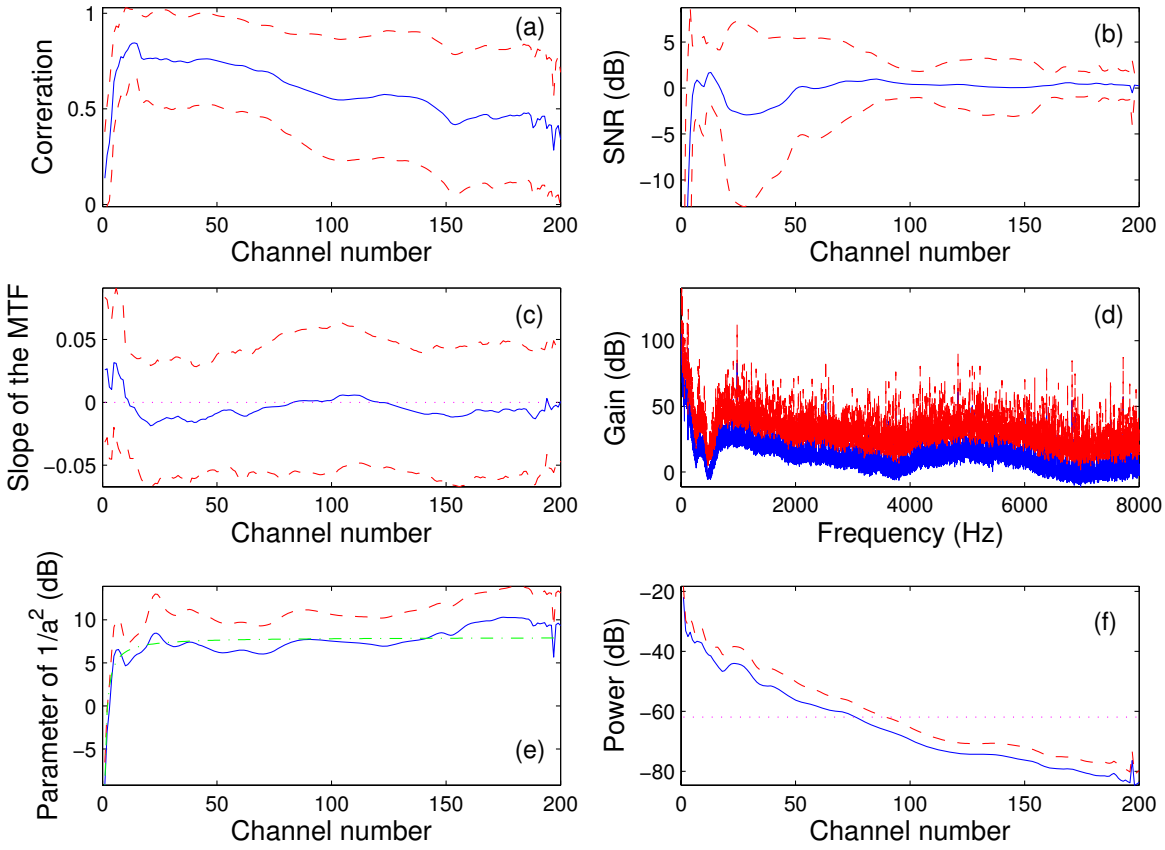

図 2.4: 観測点 2 での解析結果．体裁は図 2.2 と同じ．

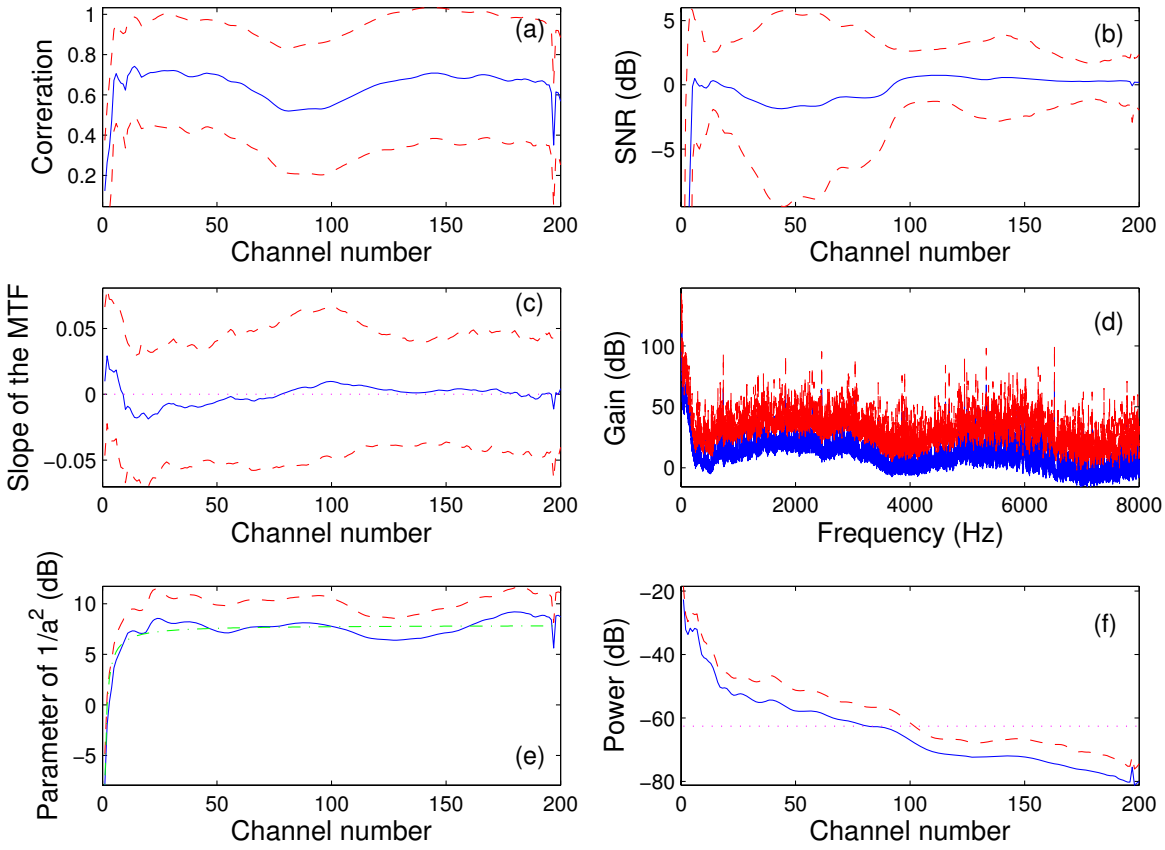

図 2.5: 観測点 3 での解析結果．体裁は図 2.2 と同じ．

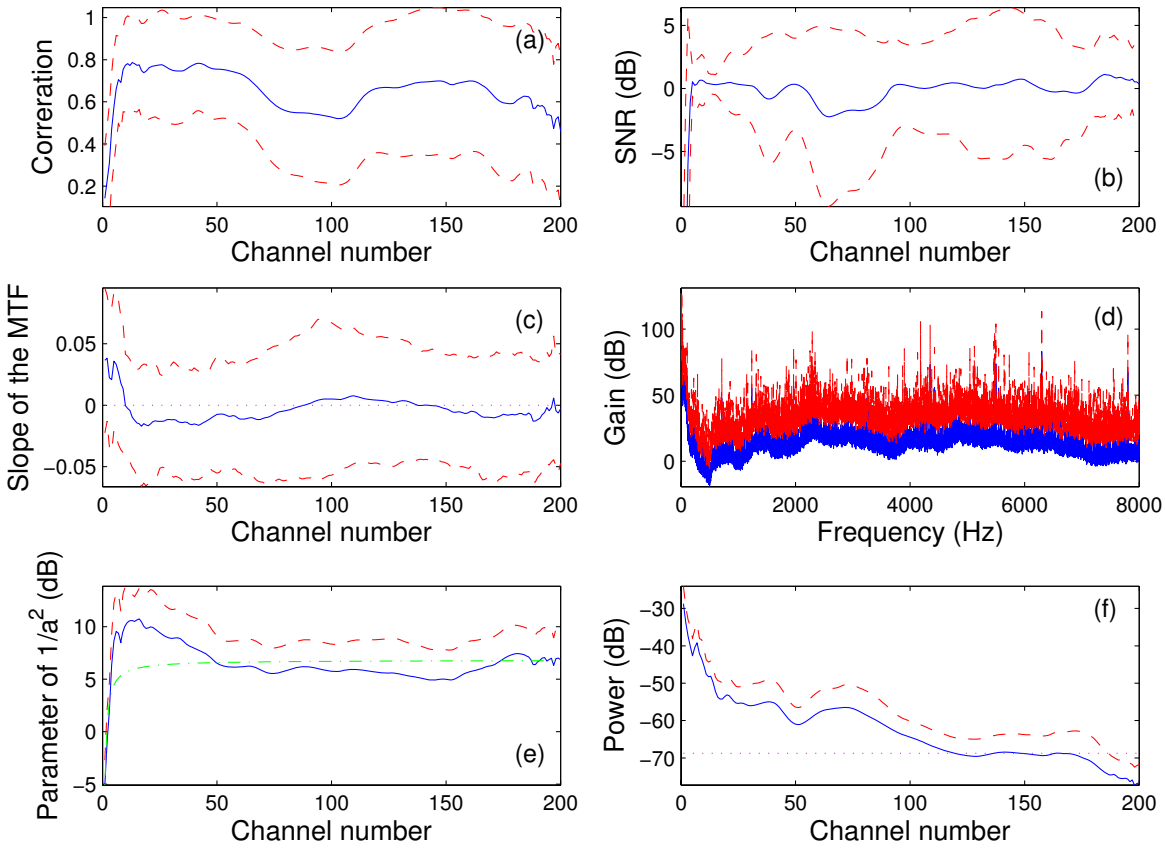

図 2.6: 観測点 4 での解析結果．体裁は図 2.2 と同じ．

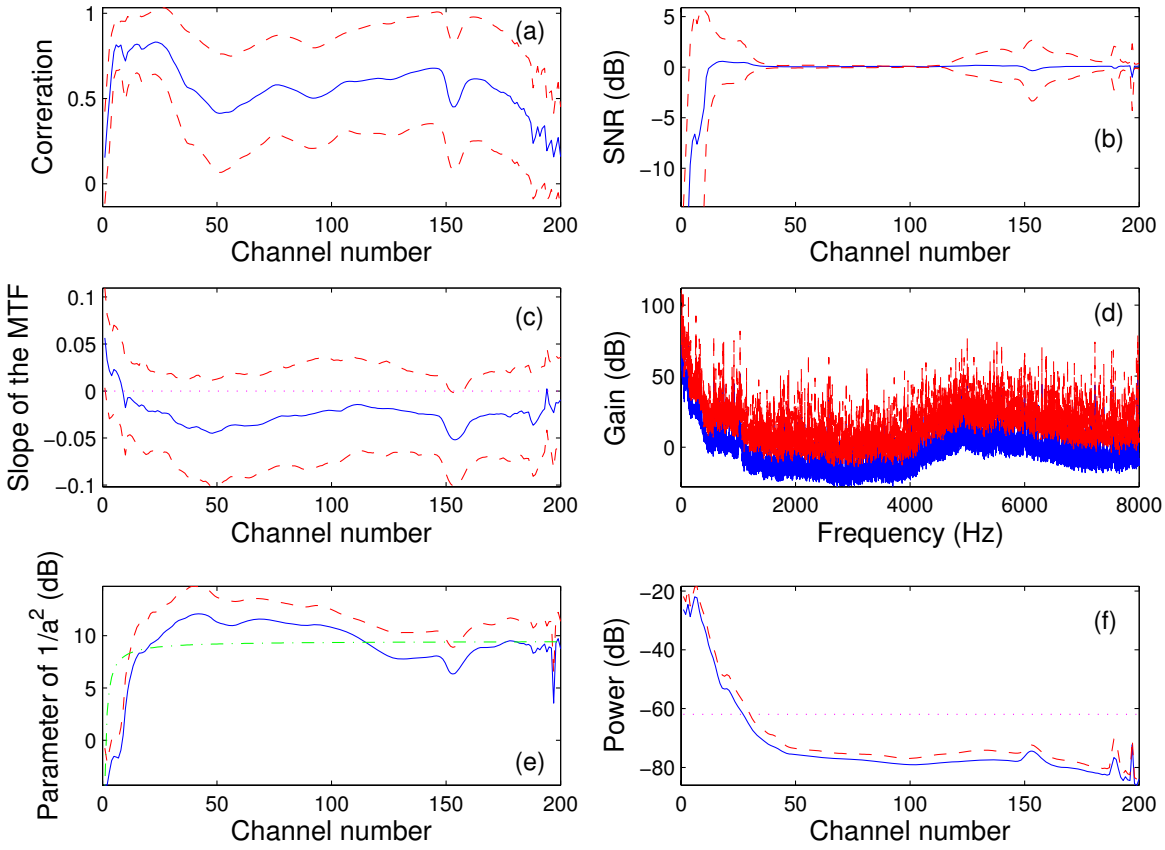

図 2.7: 観測点 5 での解析結果．体裁は図 2.2 と同じ．

2.4 気導パワーエンベロープと骨導パワーエンベロープ間のMTFのモデリング

気導パワーエンベロープと骨導パワーエンベロープ間のMTFを表現するのに最適なモデルが明らかとなっていないため，本研究では，実際のMTFを表現する最適なモデルについての考察を行う．前節での解析結果から気導パワーエンベロープと骨導パワーエンベロープの間のMTFがローパス特性であることが示唆されたため，MTFを表現するのに適切であろうと思われるローパス特性を持った3つのモデル (指数関数 $e_h(t) = at\exp(-bt)$，先行研究で用いられているモデル $e_h(t) = a\exp(-bt)$，ローパスフィルタ) を，データベースの音声から求めたMTFに対してTrust region法と共役勾配法を用いてフィッティングをかけることにより，MTFを表現する最適なモデルを求めた．図2.8は，データベースの音声から求めたMTF，骨導音声から内部雑音を除去したMTF，フィッティングを行った3つのモデルを示したものである．事前に行った解析の結果より，MTFの特性はローパス傾向にあることが分かっている．また，先行研究で利用されているMTFモデルは指数関数表現である事から，この3つのモデルを採用した．データベースの音声から求めたMTFは形状が非常に波打っている．これは，骨導音声に内部雑音が乗っている影響である．内部雑音を除去したパワーエンベロープを図2.9に，この3つのパワーエンベロープをフーリエ変換し，実部と虚部にわけて表示したものを図2.10，2.11に示す．内部雑音を取り除くとMTFの形状の揺れが抑えられているため，MTFの形状が波打つ原因が内部雑音であることがわかる．図2.8をみると，$a\exp(-bt)$ が最もデータベースの音声から求めた内部雑音を除去したMTFにフィットしていることがわかる．データベース内の音声から求めたMTFに対し，データベースの音声から求めたMTFとモデルの各変調周波数毎の誤差のRMSが最小になるようにフィッティングを行った結果，$a\exp(-bt)$ が最もデータベースの音声から求めたMTFにフィットしていることわかった．図2.12に，$a\exp(-bt)$ のモデルとデータベースの音声から求めたMTFをフィッティングした際のRMSを全データに対して求めた結果を，図2.13〜2.17にフィッティングを行った際の，モデルの回帰直線の傾きを各観測点ごとに求めた結果の平均と標準偏差を示す．実線は平均，破線は平均 ± 標準偏差を示す．RMS誤差の標準偏差が大きいのは，MTFの回帰直線の傾きが正の場合や，内部雑音の影響によりMTFの形状が非常に大きく波打っている場合があるためである．また，10チャンネルまでのRMS誤差の平均が大きいのは，フィルタバンクが振幅変調の定義を満たさない範囲であるためである．本研究では，3つのモデルの中で最もRMS誤差の小さかった $a\exp(-bt)$ が，MTFを表現できる最も適したモデルとした．このモデルを用いた逆フィルタは以下の式で定義した．

$$E_h^{-1}(z) = \frac{1}{a^2}\left\{1 - \exp\left(-\frac{2b}{f_s}\right)\right\} \tag{2.6}$$

ここで，f_s はサンプリング周波数 (本研究では16 KHz) である．MTFの回帰直線の傾きを見ると，各観測点において気導パワーエンベロープと骨導パワーエンベロープのパワー

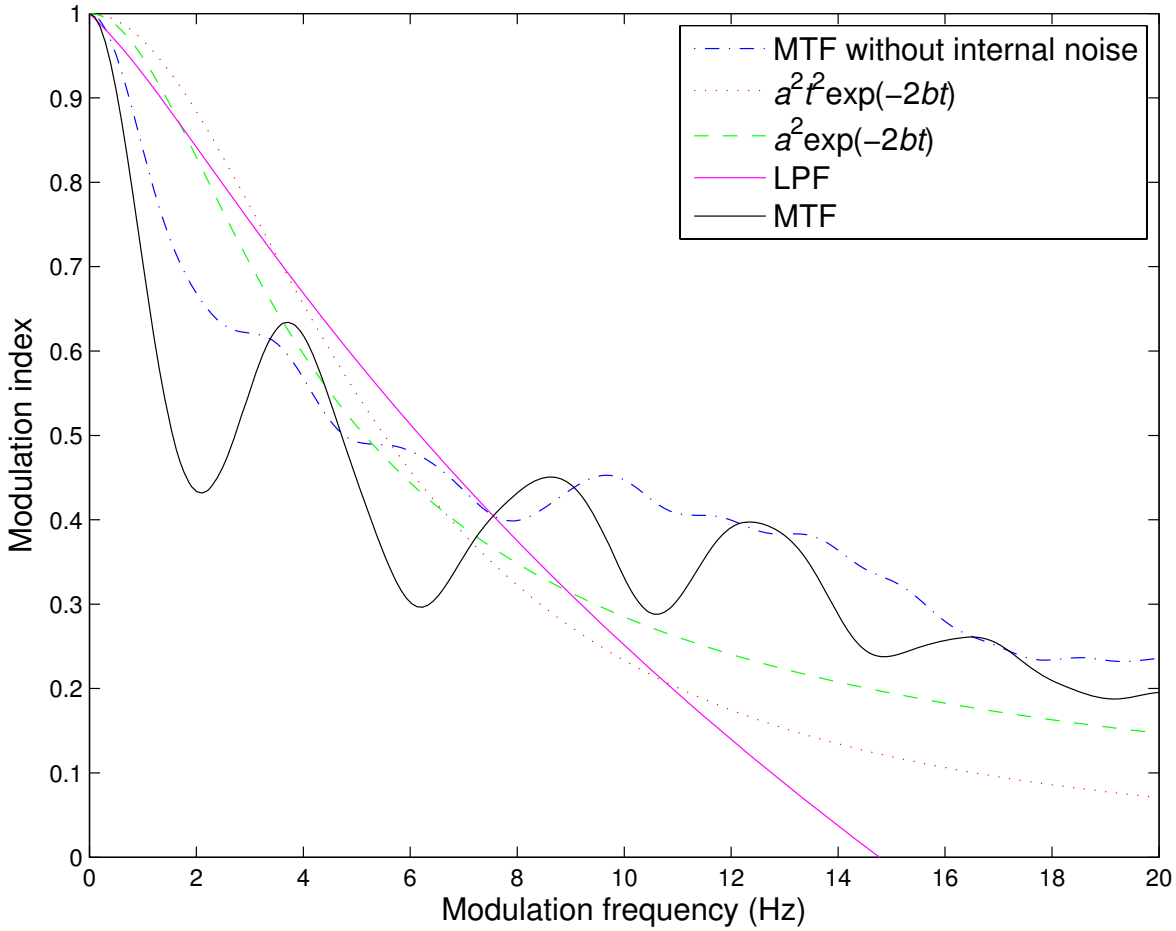

図 2.8: 実際の MTF とモデルの比較． MTF without internal noise:内部雑音を取り除いた MTF $e_h(t) = at\exp(-bt)$: 指数関数 $e_h(t) = a\exp(-bt)$: 先行研究で用いられているモデル LPF: ローパスフィルタ MTF:気導/骨導音声データベースのデータから求めたの MTF.

比と同様の傾向を示しており，骨導音声のパワーが減衰すればするほど，MTF はより大きくローパス傾向を示す事がわかる．

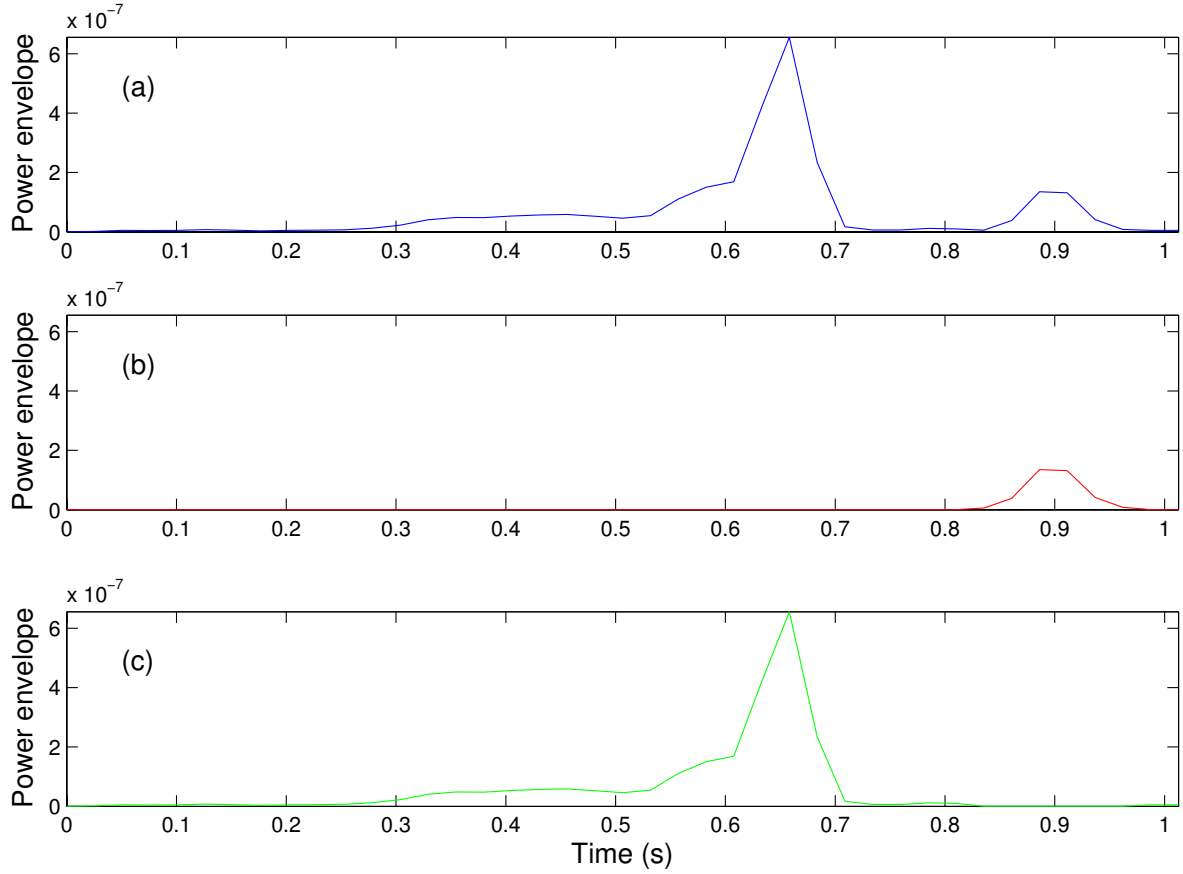

図 2.9: 内部雑音を除去したパワーエンベロープ. (a) 骨導音声 (b) 内部雑音 (c) 内部雑音除去後の骨導音声のパワーエンベロープ.

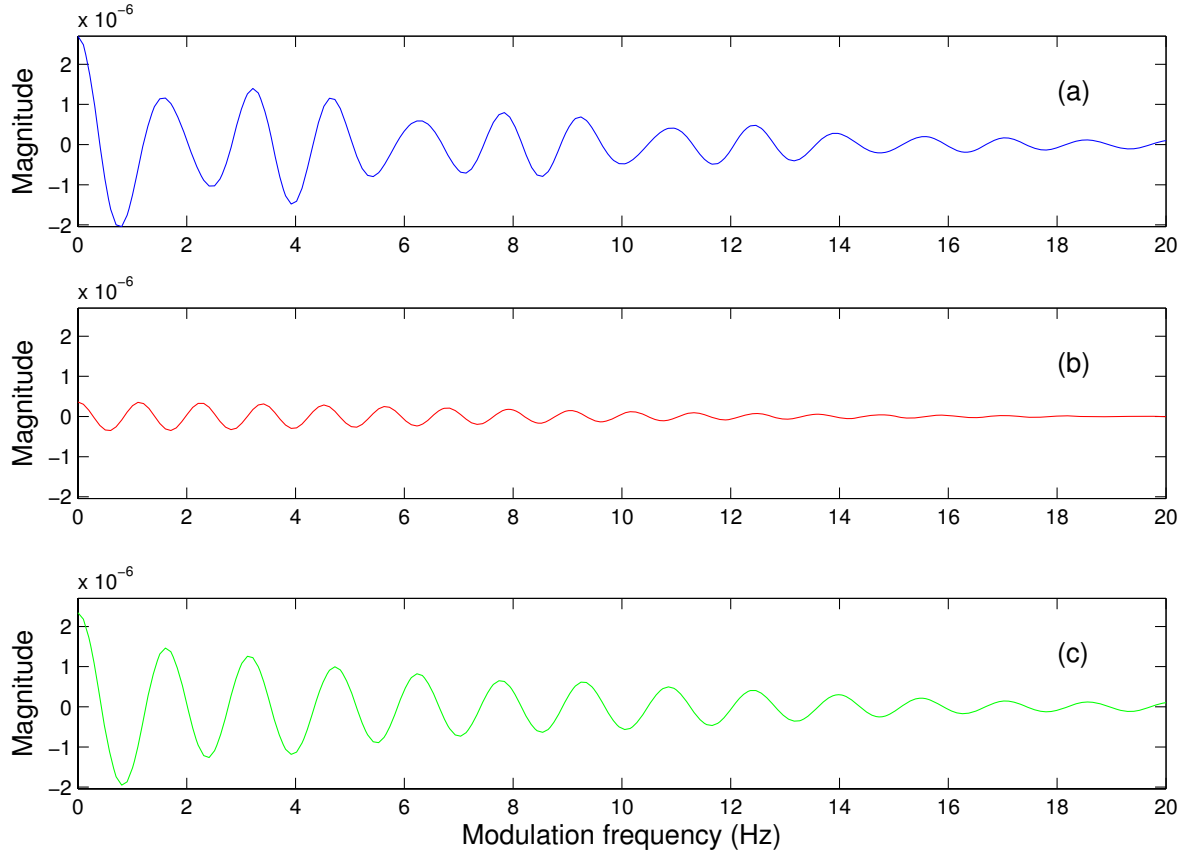

図 2.10: 内部雑音を除去した骨導音声の変調スペクトル (実部). (a) 骨導音声の変調スペクトル (b) 内部雑音の変調スペクトル (c) 内部雑音除去後の骨導音声の変調スペクトル.

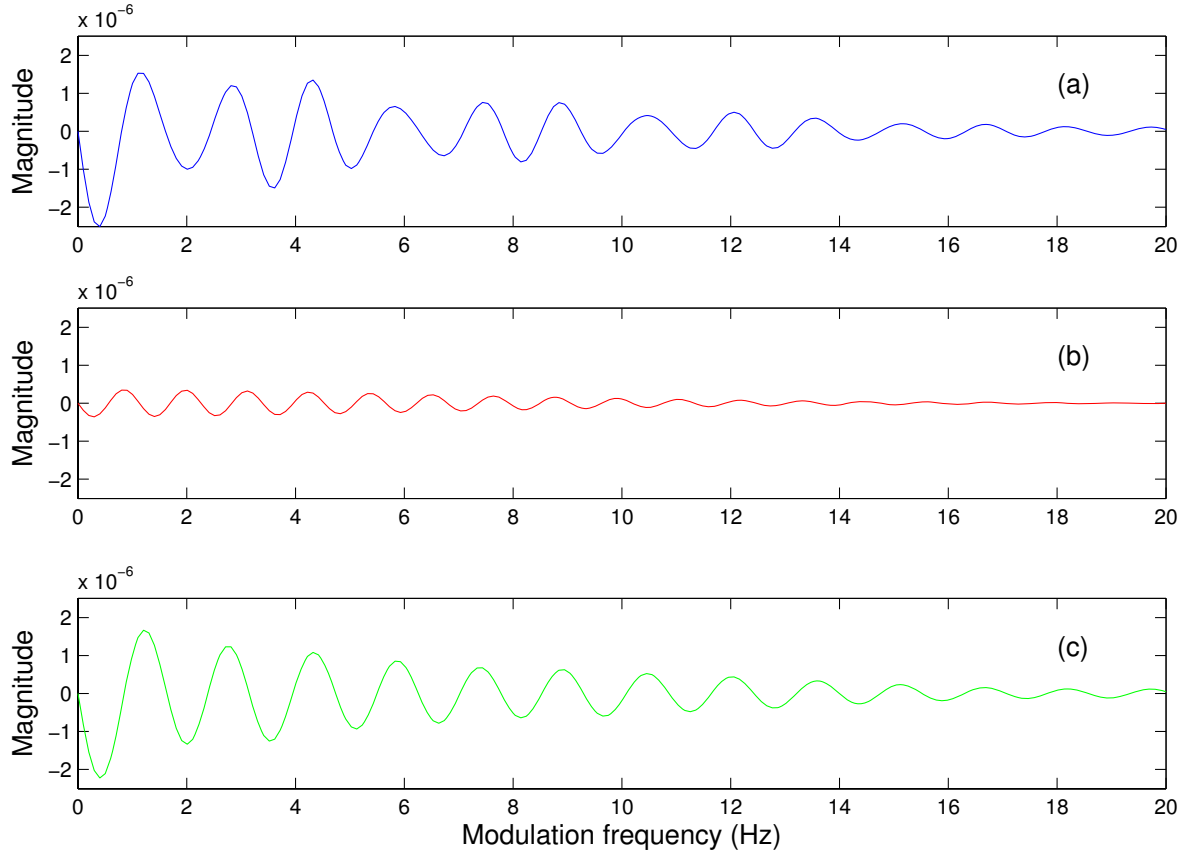

図 2.11: 内部雑音を除去した骨導音声の変調スペクトル (虚部). (a) 骨導音声の変調スペクトル (b) 内部雑音の変調スペクトル (c) 内部雑音除去後の骨導音声の変調スペクトル.

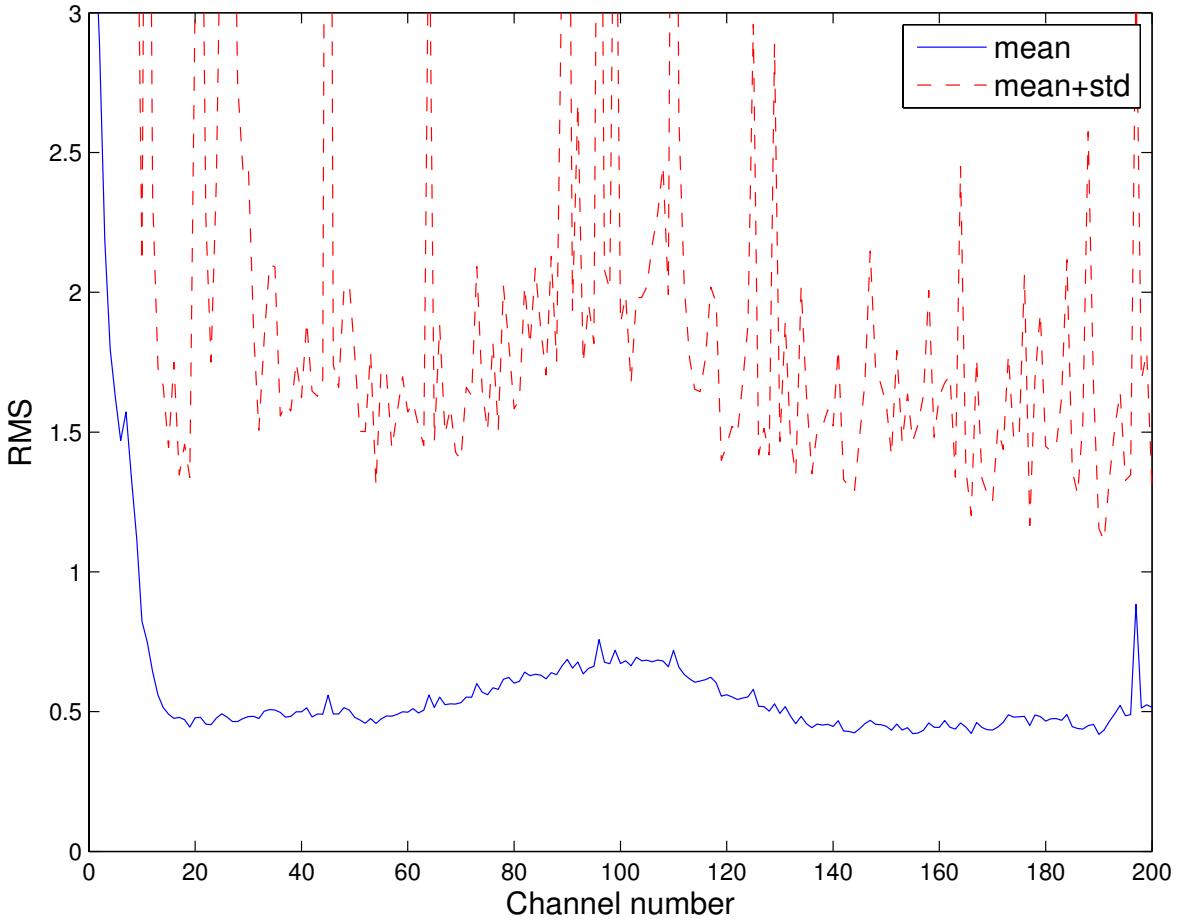

図 2.12: MTF を表現するのに最も適切なモデル $a^2 \exp(-2bt)$ と気導/骨導データベース内の全音声から求めた MTF との RMS 誤差の平均と標準偏差（実線: 平均 破線: 平均 $\pm$ 標準偏差 (std)）.

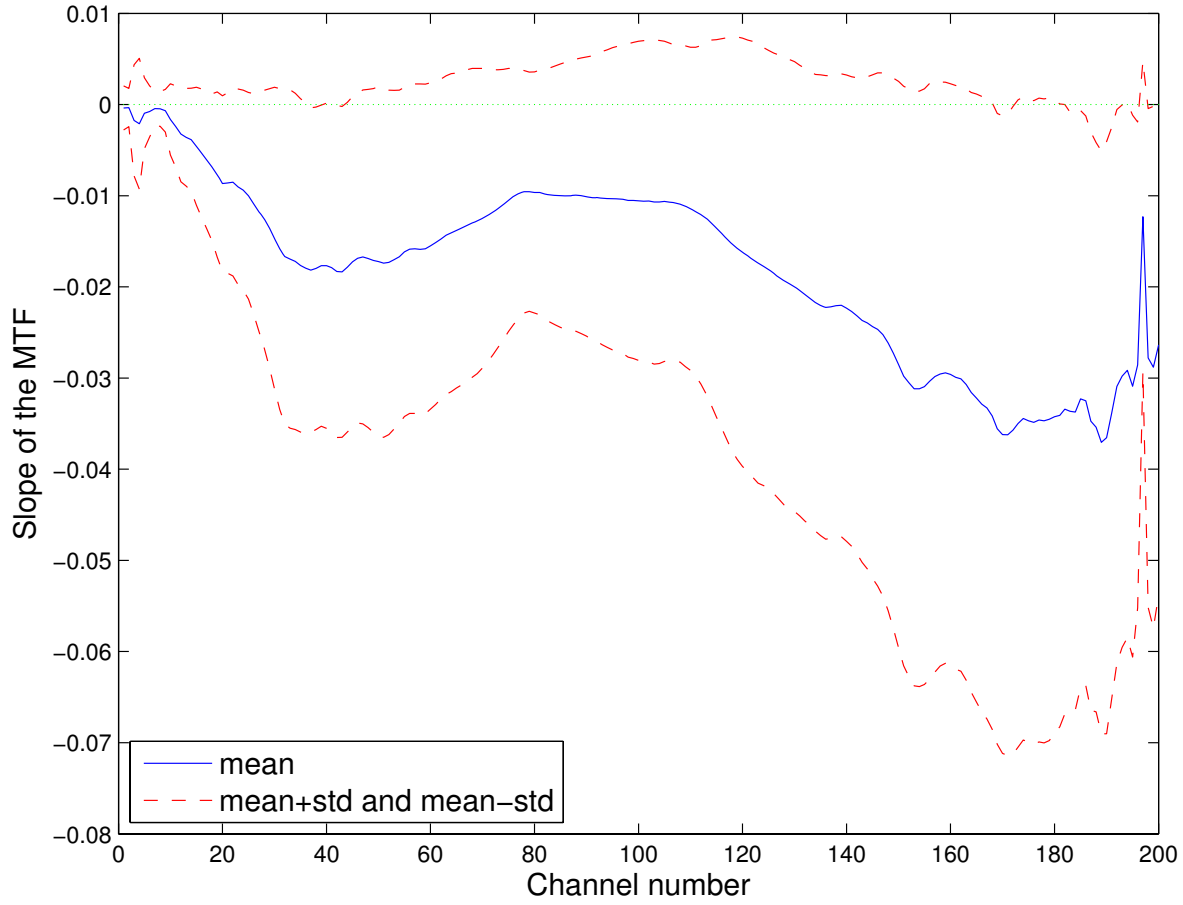

図 2.13: データベース内の音声から求めた MTF に，MTF を表現するのに最も適切なモデル $a^2 \exp(-2bt)$ をフィッティングした際のモデルの回帰直線の傾き（実線: 平均 破線: 平均 $\pm$ 標準偏差）(観測点 1).

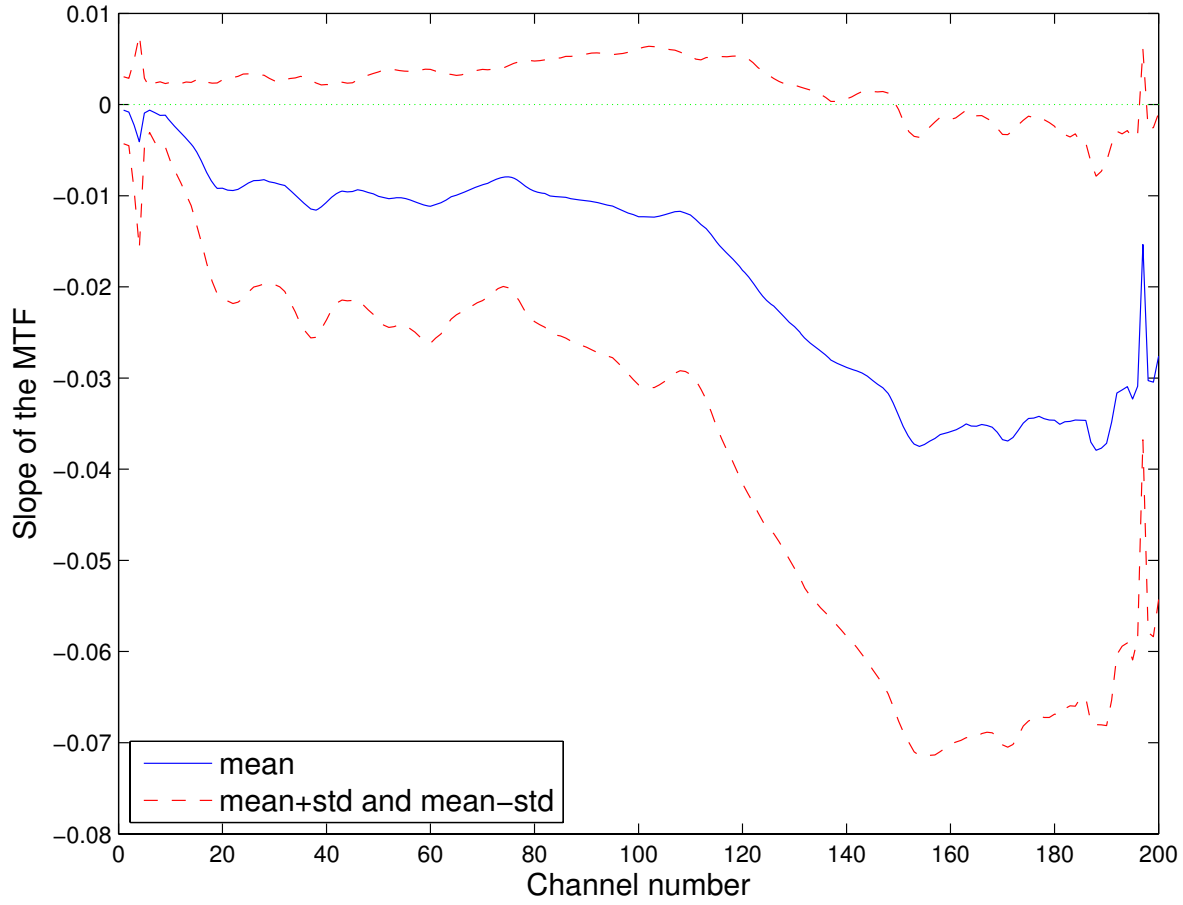

図 2.14: データベース内の音声から求めた MTF に，MTF を表現するのに最も適切なモデル $a^2 \exp(-2bt)$ をフィッティングした際のモデルの回帰直線の傾き（実線: 平均 破線: 平均 $\pm$ 標準偏差）(観測点 2).

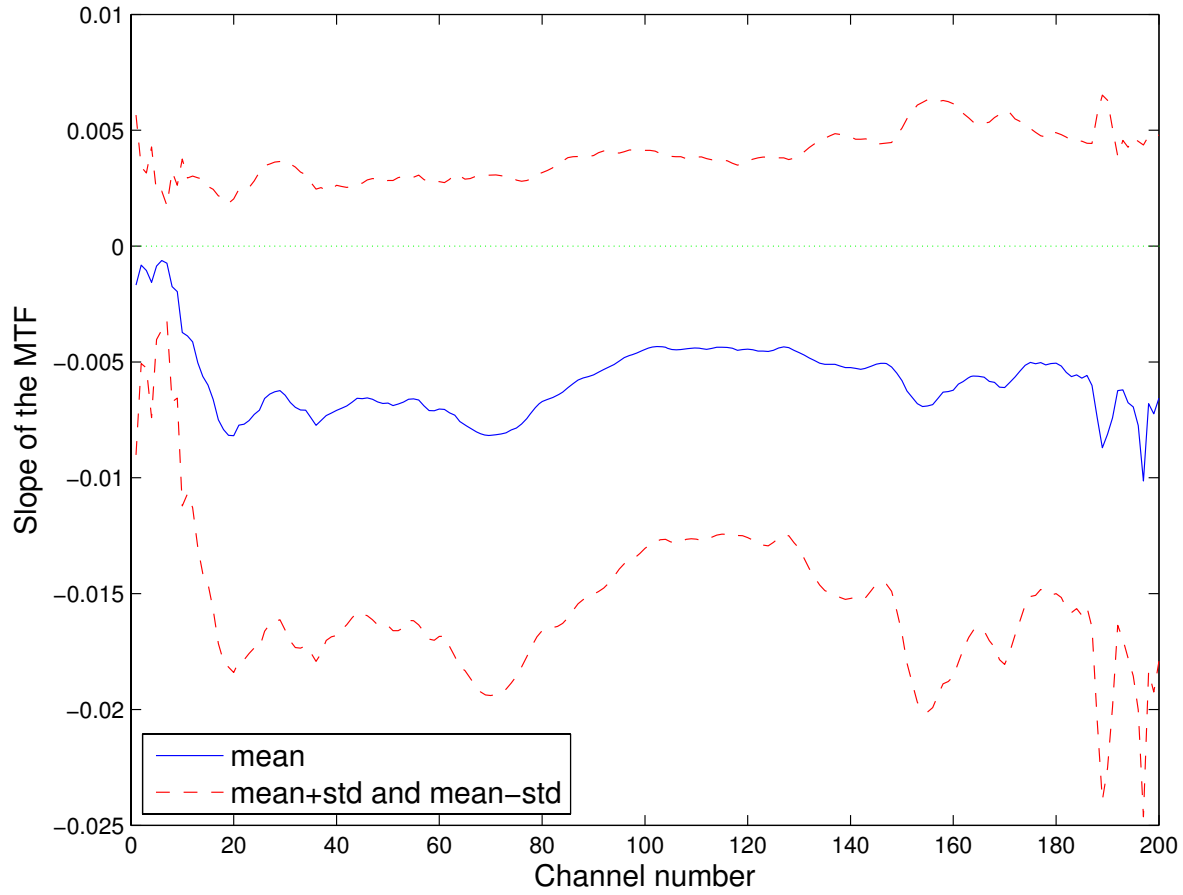

図 2.15: データベース内の音声から求めた MTF に，MTF を表現するのに最も適切なモデル $a^2 \exp(-2bt)$ をフィッティングした際のモデルの回帰直線の傾き（実線: 平均 破線: 平均 $\pm$ 標準偏差） (観測点 3).

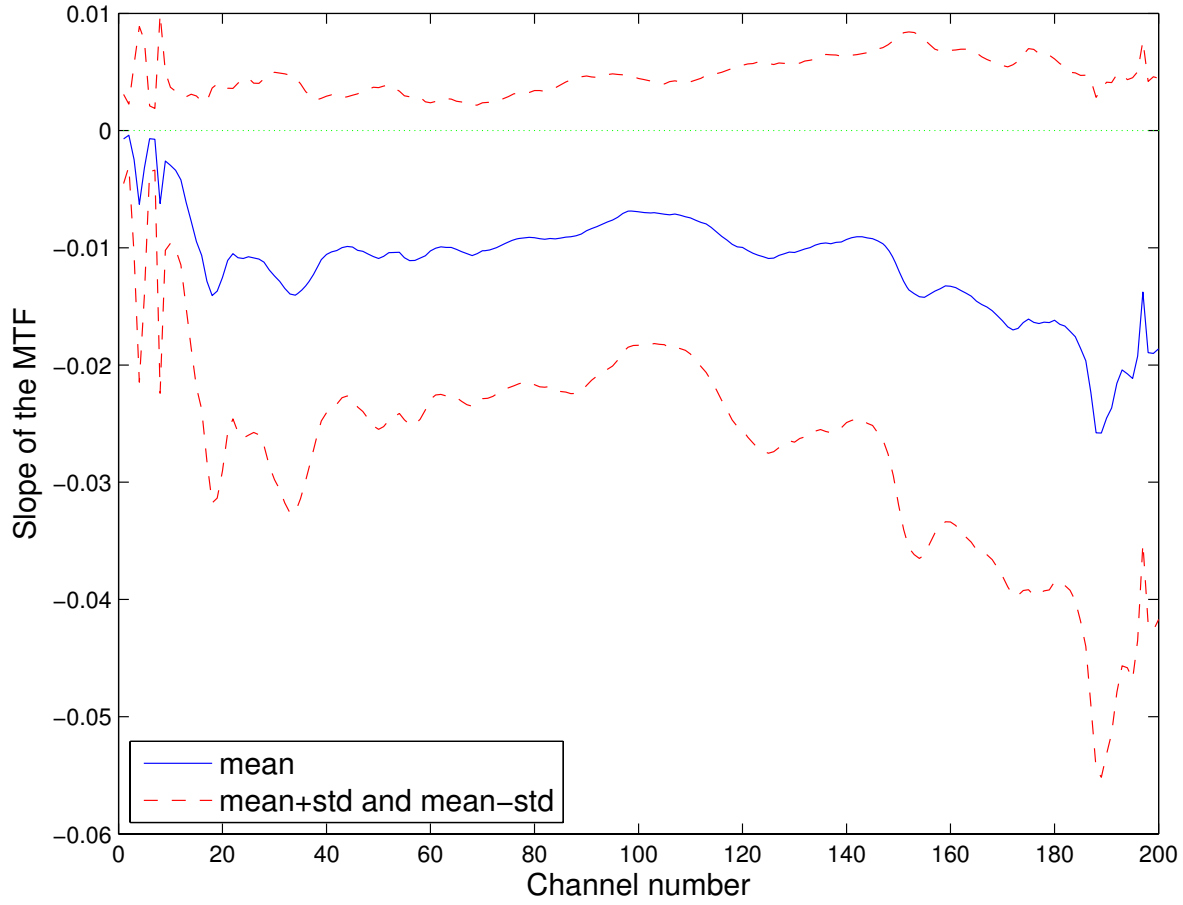

図 2.16: データベース内の音声から求めた MTF に，MTF を表現するのに最も適切なモデル $a^2 \exp(-2bt)$ をフィッティングした際のモデルの回帰直線の傾き（実線: 平均 破線: 平均 $\pm$ 標準偏差） (観測点 4).

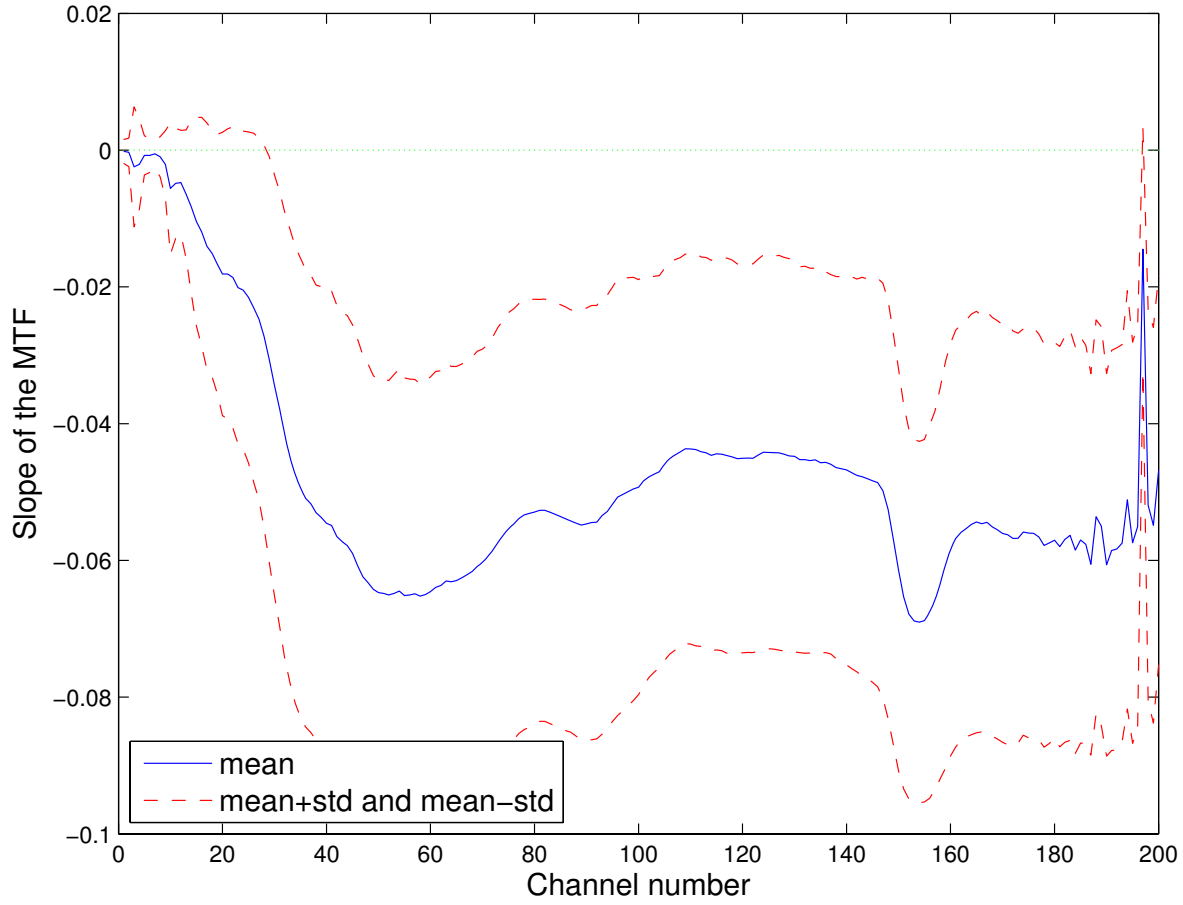

図 2.17: データベース内の音声から求めた MTF に，MTF を表現するのに最も適切なモデル $a^2 \exp(-2bt)$ をフィッティングした際のモデルの回帰直線の傾き（実線: 平均 破線: 平均 $\pm$ 標準偏差） (観測点 5).

第3章　MTFに基づくブラインド骨導音声回復法

従来のMTFに基づく骨導音声回復法では，MTFモデルのパラメータと回復条件の判定のために気導音声の情報を必要とした．本研究では，解析結果から気導音声の情報なしに上記の２点を決定できるよう手法の改良を行う．

3.1　MTFモデルのパラメータaとbの決定方法

3.1.1　パラメータaの決定方法

気導音声の情報なしに骨導音声を回復するのに最適なモデルのパラメータaとbを設定する．パラメータaについては，解析結果から図2.3〜2.7の(e)に示すように回帰曲線$1/a_n^2 = cn^{-1} + d$で表現することが可能であるため，観測点毎にデータから学習して回帰曲線を求めることで気導音声の情報なしに設定することができると考えられる．この回帰曲線が話者や発話内容によらず一意に定める事ができるかどうか，データより求めたパラメータaと回帰曲線とのRMS誤差を求め，話者及び発話内容ごとにRMS誤差の平均と標準偏差の比較を行った．図3.1〜3.5は，各観測点で収録された音声の発話内容ごとのRMS誤差を表示したものである．この図から，一部の単語を除き，各観測点において誤差に大きな差は見られなかった．また，全ての観測点において誤差が大きなな単語というものは確認されなかったため，回帰曲線は発話内容によらない可能性が示された．図3.6〜3.10は，各観測点で収録された音声の話者ごとのRMS誤差を表示したものである．この図から，一部の話者を除き，各観測点において誤差に大きな差は見られなかった．他の話者とRMS誤差の大きい話者について，パラメータaの平均をRMS誤差の小さな話者のものと比較してみた．図3.12が誤差の大きな話者，図3.11が誤差の小さな話者のパラメータaの平均である．誤差の小さな話者のパラメータaの平均は，今までに発表されている骨導音声の先行研究の結果と一致する高域減衰の形になっているのに対し，誤差の大きな話者のパラメータaの平均は40〜60チャンネル(1600〜2400 Hz)の成分を多く持ち，高域減衰とは言いがたい．図3.13にRMS誤差の大きな話者の気導パワーエンベロープと骨導パワーエンベロープのパワーの平均を，図3.14にRMS誤差の小さな話者の気導パワーエンベロープと骨導パワーエンベロープのパワーの平均を示す．この図からも，誤差の大きな話者は骨導音声の40〜60チャンネル(1600〜2400 Hz)の成分以外が大きく減

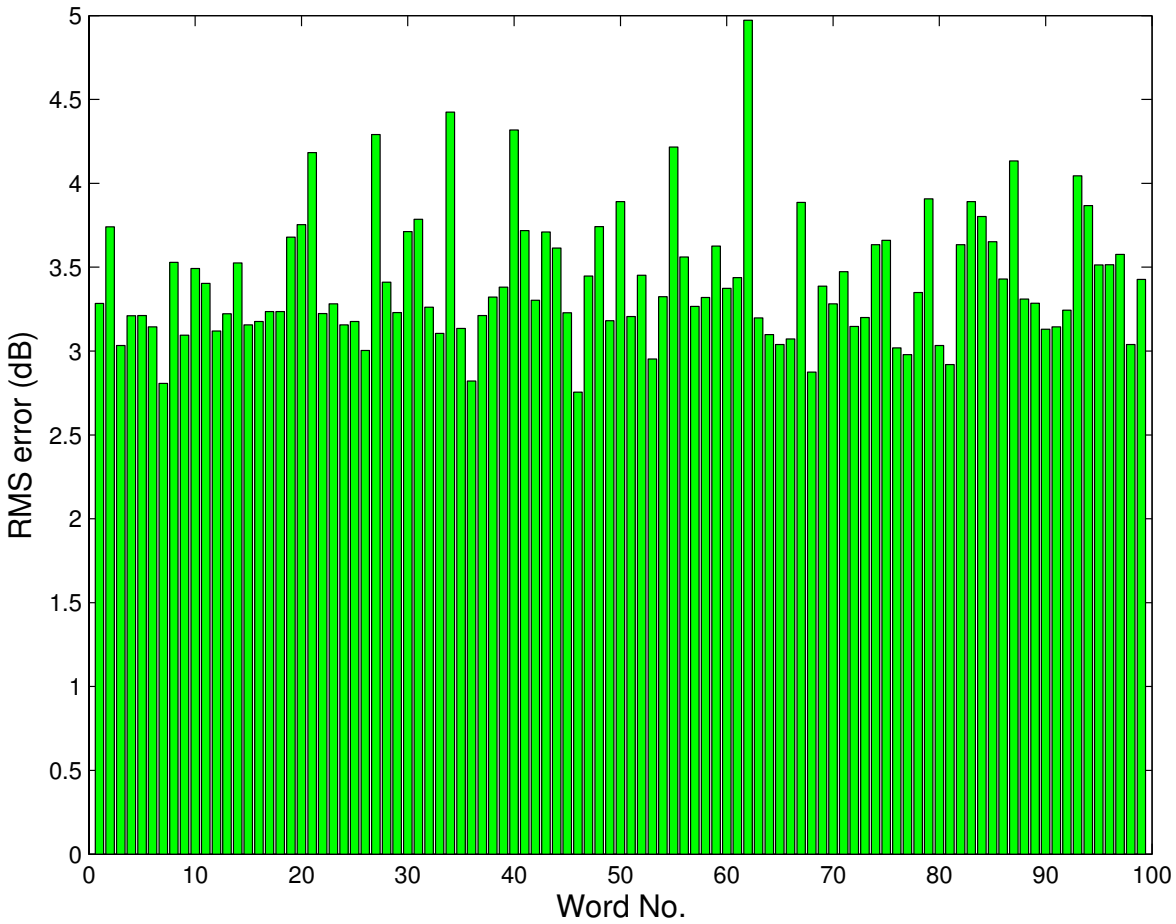

図 3.1: パラメータ a の平均の回帰曲線と，発話内容ごとの最適なパラメータ a の値と RMS 誤差の平均と標準偏差（観測点 1）.

少し，骨伝導の影響は高域減衰とはいえないように見える．このことから，骨伝導以外の影響で特定の話者のパラメータ a が他の話者と比較して大きくずれる結果となっている可能性が考えられる．今回の考察では，話者の身体的特徴に関するデータが無いため，特定話者のパラメータ a がその他の話者と大きく異なる原因は特定できなかったが，観測点ごとに設定した回帰曲線使ってモデルのパラメータ a を設定しても，ほとんどの話者に対して効果があることが明らかとなった.

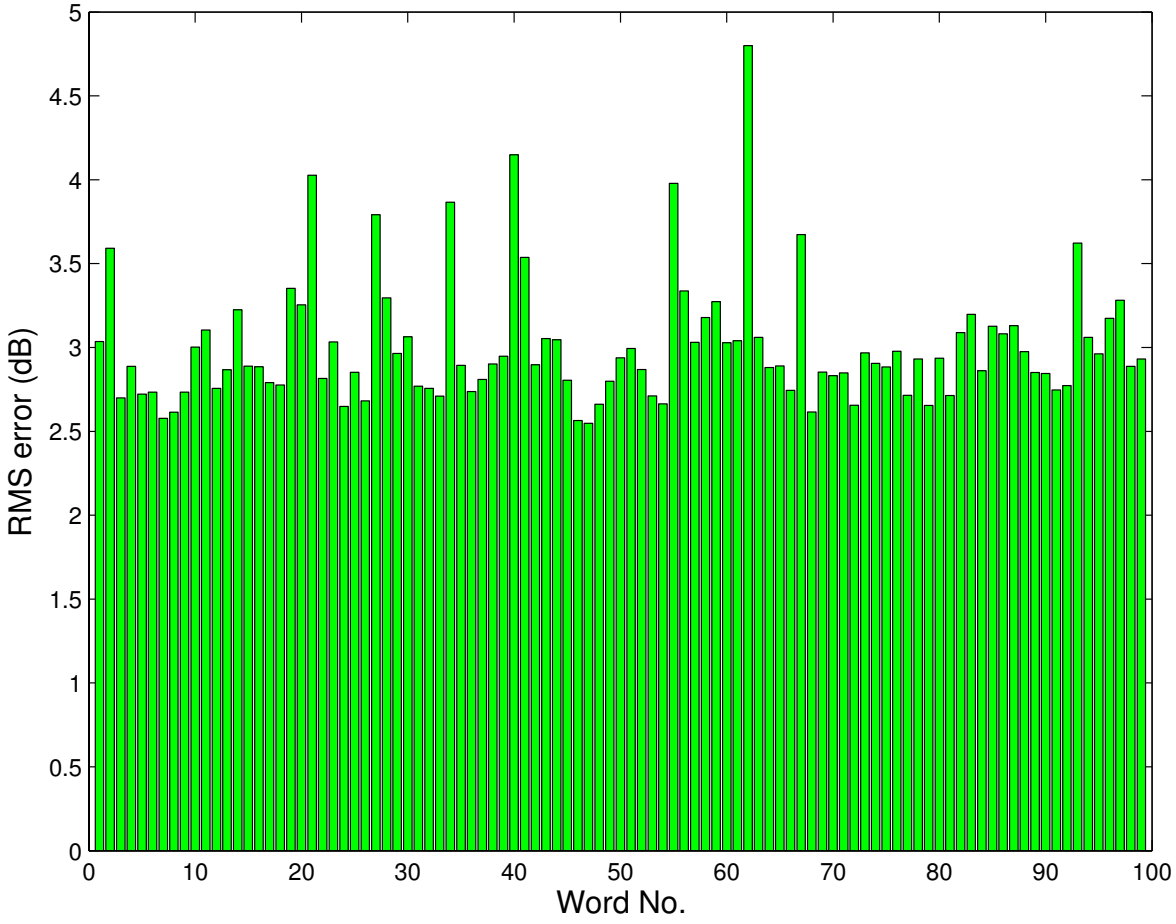

図 3.2: パラメータ a の平均の回帰曲線と，発話内容ごとの最適なパラメータ a の値と RMS 誤差の平均と標準偏差（観測点 2）.

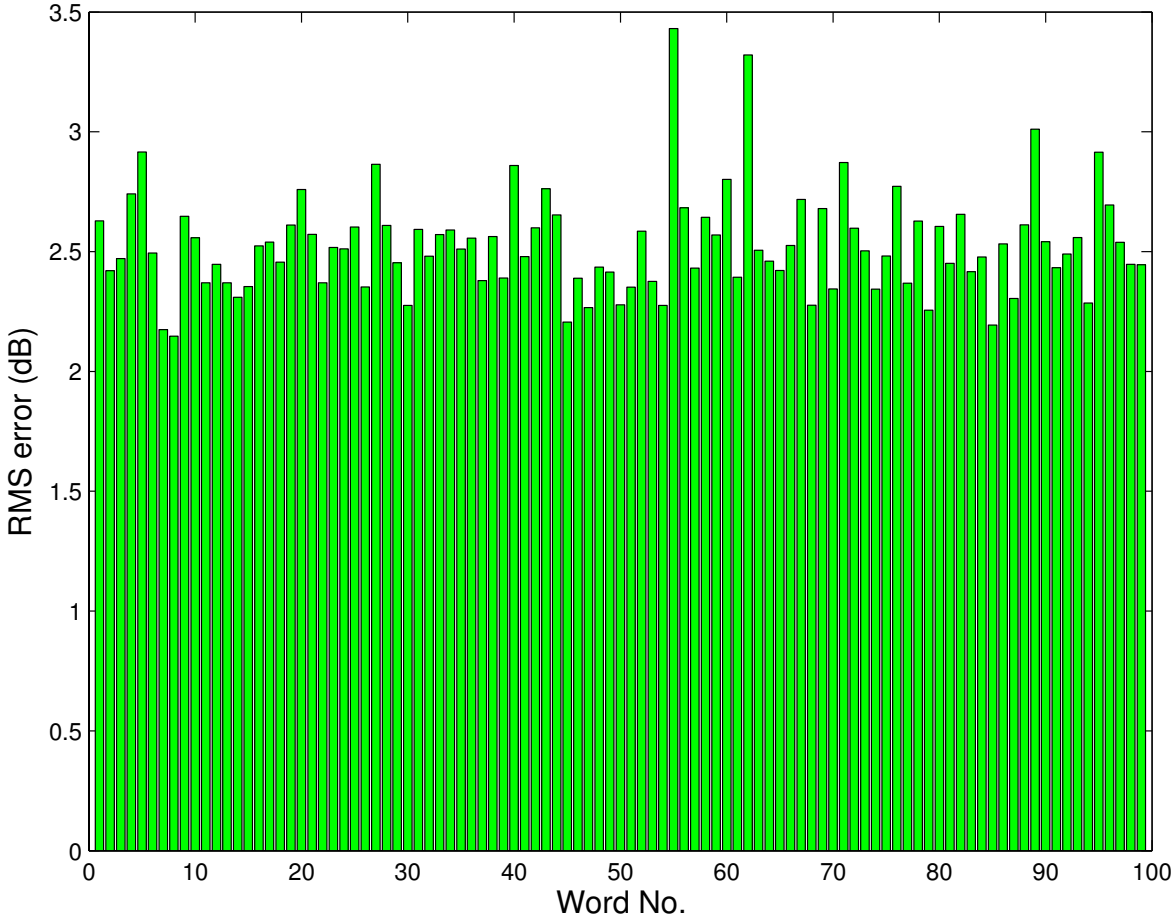

図 3.3: パラメータ a の平均の回帰曲線と，発話内容ごとの最適なパラメータ a の値と RMS 誤差の平均と標準偏差（観測点 3）.

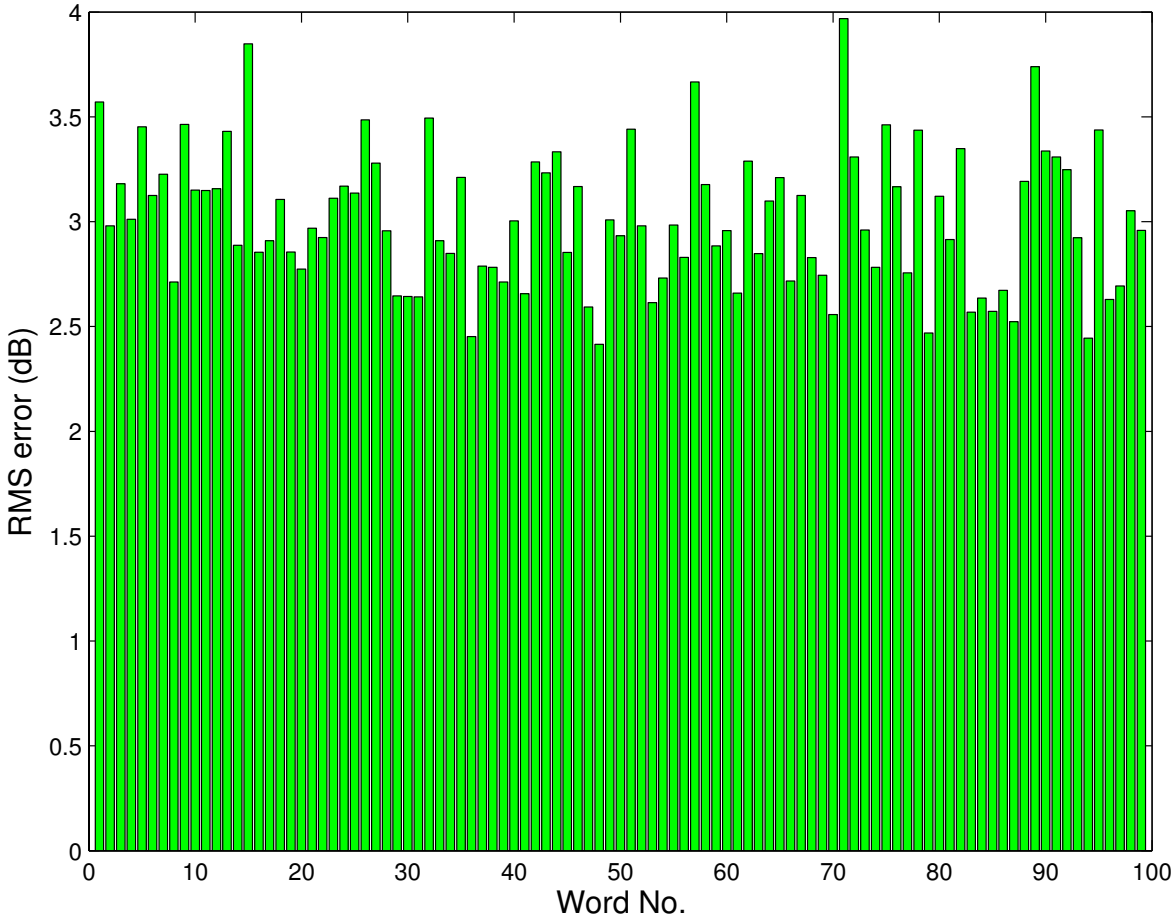

図 3.4: パラメータ a の平均の回帰曲線と，発話内容ごとの最適なパラメータ a の値と RMS 誤差の平均と標準偏差（観測点 4）.

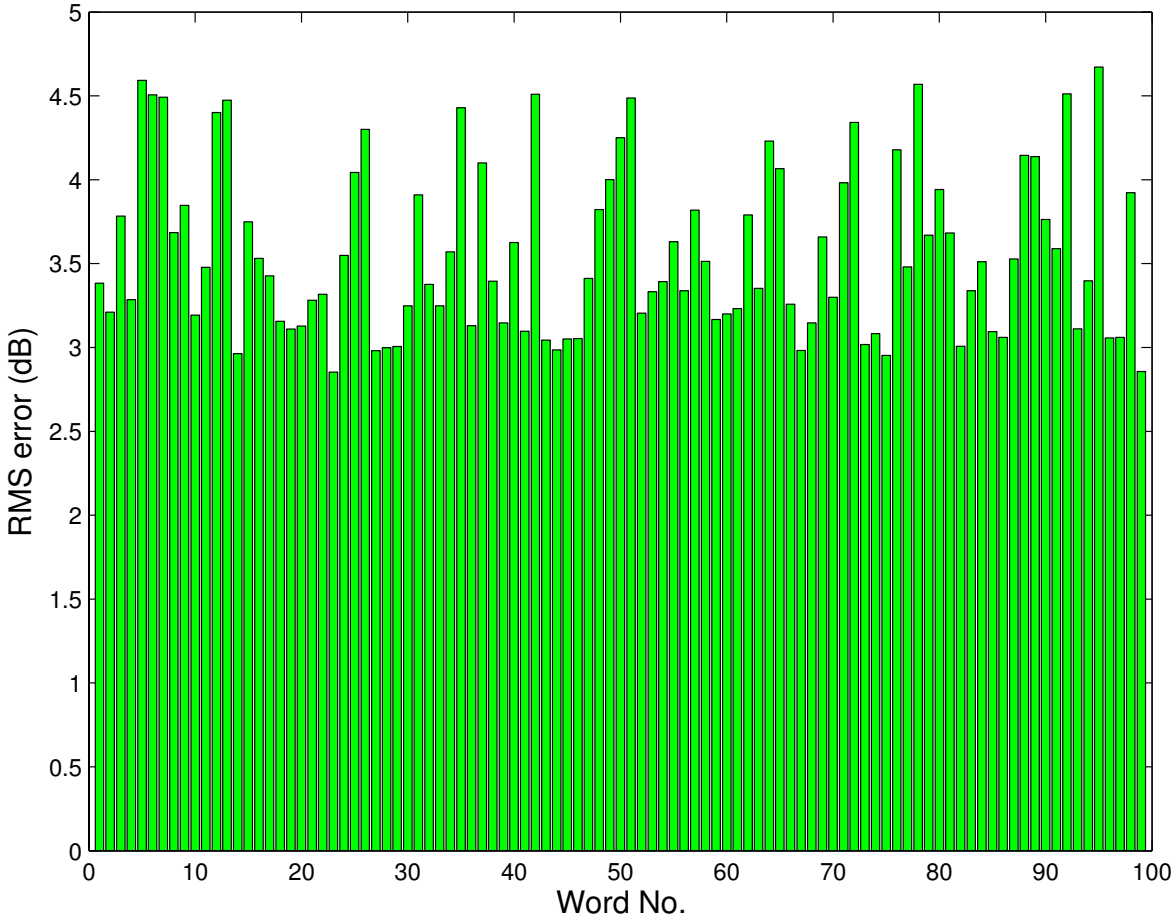

図 3.5: パラメータ a の平均の回帰曲線と，発話内容ごとの最適なパラメータ a の値と RMS 誤差の平均と標準偏差（観測点 5）.

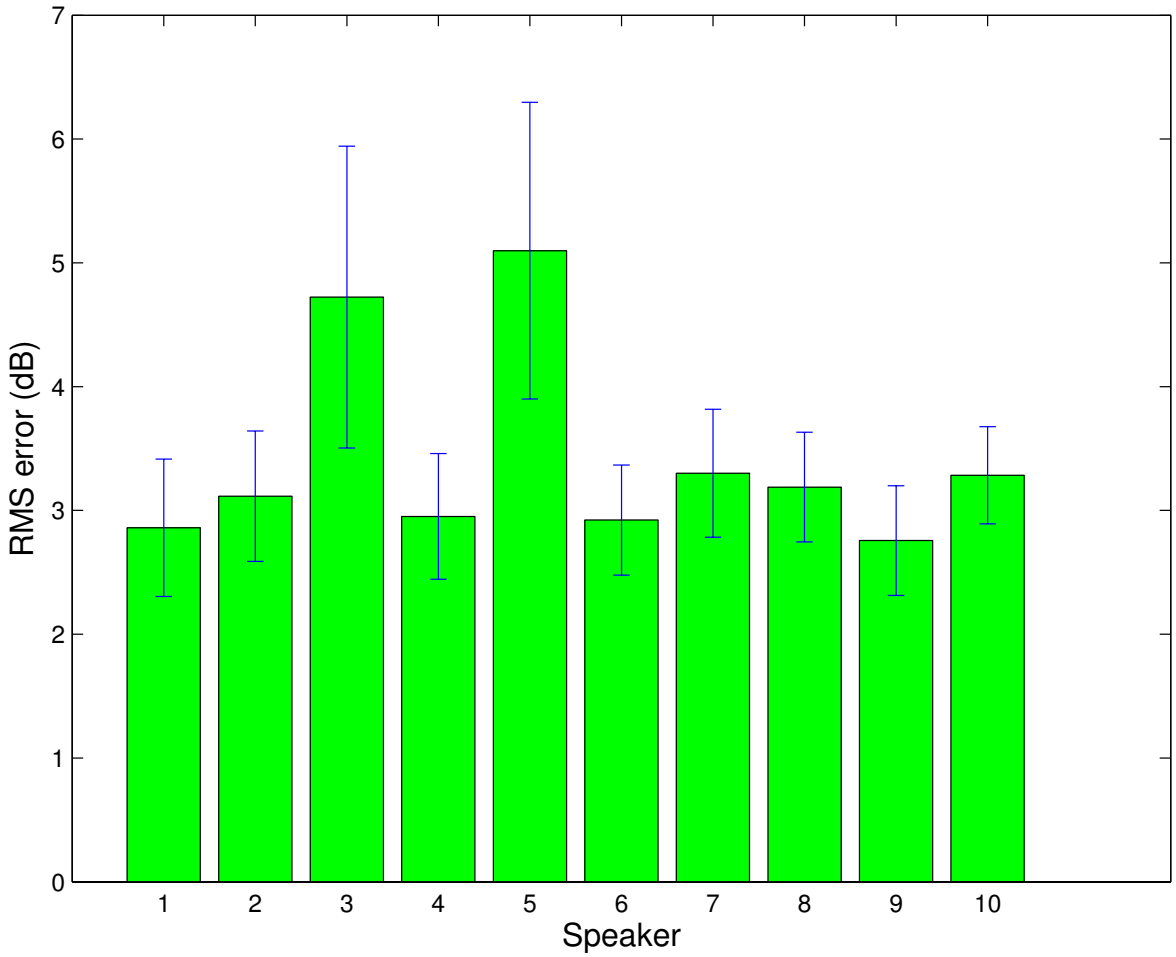

図 3.6: パラメータ a の平均の回帰曲線と，話者ごとの最適なパラメータ a の値と RMS 誤差の平均と標準偏差（観測点 1）.

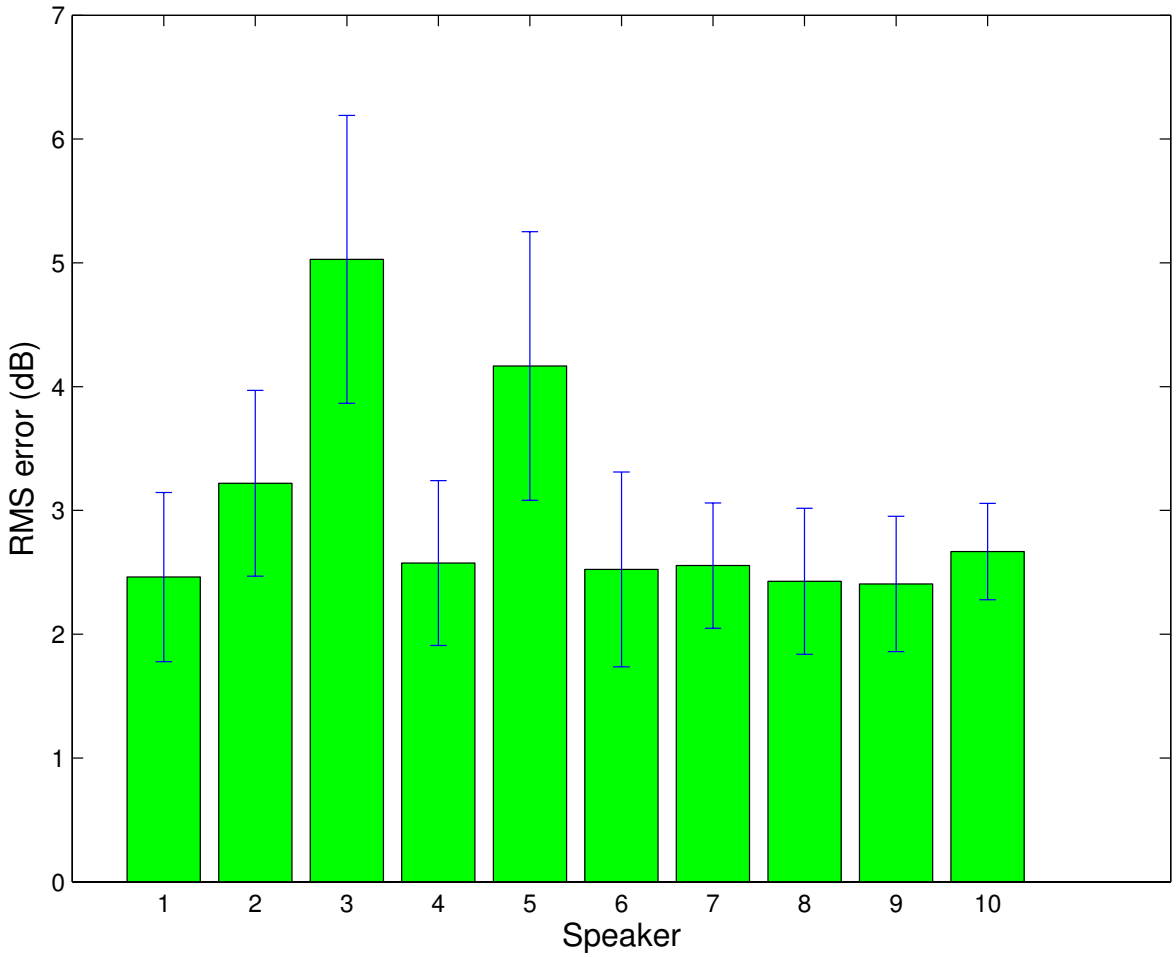

図 3.7: パラメータ a の平均の回帰曲線と，話者ごとの最適なパラメータ a の値と RMS 誤差の平均と標準偏差（観測点 2）.

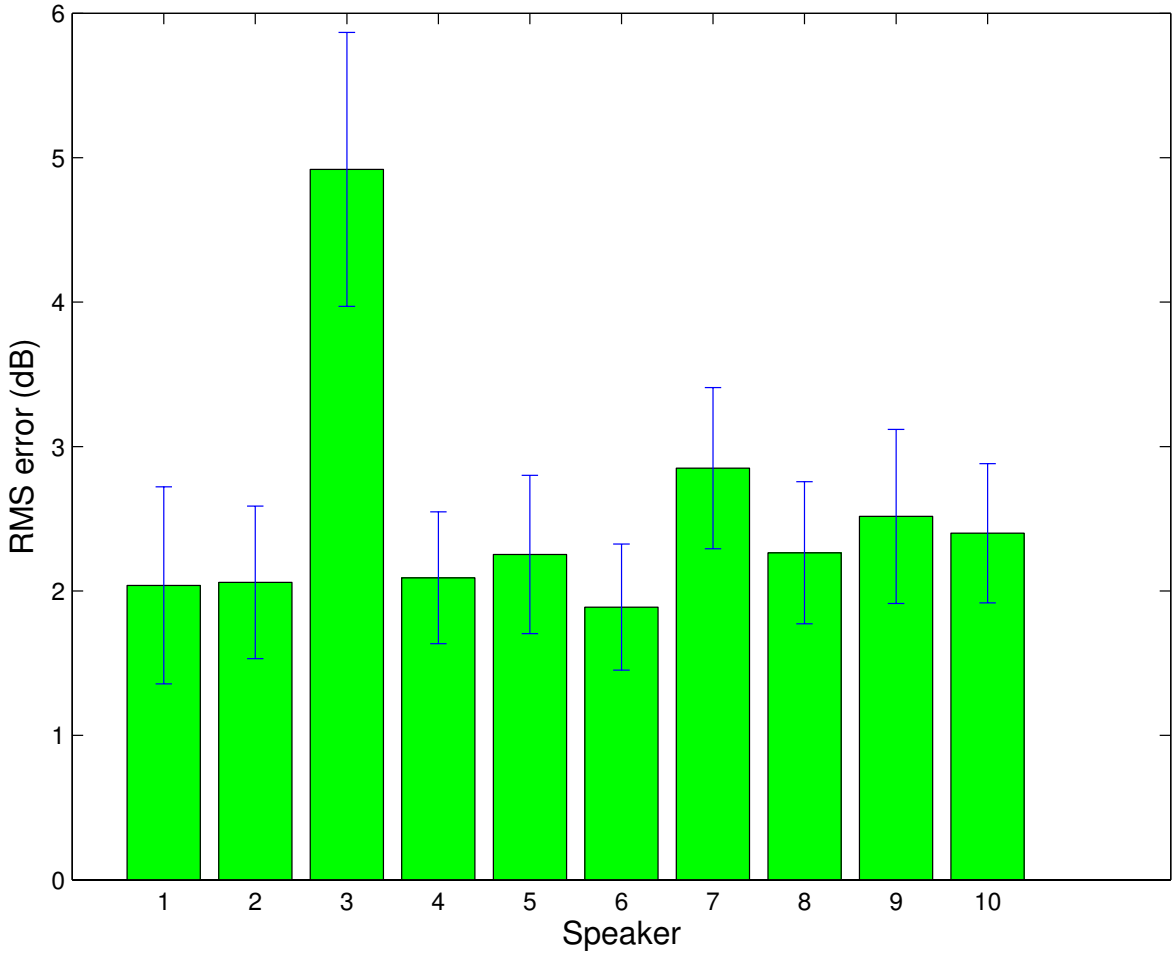

図 3.8: パラメータ a の平均の回帰曲線と，話者ごとの最適なパラメータ a の値と RMS 誤差の平均と標準偏差（観測点 3）.

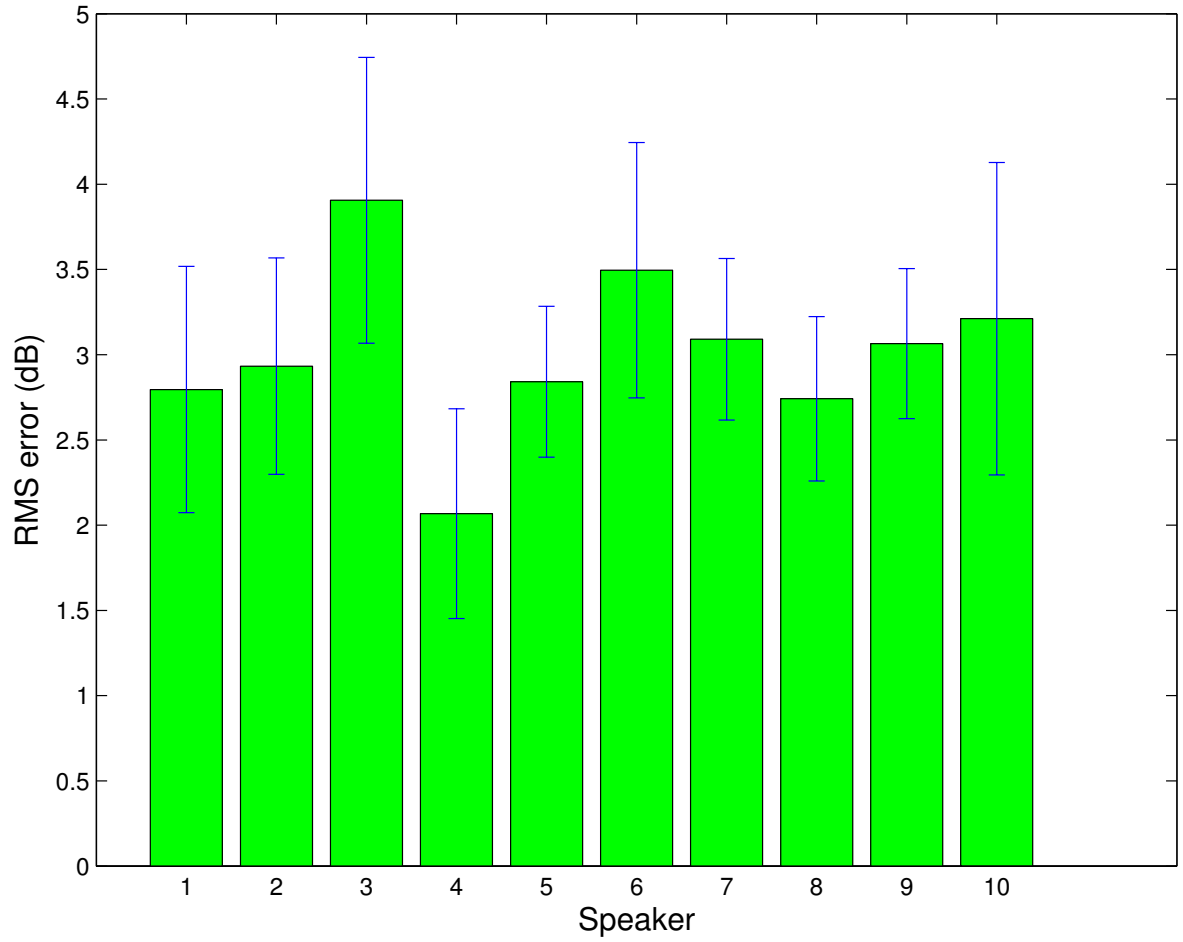

図 3.9: パラメータ a の平均の回帰曲線と，話者ごとの最適なパラメータ a の値と RMS 誤差の平均と標準偏差（観測点 4）.

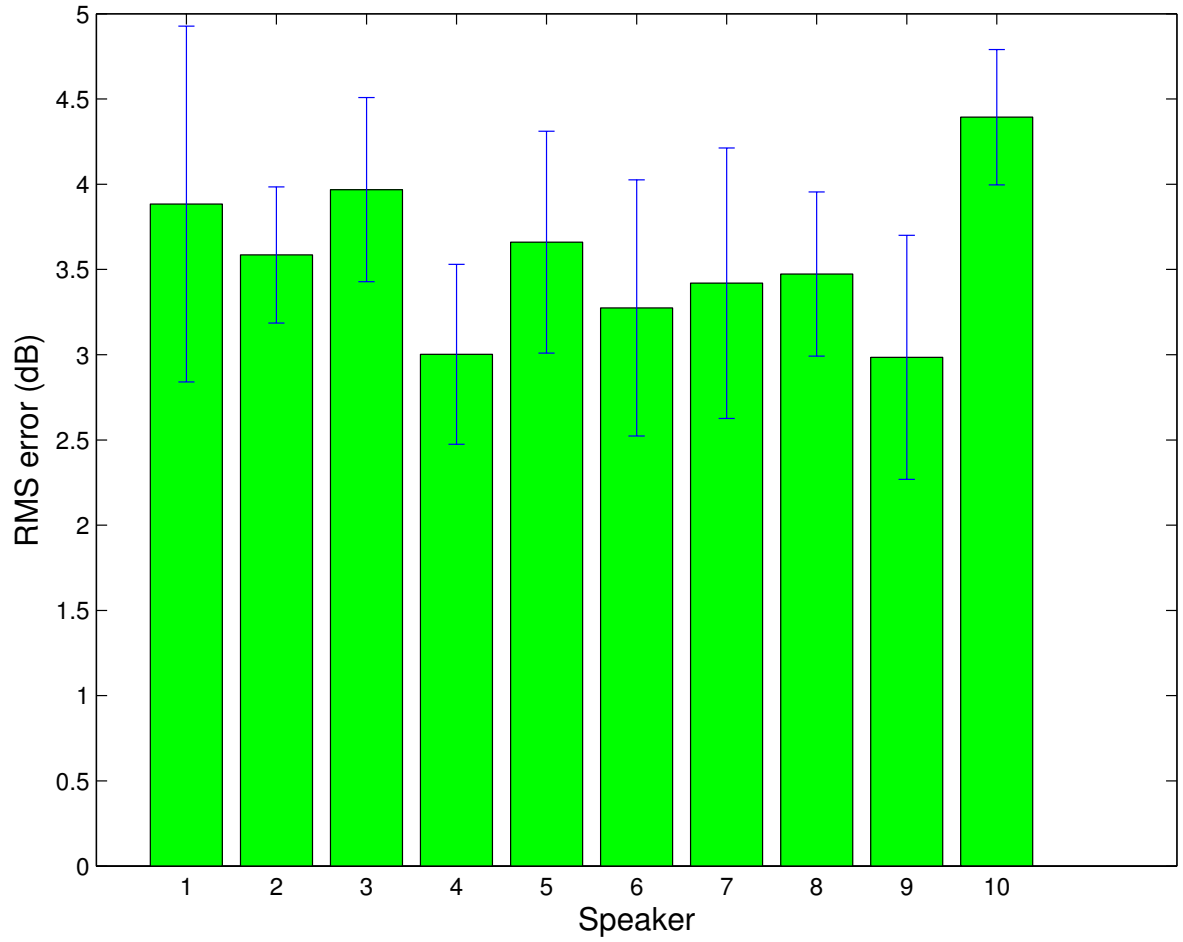

図 3.10: パラメータ a の平均の回帰曲線と，話者ごとの最適なパラメータ a の値と RMS 誤差の平均と標準偏差（観測点 5）.

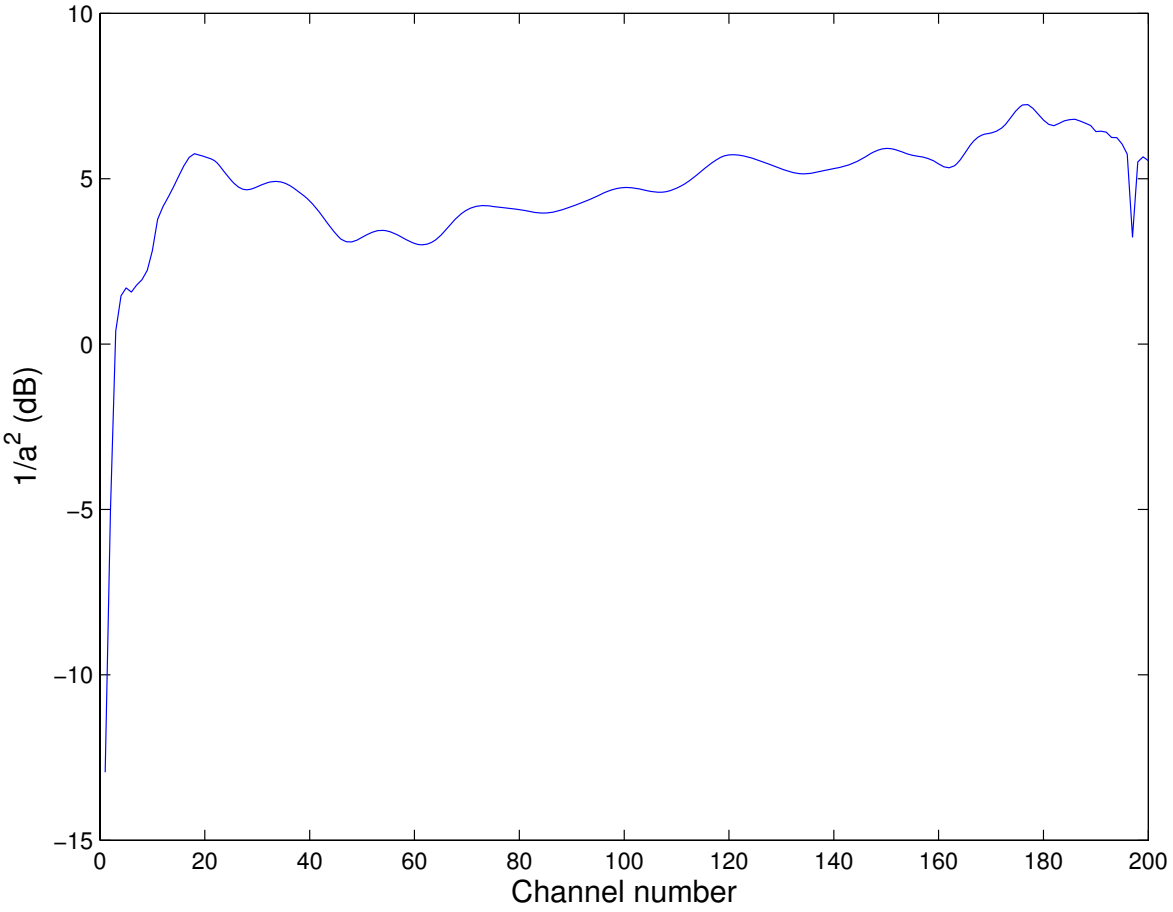

図 3.11: 実際の a の値の RMS 誤差が小さい話者（観測点 3，話者 2）のパラメータ a の平均.

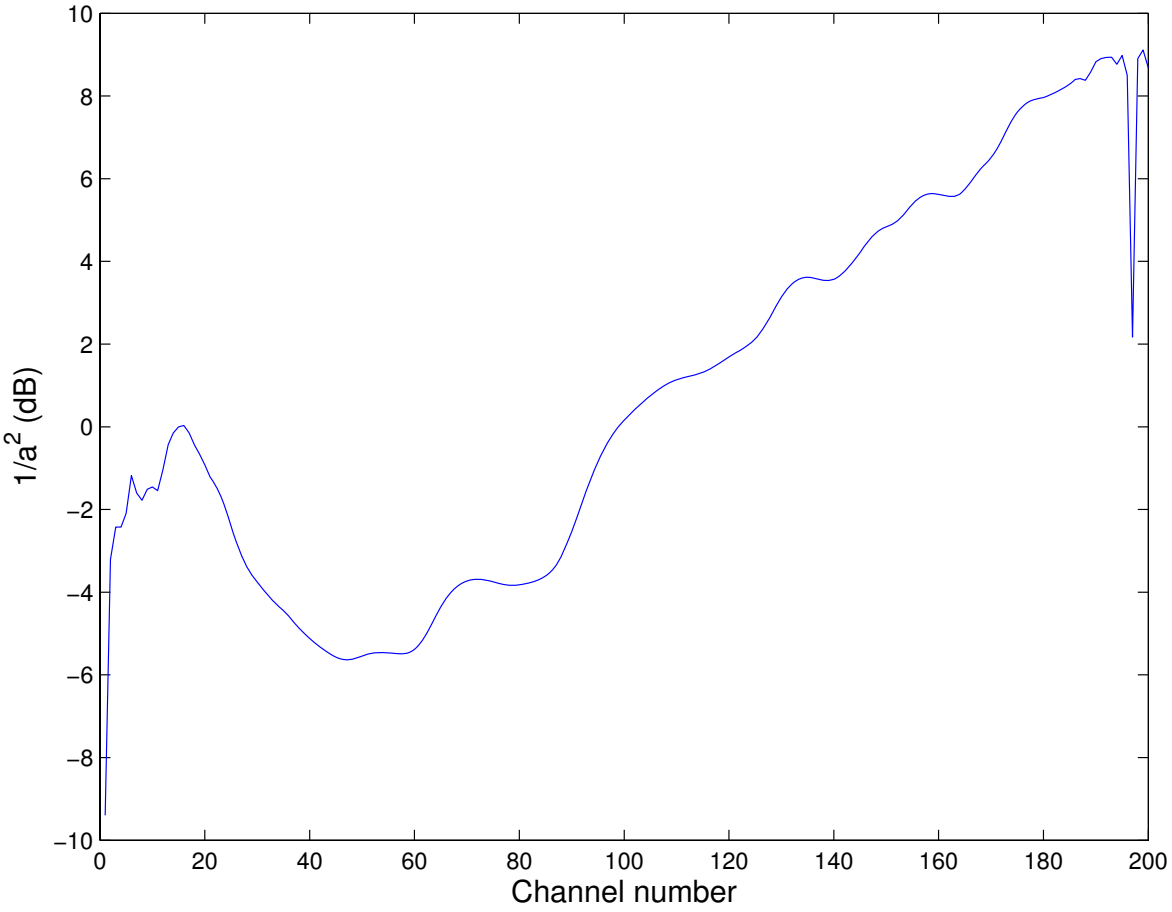

図 3.12: 実際の a の値の RMS 誤差が大きい話者（観測点 3，話者 3）のパラメータ a の平均.

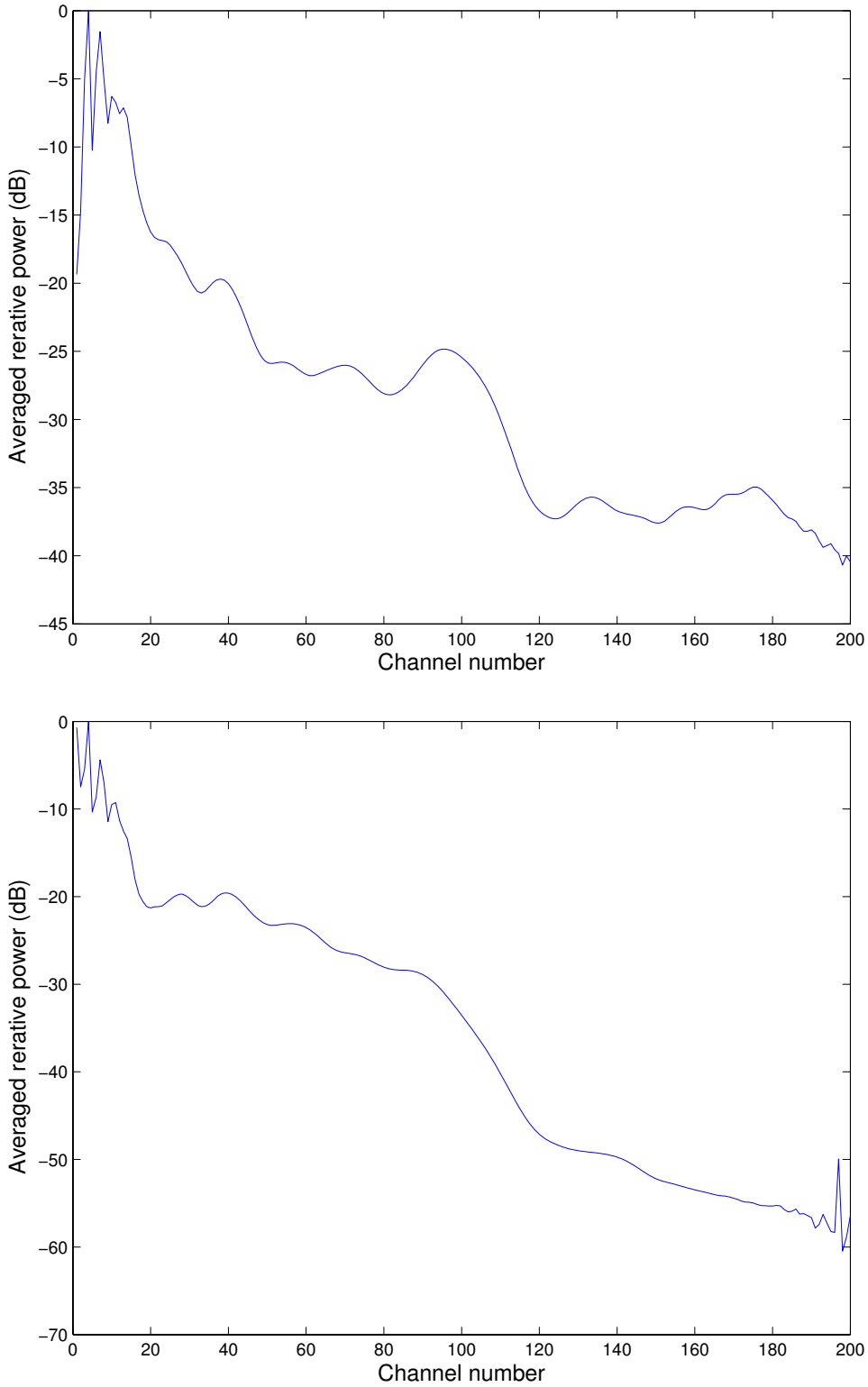

図 3.13: 観測点 3，話者 3 のパワーエンベロープのパワーの平均．上: 気導音声 下: 骨導音声．

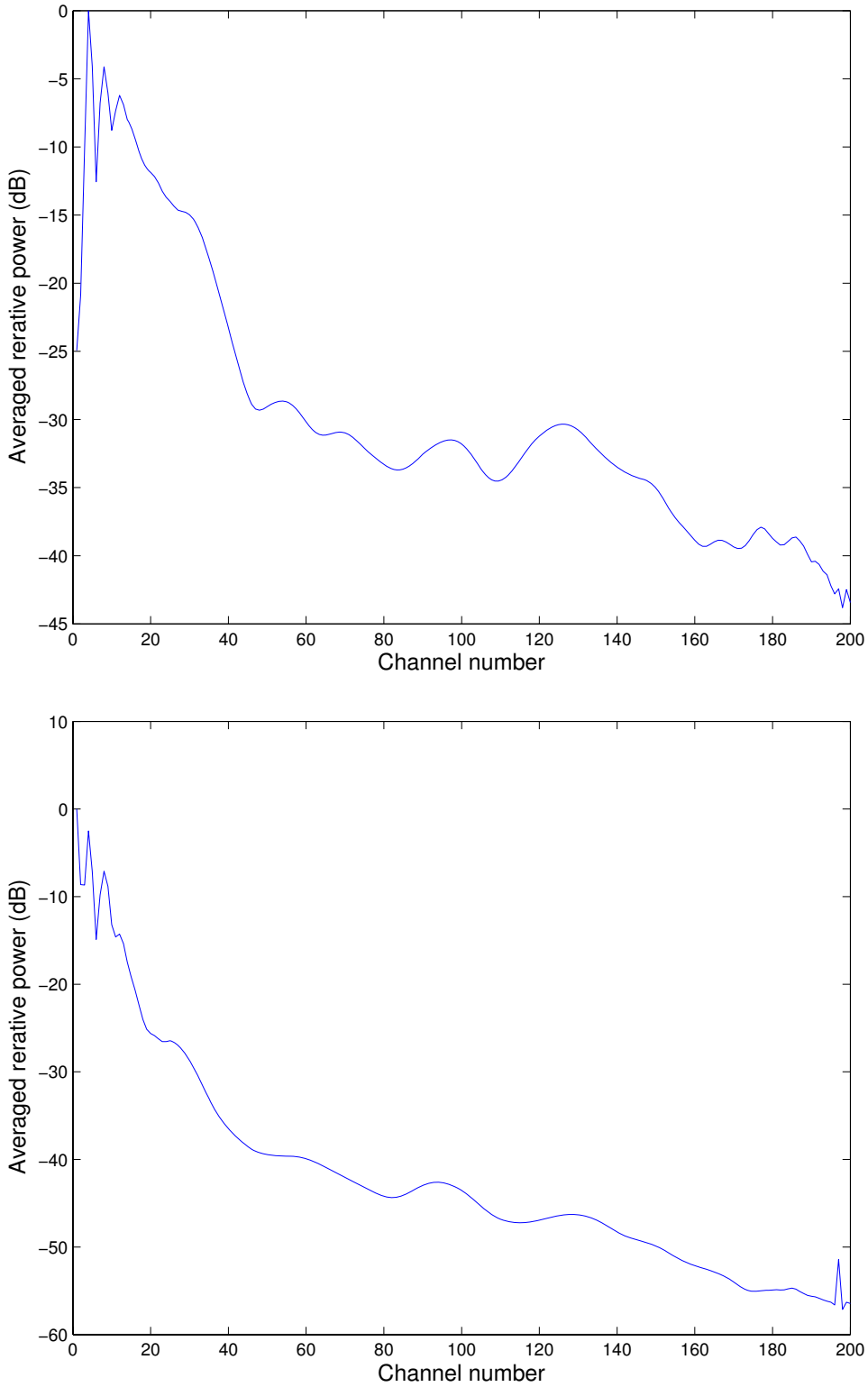

図 3.14: 観測点 3，話者 2 のパワーエンベロープのパワーの平均．上: 気導音声 下: 骨導音声．

3.1.2 パラメータ b の決定方法

パラメータ b については，従来法で用いられていた推定法より精度の良い推定法が平松と鵜木により提案されている [30]．この手法は，元々残響時間推定のために提案された手法であるが，本研究で用いる MTF モデルと平松と鵜木が用いた MTF モデルは同じモデルであるため，この手法を用いて MTF のパラメータ b を推定することが可能である．図 3.15 にパラメータ b 推定法の概念を示す．図 3.15 の上段は 4 Hz のサイン波のパワーエンベロープとその変調スペクトルの図である．この図から，主要な変調周波数と，直流成分の変調周波数が共に 0 dB であることがわかる．下段は，上段の波形をローパスフィルタにより減衰させた信号のパワーエンベロープと変調スペクトルである．この図から，主要な変調周波数の成分は減少しているが，直流成分の変調周波数は変わらずに 0 dB であることがわかる．このことから，直流成分の変調周波数と，主要な変調周波数が同じ値になるように逆フィルタで回復した際の MTF モデルのパラメータを求めることで，b の値を推定することが可能である．これより，パラメータ b の推定値 $\hat{b}$ は以下の式で表現される．

$$\hat{b} = \arg\min_{b} \left(\left| 20\log_{10}(\hat{E}_y(0)) + 20\log_{10}(M(f_{dm}, b)) - 20\log_{10}(\hat{E}_y(f_{dm})) \right| \right) \tag{3.1}$$

ここで，f_{dm} は骨導音声の主要な変調周波数を，$\hat{E}_y(\cdot)$ は逆フィルタにより回復されたパワーエンベロープである．実際に，従来の推定法で推定されたパラメータ b を用いて骨導音声の回復を行った場合と，平松と鵜木により提案されたパラメータ b を用いて骨導音声の回復を行った場合での MTF の回復量の差を比較した．図 3.16 はその結果である．平松と鵜木により提案された推定法を用いた方が，MTF の回復精度が良いことがわかる．

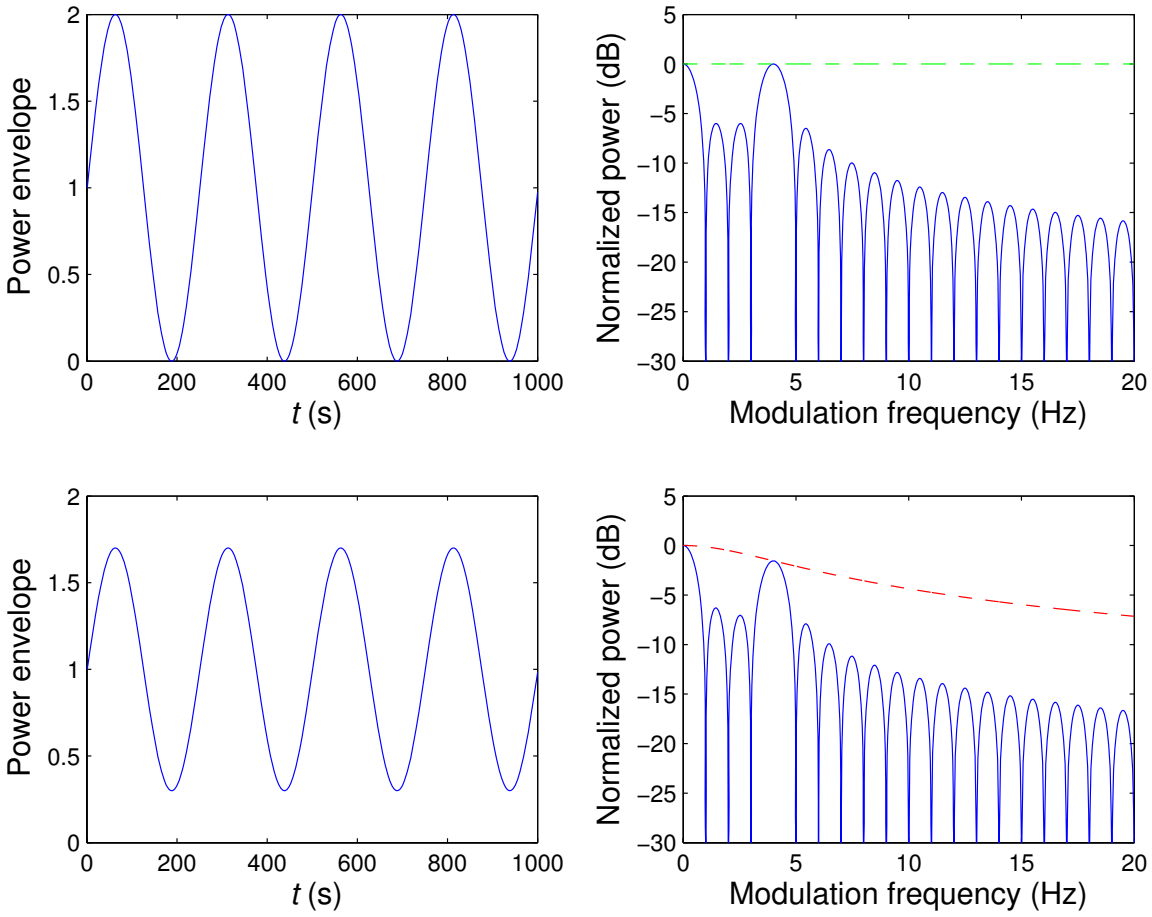

図 3.15: 周波数ドメインでのパラメータ B の推定.

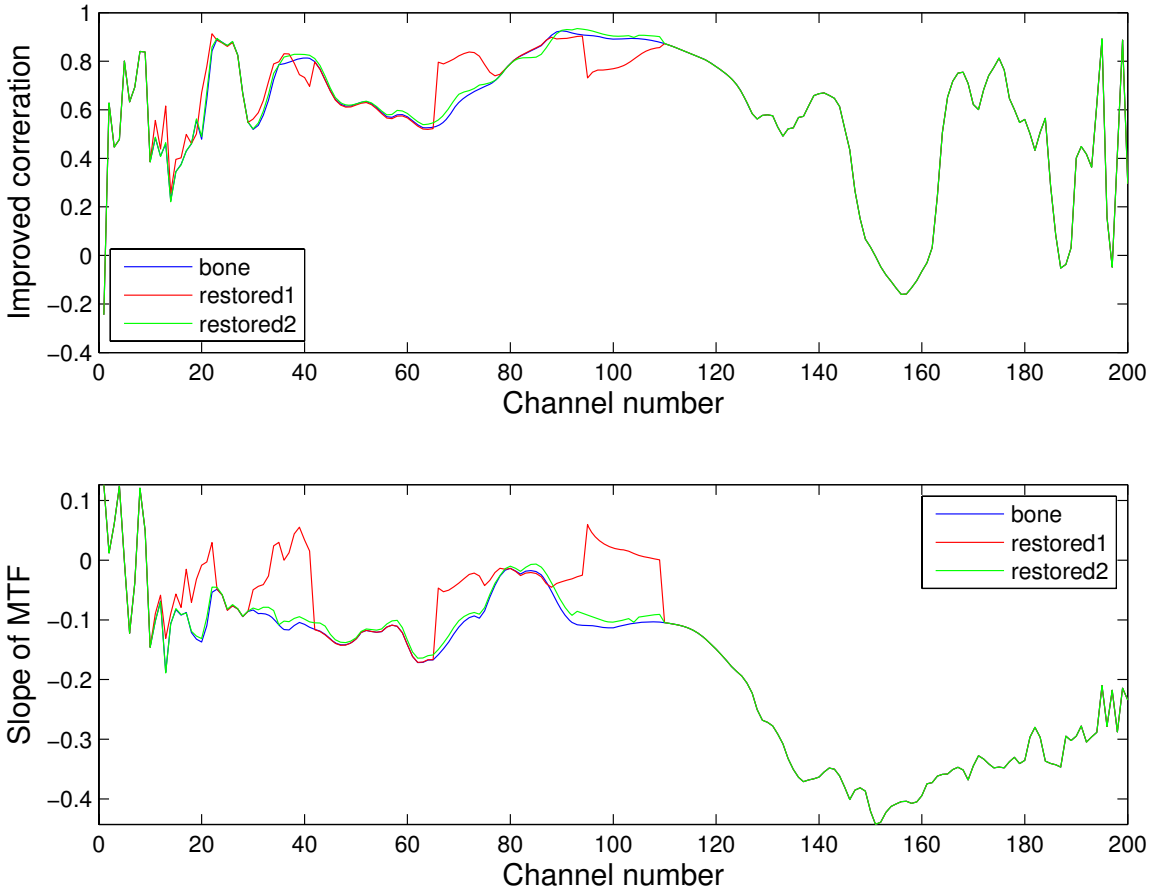

図 3.16: パラメータ B の推定法の比較．restored1: 時間ドメインでのパラメータ推定 restored2: 周波数ドメインでのパラメータ推定.

3.2 回復条件の変更

木村により提案された従来法では，気導パワーエンベロープと骨導パワーエンベロープ間の相関係数が 0.8 以上かつ気導パワーエンベロープの相対パワーが −20 dB 以内の時のみ逆フィルタによる骨導音声回復を行い，それ以外の範囲で，気導パワーエンベロープの相対パワーが −40 dB の範囲まではゲイン補正を行うと条件を定めている．これは，気導パワーエンベロープと骨導パワーエンベロープ間の相関係数が高い箇所に対して回復は必要でなく，また，音声のパワーの低い箇所は音声の成分があまりないため，これも回復の必要がないと判断したためであった．この条件を判定するためには気導音声の情報を必要とするため，骨導音声の情報のみで判定でき，骨導音声回復に適切な条件に変更する．まず，気導パワーエンベロープと骨導パワーエンベロープ間の相関係数が高い箇所については，回復を行うことによってわずかではあるが改善があるため，回復範囲から除外する必要はなかった．次に，気導パワーエンベロープの相対パワーが-40 dB 以下の範囲においても音声の成分は確かに存在するため，全帯域における回復が必要であることがわかった．しかし，骨導音声の相対パワーが 40 dB 以上減少すると，内部雑音の影響をうけて MTF の直流成分が増加してしまい，b の値の推定が妨害されてしまう．よって，本研究では回復条件を以下のように設定する．

- 骨導パワーエンベロープの相対パワーが −40 dB 以内の範囲で逆フィルタを用いて回復を行う．
- 上記の範囲以外の全範囲では，ゲイン補正のみを行う．

MTF に基づく骨導音声回復法にこの 2 点の変更を加える事により，気導音声の情報を必要とすることなく，MTF の概念に基づいて骨導音声の回復を施す事が可能となった．

第4章　提案法の評価

4.1　評価方法

気導/骨導データベース内の全ての音声 (話者 10 名，単語数 100 個，観測点 5 箇所の計 5000 単語) を用いてシュミレーションにより提案法の客観評価を行う．音声は，骨導音声，提案法で回復された骨導音声，気導音声の情報を用いてパラメータ a を求めた提案法で回復された骨導音声，木村らが提案した手法で回復された骨導音声の 4 種類を用いた．提案法で手法の回復精度を評価するために，SNR，相関係数，MTF の回帰直線の傾き，変調度 1 の MTF と回復音声あるいは骨導音声の MTF の誤差の RMS，伝達関数を用いる．また，総合評価として，音質の評価を行うため，対数スペクトル歪 (LSD), 線形予測係数を用いて求めた LSD(LP-LSD), ケプストラム距離 (CD), メルケプストラム係数 (MFCCD). 明瞭度を評価する尺度として，音声明瞭度を考慮した LSD（ILSD）[31] を用いる．総合評価の評価尺度は以下の式で表現される．

- LSD

$$\mathrm{LSD} = \frac{1}{F}\sum_{j=0}^{F}\sqrt{\frac{2}{f_{\mathrm{s}}}\sum_{\omega}^{fs/2}\left[20\log_{10}\left(\frac{\mathrm{S_x}(\omega,\mathrm{j})}{\mathrm{S_y}(\omega,j)}\right)\right]^2} \tag{4.1}$$

- LP-LSD

$$\mathrm{LP-LSD} = \frac{1}{F}\sum_{j=0}^{F}\sqrt{\frac{1}{24}\sum_{i=0}^{24}(20\log_{10}(\alpha_{\mathrm{x}}(i,j)) - 20\log_{10}(\alpha_{\mathrm{y}}(i,j)))^2} \tag{4.2}$$

- CD

$$\mathrm{CD} = \frac{1}{F}\sum_{j=0}^{F}\sum_{i=0}^{16}\sqrt{(\beta_{\mathrm{x}}(i,j) - \beta_{\mathrm{y}}(i,j))^2} \tag{4.3}$$

- MFCCD

$$\mathrm{MFCCD} = \frac{1}{F}\sum_{j=0}^{F}\sum_{i=0}^{12}\sqrt{(\gamma_{\mathrm{x}}(i,j) - \gamma_{\mathrm{y}}(i,j))^2} \tag{4.4}$$

- ILSD

$$\mathrm{ILSD} = \frac{1}{F}\sum_{j=0}^{F}\mathrm{W}(\omega)\sqrt{\frac{2}{f_{\mathrm{s}}}\sum_{\omega}^{fs/2}\left[20\log_{10}\left(\frac{\mathrm{R_x}(\omega,j)}{\mathrm{R_y}(\omega,j)}\right)\right]} \tag{4.5}$$

音声はいずれも時間フレーム分割され，各フレーム毎に求めた評価尺度の値を平均したものを各音声の値としている．Fはフレーム分割数，f_sはサンプリング周波数 (16 kHz) である．LSD は 25 ms 毎に音声をフレーム分割した．$S(\omega)$ は 400 点 FFT で算出されたスペクトログラムである．LSDLP は 25 ms 毎に音声をフレーム分割した．線形予測係数は 24 次用いた．$\alpha(i,j)$ は線形予測係数を用いて求めたスペクトル包絡成分である．CD は，サンプリング点 512 点毎に時間分割した．$\beta(i,j)$ はケプストラム係数であり，16 次までの係数を用いた．MCD はサンプリング点 256 点毎に時間分割した．$\gamma(i,j)$ はメルケプストラム係数であり，12 次までの係数を用いた．ILSD はサンプリング点 512 点毎に時間分割した．$R(\omega,j)$ は，クリティカルバンドで帯域分割され，512 点 FFT により算出されたスペクトログラムである．$W(\omega)$ は，周波数に応じて変化する重み付け関数である．

4.2 評価結果

図 4.1〜4.5 に, 提案法による回復結果の一例を示す．実線は骨導音声の結果, 破線は提案法により回復された音声の結果である．この図から, 提案法により確かに音声回復が行われていることが見て取れる．また，相関が改善されていることから，時間ドメインでのパワーエンベロープの歪を回復できていることがわかる．次に, 図 4.6〜4.9 に気導/骨導データベース内の全音声を用いた音質の総合評価の結果を示す．評価対称の音声は, 骨導音声, 提案法で回復した音声, パラメータ a を気導音声の情報を使って求め, 提案法により回復した音声，先行研究で木村により提案された従来法で回復した音声の 4 つである．リファレンスはいずれも気導音声である．全ての結果において，骨導音声より提案法の数値が低くなっていることから，提案法が骨導音声の音質を改善できていることを示している．また，従来法と比較すると，提案法はブラインド法であるにもかかわらず LSD，LP-LSD，CD において，同等の結果となった．最後に，表 4.10 に音声明瞭度の総合評価の結果として，音声明瞭度を考慮した LSD の結果を示す．骨導音声より提案法の数値が低くなっていることから，提案法が骨導音声の音声明瞭度を改善できていることがわかる．観測点 1 と 5 において, 提案法による改善度合いが気導音声の情報を用いてパラメータ a を求めた提案法と比較して大幅に小さいのは，解析の際に考察したとおり，回帰曲線と実際の a の値の RMS が他の観測点と比較し大きいためである．しかし，骨導音声の高域を良く収録できる観測点 2，3，4 については提案法により骨導音声の改善がしっかりとなされているのがわかる．この結果から，提案法は骨導音声の回復に確かな効果があることが示され, 骨導音声の高域をよく収録できる，骨導マイクを装着するのに適していると思われる点で収録した音声に対しては，事前に学習を行うことによりブラインドで非常に精度良く回復を行うことが可能であることが示唆された．

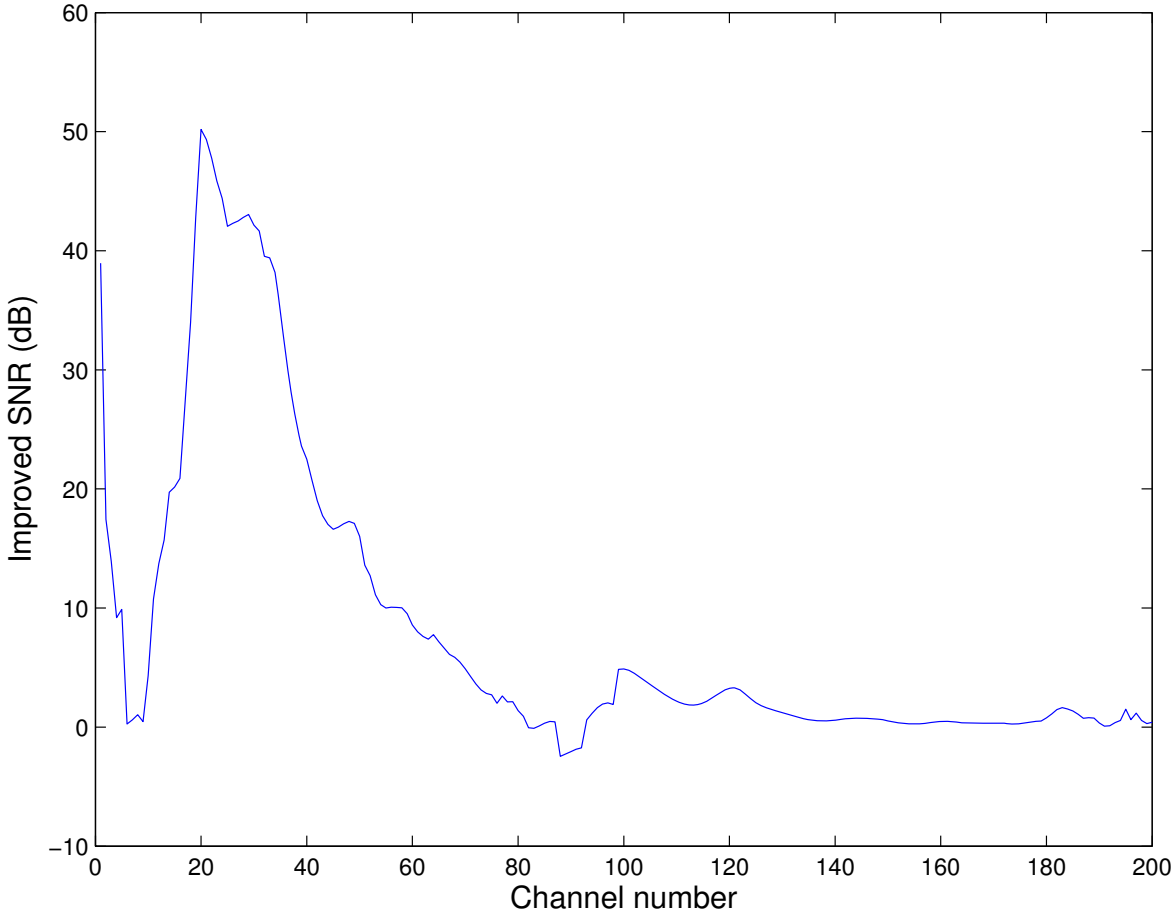

図 4.1: 提案法による SNR の改善度.

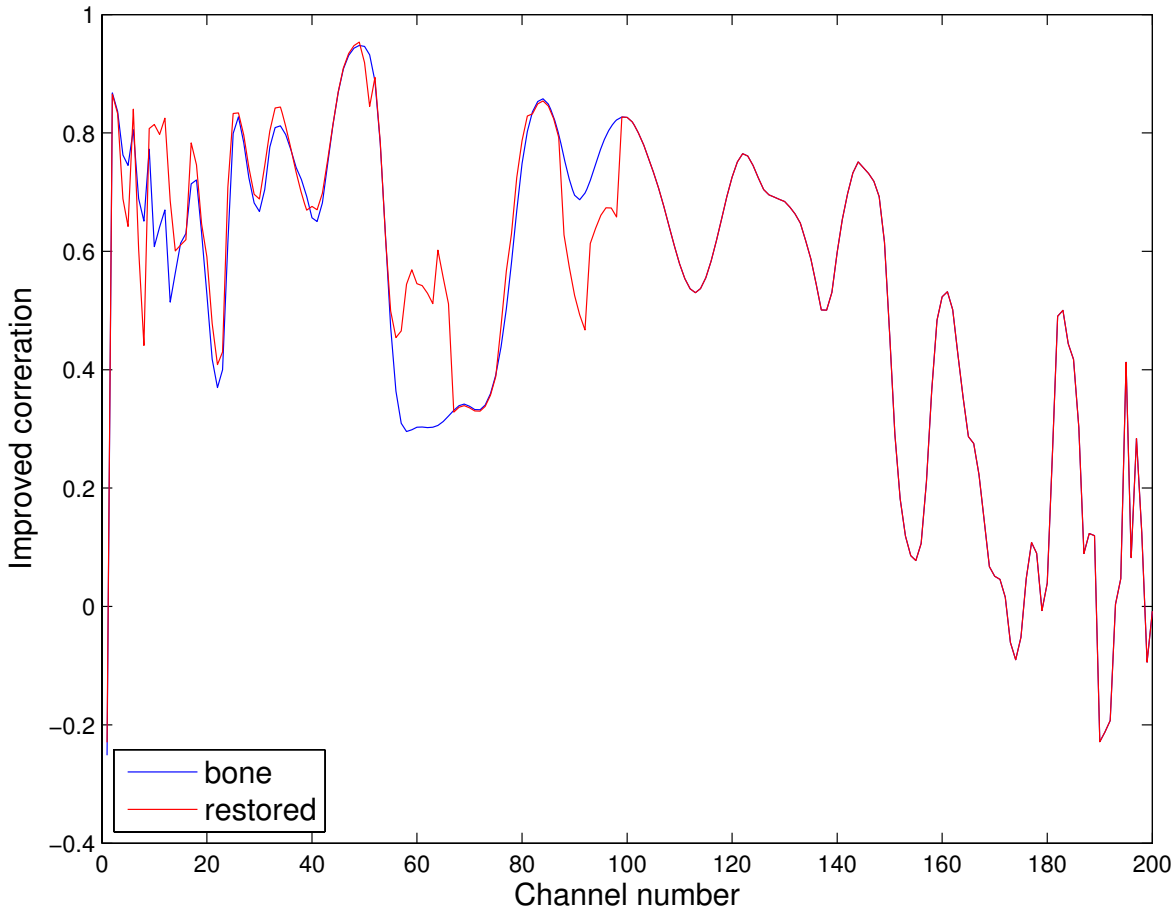

図 4.2: 提案法による相関の改善度.

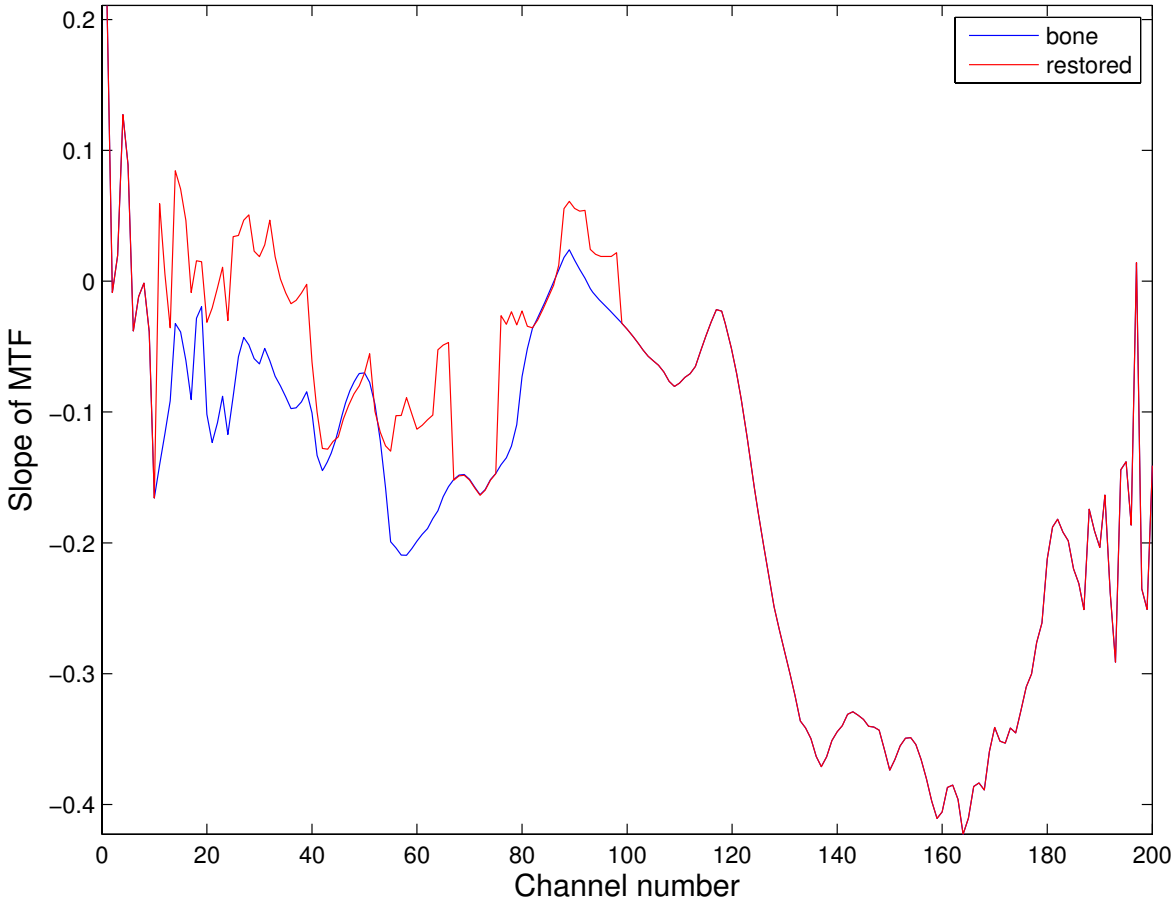

図 4.3: 提案法による MTF の回帰直線の傾きの改善度.

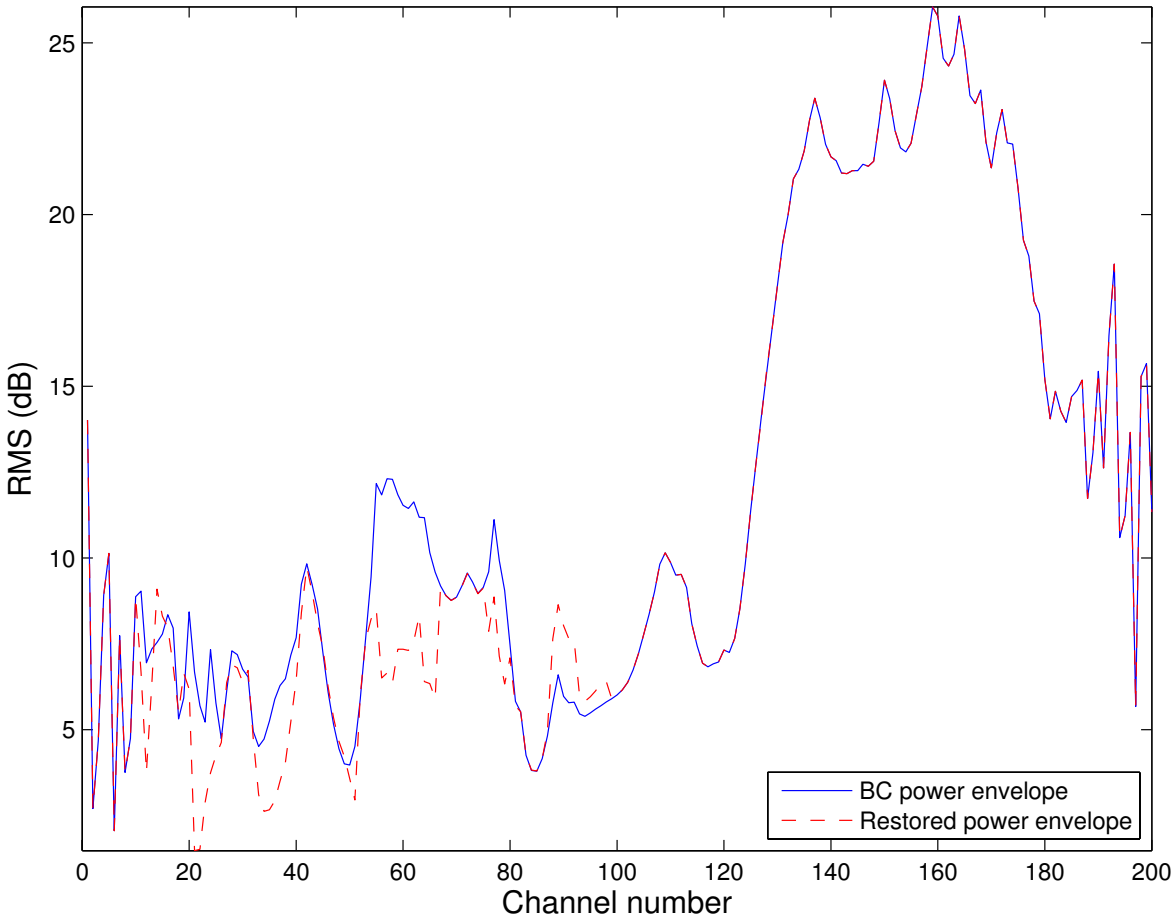

図 4.4: 提案法による変調度 1 の MTF と骨導/回復音声の RMS 誤差の改善度.

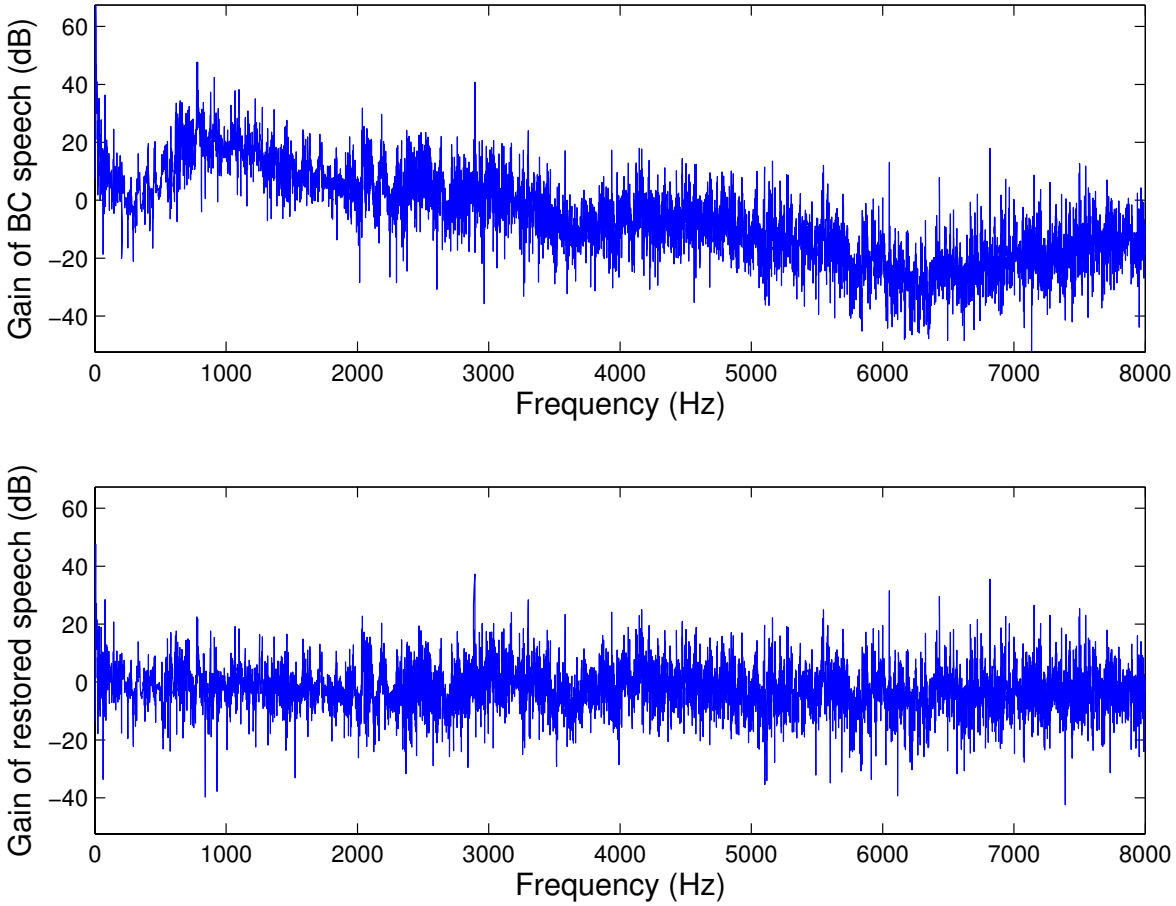

図 4.5: 提案法による伝達関数の改善度.

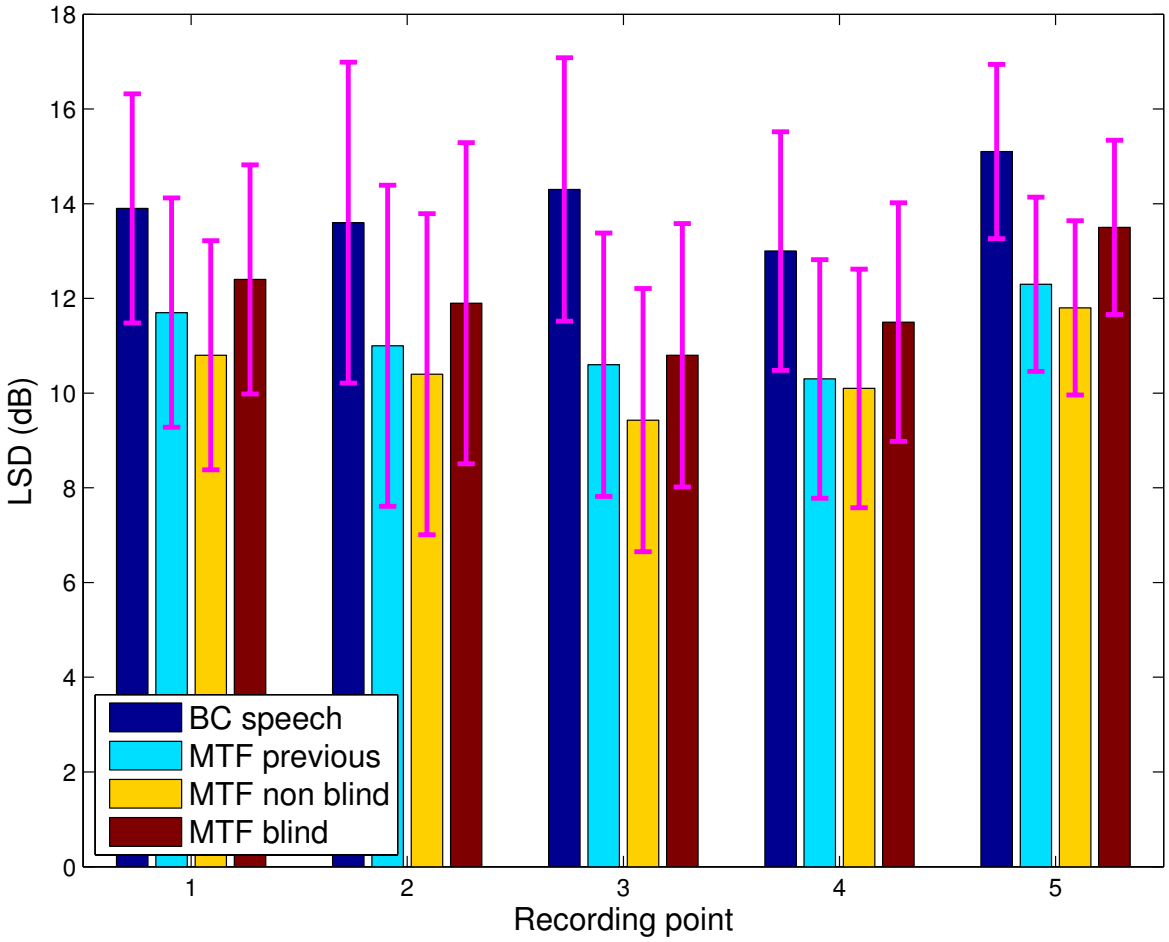

図 4.6: LSD による総合評価．BCspeech: 骨導音声，MTF previous: 従来の MTF に基づく骨導音声回復法，MTF nonblind: 気導音声の情報を用いてパラメータ a を求めた提案法，MTF blind: 提案法.

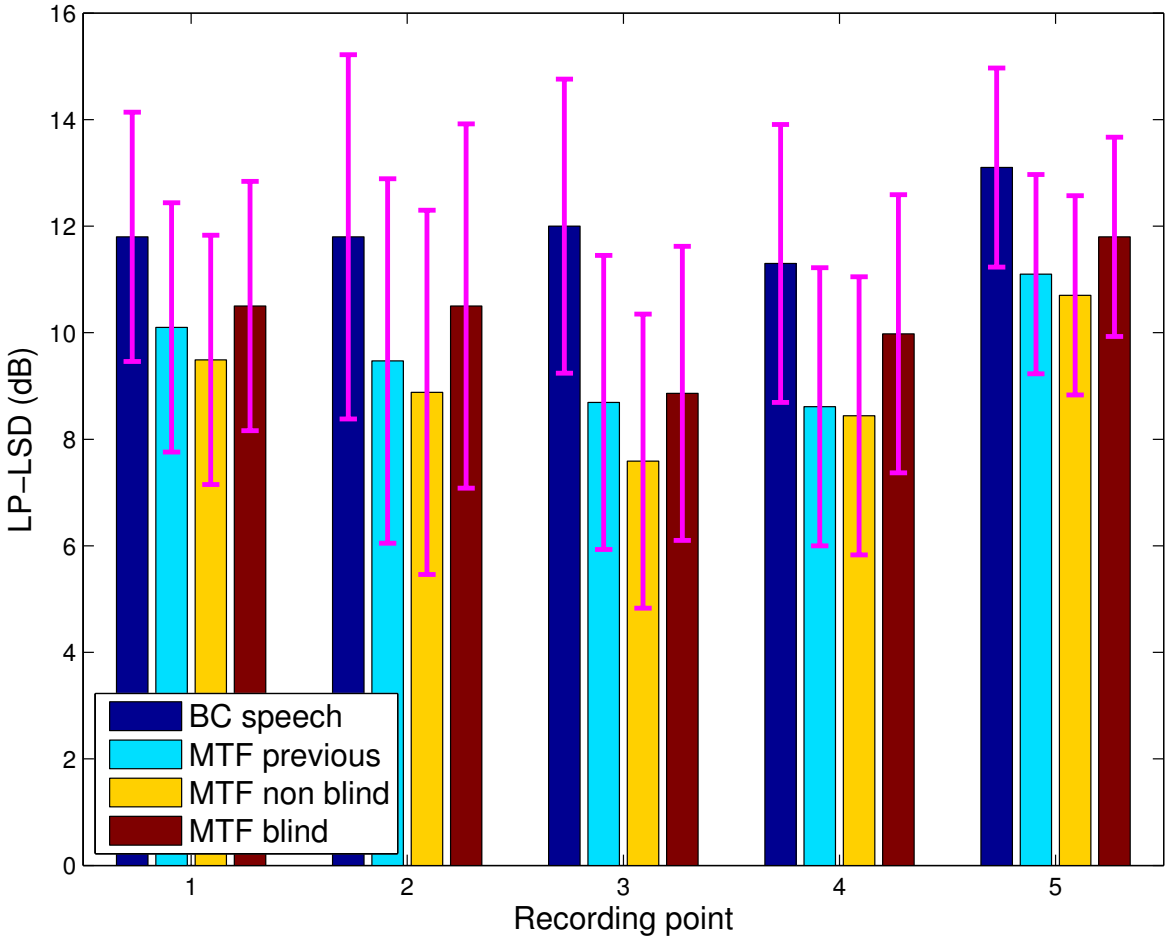

図 4.7: LP-LSD による総合評価．体裁は，図 4.6 と同じ．

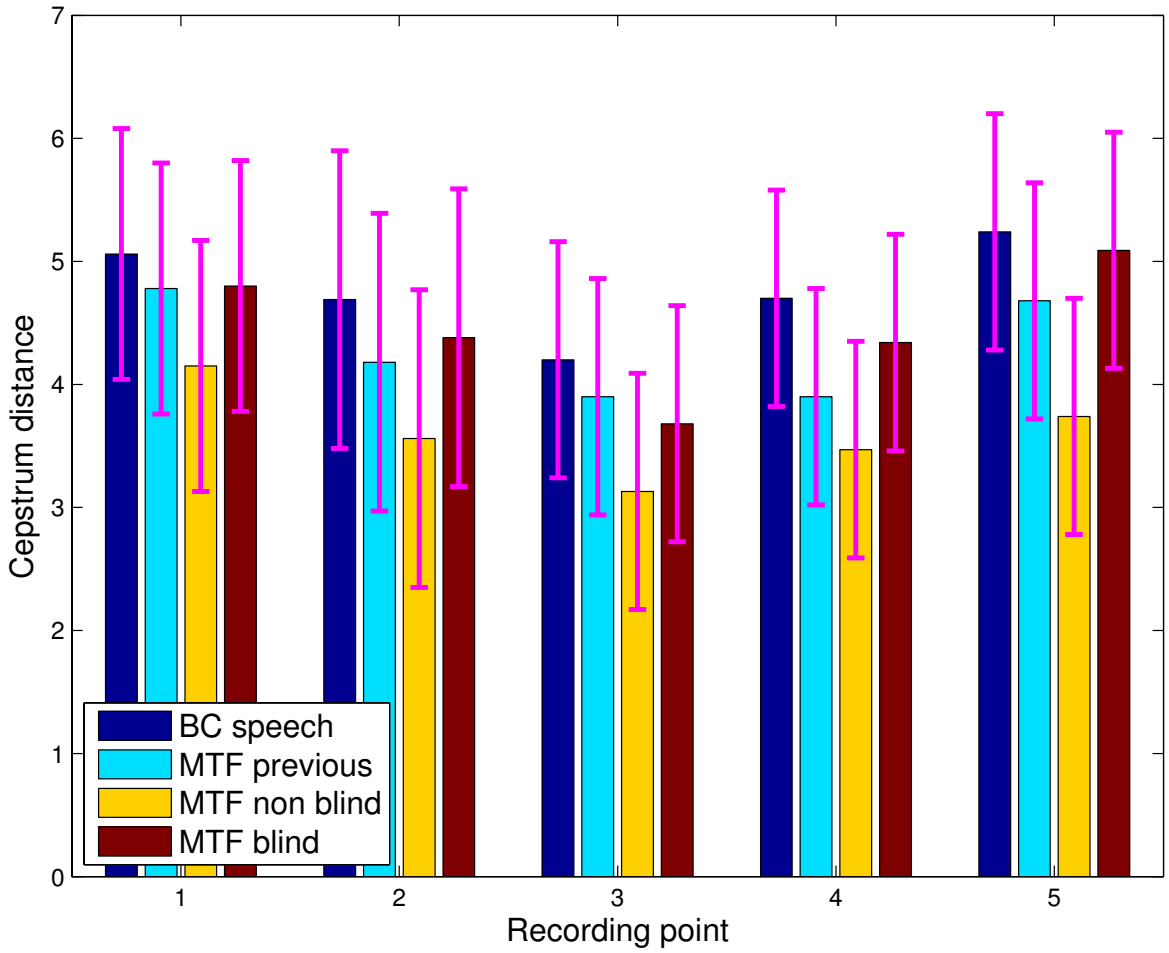

図 4.8: ケプストラム距離による総合評価．体裁は，図 4.6 と同じ．

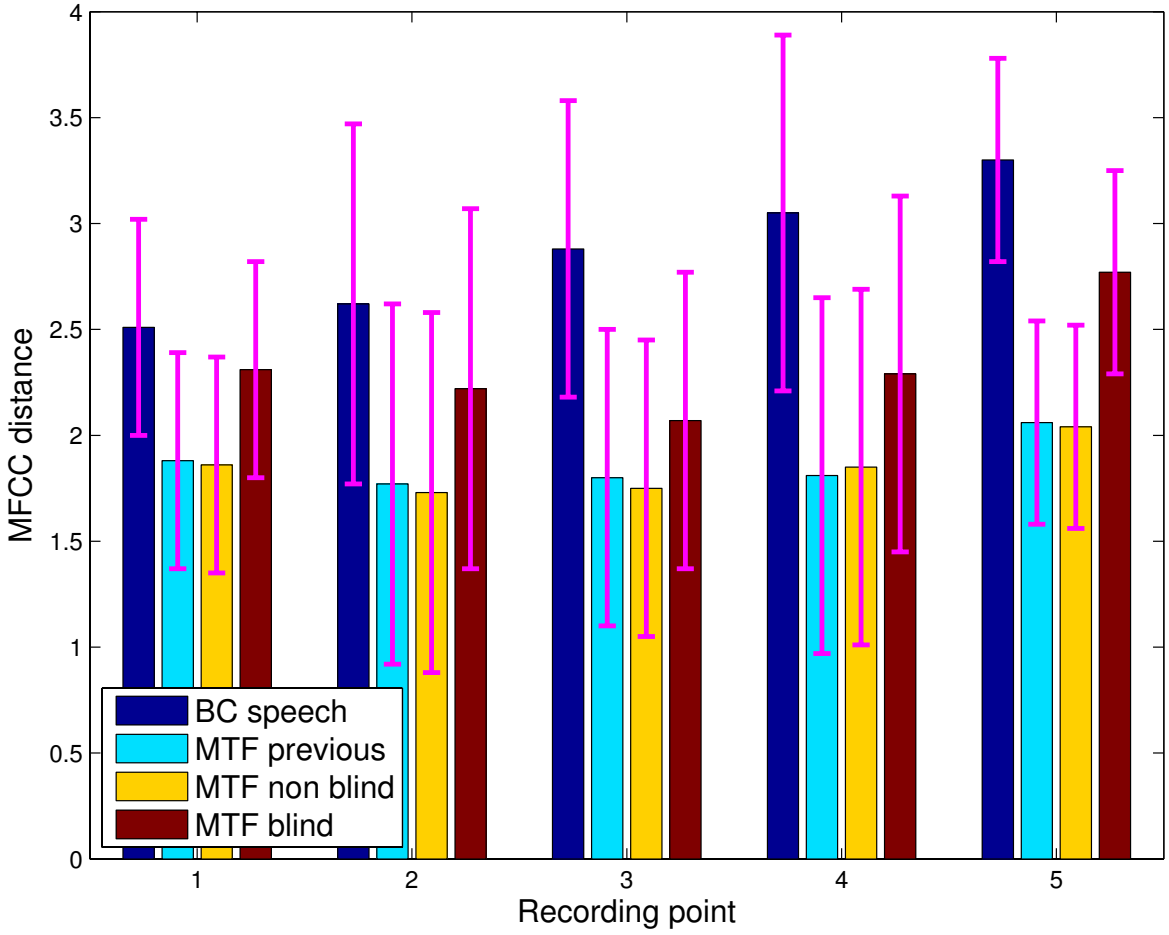

図 4.9: メルケプストラム距離による総合評価．体裁は，図 4.6 と同じ．

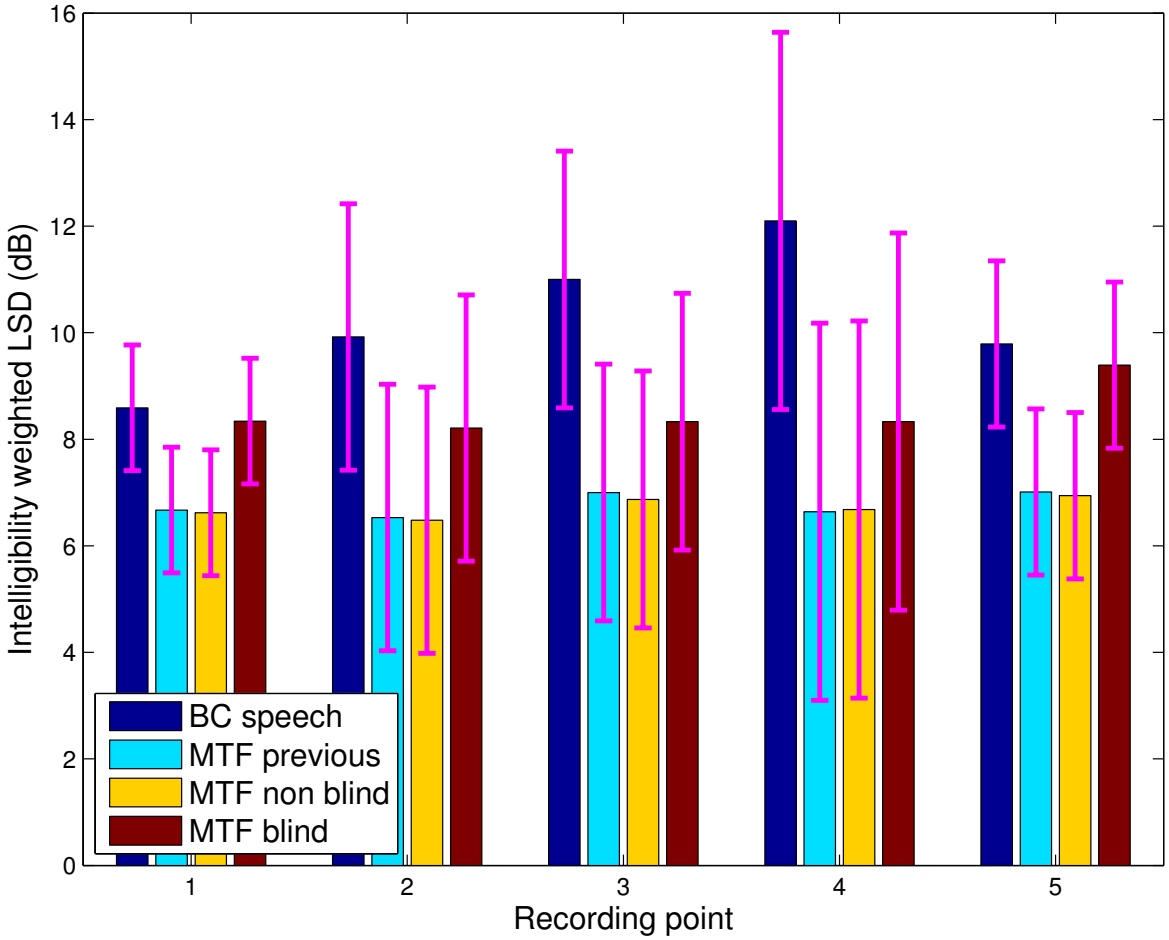

図 4.10: 明瞭度を考慮した LSD による総合評価．体裁は，図 4.6 と同じ．

第5章　結論

5.1　本研究で明らかにしたこと

本研究では，MTF に基づいた処理体系で骨導音声をブラインドで回復できる方法を提案し，その有効性を明らかにした．

気導パワーエンベロープと骨導パワーエンベロープの間の変換関係を解析し，その変換関係は，MTF で表現すると $a\exp(-bt)$ という 2 つのパラメータ a，b を持つ関数で表現できることを明らかにした．次に，観測点毎にパラメータ a がフィルタバンクのチャンネル数を従属変数として持つ回帰曲線 $1/a_n^2 = cn^{-1} + d$ で近似できることを明らかにした．この回帰曲線を用いることにより，パラメータ a を気導音声の情報を必要とすることなく決定できた．また，この回帰曲線は発話内容に依存せず，一部話者を除いた大多数の話者でパラメータ a は一致することが明らかとなった．これにより，話者や発話内容によらない MTF に基づいた骨導音声ブラインド回復法を提案した．最後に，提案法の客観評価をシミュレーションにより行い，提案法が骨導音声の音質，明瞭度の改善に確かに有効であることを明らかにした．

5.2　今後の課題

今回の検討では，話者，発話内容によらず音声を回復できる可能性が示唆されたが，これはあくまで限られたデータベース内でしか検証されていない．よって，今後別の話者，発話内容を収録した更なるデータを用いて提案法の有用性を検証する必要がある．また，提案法は回復音声の時間長の制限を設けていない，どの程度の長さの音声まで回復可能かについて，議論する必要がある．また，提案法は時間ドメインでの歪を回復する手法である．これと，Vu らにより提案されている周波数ドメインでの歪を回復する線形予測分析に基づく回復法を組み合わせることで，時間-周波数方向の歪を同時に回復できる可能性がある．

謝辞

本研究を進める上で大変熱烈なご指導を下さった指導教官である鵜木祐史准教授に甚大なる感謝の意を表します．また，研究室会議など，様々な場面でご助言くださった赤木教授に感謝の意を表します．さらに，研究に対して様々なご助言を下さいました李助教，研究室の先輩方，苦楽を共にした研究室の諸氏にも感謝の意を表します．

参考文献

[1] 熊下 正照, 島村 徹也, 鈴木 誠史, “骨導マイクを用いて収録した音声の性質,” 日本音響学会講演論文集. 2-Q-3, pp.269–270, March 1996.

[2] 北森 進, 滝沢 正浩, “明瞭度試験による骨導音声の分析,” 電子情報通信学会論文誌, vol. J72-A, no. 11, pp. 1764–1771, Nov. 1989.

[3] 衣笠 光太, ルー シュガン, ヴ タング タット, 鵜木 祐史, 赤木 正人, “線形予測分析に基づいた骨導音声ブラインド回復法の総合評価：Lombard 効果による影響について,” 日本音響学会講演論文集, 1-R-8, Sep. 2008.

[4] 齋藤 裕, 新垣 拓也, 長尾 優, 福島 学, 石光 俊介, 柳川 博文, “振動ピックアップ型マイクロホンを用いた収音音声の装着部位による変化,” 日本音響学会講演論文集，3-P-22, pp. 623–624, Sept. 2002.

[5] 加村 健一郎, 齊藤 裕, 福島 学, 石光 俊介, 柳川 博文, “振動ピックアップ型マイクによる収録音声の特性補正について,” 日本音響学会講演論文集, 1-Q-13, pp. 661–662, Mar. 2002.

[6] Maranda McBride, Meghan Hodges and Jon French, “Intelligibility differences between male and female bone conducted speech when presented in high noise environments,” *J. Acoust. Soc. Am.*. vol. 122, isuue 5, p.3064, Nov. 2007.

[7] Takeshi Tomikura. and Tetsuya Shimamura., “A study on improving the quality of voice of bone conduction,” *Proceedings 2003 spring meeting on Acoustical Society of Japan*, 2-Q-14, pp. 401–402, 2003.

[8] Shunsuke Ishimitsu, Hironori Kitakaze, Yasuyuki Tsuchibushi, Hirofumi Yanagawa, and Manabu Fukushima, “A noise-robust speech recognition system making use of body-conducted signals,” *Acoustical Science and Technology*, vol. 25, no. 2, pp. 166–169, 2004.

[9] Toshiki Tamiya. and Tetsuya Shimamura, “Reconstruct Filter Design for Bone-Conducted Speech,” *Proc. ICSLP2004*, vol. II, pp. 1085–1088, Oct. 2004.

[10] 石光 俊介, 高家 陽介, 堀畑 聡, 北風 裕教, 柳川 博文, “適応フィルタを用いた骨導音明瞭度向上の基礎研究,” 日本音響学会講演論文集, 1-Q-23, pp. 681–682, Mar. 2002.

[11] Tat Thang Vu, Germine Seide, Masashi Unoki, and Masato Akagi, “Method of LP-based blind restoration for improving intelligibility of bone-conducted speech,” *Proc. Interspeech2007*, pp. 966–969, Antwerp, Belgium, Aug. 2007.

[12] Kenji Kimura, Masashi Unoki, and Masato Akagi, “A study on a bone-conducted speech restoration method with the modulation filterbank,” *NCSP05*, pp. 411–414, Honolulu, USA, Mar. 2005.

[13] Tat Thang Vu, Kenji Kimura, Masashi Unoki, and Masato Akagi, “A study on restoration of bone-conducted speech with MTF-based and LP-based model,” *Japan Signal Processing*, vol. 10, no. 6, pp. 4070–417, Nov. 2006.

[14] Tat Thang Vu, Masashi Unoki, and Masato Akagi, “A study on an LP-based model for restoring bone-conducted speech,” the International Conference on Communications and Electronics (*Proc. ICCE’ 2006*), pp. 294–299, Hanoi, Vietnam, Oct. 2006.

[15] Tat Thang Vu, Masashi Unoki, and Masato Akagi, “A study on the LP-based blind model in restoring bone-conducted speech,” *IEICE Technical Report*, SP2007-189, pp. 53–58, Mar. 2008.

[16] T. Houtgast and H. J. M. Steeneken, “The Modulation Transfer Function in Room Acoustics as a Predictor of Speech Intelligibility,” *J. Acoust. Soc. Am.*, vol. 54, Issue 2, pp. 557, 1973.

[17] T. Houtgast, H. J. M. Steenken and R. Plomp, “Predicting Speech Intelligibility in Rooms from the Modulation Transfer Function. I. General Room Acoustics,” *Acustica*. vol. 46, 1980.

[18] T. Houtgast, H. J. M. Steenken, “A review of the MTFconcept in room acoustic and its use for extimating speech intelligibility in audioria,” *J. Acoust. Soc. Am.*, vol. 77, No. 3, pp. 1069–1077, Mar. 1985.

[19] 小椋 靖夫, 浜田 晴夫, 三浦 種敏, “音場における音声伝送品質のための MTF と STI について,” 日本音響学会誌, 40 巻 3 号, pp. 181–191, Mar. 1984.

[20] R. Drullman, “Temporal envelope and fine structure cues for speech intelligibility,” *J. Acoust. Soc. Am.*, vol. 97, pp. 585–592, Jan. 1995.

[21] 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，vol. J81-A, no. 10, pp. 1323–1330, Oct. 1998.

[22] M. R. Schroeder, “Modulation transfer function: definition and measurement,” *Acustica.*, vol. 49, pp. 179–182, 1981.

[23] M. Unoki, M. Furukawa, K. Sakata, and M. Akagi, “An improved method based on the MTF concept for restoring the power envelope from a reverberant signal,” *Acoust. Sci. & Tech.* vol. 25, no. 4, pp. 232–242, 2004.

[24] T. Arai, M. Pavel, H. Hermansky, and C. Avendano, “Syllable intelligibility for temporally filtered LPC cepstral trajectories,” *J. Acoust. Soc. Am.*, vol. 105, no. 5, pp. 2783–2791, May 1999.

[25] N. Kanedera, T. Arai, H. Hermansky, and M. Pavel, “On the importance of various modulation frequencies for speech recognition,” *Proc. Eurospeech97*, pp. 1079–1082, Rhodes, 1997.

[26] 金寺 登，荒井 隆行，船田 哲男，“変調スペクトルの重要な成分のみを選択的に用いた雑音に強い音声認識,” 電子情報通信学会論文誌，vol. 84, no. 7, pp. 1261–1269, July 2001.

[27] Database for speech intelligibility testing using Japanese word lists. NTT–AT, March 2003.

[28] Shuichi Sakamoto, Naoki Iwaoka, Yoiti Suzuki, Shigeaki Amano and Tadahisa Kondo, “Complementary relationship between familiarity and SNR in word intelligibility test,” *Acoustical Science and Technology*, vol. 25, no. 4, pp. 290–292, 2004.

[29] 北風 裕教, 村中 裕貴, 石光 俊介, “体内伝導音声認識システム構築に関する一考察,” 日本音響学会講演論文集, 1-10-20, pp. 539–540, March 2004.

[30] Sota Hiramatsu, and Masashi Unoki. “A Study on the Blind Estimation of Reverberation Time in Room Acoustics,” *J. Signal Processing*, vol. 12, no.4 , pp. 323–326, July 2008.

[31] ANSI S3.5-1997, “American National Standard Methods for Calculation of the Speech Intelligibility Index,” 1997.

研究実績

- 衣笠 光太, ルー シュガン, ヴ タングタット, 鵜木 祐史, 赤木 正人, “線形予測分析に基づいた骨導音声ブラインド回復法の総合評価：Lombard 効果による影響について,” 日本音響学会講演論文集, 1-R-8, Sep. 2008.

- Kota Kinugasa, Masashi Unoki and Masato Akagi, “An MTF-based blind restoration method for improving intelligibility of bone-conducted speech,” *NCSO09*, Honolulu, USA, Mar. 2009.