

音声の明瞭度改善を目的とした振幅レベル変調

高岡, 幸裕 / TAKAOKA, Yukihiro

(発行年 / Year)

2009-03-24

(学位授与年月日 / Date of Granted)

2009-03-24

(学位名 / Degree Name)

修士(工学)

(学位授与機関 / Degree Grantor)

法政大学 (Hosei University)

2008 年度 修士論文

音声の明瞭度改善を

目的とした振幅レベル変調

指導教員 渡辺嘉二郎 教授

大学院工学研究科

システム工学専攻修士課程

07R6124

タカオカ ユキヒロ

高岡 幸裕

目次

Abstract	3
1 はじめに	4
2 振幅レベル変調	6
2.1 スペクトル推定方式	7
2.1.1 FFT 処理と帯域レベル推定	7
2.1.2 振幅ヒストグラム	7
2.2 変調方式	8
2.2.1 周波数領域変調	8
2.2.2 時間領域変調	8
2.3 音声特徴データを考慮した振幅変調	8
3 システム概要	9
4 合成フィルタによる振幅レベル変調	10
4.1 提案する振幅変調処理システム構成	10
4.2 5ch フィルタバンク	11
4.3 変調フィルタの生成	13
4.3.1 基本フィルタ群	13
4.3.2 合成フィルタの生成	15
4.4 振幅変調処理	18
4.4.1 合成フィルタの特性制御	18
4.4.2 振幅レベル変調特性（静的変調特性）	19
5 音声特徴データによる変調レベル算出	20
5.1 変調レベル算出システム構成	20
5.2 振幅ヒストグラムに含まれる特徴データ	21
5.3 特徴データの抽出処理	23
5.3.1 相互相関係数	23
5.3.2 特徴データを考慮した振幅制御	24
5.4 特徴抽出に基づく振幅レベル制御（動的変調特性）	25

6 実験	27
6.1 物理評価実験	27
6.1.1 動的制御実験	27
6.1.2 実験結果	27
6.2 感性評価実験	29
6.2.1 高齢者を模擬した聴覚特性	29
6.2.2 評価実験	30
6.2.3 実験結果	31
7 考察	33
7.1 物理評価考察	33
7.2 感性評価考察	33
8 おわりに	34
付録	35
謝辞	42
参考文献	43

Amplitude modulation system for definition improvement of sound

Abstract: The aim of this study is to develop a speech pre-processing system used with the public address system in every environment. The pre-processing system is intended to improve the intelligibility of the public address system, without increasing the sound pressure level of the voice. Furthermore, it's aimed at performing the handling of maintained characteristic sound data. To examine the effectiveness of the system, an experimental system, which can process voice approximately in real-time, was developed. The system controls the modulation level for each spectral analysis of the input voice. In this study, the sound amplitude modulation that is validated by this system, I suggest the new method that improved the conventional method. The conventional method is band distinction modulating for independence. By quantity of calculation and complexity of processing, the detailed data of conventional frequency analysis, the burden of processor is vast. For this, the new method is a composite filter which considered reduction of quantity. It is operation for distortion that reduction of the modulation voice. The new suggestion, amplitude histogram divided in each frequency band, the extraction method of characteristic sound data is demanded. These suggestions are thought about effectively in this system. The results implied the effectiveness of the processing used in this study in respect to improving the intelligibility.

Key Words: *voice, hearing, definition, characteristic, amplitude modulation, amplitude histogram, digital signal processor*

1 はじめに

聴覚特性については、加齢に伴い劣化することが知られている。特に高齢者では周波数選択能力の低下がみられる。そのため、音韻の識別に要される音声の中の周波数成分が周囲音によりマスクされやすくなり、伝達する距離に応じて音が減衰することによって聞き取りが困難になる。このため音声聴取においては、若年者よりも難をきたしやすい。

しかしながら、音声拡声装置を必要とする若年から高齢の幅広い利用者を有する場所では、利用者全般に配慮した聴環境を整備するという観点から、特に高齢者の聴覚特性に配慮した音声拡声を行う必要があるだろう。

この単純な解決法は、音声の主観的音量感を増やすことである。拡声の出力音圧レベルを上げることは、出力される目的音声を周囲音にマスクされにくくさせる、そして音の伝達による減衰を低減させる効果的な方法である。実際、鉄道の駅のような雑音下で用いられる放送設備の一部として、周囲音の周波数分析結果に応じてマスクされそうな周波数帯域についての目的音声の音圧レベルを変化させる装置が市販されており、若年者を被験者とした雑音下での音声聴取実験において明瞭度改善に効果があったと報告されている。ただし、雑音に応じて放送の音圧レベルが上がることから、同装置により高齢者にとっての明瞭度が十分に改善されるような使い方をした場合、若年者には煩わしさを感じさせる可能性がある。また、被験者には高齢者が含まれておらず、一方で同装置の使用に伴う煩わしさの変化については述べられていない。さらに、同装置における音声の聞き取りとは、周囲音にマスクされた部分を補うための音圧レベルの変化であって、そのときの音声の特徴との関連については述べられてはいない。ここで言う”音声の特徴”とは、”音”を”その音”として認識することである。例えば知人の声を、面と向かって会話している聞こえ方も、電話越しに話す聞こえ方も、同一人物の声として判断できるということの意味している。

この改善手法として、デジタル補聴器で検討されているデジタルシグナルプロセッサ（DSP : Digital Signal Processor）の一つである多帯域振幅変調がある。振幅変調とは振幅の変化を減じる処理のことである。音声信号を振幅変調するという事は、小振幅の部分については振幅が拡大されて、大振幅の部分については振幅が縮小される。多帯域振幅変調は、音声信号を複数の周波数帯域に分けた上で、このような振幅変調を施すものである。この方法を音声拡声に応用することにより、放送の平均音圧レベルを上げることなく雑音下において高齢者にとっての明瞭度を改善できる可能性がある。またこの方法では、出力音圧レベルが不変であることから、若年者に煩わしさを感じさせる可能性も少ないと期待される。

振幅変調を施す箇所は、従来の帯域別に分かれた周波数分析の結果に基づいて、各周波数の変調割合を制御する手法ではなく、帯域別に求められた振幅ヒストグラム、つまりその音声の特徴が強く出ているデータを振幅変調する新たな手法である。この新たに提案する手法では、演算量の大幅な軽減により、従来、不可能であったリアルタイム処理の実現に期待が持てる。

本研究では、音声の特徴を保持したままで各周波数帯域の振幅変調を行うという目的で、新たな音声処理方式の実現を目的としている。

処理の物理的評価は、音の特徴データを保持した上での振幅変調が実現されているか検証する。従来の各周波数での独立的な圧縮では、単純に入力された音声の聞き取りを改善させるように高域強調を施したものが多く、その手法では音声の特徴を保持した処理とは言い難い。そこで、処理する周波数帯は人の耳の

聞こえに準ずる倍音（ローパス、ハイパス、バンドパスフィルタでのカットオフおよび中心周波数が前のその2倍となる）でのフィルタ群で分割し、各フィルタを独立的な圧縮を行うのではなく、音声全体の周波数の出現頻度から、各フィルタに振幅変調の度合いを出力する処理を施す。この処理方法で音声の特徴が保持されているかを検証するために、音声の処理前データと処理後データを、相互相関を取って比較した。

処理の感性的評価は、逐次的な目的音声の各周波数帯域の変調度合いを自動的に制御することにより、目的音声全体としての音圧レベルを上げることなく、雑音下における聞き取りを改善するような音声前処理システムを構築する。その実現のためには、周囲音の周波数分析の結果と目的音声に施す変調の度合いとの間の関数関係を記述する必要があるが、純音相互間のマスキングとは異なり、現実の雑音と音声を想定した場合、その関数関係を明示的に記述することが困難である。そこで、被験者を有する実験の結果に基づき、この関数関係を信号対雑音比率（SN比：Signal to Noise Ratio）としてデータを取り、あらゆる雑音環境下において最適制御するシステムを構築した。そして、雑音下における本システムによる聞き取り改善の評価について実験的に検討した。

なお、一般に言う“聞き取りやすさ”とは明瞭度であり、これは単語や音節の正聴率のことであるが、本研究においては単語了解度、すなわち単語についての正聴率という意味で用いている。一般的な単語や音節の正聴率と異なり、静寂な環境下で健聴者に十分な音圧レベルで呈示すれば、単語についての正聴率はほぼ100%となる。

2 振幅レベル変調

近年、デジタル信号処理技術とそのハードウェアの進歩はめざましく、特に信号処理専用のプロセッサである DSP により、アナログ方式では困難な処理でもリアルタイムで実現することが可能となった。オーディオ分野でも複雑な非線形処理をプログラマブルに実行できるフルデジタルアンプの開発が進んできた。とりわけ補聴器の分野ではその可能性が広がり、デジタル補聴器の開発が試みられている。

デジタル補聴器に用いられる DSP として、可聴しきい値レベル (Hearing Threshold Level) と不快聴取レベル (Uncomfortable Level) で挟まれた聴力ダイナミックレンジに信号レンジを適合させるための振幅レベル変調というものがある。ただし、この変調特性は周波数ごとに異なるため、入力レベルと周波数を変数とする非線形変調となる。

振幅を増大、減少させることから、振幅圧縮と定義されることもあるが、本研究では、実際の音圧レベルを不変にした上での新婦布増大、減少であることから、振幅レベル変調と定義する。

本研究で必要となる信号レンジ適合のための振幅レベル変調では、周波数ごとの変調特性が異なるため、入力レベルと周波数を変数とする非線形変調となる。³⁾ このような振幅変調処理の例として、フィルタバンクで分割された帯域ごとの信号を変調する多帯域振幅変調 (Multiband modulation) が挙げられる。しかし、帯域ごとの独立な変調となるために、音声スペクトル構造の詳細な部分が損なわれ、破裂音群や母音群内の異音の原因となることが指摘されている。このような帯域別変調に伴うスペクトルひずみを改善するために広帯域変調 (Wide-band Modulation) の考え方が示された。これは、スペクトルひずみの詳細な部分を保持しようとするものである。

例えば、主成分変調では、16ch フィルタバンク出力レベル (16 次元ベクトル) の 5ms ごとの標本群から共分散行列を求め、主成分分析を行う。この第 1、第 2 主成分に対応する固有ベクトルが短時間スペクトルの平均レベル成分と傾斜成分の近似スペクトルをそれぞれ表すことから、そのようなスペクトル概形と所要変調比で定まる利得でフィルタバンク出力の重み付け加算を行うものである。これは、帯域独立性を有する従来の多帯域振幅変調と異なり、短時間スペクトルの概形を考慮した多帯域振幅変調とみなせる。直行多項式変調では、高速フーリエ変換 (FFT : Fast Fourier Transformation) により、得られた短時間スペクトルを直行関数群で展開し、その低次の関数のみに変調を施し、逆高速フーリエ変換 (IFFT : Inverse Fast Fourier Transformation) により音声波形に戻している。

この変調の例は、短時間スペクトルの詳細な形状を保持しながら、多帯域振幅変調を行うという点では共通しているが、圧縮方式とスペクトル推定方式がそれぞれ異なっている。演算量低減の観点からその特徴を挙げる。

2.1 スペクトル推定方式

2.1.1 FFT 処理と帯域レベル推定

FFT 処理では,処理サンプル数を増やすことで詳細なスペクトル構造を得られる.しかし,プロセッサに負担がかかる.さらに,固定フレーム長であることから,過渡的な破裂音や定常的な母音のように時間変化特性が相反するスペクトル変化に追従することが難しい.

帯域レベル推定では,平滑化用フィルタの時定数の選択に自由度があることから,スペクトル変動を捉えやすいが,チャンネル数が少ないと詳細なスペクトル構造が損なわれることになる.

2.1.2 振幅ヒストグラム

音などの振動を,時間とともに変化する現象のもつ意味を適確にとらえるためには,その頻度分布や相関関係を知ることが重要となる.長時間にわたる現象の変化の中で,どの程度の量がどの程度の割合で発生していたのか,ヒストグラムは,このような情報を知るために有効なデータとなり得る.

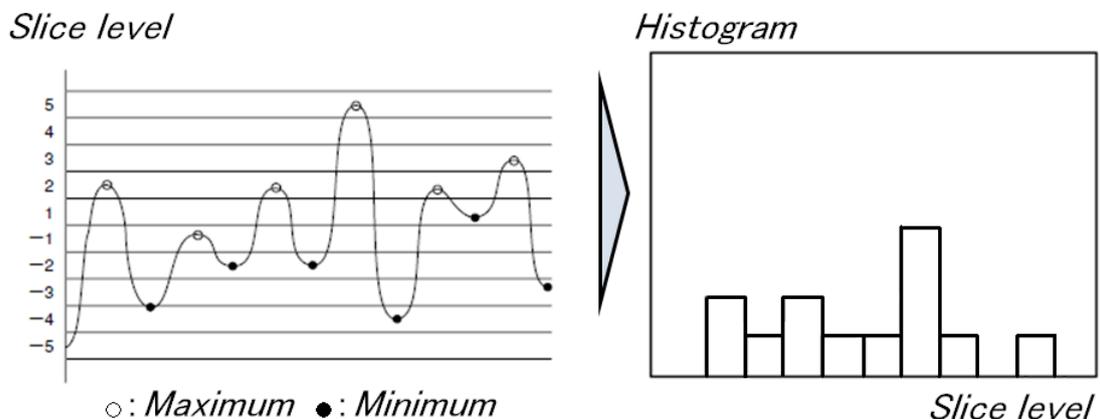


図1 ヒストグラム変換例

Fig.1 Example of histogram conversion

ヒストグラムは累積特徴であるために,信号の変形による影響を受けにくい.また,ヒストグラムによるデータ解析は,特徴を直接解析するのに比べて,それ自体の演算量が少ない上に,一定の区切られた時間単位で解析されるため,無駄な演算を省くことができる.これらの効果により,実用上十分な解析精度を保ったまま,処理の簡略化を行うことができる.

2.2 変調方式

信号に FFT を施すことによって,時間領域から周波数領域に変換することができる.フーリエ変換は信号情報を周波数毎の大きさと位相に変換する.FFT の結果に対して,各周波数の大きさ成分を二乗してパワースペクトルに変換することが多い.信号を周波数領域で分析する目的は,信号の特性の分析にある.周波数領域への変換手法として,例えばケプストラム (Cepstrum) は入力信号をフーリエ変換で周波数領域に変換し,その対数をとって,再度逆フーリエ変換を施す.これにより,非常に弱い周波数成分を強調することができる. また,自己相関からフーリエ変換によってパワースペクトル密度,またはその逆が成り立つ.

2.2.1 周波数領域変調

振幅変調においては,周波数領域で変調された音声スペクトルの IFFT では,位相特性の付加およびフレーム波形の畳み込み加算が必要なため,処理の難しさと演算量の問題が生じる.BPF 群を用いる変調の場合,変調音声スペクトルにフィルタバンクの総合振幅特性が反映され,チャンネル数やその特性の違いにより帯域間遷移部でのスペクトル歪みが生じる可能性がある.

2.2.2 時間領域変調

帯域間遷移部でのスペクトル歪みを改善する意味では,圧縮フィルタによる時間領域でのフィルタリングが適切と考えられる.しかし,そのフィルタ生成において IFFT が必要となり,演算量の問題が残る.

2.3 音声特徴データを考慮した振幅変調

多帯域振幅変調のスペクトルの詳細を得ることができる,そして広帯域振幅変調のスペクトル概形を保持することができる,という利点に着目すると,多帯域でスペクトル推定を行い,広帯域で振幅レベル変調を施すことが望ましい.

本研究では,詳細なスペクトル構造を得るための多帯域フィルタと,滑らかな振幅レベル変調を施すための広帯域フィルタ合成フィルタを用いた手法を用いる.その変調の度合いを決定するパラメータとしては,振幅ヒストグラムより音声の主要部分を算出する.この合成フィルタによる新たな手法を用いることで,音声の詳細を保持しつつ滑らかな処理を施すことができると共に,大幅な演算量の低減を実現することができる.以下で,処理の流れとシステムの詳細について説明する.

3 システム概要

本システムは、サウンド入出力機能を備えた一般的なパーソナルコンピュータと、その上で実行されるソフトウェア（処理ツールとしては MATLAB を用いた）によって構成されている。この MATLAB で行われる処理をシステム の概念モデルとして図 2 に示す。このシステムはシステム 1 の振幅レベル変調処理を行う部分（System1 : Amplitude Modulation System）と、システム 2 の変調レベルを算出する部分（System2 : Modulation Level Calculation）に分かれている。

システム 1 は、音声の特徴が保持された上で振幅変調が行われているかを検証する部分である。これは、補聴器に用いられるような、入力された音声を各周波数での望まれた振幅レベルに処理するシステムである。

システム 2 は、音声の特徴がどのように分布されているかを算出する部分である。これは、簡易的な音声認証に用いられるような、入力された音声の 3 領域（周波数、振幅、時間）の特徴的なデータを分析するシステムである。

以下で、各システムの詳細な処理の流れについてそれぞれ述べる。

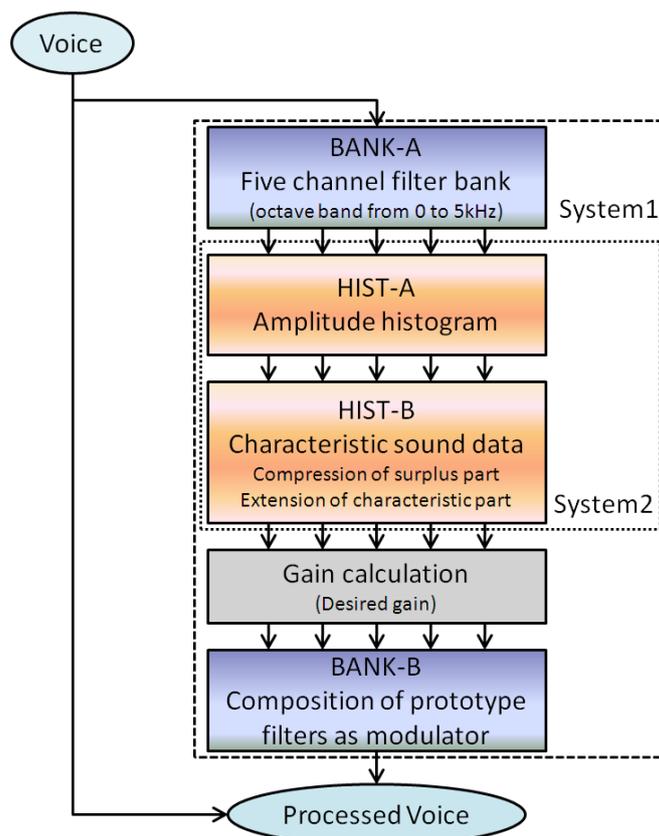


図 2 システムの概念モデル

Fig.2 Schematic model of the system

4 合成フィルタによる振幅レベル変調

本研究では,広帯域振幅変調の特徴を生かしながら,演算量低減を考慮した多帯域振幅変調処理を実現するための新たな変調方式を提案する.本処理は,スペクトル推定用の5ch フィルタバンクと変调用 FIR フィルタで構成される.この変調フィルタは,時間領域で生成可能な合成フィルタとなっている.本変調方式により,演算量低減と同時に滑らかな振幅特性をもつ変調処理が実現できる.以下で提案する合成フィルタによる変調方式について述べる.

4.1 提案する振幅変調処理システム構成

システム構成は,図 3 のように,通常の 5ch のオクターブバンドで帯域分けされたフィルタバンク (BANK-A) と音声特徴データ抽出部 (5 で後述する),ゲイン調節部および,変調フィルタとして 9ch の 1/2 オクターブバンドで帯域分けされた基本フィルタ群 (BANK-B) からなる.⁴⁾⁵⁾

本処理は,帯域レベル推定用のフィルタバンクとレベル変调用 FIR フィルタで構成される.この変調フィルタは,時間領域で生成可能な合成フィルタとなっている.本変調方式により,演算量低減と同時に滑らかな振幅特性をもつレベル変調処理が実現できる.以下で提案する合成フィルタによるレベル変調方式について述べる.ゲイン調節のための音声特徴データの抽出は後述する.

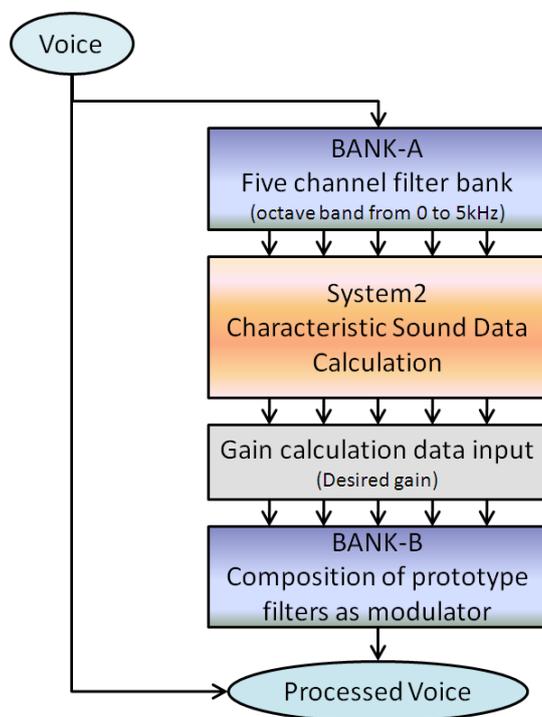


図 3 合成フィルタを用いた振幅レベル変調のブロック図

Fig.3 Block diagram of amplitude modulation using a composite filter

4.2 5ch フィルタバンク

BANK-A は,図 4 の振幅特性を有する 5ch フィルタバンクである.1ch は,カットオフ周波数 250Hz のローパスフィルタ (LPF) である.2~5ch は,中心周波数がオーディオグラム測定点に対応する 500Hz~4kHz のオクターブバンド-バンドパスフィルタ(BPF)群である.

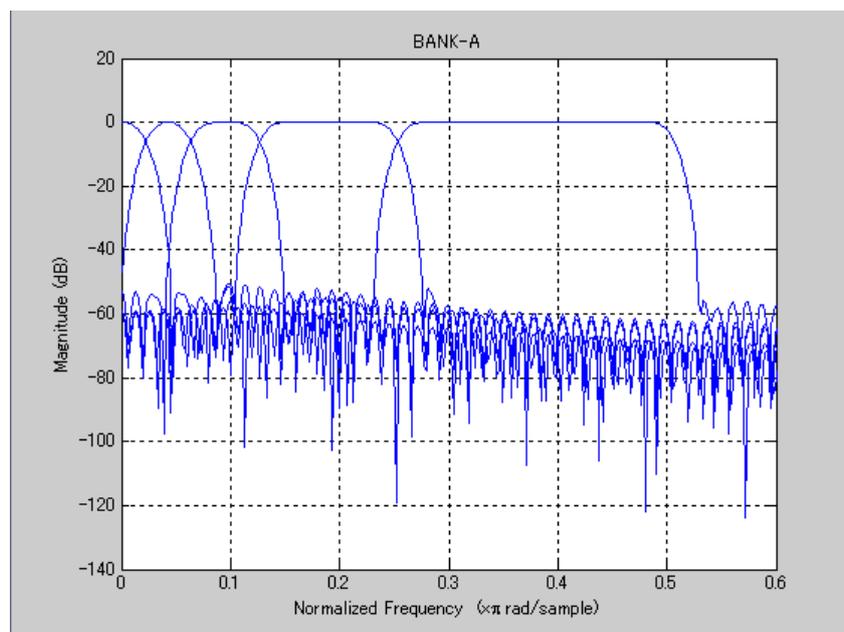


図 4 BANK-A の振幅特性

Fig.4 Amplitude responses of BANK-A

基本フィルタ群には FIR フィルタを用いた.有限インパルス応答 (FIR : Finite Impulse Reply) フィルタは入力信号のみを使うのに対して,無限インパルス応答 (IIR : Infinite Impulse Reply) フィルタは入力信号と共に,それ以前の出力信号も使う.そのため,FIR フィルタは常に安定であるが,IIR フィルタは不安定な場合がある.以下に FIR フィルタについて説明する.

FIR は,デジタルフィルタの一種である.クロネッカーのデルタ入力に対するフィルタの応答特性であるインパルス応答が「有限」であるとは,有限個の標本でゼロに安定することを意味する.対照的に無限インパルス応答 (IIR) フィルタでは,内部フィードバックがあり,無制限に応答し続ける可能性がある. N 次 FIR フィルタは,インパルスに対して $N + 1$ 個の標本まで応答が持続する.

FIR フィルタは,無限インパルス応答フィルタに比較して好ましいとされるいくつかの特性を持つ.

FIR フィルタは本質的に安定している.これは全ての極が原点にあり,したがって単位円内にあるためである.また,FIR フィルタはフィードバックを必要としない.これは,丸め誤差が反復によって蓄積されないことを意味する.同じ相対誤差が計算ごとに生じる.さらに,線形位相となるよう設計可能である.すなわち,位相変化が周波数に比例するように設計できる.これは位相が重要となるような応用で必要となる特性であり,クロスオーバーフィルタ (オーディオスピーカーで複数の音域毎のスピーカーに信号を振り分ける

フィルタ) や透過的フィルタリングが妥当なマスタリングなどがある。
入力信号と出力信号の関係を定義する差分方程式は以下ようになる。

$$y(n) = b_0x(n) + b_1x(n-1) + \dots + b_Nx(n-N) \quad (1)$$

ここで $x(n)$ は入力信号, $y(n)$ は出力信号, b_i はフィルタ係数である. N はフィルタ次数 (Filter order) である. N 次フィルタの場合, 右辺に $N+1$ 個の項がある. これらの項を一般に タップ (Taps) という.

上記の方程式は, フィルタ係数と入力信号の畳み込みとして表現することもできる.

$$y(n) = \sum_{i=0}^N b_i x(n-i) \quad (2)$$

インパルス応答を求めるには, 以下のようにする.

$$x(n) = \delta(n) \quad (3)$$

ここで $\delta(n)$ はクロネッカーのデルタインパルスである. FIR フィルタのインパルス応答は, 以下のような係数 b_n の集合である. ただし, $(0 < n < N)$ とする.

$$H(z) = Z\{h(n)\} \quad (4)$$

$$= \sum_{n=-\infty}^{\infty} h(n) z^{-n} \quad (5)$$

$$= \sum_{n=0}^N b_n z^{-n} \quad (6)$$

インパルス応答の Z 変換は, FIR フィルタの伝達関数となる.

出力は有限個の入力の有限倍の総和であるため, FIR フィルタは明らかに有界入力-有界出力 (BIBO) 安定であり, 入力の実際の最大値の $\sum |b_i|$ 倍以上にはならない.

4.3 変調フィルタの生成

4.3.1 基本フィルタ群

平均ゲイン調整用フィルタ (P-FIL₀ : 全帯域通過) と 9ch ハイパスフィルタを基本フィルタ群と呼ぶ。式 (7), (8), (9), (10) および図 5 にそれぞれの理想振幅特性を示す。図の対数周波数軸上で等間隔に並ぶ接点周波数は, 上述 9 点の中心周波数, およびカットオフ周波数にそれぞれ等しいとする。

P-FIL₀ については, 以下の式のようになる。

$$H_0(f) = 1.0, \quad (0 \leq f < F_C) \quad (7)$$

P-FIL_k については, 以下の式のようになる。

$$H_k(f) = 0, \quad (0 \leq f < f_{k-1}) \quad (8)$$

$$H_k(f) = \frac{\log f - \log f_{k-1}}{\log f_k - \log f_{k-1}}, \quad (f_{k-1} \leq f < f_k) \quad (9)$$

$$H_k(f) = 1.0, \quad (f_k \leq f < F_C) \quad (10)$$

ここで, F_C は信号帯域の最大周波数 (5,512.5Hz) である。各基本フィルタのフィルタ係数は BANK-A の BPF 群と同様の方法で求めた。左から $k=1, 2, \dots, 9$ とする。図 4-3-1 の青線がインパルス応答により計算される振幅特性の実現値である。低域において比較的大きな誤差は, 各基本フィルタのフィルタ長を増やすことで改善は可能である。

しかし, 等間隔周波数標本化でのフィルタ設計であることと, 基本フィルタの理想特性が対数周波数軸上での線形特性として与えられるため, この誤差は完全に除去できない。むしろ, それらの総合特性に着目する広域変調の必要性和演算量低減という点に考慮して, ここでは基本フィルタ群を用いた。

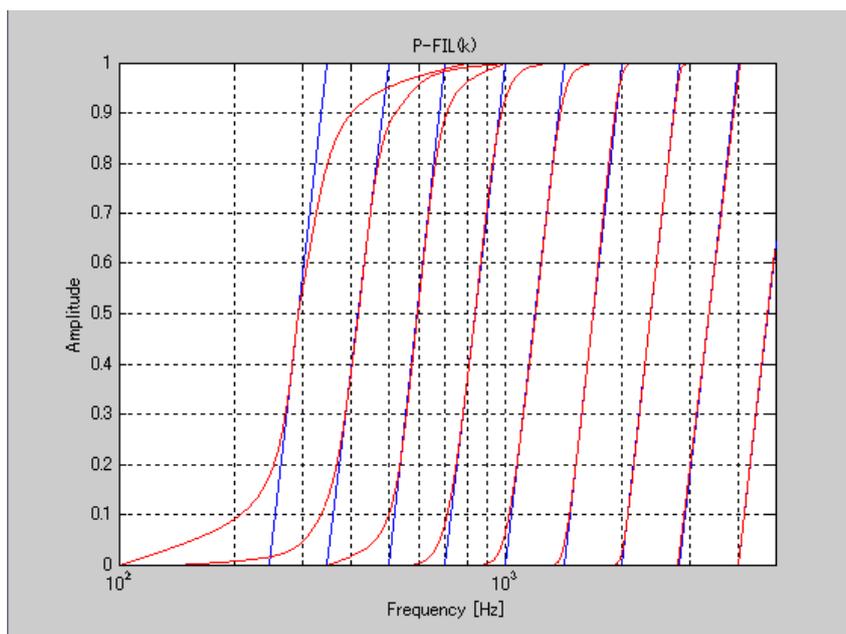


図 5 BANK-B の各フィルタの振幅特性
 Fig.5 Amplitude response for each filter of BANK-B

4.3.2 合成フィルタの生成

上記の基本フィルタ群の並列接続構成を一種のフィルタバンク (BANK-B) とみなして,その総合振幅特性 $H_c(f)$ を求めると,図 6 (青線:理想特性,赤線:実現値) に示すように,250Hz 以上で傾き 1 の高域強調特性となる.このような総合特性を有する BANK-B を合成フィルタ (C-FIL) と呼ぶ.

今,基本フィルタのインパルス応答 $h_k(n)$ を係数 a_k で重み付けした C-FIL に, $x(n)$ を入力した場合の出力応答 $y(n)$ は次式で表される.

$$y(n) = \sum_{k=0}^9 \sum_{l=-L/2}^{L/2} a_k h_k(l) x(n-l) \quad (11)$$

$$= h_c(0)x(n) + \sum_{l=1}^{L/2} h_c(l)\{x(n-l) + x(n+l)\} \quad (12)$$

ただし,

$$h_c(l) = \sum_{k=0}^9 a_k h_k(l) \quad , \quad (0 \leq l < L/2) \quad (13)$$

である.すなわち,各基本フィルタ係数群と重み係数を式 (13) にしたがって積和演算することで,帯域レベルの折れ線近似として任意の振幅特性を有する直線位相 FIR フィルタを時間領域で生成できる.

図7は,前重みを 1.0 にした場合の変調フィルタ特性であるが,算出された変調レベルによって変調フィルタは決定される.図8は BANK-A の 2ch に変調を施す場合の変調フィルタ特性例である.

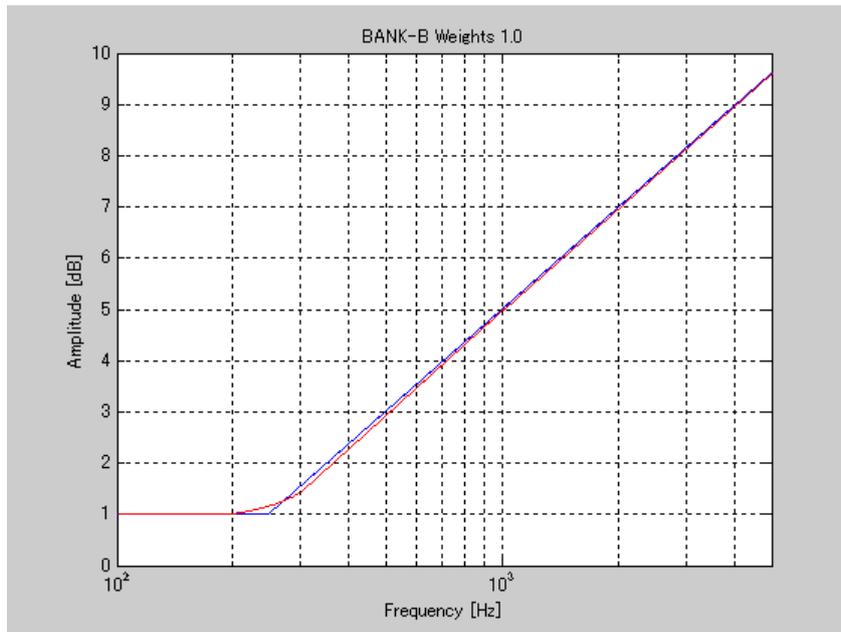


図 7 BANK-B の全重みを 1.0 としたときの振幅特性

Fig.7 Overall response of the composite filter with the same weighting coefficients 1.0

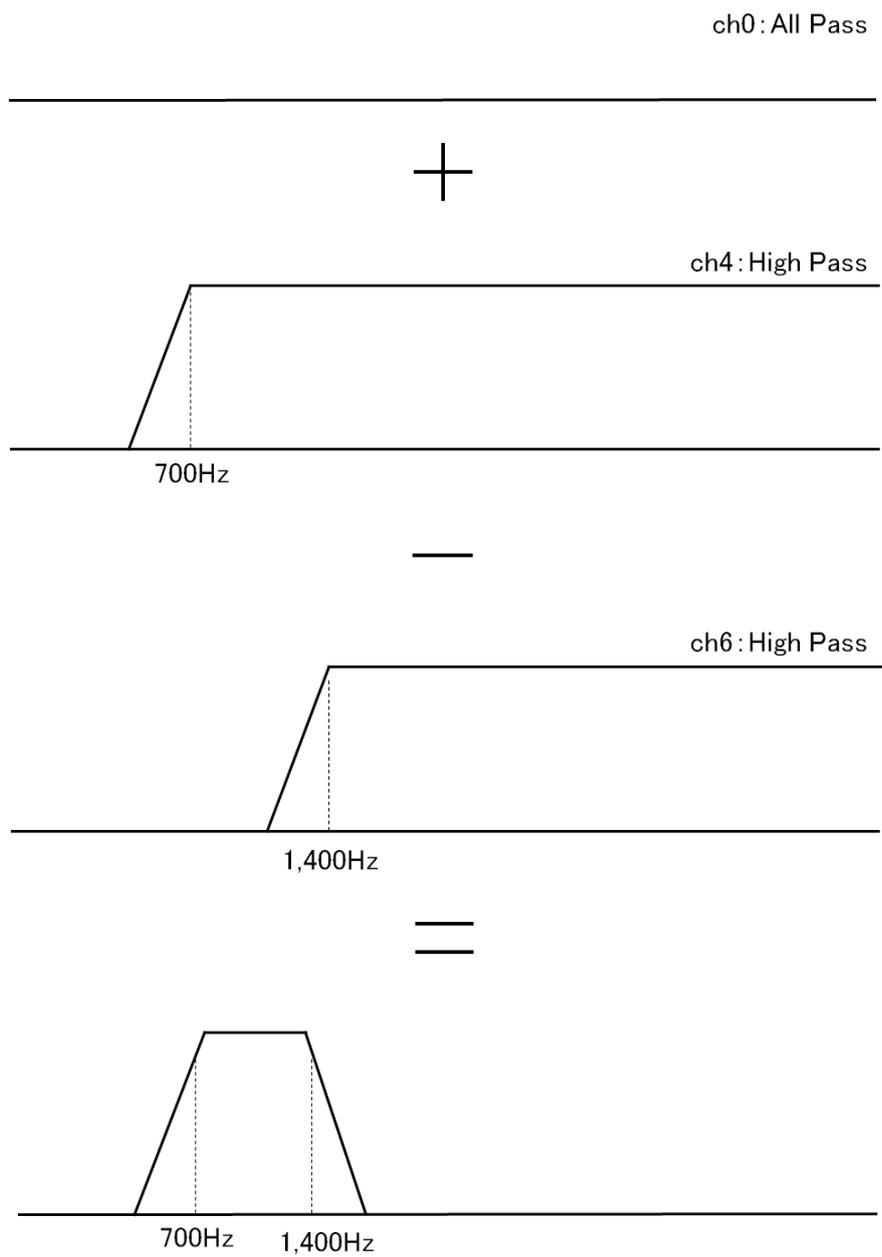


図8 合成されたフィルタの特性例
 Fig.8 Example of response of the composite filter

4.4 振幅変調処理

4.4.1 合成フィルタの特性制御

式 (13) より分かるように,C-FIL は各基本フィルタの重み係数により各区分の傾きと平均ゲインを時間領域で制御するという特徴を有する.

図 9 は,図中の緑丸で示すゲインを与えた場合の C-FIL の振幅特性を示している.

各特性は,ホワイトノイズ入力に対する出力応答の FFT スペクトルの加算平均により得たものを,2~4ch の中心周波数 4 点と各フィルタの帯域遷移部 5 点に与えたものである.図中の青線で示す同一ゲイン条件での 5ch フィルタバンク (BANK-A) の総合特性に比べて,滑らかな特性が得られている.すなわち,従来の多帯域振幅変調では遷移部の不連続性がスペクトル歪みの原因になり得るのに対して,提案する合成フィルタでは緩やかな振幅特性による変調が可能であることが分かる.

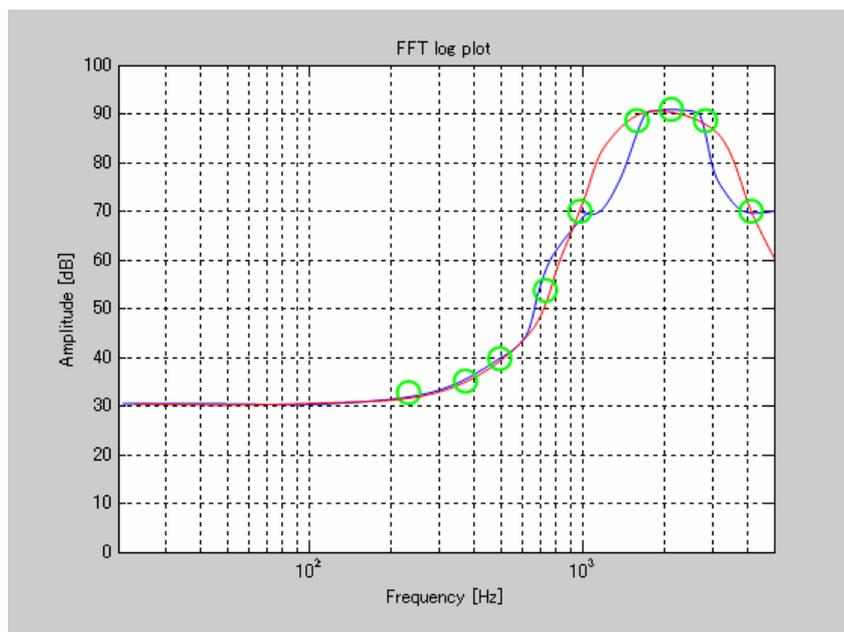


図 9 C-FIL と従来の BANK-A の振幅特性の比較

Fig.9 Comparison of amplitude response between C-FIL and BANK-A

4.4.2 振幅レベル変調特性（静的変調特性）

各帯域の振幅レベル変調に用いられる関数を図 10 に示す。図における Maximize と Minimize は、系の有するダイナミックレンジに応じて定まる。一方で、Modulation level は、当該帯域に対する変調レベルであり、後述する振幅ヒストグラムによる音声特徴データ抽出システムの出力として与えられている。同図上における Modulation level の水平方向線と太実線の垂直方向距離（圧縮レベルと入力レベルにより定まる）が、ゲイン調節量となる。この二つの線が重なっている（Modulation level が低い）状態では、入力レベルを問わずにゲイン調節量が 0 となり、完全に非変調となる。

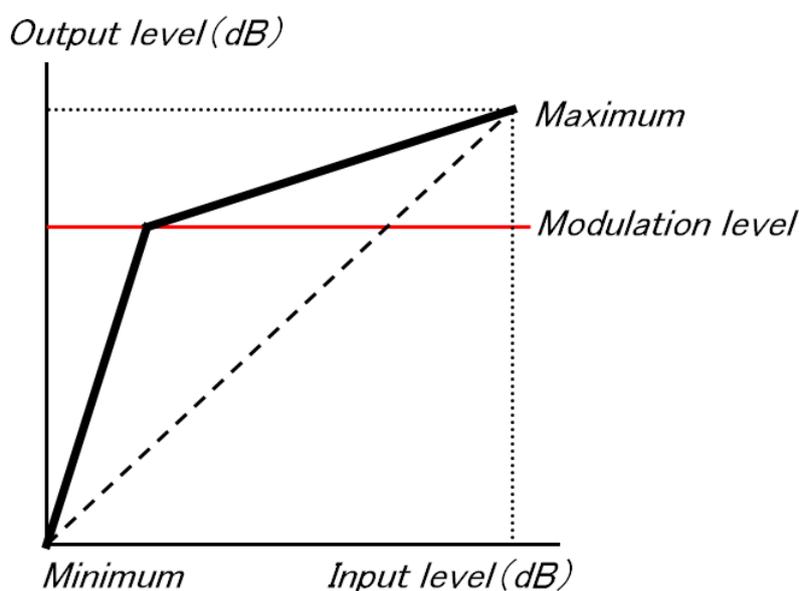


図 10 ゲイン変調関数

Fig.10 Static amplitude modulation function

この結果、望まれる特性を有する変調フィルタが生成され、入力信号のフィルタリングにより変調処理を実現する。

5 音声特徴データによる変調レベル算出

人は電話受話器などの拡声器を通した会話でも,相手を判断することができる.例えば Mp3 では,周波数帯域で分割した音声データを,振幅情報に基づいて処理を施し,人に音質の劣化をほとんど感じさせずに圧縮処理を施している.このことから,ある程度音声を簡略化しても音声の中に人の特徴が残されていることが期待できる.

ヒストグラムにより音声信号等を一定の時間窓ごとに区分けし,各々の時間窓の振幅レベルの累積値を求めている.このため,音声データの振幅レベル変調処理を行うときに,音声データの特徴を考慮しながら時間区切りで処理を行うことにより,演算量を低減できるという利点がある.

また,突発的な破裂音のような雑音に対しては,各時間窓の振幅レベルの累積値と比べると,破裂音による累積値が全体と比べて少なくなるため,振幅レベル変調処理を行う際,雑音による影響を少なくすることができるという利点がある.

本システムは,時々刻々変化する音声の分析結果に応じて,各周波数帯域の音声を,その特徴データを保持したまま,どの程度変調を施すのかを自動で決定するものとなっている.

5.1 変調レベル算出システム構成

このシステムは,図3で示したように,フィルタバンク (BANK-A) と変調フィルタとしての基本フィルタ群 (BANK-B) で挟まれた部分,ゲイン調節量を算出するシステムである.システム構成は,図 11 のように振幅ヒストグラム変換部 (HIST-A) および,音声特徴抽出部 (HIST-B) からなる.

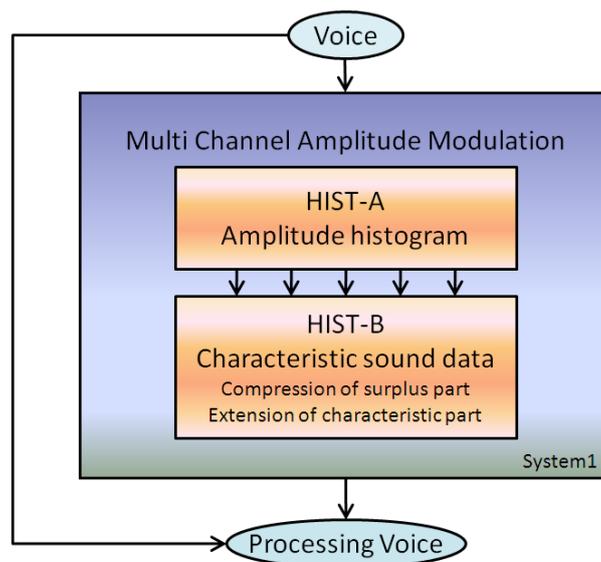


図 11 音声特徴データ抽出のブロック図

Fig.11 Block diagram of extension characteristic sound data

5.2 振幅ヒストグラムに含まれる特徴データ

音声データの簡略化を行うことで、特徴をより顕著に、余剰部分をぼかすように、処理することができる。つまり、処理の簡略化という点で、音声の特徴的な成分の抽出への汎用に期待ができる。さらに、音声明瞭度の改善という観点では、音声の特徴が強調されることにより、音韻の選別が顕著に表れることに期待できる。

HIST-A は、音声の振幅レベルの分布を出力する、振幅ヒストグラムである。**BANK-A** (5ch のフィルタバンク) で分割された音声は、各帯域出力として振幅ヒストグラムに変換する。

処理の簡略化が目的のため、ヒストグラムを求める際の各時間窓は 0.25 秒としている。人がある音声を確実にその音声であると認識するためにかかる時間が 1 秒とされている。そのため拡声音声としては、早口で話したとしても、言葉として認識するためにはその 3 倍程度 (0.3 秒前後) であれば妥当であると判断した。本研究では、拡声器から出力される音声は、時間領域での処理を施すため、目的音声の動的特性制御を考慮する必要がある。そのため、サンプリングの時間窓長を 0.25 秒と決定した。このときの各時間窓のヒストグラムの点数は 11,025 点となる。これは、音声データを 1 秒 44,100 点としている WAVE フォーマットに基づいている。

本システムでは、各時間窓を **BANK-A** によって 5 つに分割しているため、5ch 各々のヒストグラム点数は 11,025 点で、全体としては 55,125 点となる。

各々の帯域で、振幅が強い帯域はスライスレベルが大きくなり、振幅が弱い帯域はスライスレベルが小さくなる。大振幅の部分は 11,025 点の分母に対して、大きいスライスレベルで区切られた、つまり 1 点に対して大きな値の分子を持つため、変調の度合いも緩やかな振幅レベル変調が施される。対照的に、小振幅の部分は 11,025 点の分母に対して、小さいスライスレベルで区切られた、つまり 1 点に対して小さな値の分子を持つため、変調の度合いも詳細な振幅レベル変調が施される。

これは、振幅を大きいデータでは大まかに、小さいデータでは詳細に見るという点で、音声スペクトルの詳細な部分を保持している。そして、扱うデータの簡略化という点で、演算量の低減にもつながっている。

図 12 は母音の”あ”を 3 秒間録音したうち立ち上がりと立ち下りをカットした 1 秒間の振幅ヒストグラムのデータである。このデータは 0.25 秒のサンプル時間窓 4 つ分を表示しているため、データ数は 44,100 点となる。図 13 は参考用として分かりやすく視覚化した、このデータを FFT 処理したグラフである。

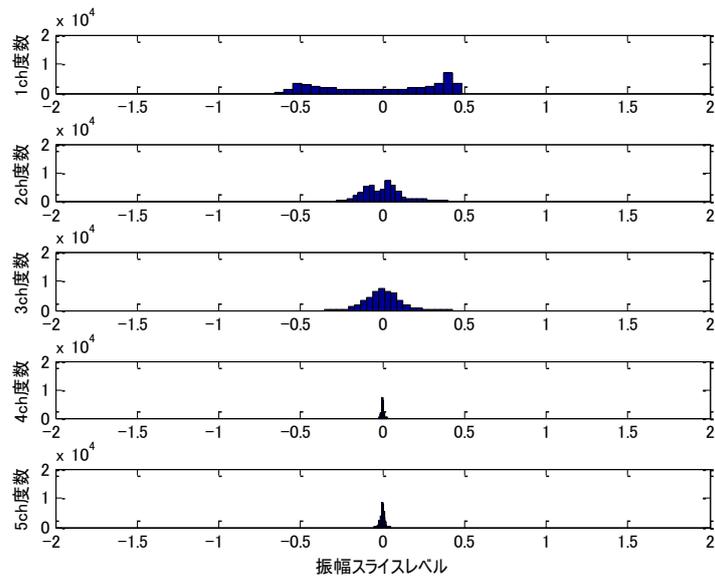


図 12 変調処理前の振幅ヒストグラム例
 Fig.12 Example of pre-modulated amplitude histogram

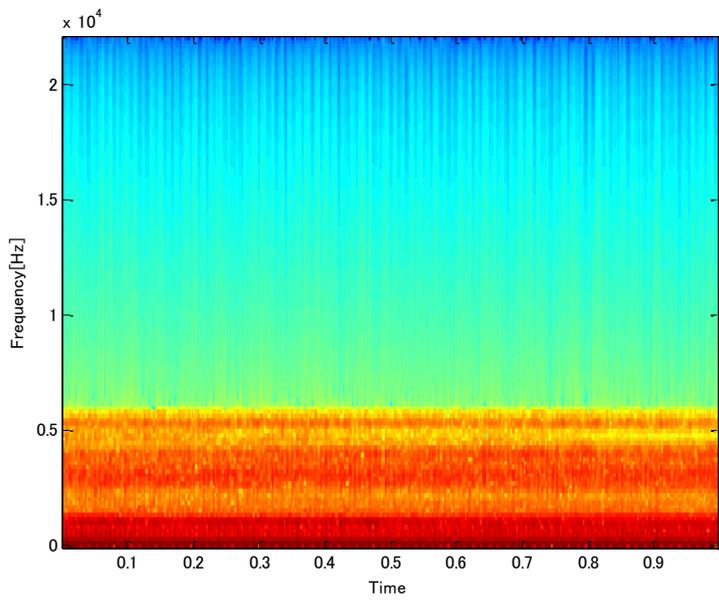


図 13 変調処理前の周波数分析例
 Fig.13 Example of pre-modulated frequency analysis

5.3 特徴データの抽出処理

5.3.1 相互相関係数

相互相関とは 2 つの信号がある場合に,それらの信号がどれだけ似通っているのか類似度を導くためのものである.

$$z(\tau) = \int_{-\infty}^{\infty} h(t) \cdot x(t + \tau) dt \quad (14)$$

(14) 式は, $h(t)$ と $x(t)$ の 2 つの関数の相関を, $z(\tau)$ で表している. $z(\tau)$ は, $h(t)$ と $x(t)$ を τ だけずらして乗算したものを積分している.

特徴を保持した値の度合いを決めるために,基本となる音声データを用意し,その音声データと同じ音声データを加工 (振幅の圧縮,伸張) したものを比較して,類似度を導き出している.

同じ音声とみなされた場合は,値は 1 となる.音声を処理前後で帯域別に周波数分析した結果の相互相関を取る.そのとき振幅のヒストグラム-レベルの相関が 0.8 以上となるような,音声の特徴データを保持した圧縮,伸張の度合いを決定している.

類似度を 0.8 とした理由は,同一話者の音声類似度が,図 14 に示すように振幅のヒストグラム-レベルの相関では,0.8 以上でも約 68%,0.7 以上に至っては約 97%の精度を示していることに基づいている.よって,目的音声をその音声として了解できるレベルの目安として 0.8 を選んだ.

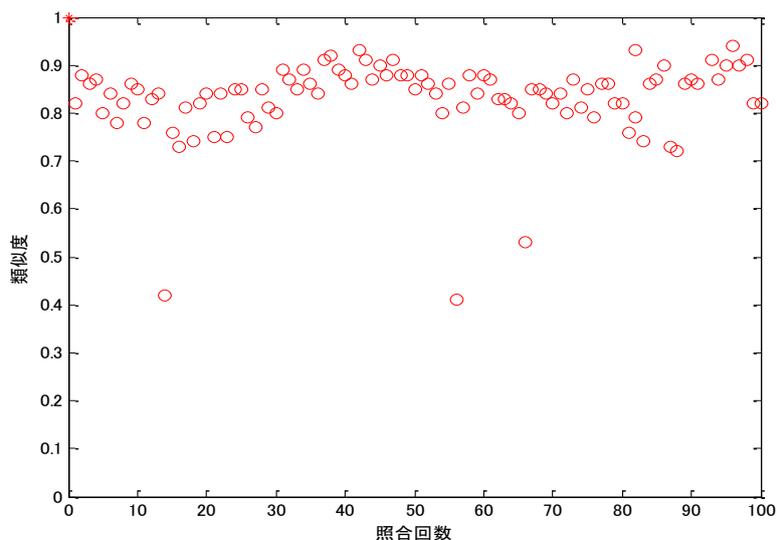


図 14 同一話者による振幅ヒストグラム-振幅レベルの相互相関係数
Fig.14 Cross-correlation coefficient of the same speaker, Histogram - Level

5.3.2 特徴データを考慮した振幅制御

振幅ヒストグラムより,入力された音声の特徴となり得ない余剰部分を圧縮,特徴となり得る主要部分を伸張させる.ここでは,HIST-B で抽出された部分を,その音声特徴データを保持した変調が可能となる値の範囲で,例えば図 15 に示すような圧縮,伸張を施す.このとき,処理の前後で平均振幅レベルは不変である.

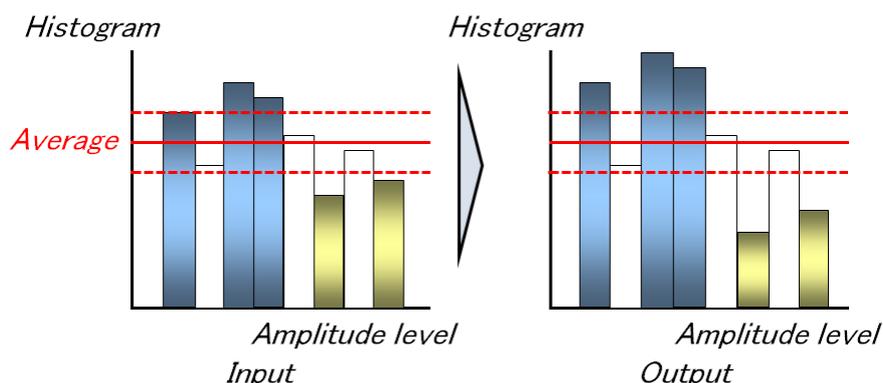


図 15 変調処理前後の振幅レベル例

Fig.15 Example of pre-/post-modulated amplitude level

この特徴を保持した値は,被験者による実験,検証によって,その度合いを決定する.加工前後の音声を,相互相関を取って,その特徴に相違がないか検証した.以下に,その値の決定方法について述べる.

この特徴となる値の算出方法を,以下 (15) 式で示す.

$v(i)$ を目的音声, $v'(i)$ を処理音声とする. $v'(i)$ は,上述の音声特徴データを考慮した上での振幅レベル変調を施す.

$$v'(i) = v(i) \cdot \left(\frac{\{x(i) \cdot y(i)\}}{\{\sum_{i=1}^l (x(i) \cdot y(i))\} / l} \right) \quad (15)$$

i は時間窓単位 (本システムでは 0.25 秒が 1 時間窓となる), $x(i)$ は振幅ヒストグラムのスライスレベルにおける大きさ, $y(i)$ はヒストグラムデータ数を示している.

5.4 特徴抽出に基づく振幅レベル制御（動的変調特性）

振幅ヒストグラムは、音声データの時間窓の区切りを 0.25 秒に設定しており、特徴の有無の判別はヒストグラムの分布から求めている。ヒストグラムに広がりがあるほど、つまり振幅が大きいほど音声データに重みを付け、広がりが少ない部分ほど音声データを削減している。重み付けは時間窓ごとに、ヒストグラムの振幅レベルの分布から平均値を求め、それに対する 5ch 各々の振幅レベルの比率を求めることによって振幅レベル変調の重み付けを施している。

この処理で、周波数などの特徴部分は変化させることなく、特徴とされる部分をより強調し、余剰部分を削減することができる。また、音声の解析を行う際に、特徴部分を強調しながら演算量を減らすことにつながる。

5.2 で用いたデータを用いて、この音声の余剰部分および、特徴部分を求め、振幅の圧縮、伸張を施した。以下、図 16 は母音の”あ”に処理を施したときの 1 秒間の振幅ヒストグラムのデータである。このデータは 0.25 秒のサンプル時間窓 4 つ分を表示しているため、各 ch のデータ数はそれぞれ 44,100 点となる。図 17 は FFT 処理したグラフである。

図 13 と図 15 を比べると分かるように、実際はこの処理の中で FFT 処理は行われていないが、振幅レベル変調された前後の周波数分析は、そのゲインに差があるのみで、スペクトルの詳細な特徴も保持されている。

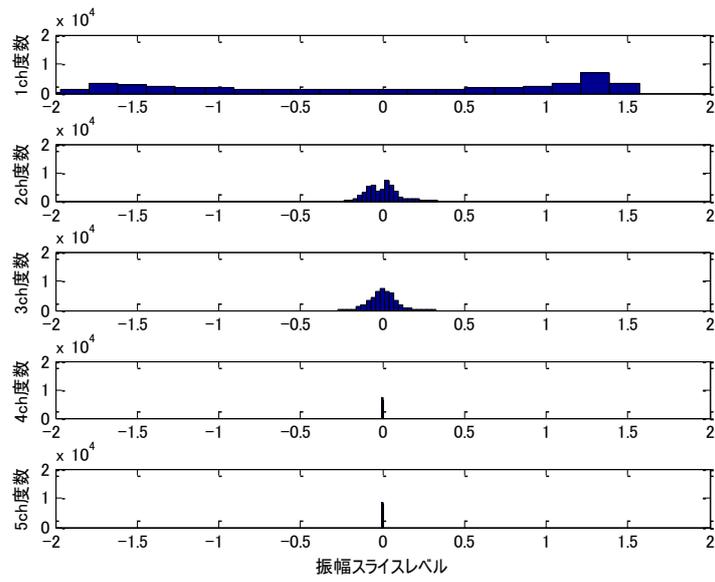


図 16 変調処理後の振幅ヒストグラム例

Fig.16 Example of post-modulated amplitude histogram

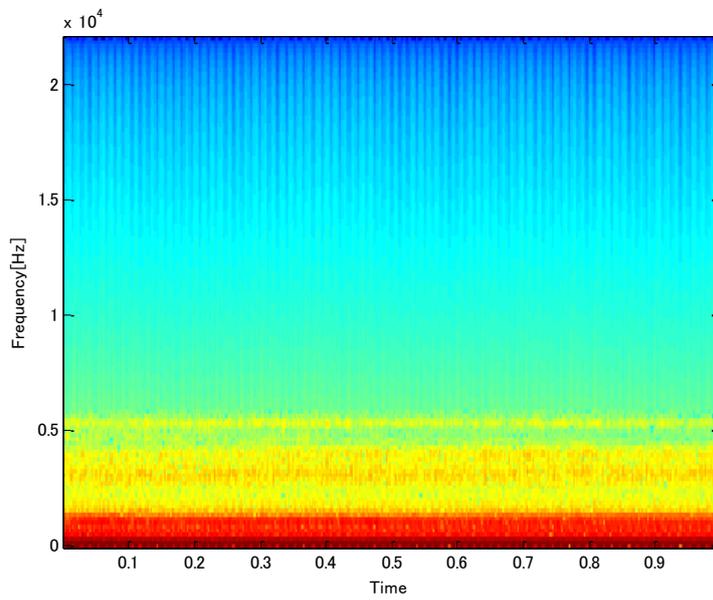


図 17 変調処理後の周波数分析例

Fig.17 Example of post-modulated frequency analysis

6 実験

合成フィルタによる音声特徴を考慮した振幅レベル変調処理の効果を確認するために、評価実験を行った。物理評価では、本システムの振幅レベル変調の動的制御の検証を、感性評価では、被験者による単語音声の正聴率と明瞭度の主観評価の検証を行う。

実験に用いた音声は、補聴器適合評価用 CD⁹⁾ に収録された 2 音節単語 (25 語) と 3 音節単語 (25 語)、計 50 語である。

6.1 物理評価実験

6.1.1 動的制御実験

本システムの処理を、上述の 2 音節単語、および 3 音節単語で行い、その特徴データの検出、および振幅変調の度合いを自動制御した。

6.1.2 実験結果

図 18 は 3 音節単語の源波形。図 19 は、処理を施したときの、特徴を保持した振幅変調の例を示している。図における色の濃淡は、当該スペクトルの振幅レベルの強弱を示している。

図より、スペクトルが設定した聴野に振幅レベル変調されていることが分かる。

また、演算量の低減により、10 秒の音声を時間追従約 0.36 秒での処理により、短い音 (単語や文節) では前変調方法では困難であったリアルタイムとみなせる処理が可能となった。

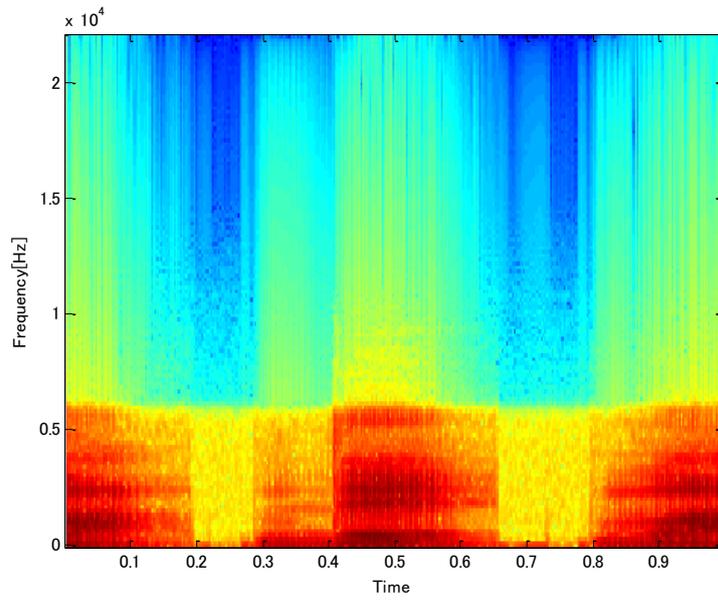


図 18 変調処理前のスペクトル分布例
 Fig.18 Example of pre-modulated spectrogram

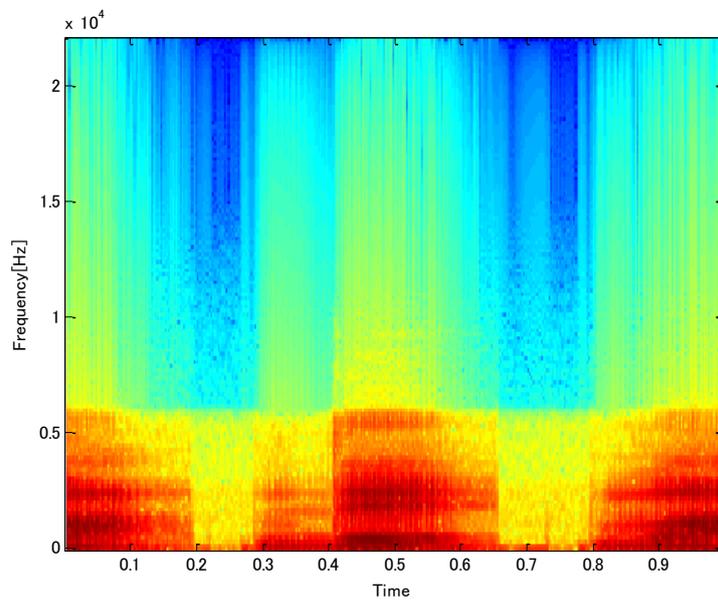


図 19 変調処理後のスペクトル分布例
 Fig.19 Example of post-modulated modulated spectrogram

6.2 感性評価実験

6.2.1 高齢者を模擬した聴覚特性

加齢に伴う聴覚機能の低下としては、純音聴力の低下、周波数分解能（周波数選択能力）の低下、そして時間分解能の低下が知られている。⁶⁾⁷⁾

これら高齢者で低下がみられる聴覚機能のうち、周波数分解能の低下と純音聴力の低下について、実験時に音声出力される手前に図 20 に示す特性を有する LPF を挿入し、若年被験者で高齢者を模擬することとした。このフィルタの特性は、聴力変化による 20 歳前半と 50 歳後半の差に準じたものである。⁸⁾

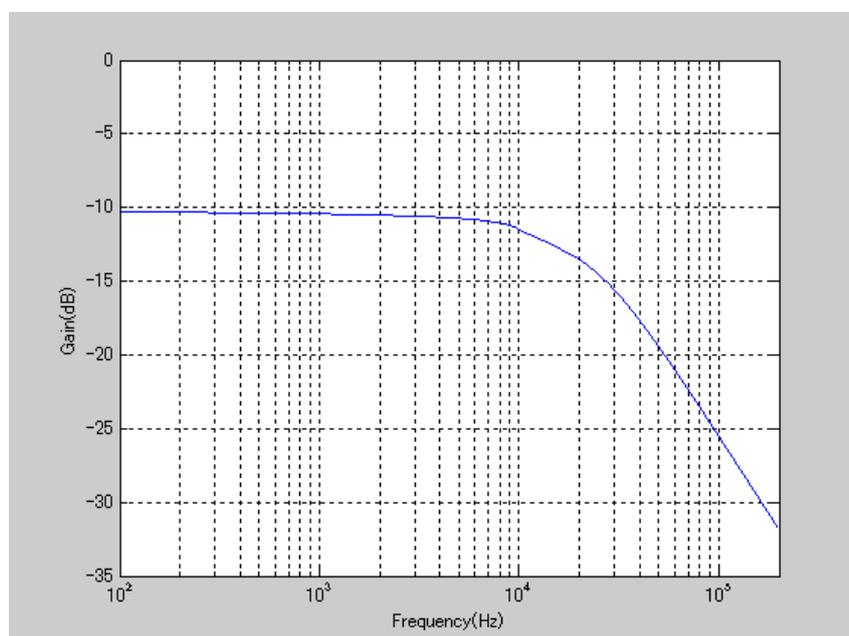


図 20 加齢に伴う聴力低下を模擬したフィルタの周波数特性

Fig.20 Frequency response of a filter simulating age related deterioration of hearing

6.2.2 評価実験

被験者は 7 名,いずれも気道純音聴力の正常な若年者である.実験に用いた音声は,補聴器適合評価用 CD⁹⁾ に収録された 2 音節単語 (25 語) と 3 音節単語 (25 語) の計 50 語を,高齢者模擬フィルタを介した上で,未加工ものと本処理で加工したもの,計 100 語を,雑音下においてランダムで呈示し,その正聴率について評価した. 実験に用いた雑音,および音声の呈示レベルは,表 1 に示す 5 種類である.¹⁾

表 1 感性評価実験で用いられた雑音

Tab.1 Noise used for experiment of sensitivity evaluation

Number	Noise source	Level (dB)	Voice level (dB)
1	A passing train	88	94
2	A standing train	81	87
3	An automobile horn	86	93
4	A bustle (i)	76	84
5	A bustle (ii)	75	84

この単語理解度に基づく実験に加え,主観評価を問う実験も行った.雑音環境下で同一内容の未加工音声,および加工音声を呈示し,“聞き取りやすさ”と“煩わしさ”について,図 21 に示す用紙に該当する箇所に印をつけさせる方法で行った.①は加工音声,②は未加工音声を呈示したが,この順序は被験者には分からないようにしてある.

初めに聞く音声を①、後から聞く音声を②とします。
①と②を聞き比べる印象を以下の項目について評価してください

	非 常 に	や や	ど も な い	ど ち ら や や	や や	非 常 に	
①のほうが聞き取りやすい	----- ----- ----- -----					②のほうが聞き取りやすい	
①のほうが煩わしい	----- ----- ----- -----					②のほうが煩わしい	
	1	2	3	4	5		

図 21 主観評価に用いた質問用紙

Fig.21 Question sheet used of subjective evaluations

6.2.3 実験結果

雑音の種類ごとに被験者間で平均した正答率を図 22 に示す。被験者と雑音の種類ごとに比較すると、未加工音声加工音声の単語正答率を上回る組み合わせはなく、いずれの雑音についても、提案するシステムによる音声加工の効果が表われていた。

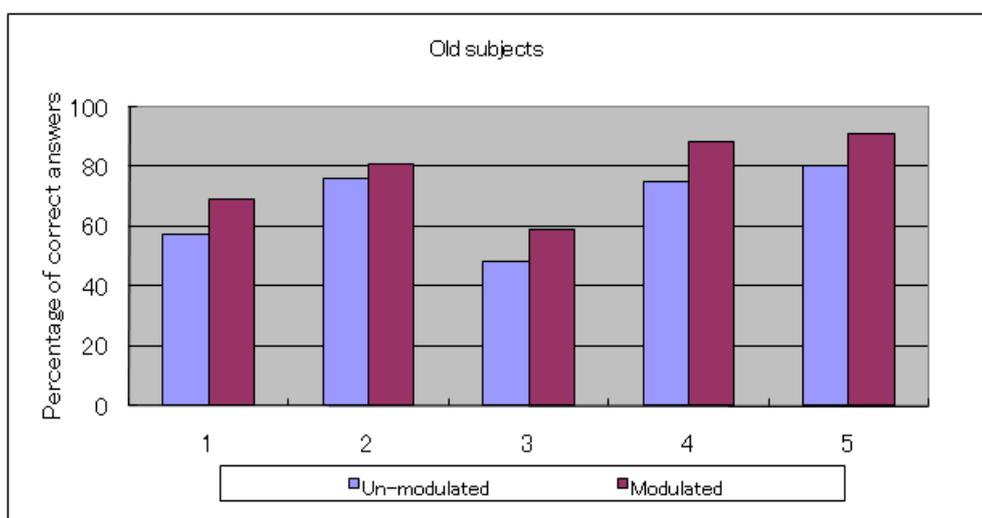


図 22 雑音環境下で呈示された未加工音声と加工音声についての単語正答率の比較
Fig.22 Comparisons of percentage of correct answers between un-modulated and modulated

実験の正答率は、高齢者を模擬したものであり、若年被験者が行った場合は、さらなる効果が見られた。図 23 は、高齢者を模擬しない状態で同若年被験者に実験を行ったものである。

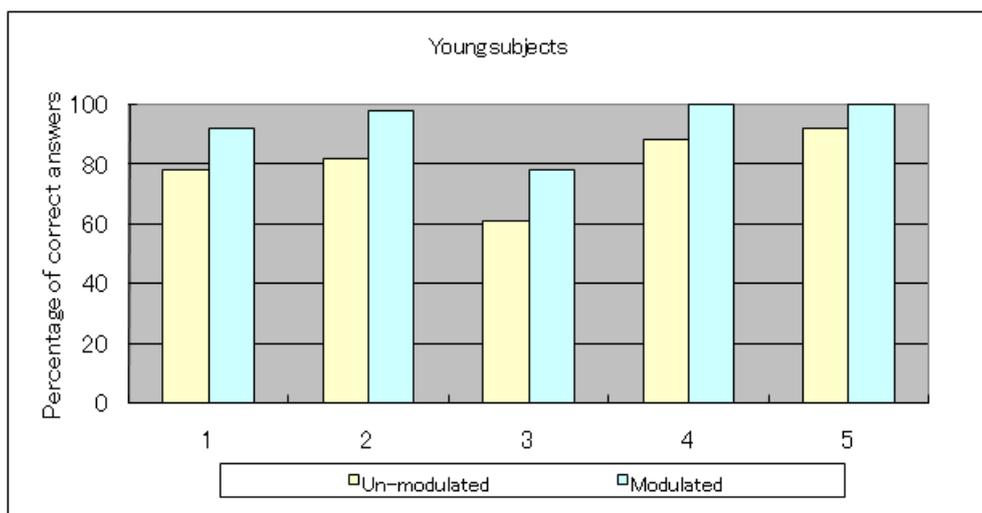


図 23 雑音環境下で呈示された未加工音声と加工音声についての単語正答率の比較（若年被験者の場合）
Fig.23 Comparisons of percentage of correct answers between un-modulated and modulated (Young subject)

表 2 に,雑音の種類ごとに被験者間で平均した主観評価の値を示す.評価を平均すると,“聞き取りやすさ”はポイントが低く,これは加工音声の方が聞き取りやすいということを示している.“煩わしさ”はポイントが高く,これは未加工音声の方が煩わしいということを示している.この結果より,“聞き取りやすさ”と“煩わしさ”のいずれについても,加工音声の効果が優れていたことが示された.

表 2 雑音環境下で呈示された未加工音声と加工音声についての
聞き取りやすさと煩わしさの主観的評価

Tab.2 Comparisons of subjective intelligibility and annoyance between
un-modulated and modulated

Noise source	Question-1	Question-2
A passing train	2.714	3.571
A standing train	1.857	4.000
An automobile horn	2.429	4.714
A bustle (i)	1.571	4.429
A bustle (ii)	1.286	4.286
Average	1.971	4.200

7 考察

7.1 物理評価考察

本研究では,音声明瞭度改善のための振幅レベル変調処理方法として,フィルタバンクと変調用フィルタからなる新たな振幅レベル変調方式を提案した.この結果,基本フィルタ群を用いた合成フィルタ生成手法により,変調フィルタ生成のための演算量低減が図れることが示せた.

また,音声の特徴データを考慮した方法でも,周波数分析によるスペクトルパターンではなく,振幅ヒストグラムによる特徴検出により,周波数の詳細を保持しながら,特徴検出のための演算量低減が図れることが示せた.

以上より,従来の多帯域振幅レベル変調で不可避であったスペクトル歪みが本変調処理では軽減されることが十分に期待できた.さらに,フィルタバンク構成の長所と振幅ヒストグラムのスライスレベルの割り振りを生かして,瞬時変調と音節変調の相反する変調手法の動的特性を制御するための帯域推定レベル手法を提案し,基礎実験においてその効果を期待できた.

7.2 感性評価考察

実験の被験者が高齢者を模擬した若者であったことから,とりわけ高齢者にとって有効なシステムであるか否かが,興味をもたれるところであった.結果として,感性評価の結果に見られるように,提案したシステムによる処理は,高齢者の単語了解度を平均して 10.4%増加させた.若年被験者では,同値が平均して 13.4%増加させた.

主観評価の実験では,聞き取りやすさ,および煩わしさは,加工音声が改善されていることが示唆された.この実験では,評価シートに実験に対する各評価の重みづけなどの考慮が必要である.

被験者数が少ない実験であったものの,提案したシステムによる処理が,単語了解度,および聞き取りと煩わしさを改善させることが示された.

8 おわりに

本研究では、音声出力の前処理システムとして、音声の特徴データを考慮した振幅レベル変調を、合成フィルタによる帯域分けされた上での滑らかな振幅制御特性を得るシステム提案した。

本システムの物理的評価では、従来の手法では困難であった帯域遷移部のゲイン変調が改善され、滑らかな変調特性を得ることができた。また、その変調の度合いを決定するパラメータとしては、従来のスペクトル推定による方法ではなく、振幅ヒストグラムによる特徴データ検出を用いたところ、データの特徴部分と余剰部分を簡易な処理で分類することができ、そのゲイン調節の度合いの自動制御が実現できた。

また、少数ではあるが、高齢者を模擬した被験者を対象として、提案したシステムによる音声出力の有効性についての感性評価実験を行った結果、本システムは若年者にとっての煩わしさを感じさせることなく、高齢者（高齢者を模擬した被験者）にとっての単語理解度を改善する傾向が見られた。

この結果より、目的音の振幅ヒストグラムから特徴得る、帯域で分割された合成フィルタによる処理は、スピーカーのような音声拡声装置のみならず、高齢者向けの音響装置として、また雑音環境下での携帯電話の会話品質の向上などにも利用可能であると考えられる。

さらに、本研究で提案した手法の展望については以下のように考えられる。

米 Xerox Palo Alto Research Center(PARC) の故 Mark Weiser 氏が最初に唱えた「ユビキタス・コンピューティング」という概念から端を発したユビキタス社会は、現在その実現に向けて本格的な技術開発が進んでいる。ユビキタスネットワーク技術の将来については、ますます一般ユーザの周りにはネットワークを介したさまざまな情報端末が存在するようになるとの展望がある。そうしたユビキタス社会に向けて多種多様な情報端末や様々なユーザを想定したとき、入力インターフェースのユーザビリティを確保する手段の一つとして、音声認識を用いたものが考えられる。音声認識を入力インターフェースとして用いる利点として、ユーザビリティを確保するというだけでなく、対話方式での情報端末の操作を複雑な処理を意識せずに行うことができる点などもあげられる。

近年、車載の情報端末や PC の入力インターフェースなどにおいて音声認識が実用化されているが、多様な情報端末のインターフェースとしての音声認識においては、あらゆる環境下で頑健な動作をする装置が必要不可欠になってくる。

現在商品化されている音声認識装置は、発話者の口元に置かれたマイクや骨伝導マイクからの入力などについては高い認識性能を示す。しかしながら、他者の音声、背景音楽、反響音などのノイズが混入した場合には認識率が大幅に低下してしまうことが知られている。特に車載の情報端末においては、走行音やカーステレオなどのノイズが影響するため、認識語彙数が限られていて、認識率が低いといった問題がある。¹⁰⁾

このような場合において、音声認識装置の前処理として、目的音の特徴を得ることが出来れば認識率の向上に大きく寄与すると考えられる。

付録

音声は,補聴器適合評価用 CD⁹⁾ に収録された 2 音節単語 (25 語) と 3 音節単語 (25 語) の計 50 語の中から参考資料として,グラフィカルに分かりやすい音声を選んだ.

グラフの窓長は,振幅ヒストグラムは音声の 1~2 秒までの 1 秒間を,スペクトログラムは音声の 0~3 秒までの 3 秒間を表したものである.

音声データ：D1_02

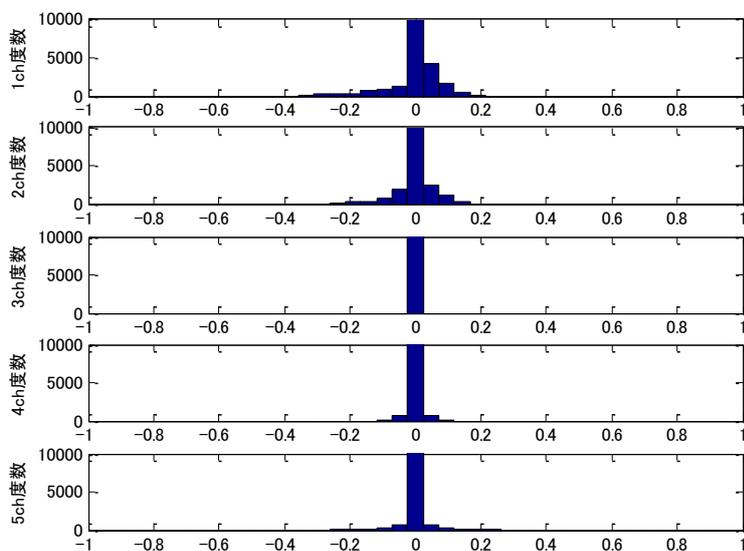


図 24 変調処理前の振幅ヒストグラム例 (D1_02)

Fig.24 Example of pre-modulated amplitude histogram (D1_02)

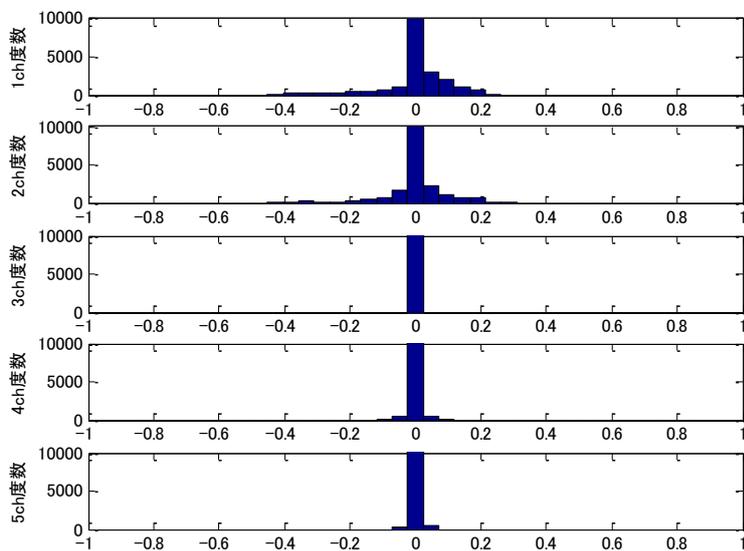


図 25 変調処理後の振幅ヒストグラム例 (D1_02)

Fig.25 Example of post-modulated amplitude histogram (D1_02)

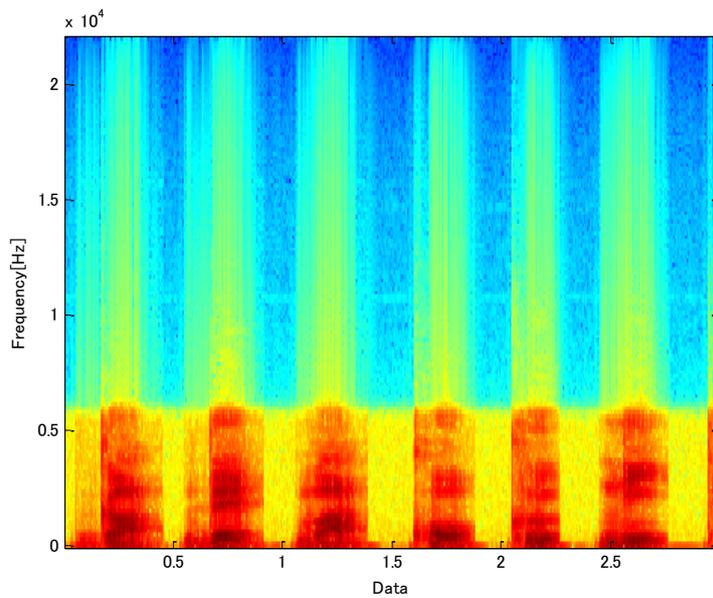


図 26 変調処理前の周波数分析例 (D1_02)

Fig.26 Example of pre-modulated frequency analysis (D1_02)

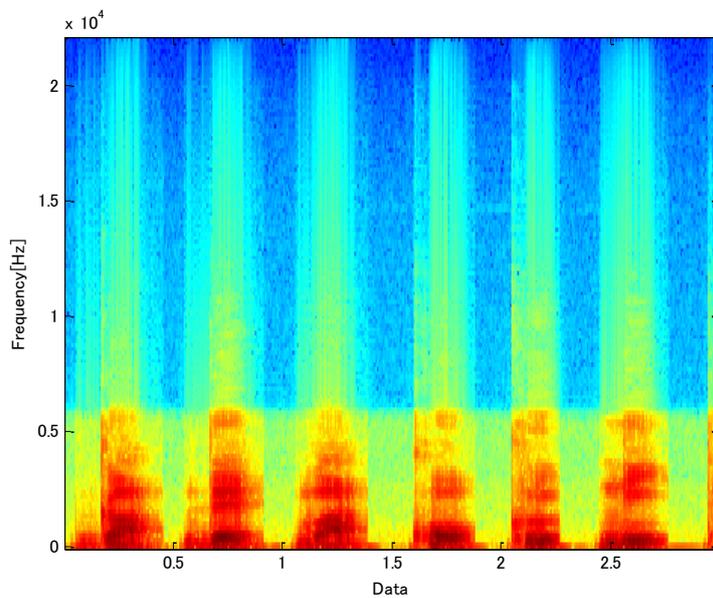


図 27 変調処理後の周波数分析例 (D1_02)

Fig.27 Example of post-modulated frequency analysis (D1_02)

音声データ : D2_02

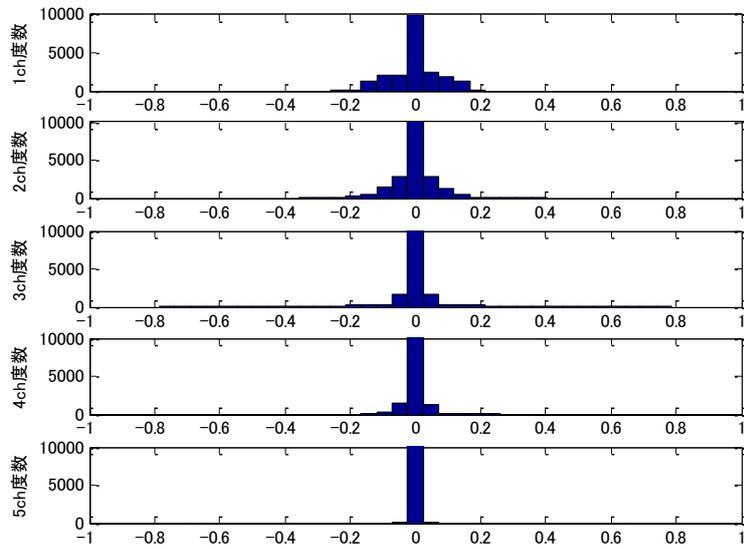


図 28 変調処理前の振幅ヒストグラム例 (D2_02)

Fig.28 Example of pre-modulated amplitude histogram (D2_02)

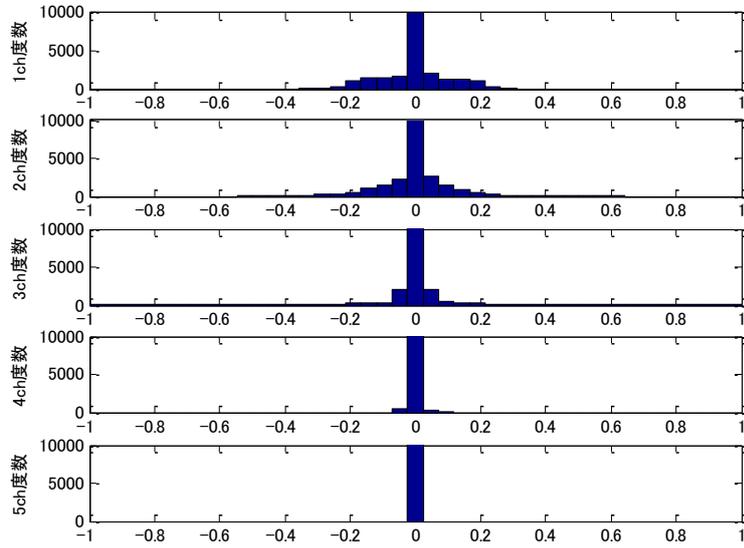


図 29 変調処理後の振幅ヒストグラム例 (D2_02)

Fig.29 Example of post-modulated amplitude histogram (D2_02)

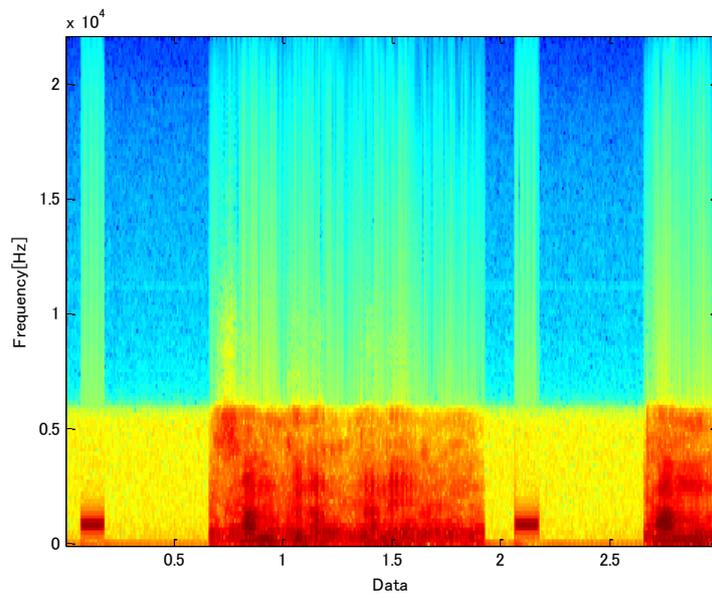


図 30 変調処理前の周波数分析例 (D2_02)

Fig.30 Example of pre-modulated frequency analysis (D2_02)

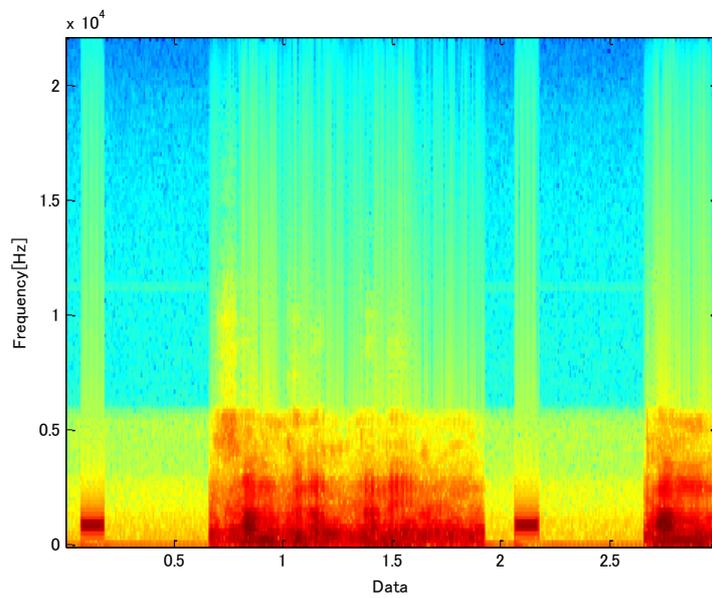


図 31 変調処理後の周波数分析例 (D2_02)

Fig.31 Example of post-modulated frequency analysis (D2_02)

音声データ : D2_05

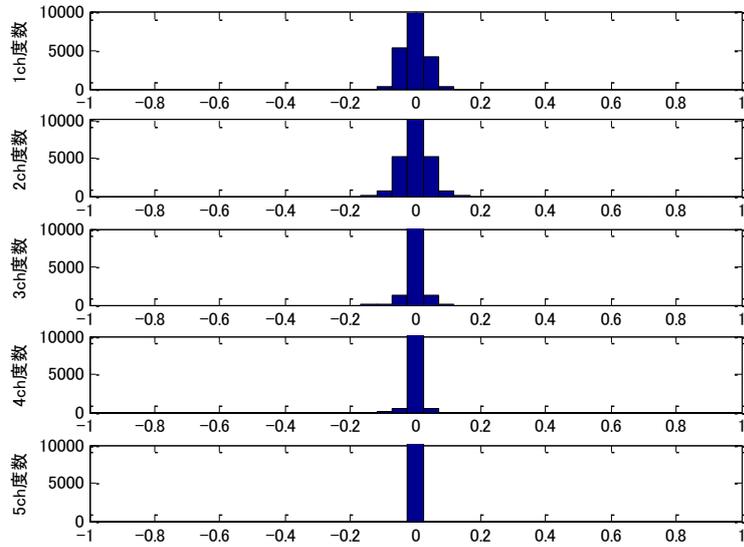


図 32 変調処理前の振幅ヒストグラム例 (D2_05)

Fig.32 Example of pre-modulated amplitude histogram (D2_05)

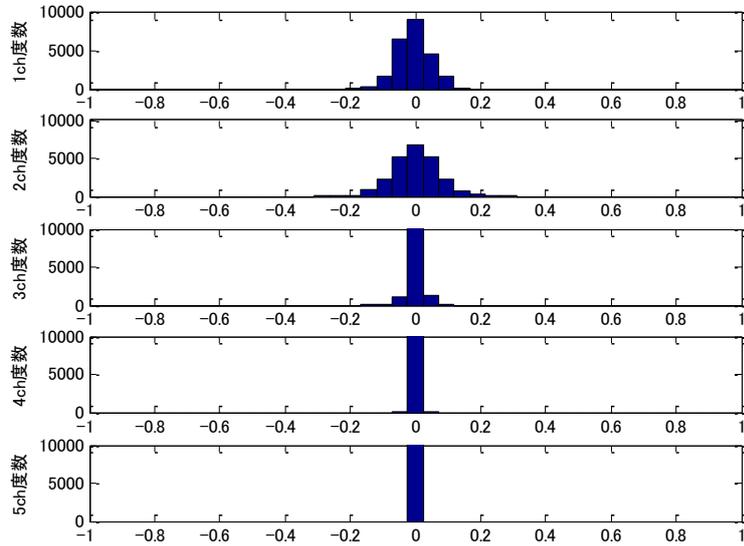


図 33 変調処理後の振幅ヒストグラム例 (D2_05)

Fig.33 Example of post-modulated amplitude histogram (D2_05)

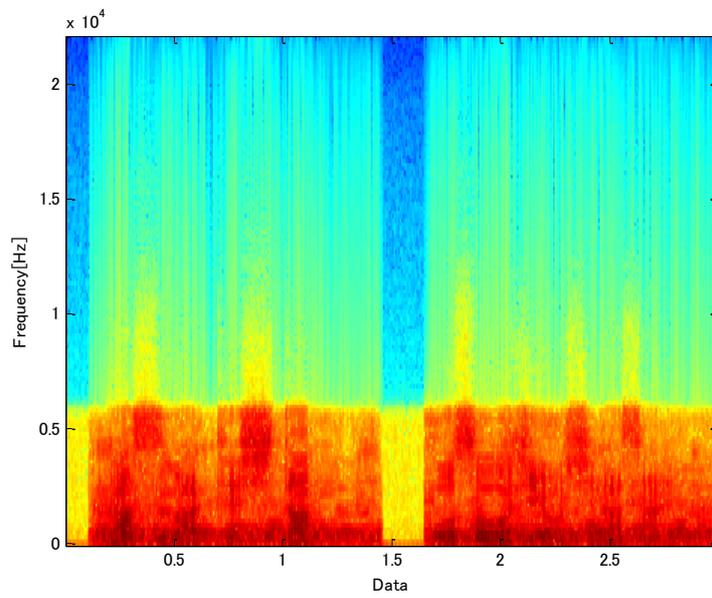


図 34 変調処理前の周波数分析例 (D2_05)

Fig.34 Example of pre-modulated frequency analysis (D2_05)

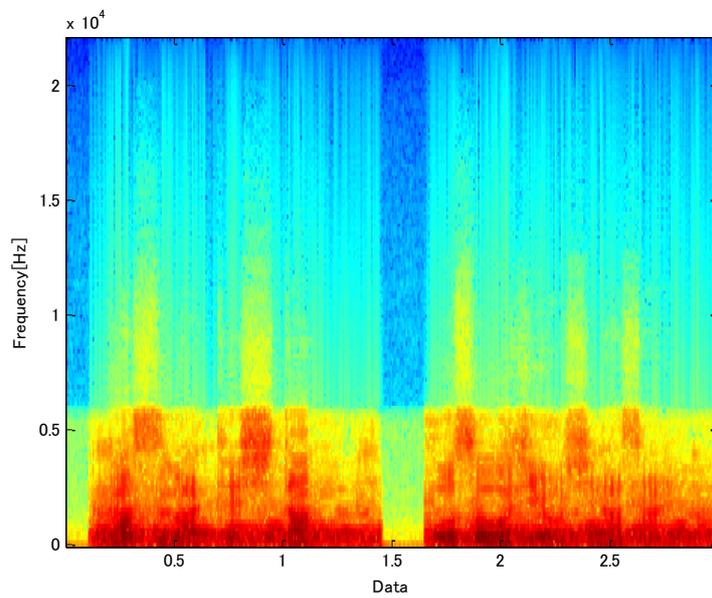


図 35 変調処理後の周波数分析例 (D2_05)

Fig.35 Example of post-modulated frequency analysis (D2_05)

謝辞

大学院の2年間、研究を進めるにあたり、全面的な支援と的確なアドバイスをしていただいた渡辺嘉二郎教授、そしてプログラムを初めとする技術的な指導をしていただいた小林一行准教授、学科の垣根を越えて協力していただいた経営学科の鈴木郁教授に心から感謝します。

また、研究を手伝ってくれた堀川照生君、そして研究室音響班の後輩、鈴木賢太君、曾根久照君、田形さちさん、竹花友希君、竹越貴成君、光岡佑実子さんの協力無しには、研究を成し遂げる事ができませんでした。

最後に、大学院まで進学させてくれた両親をはじめ家族全員に感謝します。

本当にありがとうございました。

参考文献

- 1) 鈴木郁, 林善男 : 雑音下における無意味語の若年者および高齢者における聴取について, 人間工学会誌, pp25-33 1991
- 2) 鈴木郁, 越智寛毅 : 高齢者にとっての了解度改善を目的とした, 拡声器のための音声加工処理, 人間工学会誌, pp289-301 2004
- 3) 彌寝義人, 池田宏, 熊谷幸夫 : 高齢者向けの音声加工を行うポータブルDSPシステムの開発, 電子情報通信学会技術研究報告, pp33~40 1992
- 4) 上田裕市, 阿川貴之, 青山正純, 渡辺亮 : 合成フィルタによるデジタル補聴器のための振幅圧縮処理, 電子情報通信学会論文誌, pp1728-1738 1998
- 5) 彌寝義人, 池田宏, 熊谷幸夫 : 高齢者向け音声加工を行うポータブルDSPシステムの開発, 電子情報通信学会誌, J81-A(12) pp-1728-1738 1998
- 6) 米本清, 倉内紀子, 浜田晴夫, 三浦種敏 : 音性難聴者の言語聴取能力と聴覚機能, *Audiology Japan*, 31(5) pp615-616 1988
- 7) 立木孝 : 新 難聴の診断と治療, 内外医学社, pp144-148 1986
- 8) 三浦種敏 : 新版 聴覚と音声, 電子情報通信学会, pp412 1980
- 9) 米本清 : CD(TY-89)について, 文部省 補聴器適合評価用機器の試作に関する研究報告書, pp3-24 1989
- 10) 鹿野清宏, 伊藤克亘, 河原達也, 武田一哉, 山本幹雄 : 音声認識システム, (株) オーム社 東京, 2003
- 11) 日本音響学会, 音響用語辞典, コロナ社, pp526 1988