

高齢者向け音声加工を行なうポータブルDSPシステムの開発

禰寝 義人 池田 宏 熊谷 幸夫

日立製作所 中央研究所

〒185 東京都国分寺市東恋が窪 1-280

デジタル補聴器などの、聴覚補助を目的としたシステムに用いられる各種の音声処理方式の開発や、フィールドテストでの方式検討に利用することを目的として、DSP（デジタルシグナルプロセッサ）を用いたポータブル音声処理装置を開発した。本装置は2ch14bitのAD/DA変換器および16Mbitの音声メモリを有し、単一5Vのバッテリーにて動作させることができる。本装置では高齢者の難聴特性を考慮して、一般的な周波数特性加工に加え、直前に聞いた音をメモリ録音を利用して瞬時に繰り返し聞き直す機能と、再生時の話速を変化させる機能を実現した。本報告では難聴者による評価は行なわれていない。

A development of portable DSP system for speech processing
to aid senior's hearing

Yoshito Nejime Hiroshi Ikeda Yukio Kumagai

Central Research Laboratory, Hitachi Ltd.

1-280 Higashi-koigakubo, Kokubunji, 185, Tokyo

In this paper, we propose a portable speech processing system with a DSP (Digital Signal Processor) chip. This system can be used in developments and field tests of various kind of speech processing technique that aid hearing impaired people. This portable system has 2 channel 14bits AD/DA converters and 16 Mbits memory for speech recording. This system can work only with a 5V battery. At this point, considering the characteristics of senior's hearing, we developed a function to repeat the speech given just before and a function to make the repeated speech slow, in addition to a standard spectrum enhancement process. This system has not been evaluated by hearing impaired subjects.

1. はじめに

近年、デジタル信号処理の進歩やDSP（デジタルシグナルプロセッサ）の高速化に伴い、衰えた聴覚を補助するための音声信号処理に関する研究が盛んに行なわれている[1][2][3][4][5]。これらの研究においては、周波数空間における音声の特徴を強調加工する、新しい信号処理方式の開発に主眼が置かれており、デジタル補聴器での使用を主な目的としている。

一方、このような複雑な特徴強調処理を必要とすると考えられる感音性難聴の多くは、高齢者によく見られる老人性難聴であり、真に有効な補聴システムの開発のためには、使用者の多くが高齢者であることを考慮することが望まれる。

一般に老人性難聴の場合、聞こえの劣化は単に聴覚末梢系での処理の劣化のみならず、さらに上位の中枢における言語処理能力の衰えや、短期記憶能力の衰えにも強く依存していると考えられる。このため、高齢者を対象とした補聴システムには、音声加工以外にこれらの機能劣化を補助するための工夫も必要であると考えられる。

実際、高齢者の方々とのコミュニケーションにおいては、大きな声を出すことよりも「ゆっくり、はっきり」話すことが、求められることが多い。さらに、同じことを何度か繰り返して話すことで、意志を伝えることができる場合も多い。

一方、従来のデジタル補聴器の研究では、パソコン内部または周辺に取付けたDSPボードを用いてリアルタイム処理を実現している場合が多い。しかし種々の音声加工方式の評価には、実際の使用環境におけるフィールドテストも必要であり、DSPボードのポータブル化、スタンドアロン化も必要であると考えられる。

そこで筆者等は、リアルタイム処理による音声の周波数特性加工を行うと共に、直前に聞いた音声をメモリ録音を利用して瞬時に繰り返し聞き直すことができ、さらに再生時の話速を変化させることが可能なDSPシステムを試作した。本システムは、様々な音声加工方式を実行して比較検討できるだけの「汎用性」と、フィールドテストでの使用を可能とする「ポータブル性」を同時に実現している。

本報告では、この試作したシステムの構成、機能および性能について述べる。なお、現時点では難聴者による補聴効果の評価には至っていない。また、本システムは補聴のための音声処理を検討

するツールとして使用することを主な目的としており、実際に補聴器として使用するために不可欠な技術、たとえば超小型化など、についてはあまり考慮されていない。

2. システム構成と機能

(1) システム構成

図1に本システムの構成を示す。本システムは実際の音声処理を行なうスピーチプロセッサと、このスピーチプロセッサに接続して使用するエミュレーションシステム、および個人の聴力特性から音声処理に必要なパラメータを設定するためのソフトウェアからなる。

通常の音声加工方式の開発や研究室レベルの評価実験では、スピーチプロセッサとエミュレーションシステムとを接続し、スピーチプロセッサをいわゆる「マスター補聴器」として使用する。

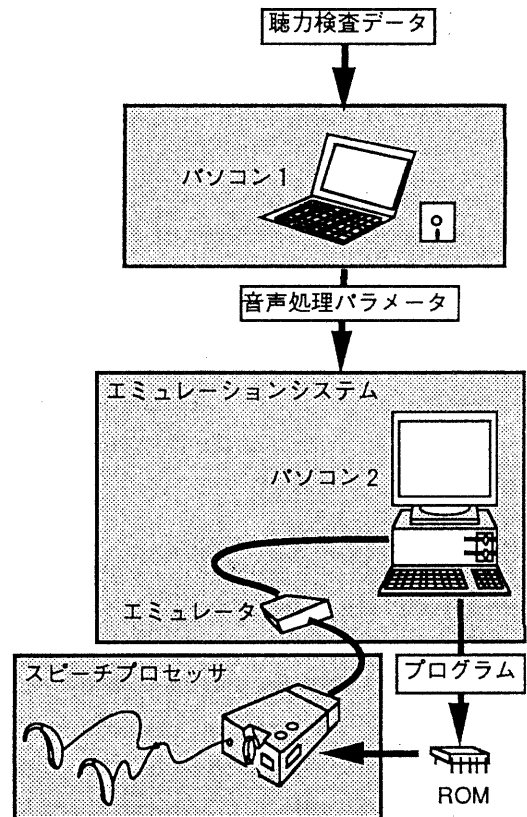


図1 システム構成

エミュレーションシステムには、スピーチプロセッサに使用したDSP専用の市販システムを利用した[6]。ここでは、パソコン2よりエミュレータを通してスピーチプロセッサ内のDSPへのプログラム転送や、DSP動作のモニタなどが行える。一方、音声加工処理とくに周波数特性加工処理では、後に述べるように使用者個人の聴力特性を十分考慮しなくてはならない。この場合、聴力検査で得られたオーディオグラムなどのデータから、信号処理に必要なパラメータを設定するための処理が必要となる。図1のパソコン1上のソフトでは、このような個人特性への適合を目的とした処理を行なっている。

スピーチプロセッサをフィールドテストなどで使用する場合には、評価したい音声処理プログラムおよび各パラメータをROM化してスピーチプロセッサに内蔵し、バッテリーを電源としてスタンドアロンで動作させる。

(2) 機能

本システムは、様々な音声加工処理の比較検討に使用することを利用目的の1つとしているが、現時点ではプロトタイプとして、以下のような機能を搭載している。

(a) 帯域別振幅圧縮処理

・音声情報を使用者の可聴域に入れるための周波数特性の加工手段として、一般に広く用いられている帯域別振幅圧縮処理(マルチチャンネルコンプレッション)を採用した。本システムでは4バンドに帯域を分割し、各帯域毎に使用者の聴力損失に合わせた振幅の増幅およびダイナミックレンジの圧縮を行なう。

(b) メモリ録音/再生

・直前に話された内容を、繰り返して表示する

機能を実現するため、メモリに入力音声を常時録音し、再生ボタンが押される毎に数秒前の音声も瞬時に再生する機能を設けた。

(c) 再生音の話速変換

・ゆっくりと音声を表示することを目的として、メモリ録音再生時に、音声の無音区間を検出し無音区間部分のみを伸張することで、ピッチを下げずにゆっくり話したように再生する処理を行なうようにした。

(d) 2ch入出力

・一般に、方耳のみで音声を聴取するのに比較して、両耳で聴取する方が会話明瞭度が高くなる傾向にあることは良く知られている。本システムではスピーチプロセッサに2つの入出力系を用意し、1つのDSPで2ch処理を行なうことで、両耳分の処理を行う機能を持たせた。

以下、これらの機能を実現するために開発したハード及びソフトウェアの詳細について述べる。

3. スピーチプロセッサのハードウェア

(1) 全体構成

図2に本スピーチプロセッサの回路構成を示す。2ch処理に対応するため、アナログ/デジタルのインターフェイス部分には、すべて2系統の回路が用意されている。

マイクロホンから入力された音声は、後述するアナログインターフェイス回路(AIC)により14bのデジタル信号に変換され、DSP内のシリアルポートを經由してDSPに送られると共に、リングバッファ状に構成されたアドレスを有する16Mbの音声メモリに記録される。録音時間は最大で40秒である。なお、現段階では音声データの圧縮は特に行っていない。

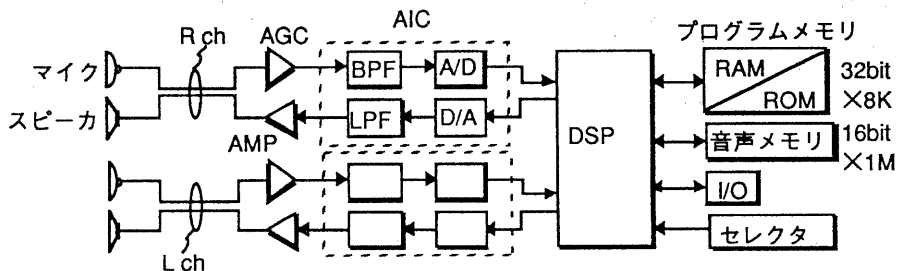


図2 スピーチプロセッサの構成

32×8Kbのプログラムメモリには、DSPで実行させる各種音声処理プログラムが内蔵される。スピーチプロセッサをスタンドアロンで使用するときには、このプログラムメモリにEPROMを使用するが、エミュレーションシステムに接続して使用する場合にはSRAMを使用し、エミュレーションシステムから転送されるプログラムを格納する。

I/O回路はエミュレータとの接続に使用される。本システムで使用したエミュレータは、DSP内部に用意されている12ピンの高速シリアルスキャンバスを通じてDSPの動作を制御するタイプのもので、ボード上のDSPを取り外すことなく、メモリへのプログラム転送やDSPの動作制御、およびDSP内部レジスタのモニタができる。

セクタ回路は各処理の実行時に必要なパラメータを、ロータリースイッチを切り替えて選択/変更するのに用いる。現時点では、後述する帯域別振幅圧縮処理における4つのパラメータセットの選択/変更と、話速変換のための4つの閾値の選択/変更に使用している。

図3には本回路基板の外観を示す。回路実装は4層基板上の片面実装であるが、両面実装などを利用することで、さらに高密度化することも可能である。

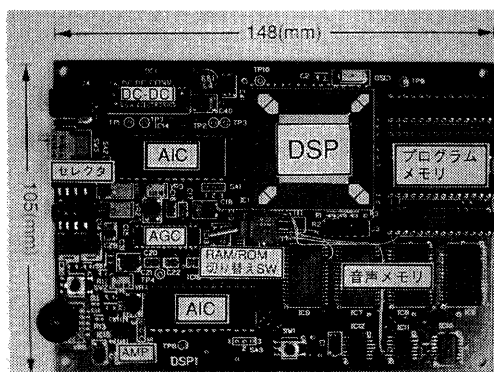


図3 スピーチプロセッサ回路基板

(2) 使用したデバイス

表1に本スピーチプロセッサで使用した主要デバイスの性能を示す。

DSPには、テキサスインスツルメンツ(TI)社の高速浮動小数点型DSP(TMS320C30)を採用した。

このDSPにはデータ入出力用のシリアルポートが2系統用意されており、2chの入出力系を構成するのに適している[6]。

ポータブル化を考える際に、最も問題となるのは消費電力であり、固定小数点型のDSPを使用する方がこの点では有利である。しかし本システムでは、種々のアルゴリズムを比較検討するための汎用性を重視して、信号処理アルゴリズムに対して精度面から生じる制約が小さい、浮動小数点型のDSPを採用した。

AD/DA変換部には、同じTI社のアナログインターフェイス回路(AIC)TLC32044を用いた。本ICは同社のDSPに接続して使用するように設計された専用ICで、DSPからシリアルポートを通じてデータの入出力およびコントロールがなされる[7]。AIC内部のAD/DA変換器の分解能は14bitで、変換周波数を13.3kHzにして使用している。AD変換のためのアンチエイリアシング用のバンドパスフィルタ(BPF)とDA変換後のスムージング用のローパスフィルタ(LPF)は同ICに内蔵されている。ただしAD変換器だけでは、入力音声の広いダイナミックレンジをカバーしきれないため、本システムにはアナログのフロントエンドに34dB程度のレンジを持つAGCアンプを設けている。

また、上述のAICは±5Vの2種類の電源を必要とするが、ポータブルシステムとして使用するには、単一電源が望ましい。そこで本スピーチプロセッサでは、DC-DCコンバータを利用して、電源の+5V単一化を図っている。

表1 主要デバイスの性能一覧

DSP (TMS320C30)	浮動小数点演算型(40bit) 32MFLOPS (MAX:33.3MFLOPS) 16MIPS (MAX:16.7MIPS) 内部RAM2Kward メモリ空間 64Mward
AIC (TLC32044)	AD/DA 変換周波数 13.3KHz (MAX:19.2KHz) 分解能14 bit アンチエイリアシングフィルタ スムージングフィルタ内蔵
プログラム メモリ	4×8bit×8K (ROM : WS57C49B) 4×8bit×8K (RAM : HM6264ASP)
音声メモリ	4×8bit×512K (HM658512)

4. 音声加工のためのソフトウェア

(1) 帯域別振幅圧縮処理

難聴者の可聴域に音声情報を入れるための周波数特性加工は、補聴を目的とするシステムには不可欠な処理であると考えられる。そして、このような音声の周波数特性加工方式として、従来より様々な方式が提案されている[1][2]。このうち、フィルタにより音声帯域分割した後、各帯域毎に信号振幅の増幅とダイナミックレンジの圧縮を行う帯域別振幅圧縮処理（マルチチャンネルコンプレッション）は、FFTなどの周波数空間への信号写像を行わなくても、時間領域でのフィルタリング処理と瞬時振幅値の変換で実現でき、処理量も比較的少ないため、DSPを用いたリアルタイム処理に向いている。またこの種の処理は従来のアナログ補聴器や最近市販されたデジタル補聴器

（リオン社製DIGITALIAN）にも使われ[8]、補聴効果に関する評価も数多く行われていることから、本システム上での実現を試みることにした。

本システムでは、フィルタを用いた帯域分割に41次のFIRフィルタを使用した。図4に実測したフィルタ特性の一例を示す。フィルタ係数は良く知られたカイザー窓関数を用いた窓関数法で求め、阻止域減衰量がAD変換器のダイナミックレンジ程度になるようパラメータを調節した。

リアルタイム処理を考えた場合、FIRフィルタの次数は、フィルタ特性だけでなく、処理時間からの制約とDSP内部のメモリ容量からの制約を加味して決めなくてはならない。本システムの場合、メモリ容量の制約から、2ch4バンドの場合で1バンドあたり60次程度まで増やすことができる。

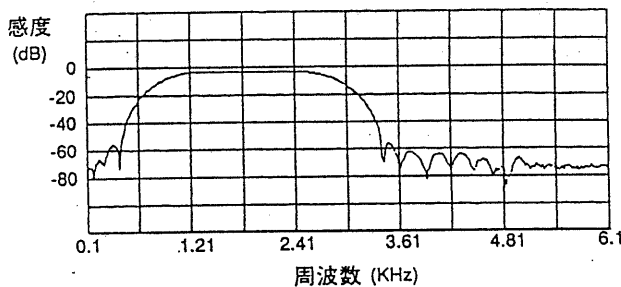


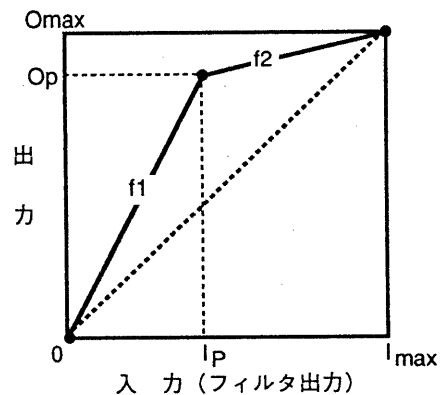
図4 帯域分割フィルタの一例

一方、帯域分割後のフィルタ出力の振幅処理方式に関して、多くの方法が提案されている

[1][2][9][10]。本システムでは、図5に示すような振幅変換を各帯域分割フィルタの出力に施し、小振幅信号の増幅と大振幅信号の圧縮によるダイナミックレンジの圧縮を行なったのち、全帯域の信号を加え合わせて出力を得ている。

帯域別振幅圧縮処理に於て、各帯域の振幅増幅度や出力のダイナミックレンジは、聞き手の聴力損失に適合するよう選ぶのが一般的である。本システムでは、図1で示したとおり、オーディオグラムなどの聴力検査データを元に、本処理に必要なパラメータをグラフィック上で調整するソフトウェアを用意し、このソフトで得られた信号処理用パラメータをDSPプログラムの中に取り入れる形で、個人特性への適合を実現している。図5の折れ線の折れ曲がり点や出力の最大値等は、各帯域毎にこのソフトウェア上で対話形式で行なう。

また、個人の聴力特性にパラメータを合わせる場合でも、いくつかのパラメータセットを用意し、使用環境に応じて選択できる機能が評価実験などでは必要になる。本システムでは、4通りのパラメータセットをプログラム内に記憶しておき、実行時にセレクト回路を用いて4種類のパラメータセットから選択できるようにした。



$$f1(I) = \frac{O_p}{I_p} \cdot I$$

$$f2(I) = \frac{(O_{max} - O_p)}{(I_{max} - I_p)} \cdot (I - I_p) + I_p$$

図5 振幅変換処理の入出力関係

(2) 録音再生時の話速変換

スピーチプロセッサ上の録音メモリには、常に約40秒の入力音声記憶される。そして、使用者が再生ボタンを押すと数秒前の音声から再生が開始され、再生ボタンを押す毎に同じ秒数ずつ進むことができる。聞き直し動作に最適な戻り量についての詳細な検討は行っていないが、本システムでは5秒に設定した。また5秒分の再生が終了した直後には約1秒の待ち時間を設け、この間にもう一度再生ボタンを押すと、同じ部分を繰り返すようにした。

前述したとおり、本システムではこの録音再生による聞き直しを行う時に、無音区間伸張を利用した話速変換を行っている。無音区間部分のみを伸張することで、ピッチを落とさずに話速を遅くすることができる。

この無音区間の伸張は、図6に示すように、再生音声の時間平均で算出されるエンベロープ波形の閾値判定を行い、閾値以下の区間が40ミリ秒～600ミリ秒連続した場合を無音区間とみなし、無音区間部分のデータをN回繰り返して出力することにより実現した。

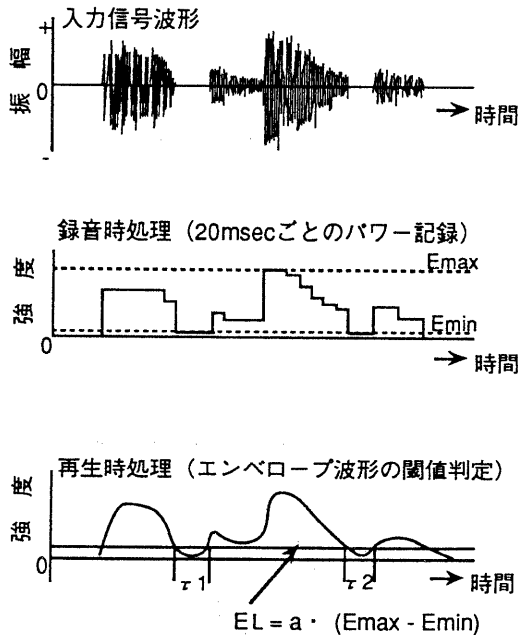


図6 無音区間検出の概念

本方式の場合、入力レベルの変動や背景雑音によらず、安定した無音区間検出ができるように閾値を設定することが重要である。そこで本システムでは、図6に示したように、入力音声をメモリ録音する際に、20msecごとのパワーも同時に計算して記憶しておき、再生を行う毎に録音されている区間内のパワーの最大値 E_{max} と最小値 E_{min} に基づいて閾値 E_L を設定することで、入力音声のレベル変動があった場合でも、閾値が極端に外れないようにしている。また、閾値設定のための係数 a は、4種類用意され、実行時にセレクタ回路によって選択/変更することができる。

無音区間データの繰り返し量 N については、詳細な検討はまだ行っていない。ゆっくり話したと感じられるためには、 $N=10\sim 20$ 程度の値を設定する必要があるが、吃音部などに言語としての不自然さが生じるため、本方式での話速変換には限界がある。しかし、感音性難聴者では継時マスキングの影響が、健常者に比較して大きいことも報告されており[11]、本方式のような無音区間の伸張により、この継時マスキングの影響をある程度軽減して、聞きやすさを向上できる可能性もある。

5. 性能

図7には、ポータブルシステムとして使用するためのケースに収めたスピーチプロセッサの外観を示す。この場合の大きさや重さは表2に示すとおりで、システム手帳程度の携帯可能な大きさに収めることができた。電源には1.2Vの充電式ガム型電池を4本使用しており、約2時間の連続動作が可能である。

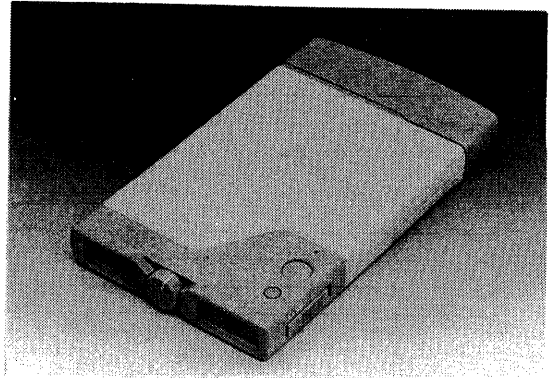


図7 スピーチプロセッサの外観

表2 性能一覧

大きさ	システム手帳サイズ 11.2(cm)×19.5(cm)×2.2(cm)
重さ	430(g) 本体：290(g) 電池部：140(g)
携帯可能時間	2時間弱

表3には、各部分回路毎に測定した消費電力を示す。DSPの消費電力は実行するプログラムにより大きく変動するが、本報告で述べた処理を行うのに必要な電力は0.6W程度であった。表3より分かるとおり、AICの消費電力がDSPでのそれとはほぼ同じであり、ポータブル性を一層高めるためには、アナログインターフェイス部分の低消費電力化が、DSPの低電力化と同様重要になる。

表3 消費電力一覧

部分回路	消費電力 (W)
DSP	0.6
AIC	0.6
プログラムメモリ	0.2
音声メモリ	0.05
その他	0.25
合計	1.7

また表4には、DSPプログラムにおける各処理のプログラム長および処理時間を示した。DSPでは、命令によって必要なマシンサイクルが大きく異なるため、実際にプログラムを実行し、エミュレーションシステムでモニタしたマシンサイクルから、各処理に要した処理時間を概算した。

イニシャルプログラムには、プログラムメモリ上から各処理に必要なパラメータをDSPの内部RAMに転送する処理と、AICの条件を設定する処理などが含まれる。これらの処理は、電源投入時はもちろん、セレクト回路によってパラメータの選択/変更が行なわれる際にも実行される。またメインプログラムには、帯域別振幅圧縮処理と話速変換処理が含まれる。

帯域別振幅圧縮処理に必要な時間は1バンドあたり4.27μsecで、2ch4バンドの処理を行なうのに約38μsecの時間を要する。サンプリング間

隔は75μsecなので、リアルタイムに行なう処理として十分な速度である。

一方、録音再生/話速変換処理では、再生ボタンが押されてから再生が開始されるまでに、無音区間検出のための計算/設定が行なわれるが、これに要する時間は290μsecであった。また、無音区間伸張処理は、無音区間の検出に条件判定などが多いためプログラム長は大きくなるが、無音区間部分の検出から出力データのアドレス区間を決定するまでには45.7μsecの時間しか要さない。

表4 プログラム長と実行時間

処理内容	プログラム長 (step)	実行時間 (μsec)
イニシャルプログラム	180	
パラメータ転送	40	344.0
AIC設定	35	516.3
その他	105	—
メインプログラム	400	
帯域別振幅圧縮	55	37.7
録音再生/話速変換	280	
閾値設定	35	290.0
無音区間伸張	245	45.7
その他	65	—

(AD/DAサンプリング間隔=75μsec)

6. おわりに

本報告では、従来より用いられている帯域別振幅圧縮処理に加えて、メモリ録音を利用した聞き直し動作と、無音区間伸張による話速変換を併用する補聴方式を提案し、これを携帯可能な大きさを実現したスピーチプロセッサの構成について述べた。また、種々の音声処理方式の比較検討に利用するための、エミュレーションシステムとのシステム化方法についても述べた。

本報告で述べた各種音声処理については、難聴者による評価が不可欠である。特に聞き直し動作や話速変換の難聴者に対する効果については、従来より検討がなされておらず、評価を踏まえたパラメータの最適化や処理方式の改良が必要である。

また、帯域別振幅圧縮処理では、フィルタ出力の比線形増幅を行うことになり、高調波歪の発生は避けることができない。本方式の場合、図5から明らかなように、フィルタ出力信号のレベルが入出力関係を表す2つの直線にまたがって変化するような場合に高調波歪が発生する。この点で時間領域での振幅圧縮処理には限界があり、高調波

歪を完全に避けて高い音質を得るためには、周波数空間での信号処理を行う方が有利であろう。

一方、無音区間伸張による話速変換には、前述したとおり言語としての不自然さがあり、背景雑音の影響も受けやすいため、実用面での問題が大きい。従って、ピッチ単位での波形伸張など[12]の、話速変換を実現する他の方法との比較検討をおこなう必要がある。

さらに、本報告のシステムでは2ch処理を実現したが、それを補聴に対して積極的に利用してゆく方法は、今後の検討課題として残されている。

参考文献

- [1]伊福部「聴覚障害補償へのデジタル技術の応用」音響学会誌, Vol.47, No.10, 760-765 (1991).
- [2]Working Group on Communication Aids for the Hearing-Impaired: "Speech-perception aids for hearing-impaired people: Current status and needed research" J.Acoust.Soc.Am. 90(2),Pt.1 (1991).
- [3]浅野、他「ラウドネス補償特性を有するデジタル補聴器の一構成法」音響学会誌, Vol.43, No.6, 373-379 (1991).
- [4]H.Levitt et al.: "Evaluation of orthogonal polynomial compression", J.Acoust.Soc.Am, 90(1) (1991).
- [5]上田、他「合成母音の聴取評価に基づくデジタル補聴器設計システム」音声研究会資料, SP90-94 (1991).
- [6]Texas Instruments, "TMS320C30ユーザーズマニュアル", 日本テキサス・インスツルメンツ(株), (1990)
- [7]J.Bradley: "TMS320C30 Hardware Applications" in 「Theory, Algorithms, and Implementations Vol.3」, Texas Instruments (1990).
- [8]尾野「デジタル補聴器の現状と将来」音響学会誌, Vol.47, No.10, 778-784 (1991).
- [9]池田、他「単共振分解・合成方式とその補聴器への応用」音声研究会資料, SP92-1, (1992).
- [10]B.C.J.Moor et al.: "Optimization of a slow-acting automatic gain control system for use in hearing aids", British J. of Audiology, 25,171-182 (1991).
- [11]林、他「破裂音の有声・無声知覚への継時マスキングの影響」音声研究会資料, SP90-97 (1991).
- [12]D.Malah: "Time-Domain Algorithms for Harmonic Bandwidth Reduction and Time Scaling of Speech Signals", IEEE Trans. Acoust., Speech, Signal Processing, Vol.ASSP-27, NO.2 (1979).