

ホルマント情報を伝達する聴覚障害補償システムの開発と
音声の認知に関する研究

2002年3月

熊本大学大学院自然科学研究科

池田 隆

目次

第1章 緒言	1
1-1. 音声コミュニケーションと聴覚障害	1
1-2. 補聴器とその歴史	1
1-3. 音声認知のための補聴処理	3
1-4. 音声情報としてのホルマント周波数	3
1-5. 単共振分解型補聴器の提案と音声認知の可能性	4
第2章 音声ホルマント情報を難聴者の可聴領域に圧縮伝送する 補聴方式の開発	7
2-1. 単共振分解型補聴器の原理	8
2-1-1. 音声信号の単共振分解・合成法	8
2-1-2. 単共振信号の振幅圧縮法	11
2-2. 電話帯域音声のための単共振分解型補聴器のシミュレーション	14
2-2-1. 構成と機能	14
2-2-2. 男女声により異なる第3ホルマントの概略的推定法	17
2-2-3. フレーム実効値に基づく圧縮係数の決定	21
2-3. 本補聴方式における振幅圧縮の効果	26
2-3-1. 母音部の振幅圧縮例	26
2-3-2. 子音部の圧縮性能	28
2-3-3. 難聴状態を模擬した健聴者による予備的な圧縮音声認知実験	32
2-3-3.1 模擬難聴状態の生成法	32
2-3-3.2 聴力特性と入力音声レベルの変化に対応する圧縮音声の認知実験	32

2-3-3.3 合成音声の母音カテゴリー判断に対する振幅圧縮の効果	35
2-4.まとめ	40
第3章 音声画像で強化する単共振分解型補聴器による聴能訓練と その効果	41
3-1. 音声パラメータの画像表現とその特徴	41
3-2. 聴能訓練システムの構成と機能	47
3-3. 訓練方法	49
3-4. 訓練実験に参加した感音性難聴者の聴力特性と補聴器の 振幅圧縮特性	52
3-5. 訓練の効果と聴力特性の関係	55
3-5-1. 最小可聴閾値 (HTL) による平均聴力レベルとの関係	55
3-5-2. 提案する平均ダイナミックレンジの定義による訓練効果の解釈	64
3-6. まとめ	68
第4章 難聴者の可聴領域内ホルマント情報と語音聴取能力の関係	69
4-1. 難聴者の平均ダイナミックレンジと振幅圧縮音声による 語音聴取能力の関係	69
4-2. 各難聴者の可聴領域内ホルマントピーク分布の推定法	69
4-3. 難聴者における可聴領域内ホルマントピーク分布と 語音聴取能力の関係	71
4-4. 難聴者の語音聴取能力に関する一仮説	73
4-5. まとめ	74

第5章 健聴者の模擬難聴実験による難聴者の語音聴取能力 に関する仮説の検証	75
5-1. 仮定された難聴の聴力特性と振幅圧縮の制御	75
5-2. 模擬難聴状態の生成方法	77
5-3. 実験方法	77
5-4. 実験結果と仮説に関する検討	78
5-5. まとめ	81
第6章 単共振分解型補聴器の振幅圧縮性能の改善に関する検討	82
6-1. 振幅圧縮特性が真の難聴と模擬難聴に与える影響の違い	82
6-2. 振幅圧縮特性の改善法	82
6-3. 特性の改善による聴取能力の向上の推定	88
6-4. まとめ	91
第7章 結 言	92
謝 辞	95
参考文献	96

第1章 緒言

1-1. 音声コミュニケーションと聴覚障害

発話と聴取により相互に意志の伝達を図る音声コミュニケーションは、ヒトを特徴づける行動のひとつである。呼吸に伴う空気の流れを発声器官の運動で制御し声を発し、聴覚を通して得た音響信号を脳で処理して言葉として理解する。これは「言葉の鎖」[1]と呼ばれ、音声コミュニケーションにおける情報の流れを特徴づけたものである。もし、聴覚に何らかの障害が生じると、聞き手側の鎖の一部が途切れるため、円滑な音声コミュニケーションが阻害される。さらに、このような聴力損失の度合いが大きくなると自己の発話の聴取（聴覚フィードバック）にも支障を生じ、話し手側で発声そのものも正確に行えなくなる恐れがある。先天的要因や後天的要因に起因するこのような聴力損失は、徐々にあるいは急速に進行し、失聴に至ることも多い。特に、音声言語の習得以前の幼児期に高度の聴覚障害が出現している場合は、聴力損失に対し早期に適切な補償を講じることがその後の順調な発育をもたらすと言われている[2]。また、健聴者でも加齢と共に聴力が低下するため高齢人口の増加に伴い、最近では老人性難聴への有効な補償も重要な課題となっている[3]。

1-2. 補聴器とその歴史

聴覚障害を生じた時、可聴領域が残っていれば聴力損失を補うための補聴器が必要となる。補聴器の歴史は古く、電気技術以前の集音型ラッパ補聴器に始まり、音声の電気信号を増幅するアナログ電気回路による補聴器[4]、更には半導体技術の進歩に併せて、箱型、耳かけ型、耳穴型、インナー型補聴器へとその小型化と性能向上

が促進され装用性も高くなってきた[5]。これらの補聴器で用いられている信号処理方式を大別すると、音全般を対象とする音響信号処理と音声情報伝達を意識して音声信号処理に特化した方式がある。前者の例では、高域、中域、低域などの周波数帯域毎での増幅度を制御する方式[6]、損失聴力を補う周波数応答を持った増幅器（オーディオグラムの逆特性を有するミラー増幅器）[7]などが挙げられる。一方、後者では、音声の周波数スペクトル分析に基づいて増幅特性を決定する方式[8]、音声における子音部やホルマント成分など特定要素の強調を狙った方式[9][10]、大まかな処理によって得たホルマント信号成分を強調する方式などがある[11]。

このように、補聴器のための信号処理は単なる線形増幅から各種フィルタ処理や複雑な処理を応用した非線形増幅へ進み、1986年頃にはデジタル補聴器の提案もなされ[12]、音声情報の伝達に特化した信号処理の可能性も見えてきた。1990年代に入りマイクロコンピュータやDSP（Digital Signal Processor）の活用も広がって、デジタル補聴器が具体化し[13]、技術革新の進展により複雑な信号処理方式もリアルタイムで実現し得るようになった。国内でも固定小数点DSPを使用したデジタル補聴器が開発された[14]。更に今日では、浮動小数点演算型のDSPの性能充実でハードウェア上の問題は解決されつつある。一方、ソフトウェア上の問題、すなわち、より効果の高い補聴方式の実現、言い換えると補聴器で目標とすべき理想的な信号処理はどのようなものか、を明らかにすることが強く求められるようになった。このような補聴処理方式の問題を解決するためには、音声に含まれるどのような音声情報を難聴者の可聴領域の中に伝達するのが最も効果的であるかという、音声認知における本質的な部分に関わる事項の究明とその応用技術の開発が望まれている。

1-3. 音声認知のための補聴処理

日常生活での音声コミュニケーションを回復するには音声の認知を十分に補助する補聴器が期待されるが、その補償方法は聴覚障害の特性によって異なる。即ち、音の伝音系（外耳, 中耳）に障害があり最小可聴音圧レベルの上昇と共に、不快に感じる音圧レベルも同様に上昇し、聴取できる音圧のダイナミックレンジが健聴者と同じ程度に保たれる伝音性難聴では、ミラー増幅などによる補償効果が期待できる。一方、内耳以降に障害の原因がある感音性難聴の場合は音を不快に感じる大ききのレベルは健聴者とあまり変わらないにもかかわらず、聴き取ることのできる音圧の最低レベルが上昇する。その結果、聴取可能な音圧のダイナミックレンジが狭められてしまうので、巧妙で緻密な信号処理が必要となる。このような聴力特性の補償には一般に、信号レベルのダイナミックレンジを聴取可能な音圧のダイナミックレンジに入れ込む振幅圧縮型の補聴器が必須となるが、その性能は、音声の聴き取りが自然にでき、装用していることを意識しない程度に歪が少なく、使用者の聴力に適合する特性が達成されねばならない。可聴領域が狭くなった難聴者に対し、いかにしてこれらの要求を満たす信号処理を行うか、とりわけ、日常生活に不可欠な音声の聴き取り改善に、どのような方式がよいのかということは聴覚障害の補償システムを構築する上で大変重要な課題である。これらの課題に解決を見出すことができれば、音声情報を効率良く伝達する方法、補聴器による音声認知に対する補償効果の評価、補聴器設計や人工内耳の研究などにも有効な指針を与え得ることになる。

1-4. 音声情報としてのホルマント周波数

音声生成モデルによると、音声生成過程は、音源, 声道, 放射の要素によって形成

され、声道の伝達関数を全極モデルとして記述するとき、その共振周波数は、周波数の低いものから順に第1, 第2・・・第nホルマント (F_n) と名づけられる[15]。音声分析により抽出した第1, 第2ホルマント周波数 (F_1, F_2) を平面にプロットするとそれらの点は各母音ごとのまとまりを示すことが知られている[16]。これらのホルマント情報は補聴器で伝達する音声情報として重要なパラメータであるにもかかわらず、これまで音声信号から安定で精度良く抽出することが容易でなく実用装置に応用される例はほとんどなかった。しかし近年、提案、確立された逆フィルタ制御方式によるホルマント周波数抽出法[17]は、ホルマント情報を実用装置で利用する上で高品質と高精度が期待できるものである。

発声の際に話者は無意識の内に発話したいと思う音素に対応して、音源を制御したり、その他の発声器官を運動させる。たとえば声道内にせばめが存在すると、音声信号の中に共振成分（ホルマント周波数）を生じ、スペクトルピークとして観察される。更に、せばめの位置変化等によりホルマント周波数（複数）やレベルが変化する。

1-5. 単共振分解型補聴器の提案と音声認知の可能性

本研究では音声情報を感音性難聴者の狭い可聴領域へ効率的に伝達するため、発声者の意思に直結した発話器官の運動や音声の認知と関わりが深いと考えられるホルマント成分そのものに着目した新しい補聴方式として、単共振分解型補聴器[18][19]を提案する。この補聴器の音声により聴取実験とその分析によって、ホルマント情報と音声認知との関連を難聴者、模擬難聴状態の健聴者双方で明らかにし、ホルマント情報の積極的応用への足がかりを示すことを目的としている。

この音声認知とホルマント情報の中に存在する一定の関連を明らかにすること

ができれば、聴覚障害の補償システムの開発や評価への活用が期待できる。また、かつて電話が A.G.Bell により、当初聴覚障害者とのコミュニケーションを目的に開発され、その後社会全般で不可欠な機器として普及したように、障害の補償に関する分野のみならず、現代の多様な情報機器における音声コミュニケーション性能評価など、適用分野は極めて広く意義が深いと言える。

提案する単共振分解型補聴方式は、逆フィルタ制御方式により時間領域で分析・抽出したホルマント (F_1 , F_2) のピークレベルを、それぞれの実効値と周波数での可聴領域から決定した係数で圧縮後加算し難聴者の可聴領域に入れるので、歪みの少ない振幅圧縮音声を提供する。この補聴器によって可聴領域に入ったホルマント情報と音素を結びつけ音声認知の改善を図るべく、音声画像[20]を併用し効率化した聴能訓練を5人の難聴者に対し実施しその結果を分析する。これらより音声認知と可聴領域内のホルマントピーク分布との関連について考察しその関係について仮説を得る。次に健聴者による模擬難聴下での振幅圧縮音声を含めた音声の認知とホルマント情報の関わりについて仮説を実証し、最後に単共振分解型補聴器の特性改善について考察する。

本論文の構成は次の通りである。第1章は本章緒言である。第2章はホルマント情報の伝達を目的として新しい補聴方式（単共振分解合成方式）を提案し、その原理、信号処理について述べる。

第3章では、ホルマント情報を認識手法を全く用いずカラー情報に変換した音声画像を併用した聴能訓練で難聴者の音声に対する認知能力を高め、音声聴き取りに関わりの深い要素について検討する。

第4章では、難聴者の可聴領域内ホルマント情報と語音聴取の関係について議論し、

音声認知と音声ホルマント情報との関連について仮説を得る。

第5章では前章で得た仮説を模擬難聴状態の健聴者により検証し、模擬難聴下の健聴者及び難聴者の音声認知から仮説を立証する。

第6章では仮説による知見を基に、提案した単共振分解型補聴器の振幅圧縮性能の改善について検討し、難聴者での更なる音声認知能力の改善について検討を行う。

第7章は結言である。

第2章 音声ホルマント情報を難聴者の可聴領域に圧縮伝送する 補聴方式の開発

この章では、音声を時間領域で分析し、歪みの少ない補聴音声を提供する単共振分解型補聴器を提案する。一般に、聴覚障害の性質は障害部位の違いにより、伝音性、感音性、および両者の混在した混合性の3種に分類できると言われる。鼓膜から中耳にかけての障害による伝音性難聴の典型的な特性は、健聴者の特性に比べて最小可聴音圧(HTL:Hearing Threshold Level)の上昇とともに大きすぎて不快と感ずる音圧(UCL:Uncomfortable Level)も上昇するので可聴領域のダイナミックレンジは健聴者とあまり変わらない。したがって、基本的には、全周波数帯域に関して一様または最小可聴音圧の周波数特性を反映したミラー形の特性で増幅する線形増幅の補聴器によって補償できる[7]。それに対して、内耳に障害のある感音性難聴では、最小可聴音圧(HTL)が上昇するのに対し、不快と感ずる音圧は健聴者の場合とあまり変わらないことから可聴領域のダイナミックレンジが狭くなる特性が現れる。この特性に音声を適合させる補聴器は、音声の周波数スペクトルに応じて利得の周波数特性を変えることの可能なものであり、振幅圧縮型補聴器と呼ばれる。本章で提案するのは、感音性難聴の補償に対する新しい振幅圧縮型補聴器である。これまで提案された振幅圧縮型補聴器には、帯域フィルタ群を用いてそれぞれの出力成分が可聴領域に入るように圧縮して再合成する方式[6]、同様に、フーリエ変換・レベル圧縮・フーリエ逆変換を用いてスペクトルを圧縮する方式[8]などがある。これらの方式はスペクトルの各成分がすべて可聴領域に入るように圧縮するので、可聴領域が狭くなるにつれてスペクトル歪みが大きくなる傾向は避けられない。それに対して、提案方式は、音声信号をホルマント周波数と逆フィルタで単共振成分に分解

し、それぞれの成分のホルマントピークが可聴領域に入るように増幅したのち再合成(加算)することにより歪みの少ない振幅圧縮を実現することを目的としている。この方式は、信号処理がやや複雑になることから現段階では超小型の補聴器とは為し得ないが、まず電話機に付属した補聴器として実用可能なものにするを考えている。次に、その原理、補聴器のシミュレーションと特性、補聴効果の主観評価などについて述べる。

2-1. 単共振分解型補聴器の原理

2-1-1. 音声信号の単共振分解・合成法

まず、音声信号を単共振波に分解する原理について述べる。

一般に、 n 個の共振成分で合成されている音声信号を個々の成分に分解するには、1 個を残して他の $(n-1)$ 個の共振周波数を零点の周波数とする 2 次系の逆フィルタが縦続接続されたシステムに音声信号を入力して出力を得る操作を n 個の成分に順次繰り返せばよい。この原理を簡単に述べれば次のようになる。

通常、音声の生成原理によれば、音声信号 $x(n)$ の z 変換 $X(z)$ は次式で表現される。

$$X(z) = G(z) \cdot H(z) \cdot R(z) \quad (2.1)$$

ここで、 $G(z)$: 音源特性、 $H(z)$: 声道特性、 $R(z)$: 放射特性を示す。

今、周波数帯域制限により、2 個の共振成分のみを含む信号が得られているとすると、

$$H(z) = H_1(z) \cdot H_2(z) \quad (2.2)$$

となり、何らかの方法で 2 個の逆フィルタのシステム関数、 $1/H_1(z)$ 、 $1/H_2(z)$ を推定できたとすると、それぞれに音声信号を入力して

$$X(z)/H_1(z) = G(z)H_2(z)R(z) = X_2(z) \quad (2.3)$$

$$X(z)/H_2(z)=G(z)H_1(z)R(z)=X_1(z) \quad (2.4)$$

を得る。一般に、 $H_1(z)$ 、 $H_2(z)$ の帯域幅（ホルマント帯域幅）の正確な推定は困難であるので、実際には、逆フィルタの帯域幅は固定して零点の周波数のみを一致させることにより、近似的な単共振成分 $X_1(z)$ 、 $X_2(z)$ を得る。フィルタ出力は $X_1(z)$ 、 $X_2(z)$ のそれぞれの逆 z 変換と一致する時間信号 $x_1(n)$ 、 $x_2(n)$ となるので、最後に、両成分を係数 a_1 、 a_2 を乗算して増幅したときにホルマントピークが可聴領域に入るようにしたのち両者を加算する。すなわち、

$$a_1 \cdot x_1(n) + a_2 \cdot x_2(n) = x'(n) \quad (2.5)$$

以上の処理を Fig.2-1 の原理図に示した。

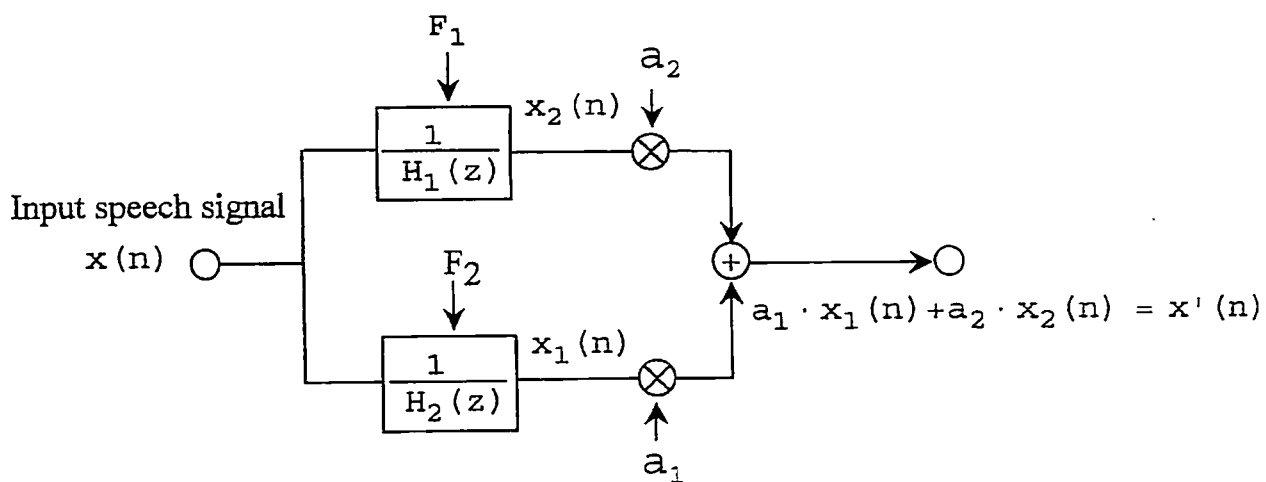
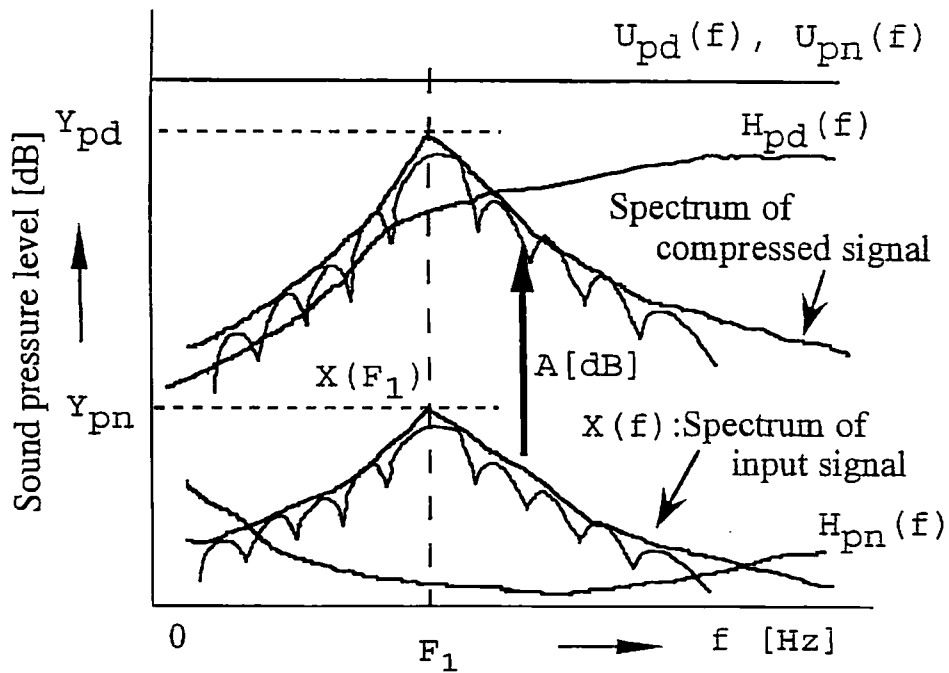


Fig. 2-1 Concept of single resonant analysis-synthesis (SRA) system.

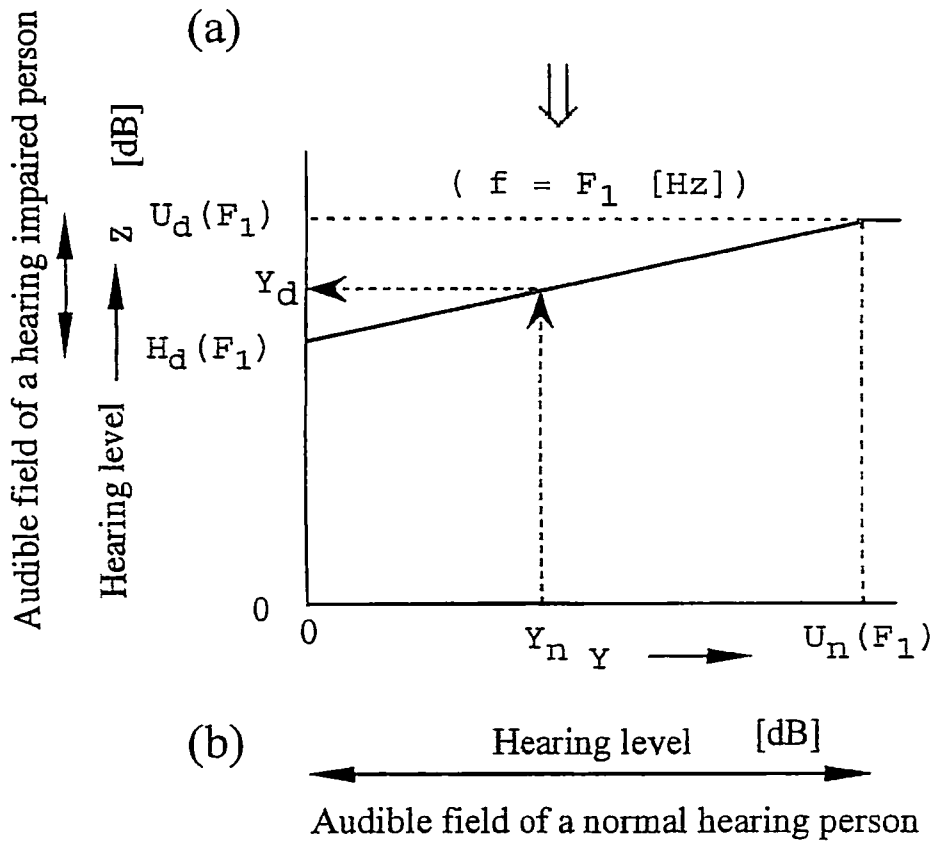
2-1-2. 単共振信号の振幅圧縮法

前節の原理において、増幅係数、 a_1 、 a_2 を難聴者の聴覚特性からどのように定めるかが一つの問題になる。今、Fig.2-2(a)のように、難聴者と健聴者の最小可聴閾値がそれぞれ音圧レベルの周波数特性で $H_{pd}(f)$ と $H_{pn}(f)$ として与えられ、かつ、不快と感ずる音圧レベルが $U_{pd}(f)$ 、 $U_{pn}(f)$ であると仮定し、単共振分解された 1 成分（図では、 F_1 成分）の振幅スペクトルが $X(f)=20 \log |X(e^{j2\pi f})|$ [dB] として示されているものとする。一般に、単共振分解されたスペクトルは難聴の $H_{pd}(f)$ より低音圧レベルのため聴取できないから、 A [dB] 増幅してホルマント近傍のスペクトルピークが可聴領域に入るようにしなければならない。この方式の特徴は、単共振スペクトルの形状は変えずに信号を増幅するので歪みを軽減できることである。ここで、利得 A は、ホルマント周波数（図の F_1 ）における $H_{pd}(F_1)$ 、 $U_{pd}(F_1)$ およびホルマントピーク値、 $X(F_1)$ の関数として定めることになる。Fig.2-2(b)は、Fig.2-2(a)のホルマント周波数 F_1 における健聴者の最小可聴音圧を 0 [dB] として定めた難聴者の聴力閾値レベル (HTL) と不快レベル (UCL) を健聴者聴力レベル (HL) の 2 次元空間に表したものである。図のように、難聴者の可聴音圧の範囲を健聴者のそれにデシベル ([dB]) のスケールで線形に対応させれば、それを変換関数として単共振成分のホルマントピーク音圧レベルから難聴者が聴取可能な音圧レベルへ変換するのに必要な増幅度 ([dB]) を算出することが可能となる。

すなわち、聴力レベルとして表現された難聴者の HTL, UCL をそれぞれ、 H_d 、 U_d 、健聴者の UCL を U_n とし、同じ基準の Y_n 、 Y_d をそれぞれ健聴者と難聴者に与えられるホルマントピークレベルとすれば、



(a)



(b)

Fig. 2-2 Principle of amplitude compression in SRA system.

$$Y_d = Y_n(U_d - H_d)/U_n + H_d \quad ; \quad 0 \leq Y_n \leq U_n \quad (2.6)$$

$$Y_d = U_d \quad ; \quad Y_n > U_n$$

したがって、 $0 \leq Y_n \leq U_n$ における増幅利得は、

$$\begin{aligned} A &= Y_d - Y_n \quad [\text{dB}] \\ &= Y_n \{ (U_d - H_d) / U_n - 1 \} + H_d \end{aligned}$$

ここで、可聴領域を $U_d - H_d \equiv B_d$ と定義して、

$$A = Y_n \{ B_d / u_n - 1 \} + H_d \quad (2.7)$$

このようにして瞬時、瞬時のフレームにおける単共振スペクトルのピーク（ホルマントピーク）、 Y_n が与えられると(2.7)から利得を計算できる。今、音声信号を2つのホルマント (F_1 , F_2) により単共振信号に分解したとすると、 Y_n のみならず B_d , U_n もまたそれぞれのホルマント周波数によって決定され、(2.7)はそれぞれの成分に対して、 $A(F_1)$, $A(F_2)$ が計算される。従って、(2.5)の時間信号、 $x_1(n)$, $x_2(n)$ への係数値 a_1 , a_2 は

$$a_1 = 10^{A(F_1)/20}, \quad a_2 = 10^{A(F_2)/20} \quad (2.8)$$

として定めることが可能になる。しかし、上記の処理で単共振分解後のホルマントピーク値を直接使用するには信号の FFT 処理などを必要とする。したがって、係数値の決定をより簡単に行うために、ホルマントピーク値と単共振分解信号の実効値の関係を多数のデータから統計的に推定しておいて、実効値から計算する方法を用いる。連続音声の振幅を表すフレーム実効値の変化は緩やかであり、実効値による振幅圧縮では、瞬時振幅圧縮での非線形変換による高調波歪を生ずることはないという特徴を有する[18]。この実効値を利用した振幅圧縮への拡張については次節で述べるシミュレーションの具体例(2-2-3)の中で示す。

2-2. 電話帯域音声のための単共振分解型補聴器のシミュレーション

上記の原理を応用した実用化システムとして、電話用の補聴器を取り上げ、シミュレーションによって得た特性について述べる。電話音声は通話帯域の制限により、声道の共振点も 2-3 個に限定されるので、システムの簡略化とリアルタイム処理が可能になると考えられる。

2-2-1. 構成と機能

Fig.2-3 は、電話音声用補聴器の構成を示している。図において実線は信号の流れ、破線は信号から推定されたパラメータの流れを示し、次のように信号処理が行われる。まず、電話の伝送周波数帯域から、伝送経路によらず伝送の周波数特性が比較的平坦である帯域として、300-3000[Hz]を選択し入力信号を帯域フィルタで制限した後、A/D 変換器の入力範囲を適正に利用し得るようにアナログ増幅する。その信号を 12[kHz]で標本化、A/D 変換して直接ピッチ抽出ルーチンへ入力し、1 サンプルおきに間引いた 6[kHz]サンプリングの信号は、第 1, 第 2 ホルマント周波数 (F_1, F_2) の推定ルーチンへ入力する。ホルマントの推定は、電話帯域用に開発した逆フィルタ制御法を、ピッチ抽出は信号振幅の非線形処理による基本波の復元と F_1, F_2 成分を消去する逆フィルタによるスペクトル平坦化、更に、複数候補の論理処理に基づいた最終推定を用いて行う[21]。この処理におけるひとつの問題は、男声の F_3 が帯域内に存在することが多く、女声のそれは 3[kHz]より高いことが多いことである。このことより、逆フィルタによる F_1, F_2 の推定に対する高次極の影響を逆フィルタ IF3 で除去するのに、 F_3 を男声と女声で切り替える必要があり、それを自動化しなければならない。

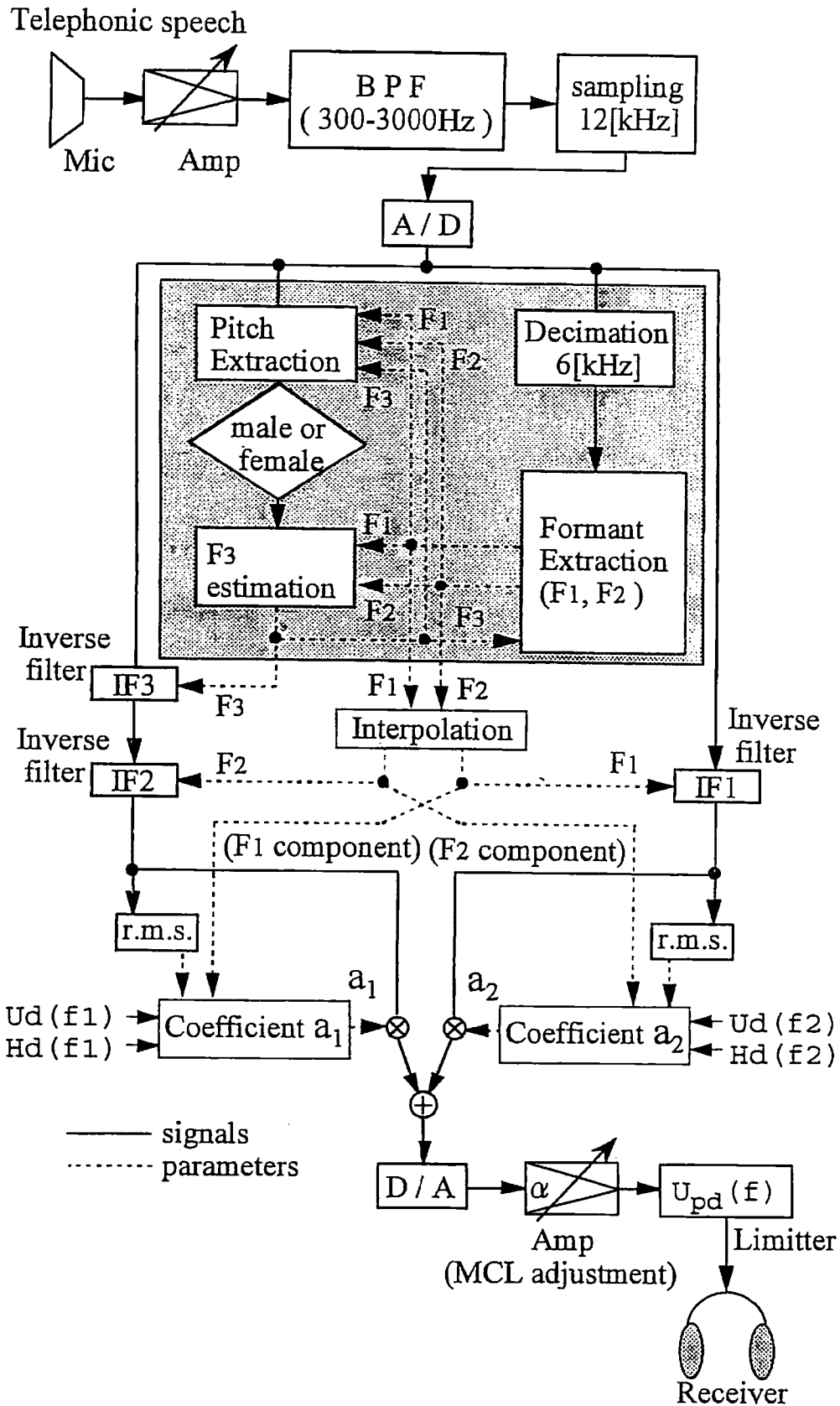


Fig. 2-3 Block diagram of hearing aid for telephonic speech by SRA method.

この問題に対する処理方式として、まず、スペクトル平坦化が不完全であってもピッチ抽出ができるだけ誤差を小さくできるようなものとし、図のように、抽出されたピッチから次節で述べる方法によって、男／女声の判定とそのフレームの F_1 , F_2 から統計的に前もって得られている回帰関数[22]を用いて F_3 を定め、それを次フレームのホルマント推定の高次極除去特性として与えることとした。このようにしてフレーム毎に抽出した F_1 , F_2 をフレーム間で個別に内挿し、それによって図の左右に置かれた逆フィルタ, IF1, IF2 の零点の周波数を制御しながら音声信号の逆フィルタ処理をおこなう。IF2 と縦続接続されている前段の逆フィルタ IF3 は推定された F_3 で制御される。その推定値が 3[kHz] を越える場合は IF3 の零点周波数を 3 [kHz] に固定して帯域幅を制御することにより 3[kHz] 以下の周波数特性 (スペクトル傾斜) を変化させる。このように逆フィルタ処理で2つの単共振分解された時間信号(F_1 成分および F_2 成分)を得る。(IF1の逆フィルタには IF3 を接続せず、 F_2 成分に F_3 の成分の部分的な混入を許容している。) さらに、(2-1-2)に述べた方法によって、それぞれの成分の実効値と使用者の聴力特性, HTL ($H_d(f)$) と UCL ($U_d(f)$) の F_1, F_2 におけるレベルから可聴領域まで増幅する係数, a_1 , a_2 を求める。最後に、 F_1 成分、 F_2 成分にそれぞれ a_1 , a_2 を乗じて加算し、振幅圧縮音声信号を得る。その信号を D/A 変換して得たアナログ信号を可変の利得 α で線形増幅する。利用者は、聴取音声最も聞き取り易くなるように、MCL(快適レベル)に α をセットして使用する。更に、その信号が $U_{pd}(f)$ を越えることのないようにリミッタで押さえる。この一連の処理によって、可聴領域と入力音声信号レベルの関係が音圧レベルで絶対的な関係を満足せずに別々に定められていても、通常は、ホルマントピークレベル ((2.7) の Y_n) が $H_{pn}(f)$ と $U_{pn}(f)$ の間 (聴力レベルで 0[dB] と U_n [dB] の間) にあるので、適正な圧縮信号に変換することが可能になる。

2-2-2. 男女声により異なる第3ホルマントの概略的推定法

上記の方式においては、処理を簡略化するために、 F_3 を男声と女声で別個の回帰関数に基づいた推定値として与える。回帰関数は、 F_1 、 F_2 を変数とする 2 次関数であり、男女声の別で係数値が異なるので、入力音声の男女判定によって関数の係数値を選択し、その関数に F_1 、 F_2 を代入して次フレームの F_1 、 F_2 推定のために必要な F_3 の値を得ることになる。男女声判定を厳密に行うことはやや難しい問題であるが、本方式の場合は、男女声の誤識別、すなわち、 F_3 の誤推定が敏感に F_1 、 F_2 の誤差に反映される訳ではないので、ピッチ (F_0) から大まかに男女声の別を判定して F_3 を推定する簡便な方法を用いた。Fig.2-4 は、その処理を含んだパラメータ推定の流れ図 (Fig.2-3 の網掛け部分) を示す。まず、あらかじめ、多数の男女別に収集された単母音から逆フィルタ制御法でホルマント周波数を抽出し、 F_1 、 F_2 から F_3 を推定する回帰関数[22]が得られているものとする。単語のようなホルマントが連続して抽出される音声ブロックを想定するとき、先行フレームの F_1 、 F_2 を用いて、男声、女声それぞれを仮定して、回帰関数から F_3 (図中の F_{3M} と F_{3F}) を推定する一方、単母音のピッチ周波数の測定による男声と女声の境界、192.4[Hz]を基準として、各フレームのピッチがその基準を越えた場合は女声フレーム、0 でない他のフレームは男声フレームと仮判定しながら両者の累積フレーム数を逐次カウントする。つぎに、各時点での男声と女声のフレーム数の割合から F_{3M} と F_{3F} についての比例配分の形で F_3 を決定する。 F_1 が検出されてピッチが抽出されない ($F_0=0$ の) 音声は、無声子音である可能性が高く、高い F_3 を想定した方がよいので、常に女声の回帰関数から推定値を得るものとした。 F_3 成分を消去する逆フィルタ IF3 の帯域幅、BW3 を F_3 の上昇によって増加する線形関数で与え、零点周波数 FF3 は、3000[Hz]以上にはシフトさせない。この処理によって、零交差周波数の加重平均でホルマントを推定する

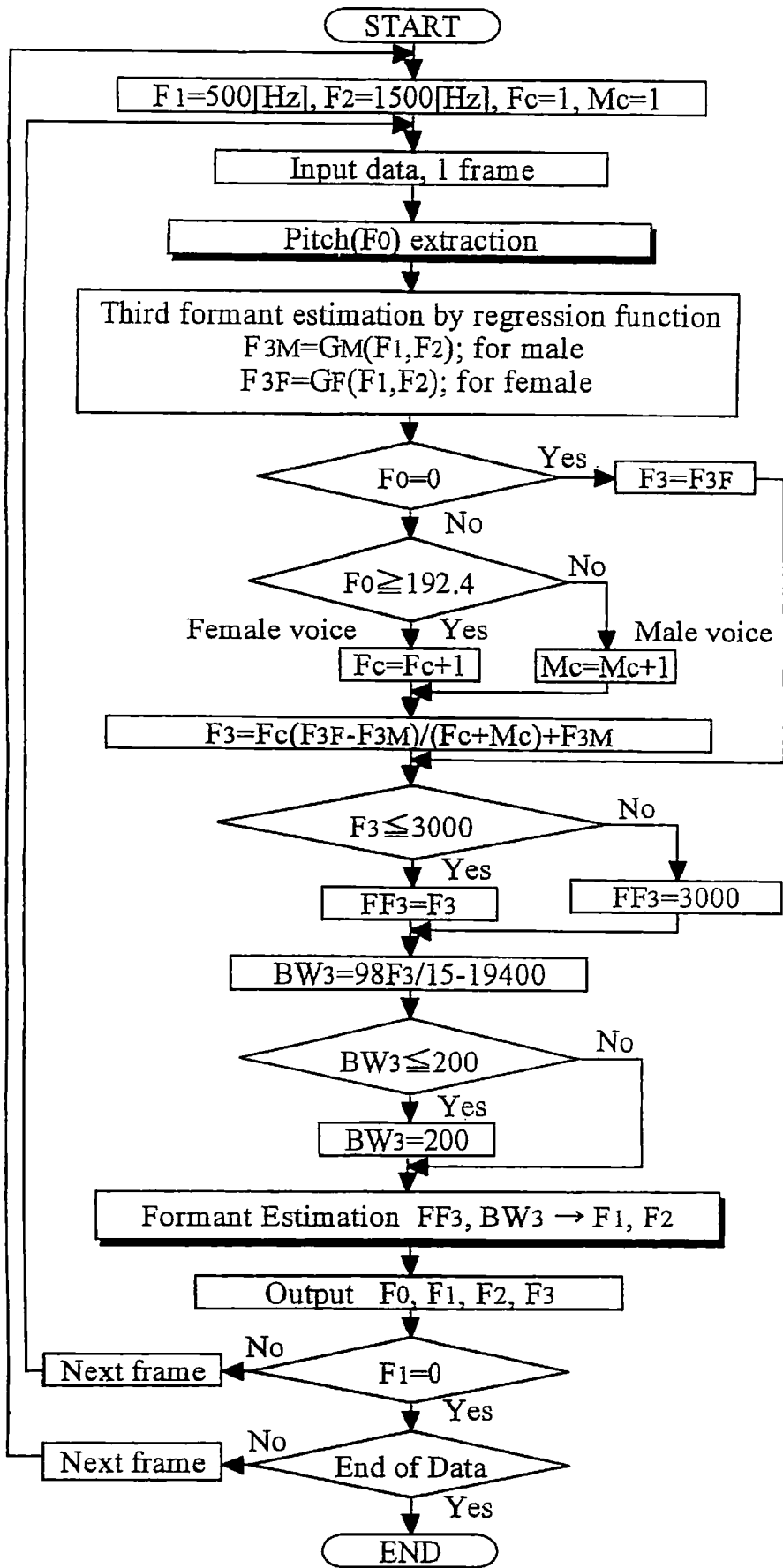
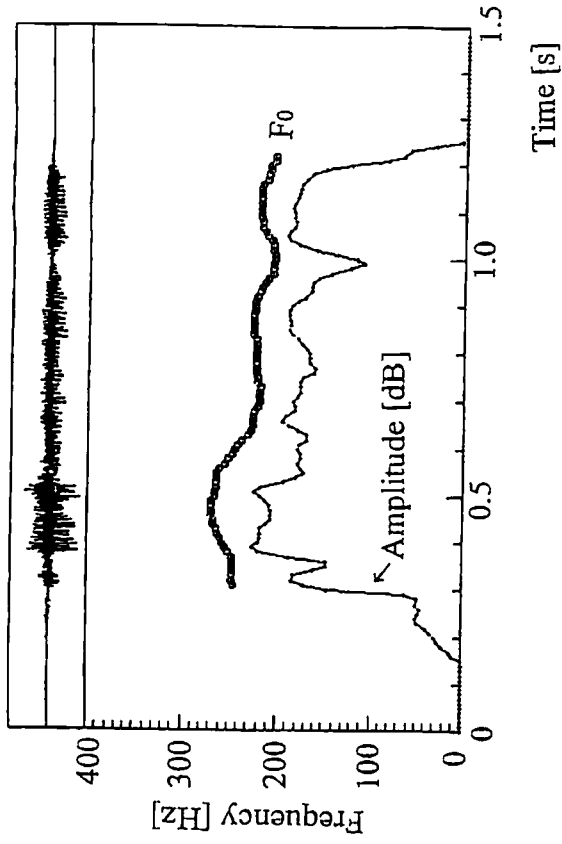


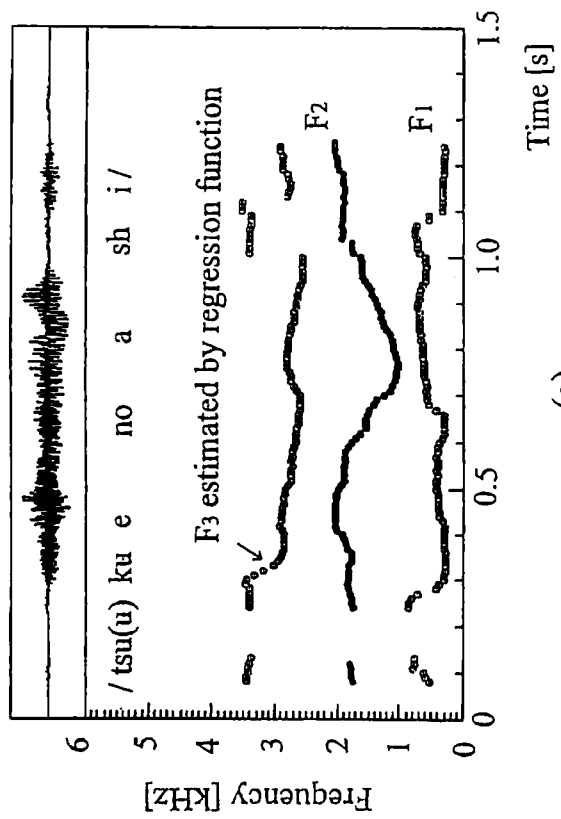
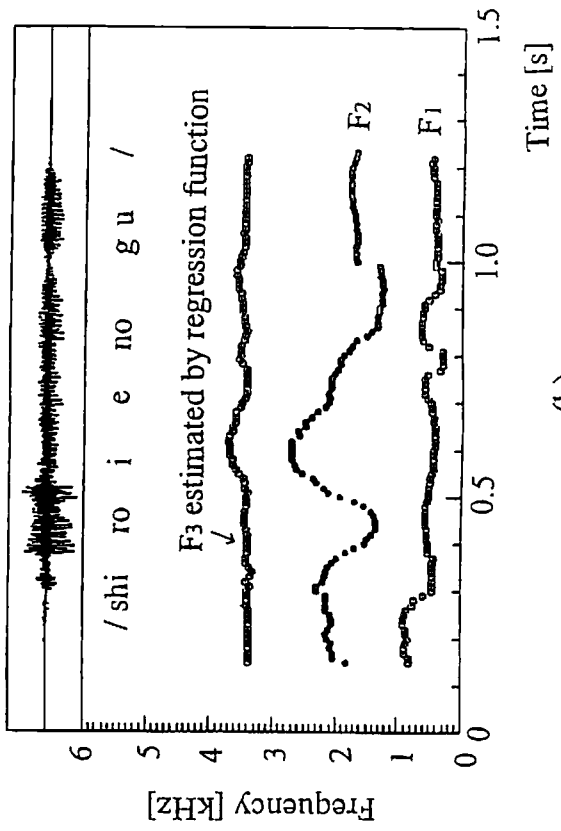
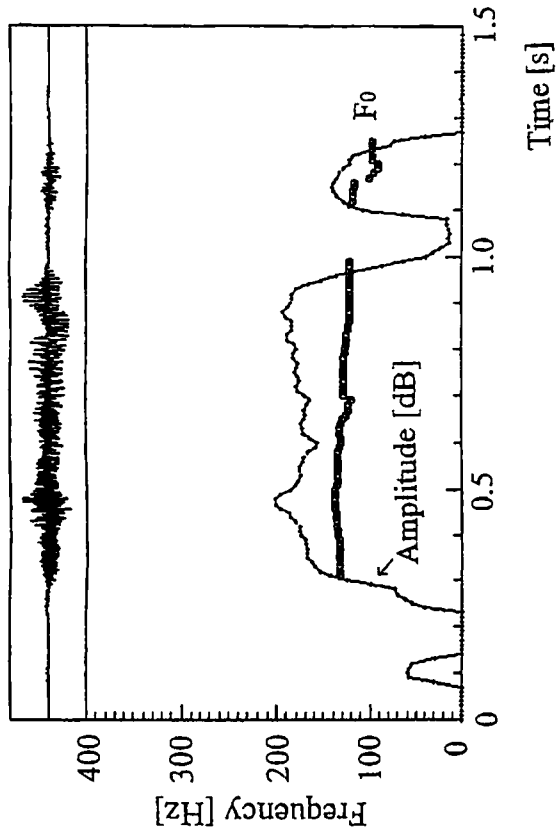
Fig. 2-4 Algorithm of the third formant estimation for formant analysis by inverse filter control.

のに必要な, 2つの共振成分を残すことが可能となる。Fig.2-5の例は, 男声と女声のホルマントおよびピッチの推定結果を示したものである。男声, 女声とも, 300[Hz]以下の基本波成分が復元され, それらしいピッチパターンが抽出されている。更に, 男声 (Fig.2-5(a)) の語頭は無声子音から有声音へ変化する部分でピッチ周波数が0なので女声と判定され, 女声用の回帰関数によって IF3 の零点が設定されているが, F_1, F_2 の推定に誤りを生ずるような影響はない。女声 (Fig.2-5(b)) の例では, 累積フレーム数による F_3 推定の荷重は圧倒的に女声の方が大きく, 結果として F_1, F_2 は正しく推定されている。

< Female (Age : 20) >



< Male (Age : 21) >



(b)

(a)

Fig. 2-5 Pitch and formant trajectories extracted by SRA system.

2-2-3. フレーム実効値に基づく圧縮係数の決定

推定された F_1 , F_2 と回帰関数で与えた F_3 を用いて, Fig.2-3 の3つの逆フィルタ, IF1, IF2, IF3 によって2つの単共振成分を抽出し, それぞれに(2.7)で求めた圧縮係数 a_1 , a_2 を乗じて加算する。しかし, 2-1-2 で述べたように, ホルマントピークレベルを推定することは処理を複雑にするので, その代わりに単共振成分の実効値からホルマントピークレベル値を推定して(2.8)に代入する方法を用いる。Fig.2-6 は, Fig.2-3 のシステムを用いて多数の単語データから2つの単共振成分を分解, 抽出した後, それぞれの成分 (F_1 , F_2) の実効値とホルマントピークレベル値の関係を示したものである。用いた音声試料は, ATR データベースの 520 語の重要語[23]から, 男女各5人の話者のそれぞれについてランダムに50語ずつ抜き出したものである。合計 500 単語の全フレームを分析に使用した。ホルマントピークレベルは単共振成分の各フレームをフーリエ変換 (512 点 FFT) し, ホルマント周波数の近傍に存在する高調波ピークを探索して得た。ここでは, Fig.2-6 の音圧レベルの関係を1次関数で近似することとした。即ち, 実効値を Z_{pn} として, (2.7) のホルマントピーク音圧レベル Y_{pn} が

$$Y_{pn} = cZ_{pn} + d \quad (2.9)$$

で表されるものとする。さらに, (2.7) の聴力レベルで示された Y_n との関係は,

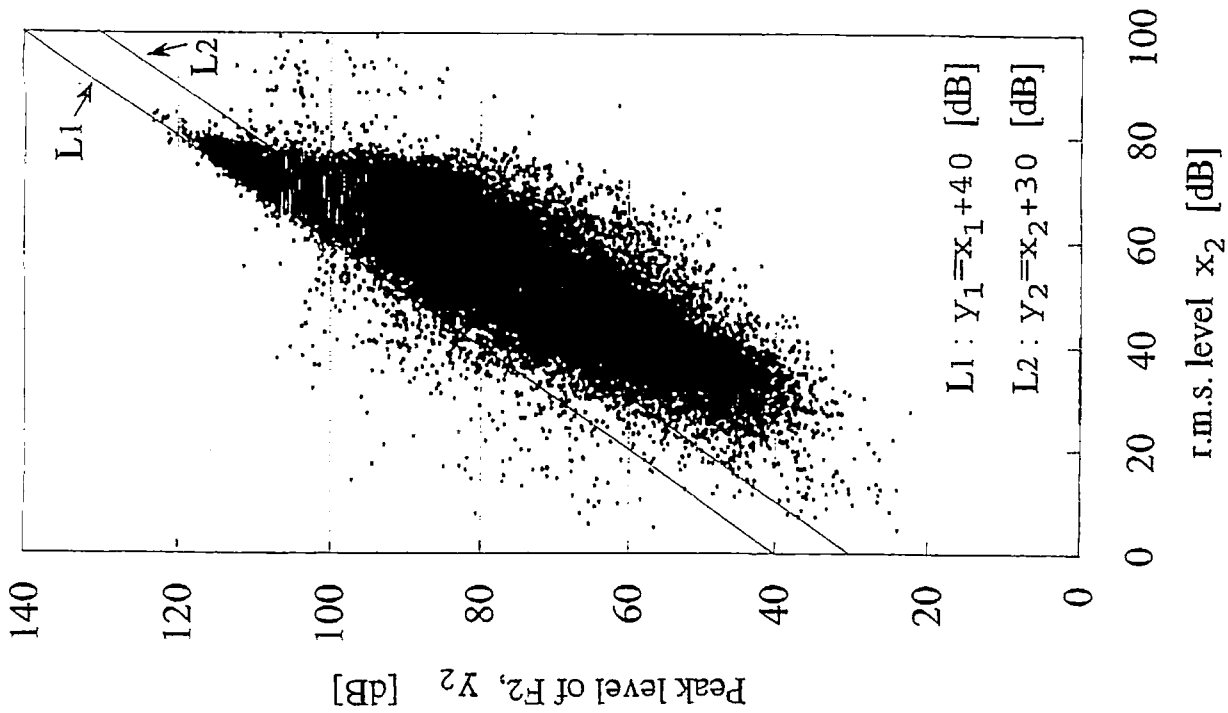
$$Y_n(f) = Y_{pn} - H_{pn}(f) \quad (2.10)$$

となり, 最小可聴音圧(MAF)の周波数特性 $H_{pn}(f)$ [dB] を, 代表的な測定結果から直線近似によって,

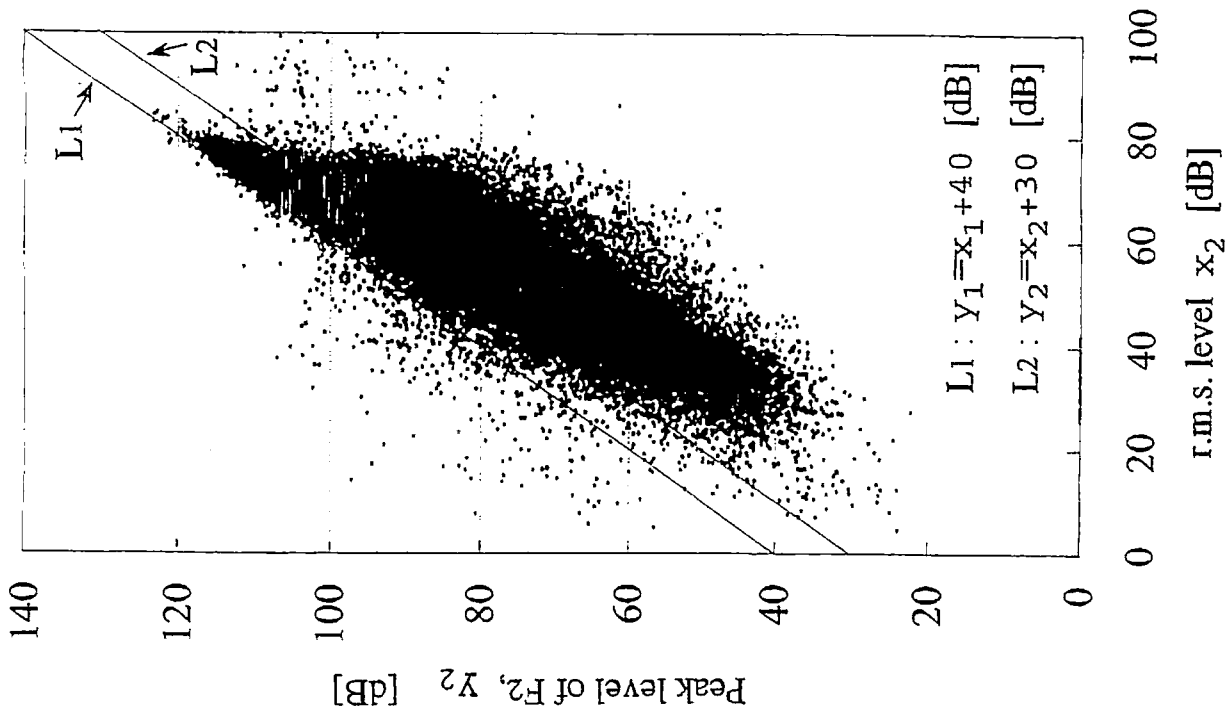
$$H_{pn}(f) = -10 \log(f) + 34 \quad 300 \leq f \leq 3000 [\text{Hz}] \quad (2.11)$$

と表現すればよい。Fig.2-7 に(2.11)の近似直線を示す[24]。(2.9) の定数である c と d は, F_1 成分, F_2 成分のそれぞれについて c_1 , d_1 , c_2 , d_2 があり, それらの値によ

り圧縮の状態が変わる。Fig.2-6 に示したと同じデータの(2.8)による圧縮の結果がどのようになるかを感音性難聴の実際の特徴を例として(2.9)の2つの条件についてFig.2-8 に示した。Fig.2-8(a)は、難聴者のHTL, UCL と原音声のホルマントピークの相対的關係を示し、1000[Hz]付近の可聴領域が 20[dB]程度であるのに対し音声のダイナミックレンジは 70[dB]に広がっている。それに対して、Fig.2-8(b)は $c_1=c_2=1.0$, $d_1=d_2=40.0$ として圧縮した場合であり、音声のホルマントピークレベルは約 30[dB]のレンジに縮小され、 F_1 , F_2 ともにほぼ良好な振幅圧縮特性を与えている。しかし、音声の実効値を用いて増幅率を定める方式では、ホルマントピーク値のダイナミックレンジを 30[dB]より狭くすることは困難で、可聴領域の極端に狭い聴力特性の圧縮補償には限界が存在する。



(a) F1 component



(b) F2 component

Fig. 2-6 Relationship between formant peak levels and r.m.s. value of each speech frame.

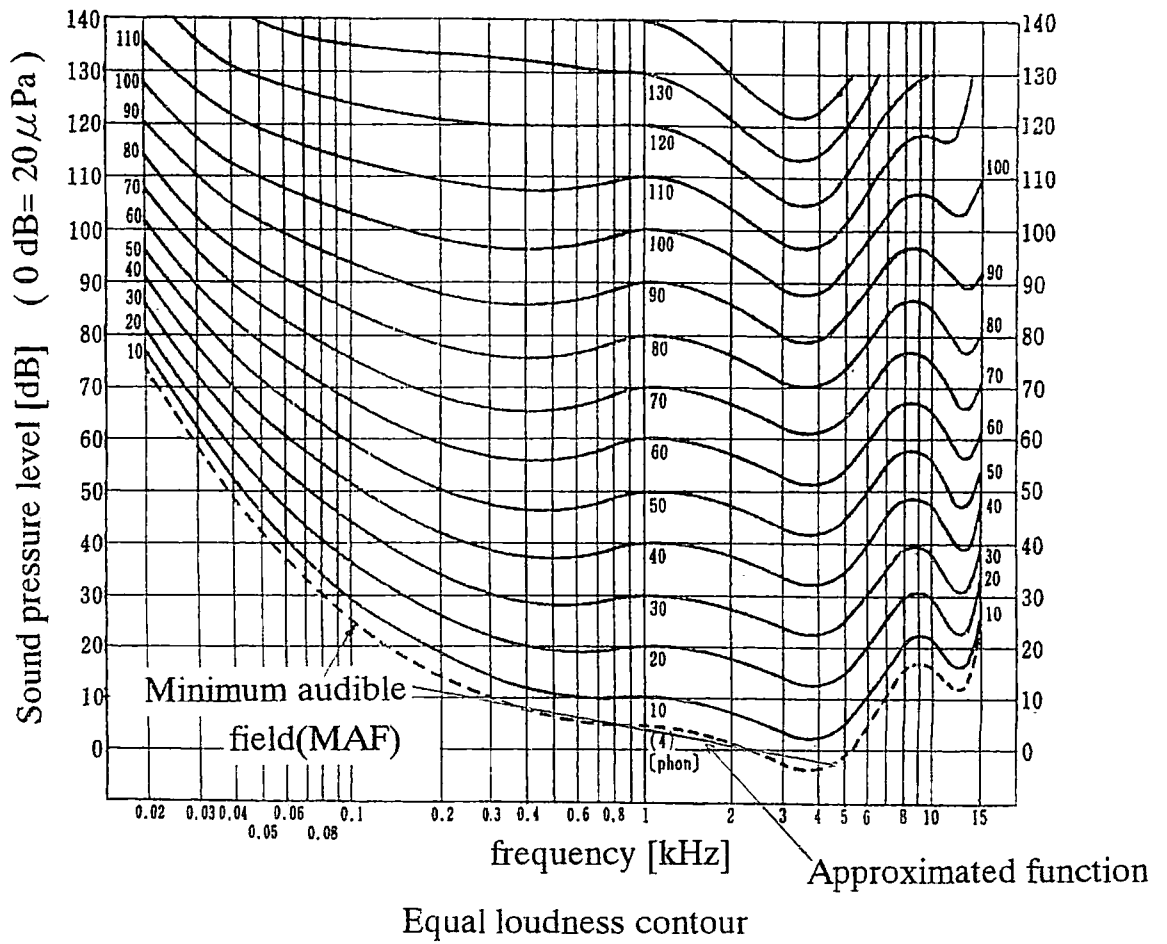
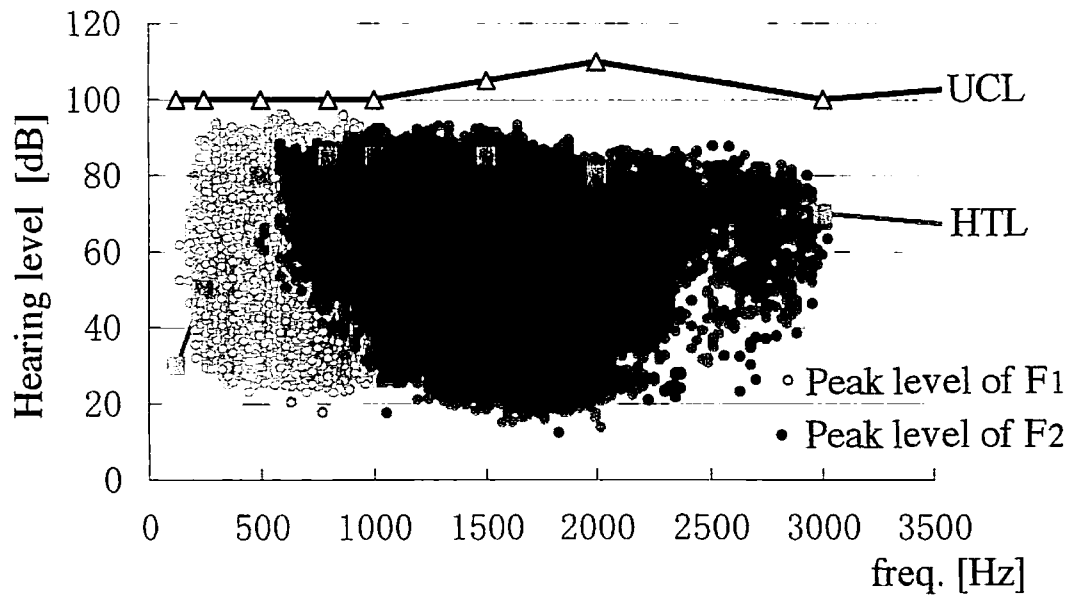
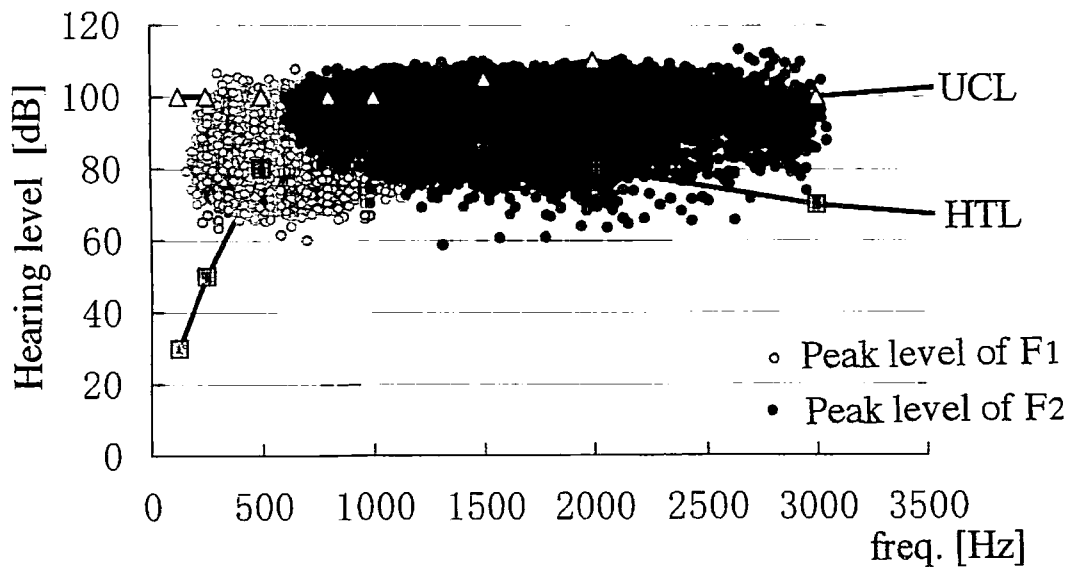


Fig. 2-7 Approximation of SPL to HL conversion(300-3000[Hz]). [24]



(a) input speech



(b) compressed speech

Fig. 2-8 Distributions of formant peak levels before and after the amplitude compression.

2-3. 本補聴方式における振幅圧縮の効果

前節までに述べた振幅圧縮処理が実音声に対してどのように行われているかを、いくつかの具体例を通して調べる。仮定された聴力の周波数特性は、聴力レベルで

$$H_d(f) = f / 135 + (40 - 300 / 135) \quad [\text{dB}] \quad (2.12)$$

$$U_d(f) = 90 \quad [\text{dB}]$$

即ち、300[Hz]で 50[dB]、3000[Hz]で 30[dB]の可聴領域があり、その間で HTL が周波数に対して直線的に変化する高音漸傾型のオーディオグラム特性と仮定している。また、(2.9)に示した単共振分解された信号の実効値とホルマントピーク値の関係を定める係数は、 $c_1=c_2=1.0$ 、 $d_1=d_2=40.0$ としている。

2-3-1. 母音部の振幅圧縮例

Fig.2-9 は、成人男性話者による、文節 / shiroienogu / (白い絵の具) の音声を処理した例を示している。(a)は抽出されたホルマント軌跡、(b)は(a)の2箇所(①, ②)から選択されたフレームのスペクトルが圧縮の前後で変化する状態を示す。原音声に存在する2つのホルマントピークレベルの差は振幅圧縮後に減少して、 F_1 、 F_2 が可聴領域の中央部に入り、スペクトル微細構造も忠実に保持されて傾斜のみが変化していることが分かる。また(b)では F_2 の単共振成分中に残留していた F_3 成分も圧縮により可聴領域内に現れている。

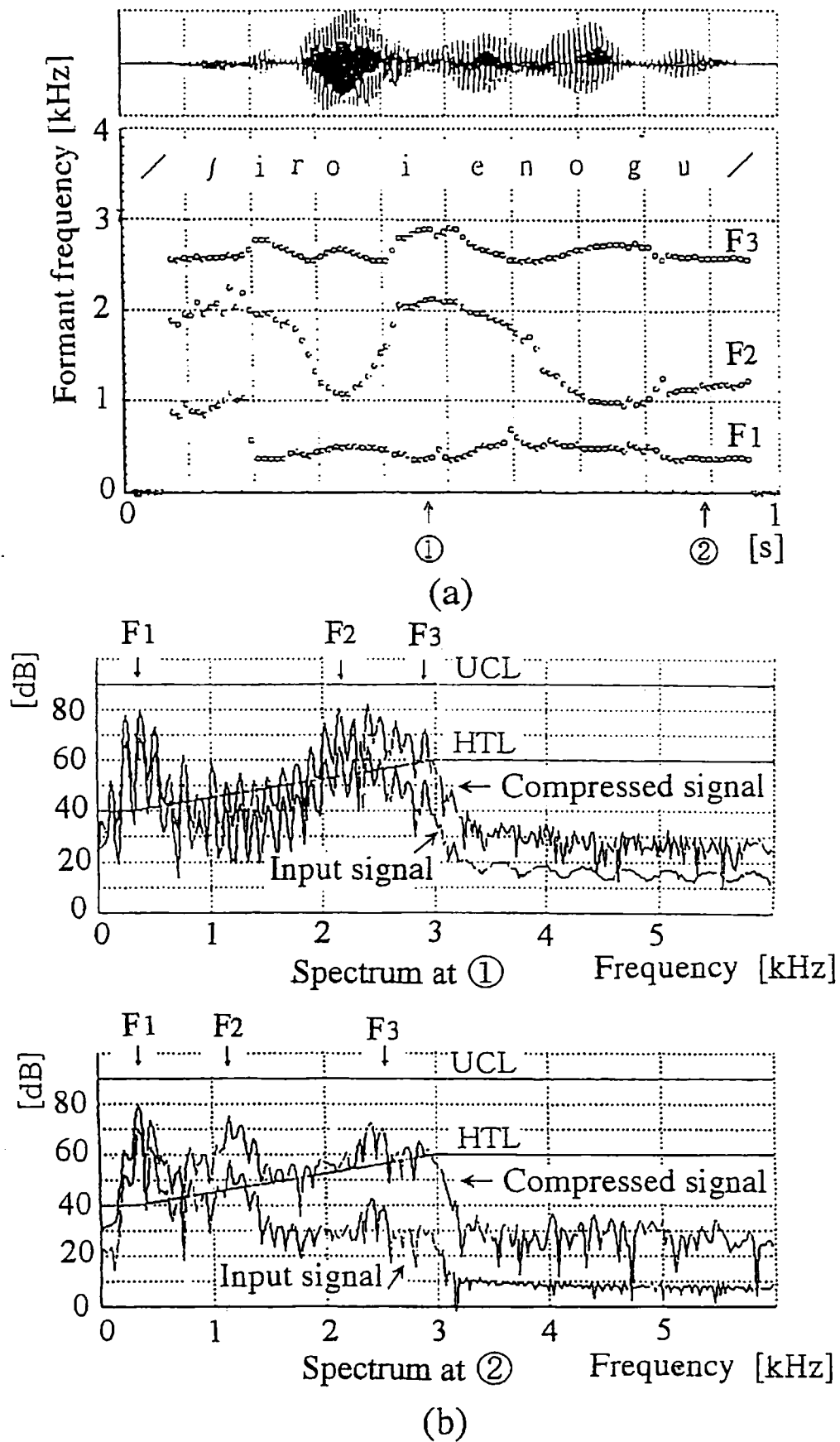


Fig. 2-9 Examples of extracted formants and amplitude compressed spectrum.

2-3-2. 子音部の圧縮性能

連続音声の振幅レベルの変化は大きく、ダイナミックレンジは 70[dB]程度に及ぶことは 2-2-3 で述べた。その中で、小振幅子音が聴取できるか否かが補聴器の良さに関わるので、ここでは、特に、小振幅子音として、破裂音、摩擦音の圧縮効果を調べる。音声試料は、英語を母国語とする成人男性話者 1 人による英単語 (27 語) と日本人の成人男性 3 名による /VCV/ であり、すべてを本提案方式で処理した。各単語の音声波形とサウンドスペクトログラムによって、有声破裂音 /b/, /d/, /g/, 無声破裂音 /p/, /t/, /k/(英語単語), 有声摩擦音 /z/, 無声摩擦音 /s /, /sh/ (/VCV/)のセグメントを指定し、その中のフレームを FFT してホルマントピークレベルを求めた。振幅圧縮前後のホルマントピークレベルが可聴領域のどの部分に位置するかを規格化して示すために、次の指数を用いる。

$$P_k = 100 \times (Y_p(F_k) - H_{pd}(F_k)) / (U_{pd}(F_k) - H_{pd}(F_k)) \quad (2.12)$$

ここで、 F_k は第 k ホルマント周波数 ($k=1, 2$), $Y_p(F_k)$ は第 k ホルマント近傍でのホルマントピーク音圧レベル、 $H_{pd}(F_k)$, $U_{pd}(F_k)$ は、それぞれ第 k ホルマントの周波数に対応する難聴者の最小可聴音圧および不快と感ずる音圧を示す。また、(2.7) で表されたように、

$$U_{pd}(F_k) - H_{pd}(F_k) = B_d(F_k)$$

は相対レベルであるので、難聴者の UCL と HTL の dB 差として定義され、同様に、(3.3) の分子も聴力レベルとしての $Y(F_k)$ と $H_d(F_k)$ (難聴者の聴力閾値レベル, HTL) の差として算出できる。

$$P_k = 100 \times (Y(F_k) - H_d(F_k)) / B_d(F_k) \quad (2.13)$$

圧縮前の $Y(F_k)$ による P_k と (2.8) を用いて圧縮した後の $Y(F_k)$ から P_k を求めて比較する。そのとき、(2.13) より、指数 P_k が周波数に関わらず、0-100 で可聴領域を示

し、 $P_k < 0$ は聴取できないこと、 $P_k > 100$ は過増幅を示している。圧縮処理の前後で P_k の分布が負の領域から正の領域へ移動すれば聴取が可能になったことを意味する。 $H_{pd}(F_k), U_{pd}(F_k)$ を先の Fig.2-8 に示した感音性難聴者の特性例を近似したのと同じに定めて圧縮処理した圧縮音声の分析を行う。Fig.2-10 , Fig.2-11 は、可聴領域のレベルを [dB] で 10 等分してその中に入る頻度分布を示したものである。図より、ここに示されたすべての場合について、圧縮処理により、適切なレベルに変換されていることが分かる。

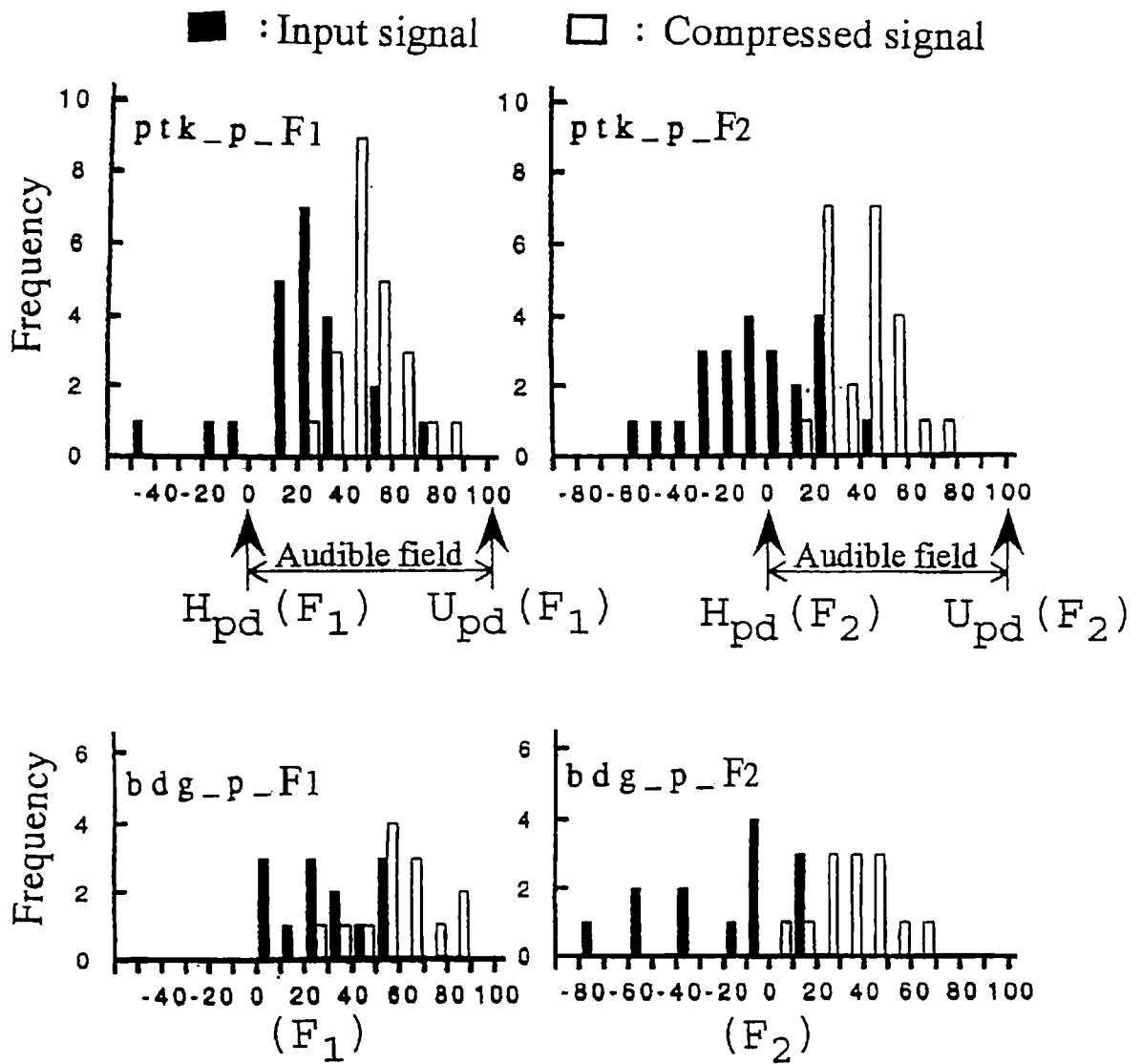


Fig. 2-10 Changes of formant peak levels caused by the amplitude compression in consonants. (/p/, /t/, /k/, /b/, /d/, /g/)

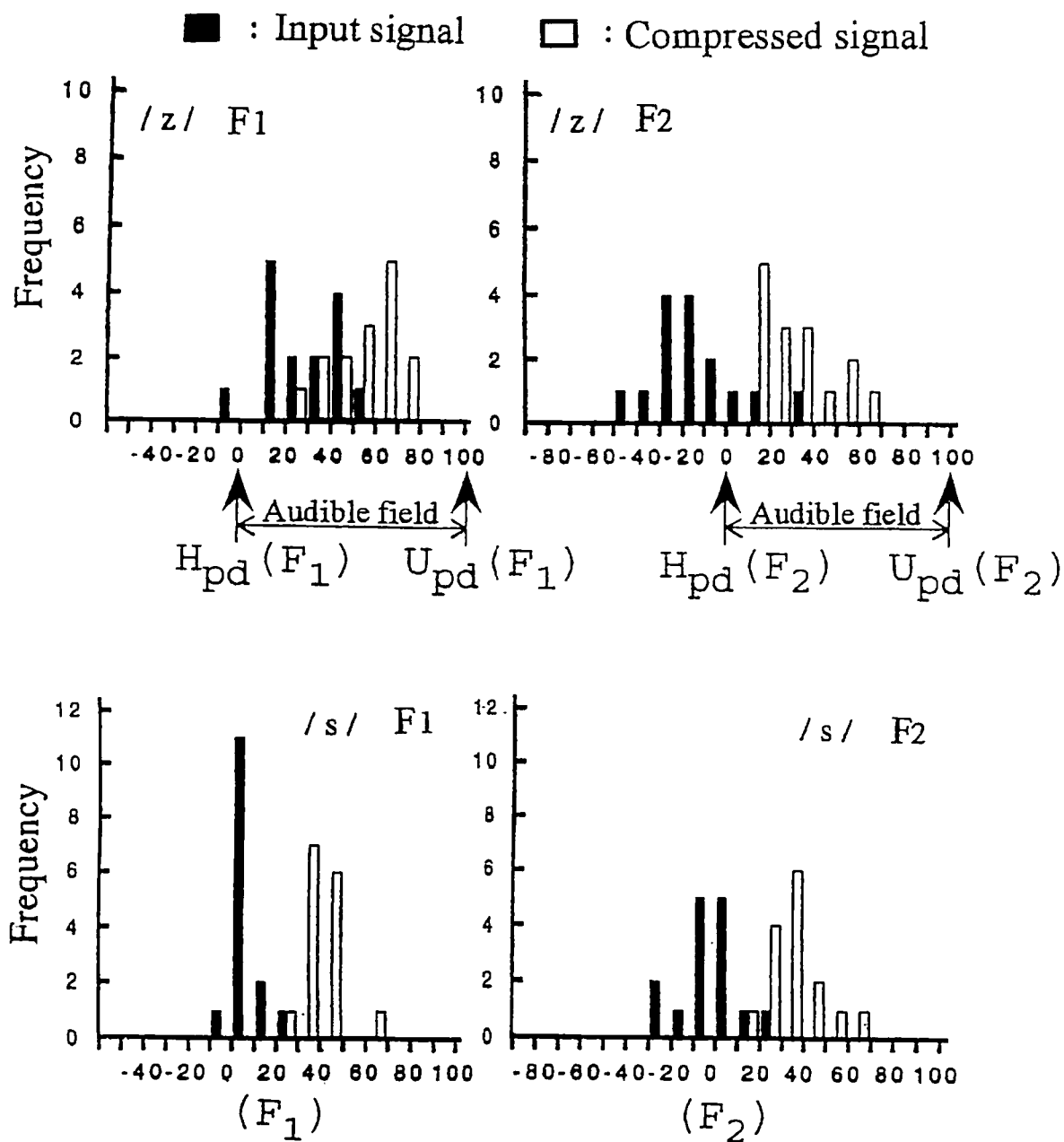


Fig. 2-11 Changes of formant peak levels caused by the amplitude compression in consonants.(/z/, /s/)

2-3-3. 難聴状態を模擬した健聴者による予備的な圧縮音声認知実験

ここでは、健聴者の模擬難聴状態による聴取試験で、提案した補聴器における振幅圧縮処理の効果を確認する。

2-3-3.1 模擬難聴状態の生成法

模擬難聴状態は HTL に相当するレベルと周波数特性を持ったマスクノイズを常時間かせることで生成する。聴取試験に際しては音声にそのマスクノイズを加算するために、次の方法を用いる。即ち、まず、パソコンの中で、適当な相対基準の下での HTL と UCL が仮定され、音声の圧縮が行われる。また、同基準において、HTL の 1[kHz] に相当するレベル（実効値）の純音が作られる。一方、HTL の関数を逆フーリエ変換して FIR フィルタのインパルス応答とし、適当なレベルのホワイトノイズと畳み込んで有色雑音とし、D/A 変換したものを DAT に録音する。パソコンから作成された 1[kHz] の純音を D/A 出力したときのレベルを FFT アナライザでモニターし、それとノイズの 1[kHz] の成分が同じレベルに達するように調整すれば、HTL の特性、 $H_0(f)$ を持つマスクノイズを得る。このマスクノイズとパソコンから D/A 出力した圧縮音声をミックスすることにより模擬難聴状態の音声を作成する。

2-3-3.2 聴力特性と入力音声レベルの変化に対応する圧縮音声の認知実験

振幅圧縮効果を調べるために与えた模擬難聴状態は Fig.2-12(a) に示す 2 種類である。図の①は信号の分析で圧縮効果を調べたときの(3.1)と同じ特性、②は①よりもすべての周波数で 20[dB] だけ HTL を上昇させて聴力ダイナミックレンジを狭くした特性である。更に、②の条件で入力信号のみを 20[dB] 小さくした状態を条件③とする。この 3 条件のそれぞれに対し、さらに、次の 3 種の状態を与える。

(A) マスクノイズを用いなくて、健聴耳による補聴器の圧縮信号歪の状態を調べる場合。

(B) 補聴器を用いなくて原音声とマスクノイズの加算音声を聴取する場合。

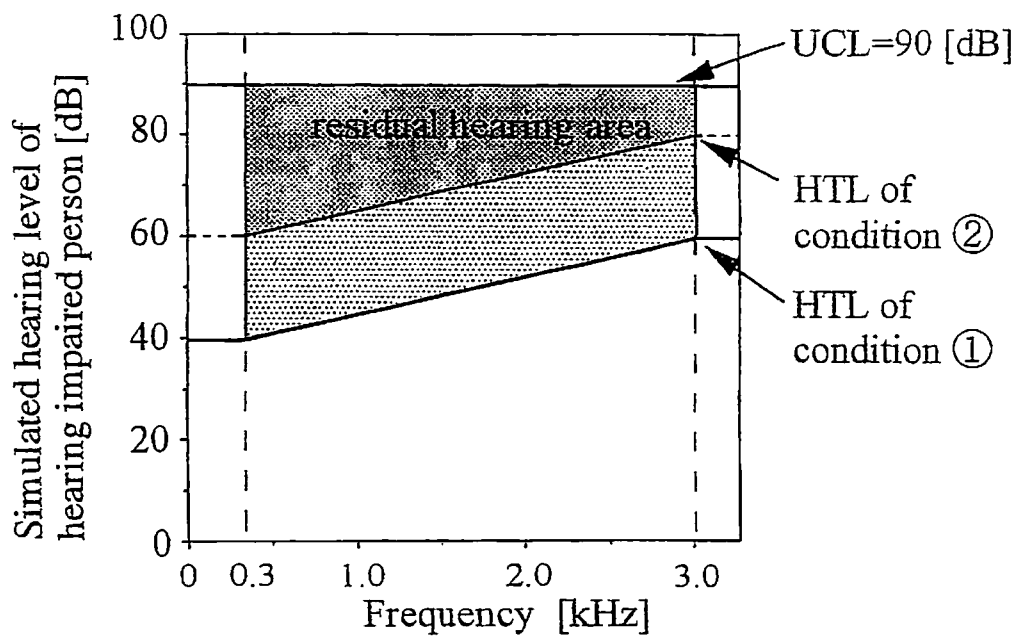
(C) マスクノイズで難聴状態を模擬し、補聴器で圧縮した音声を聴取する場合。

このように、合計9種の条件を設定して聴取実験をおこなった。

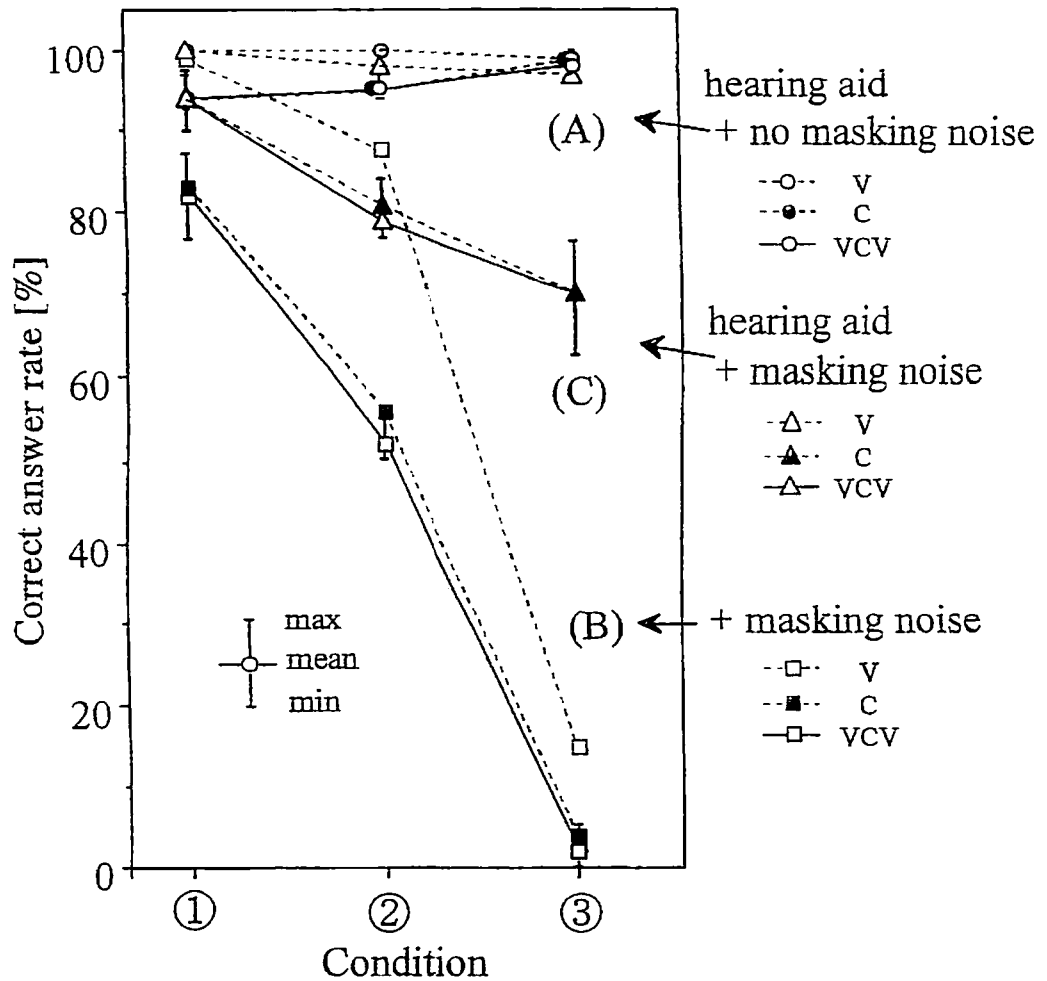
音声試料は成人男性話者1人による62 /VCV/音節、被験者は健聴者(大学生)3人である。ヘッドホンを通して片耳に提示された。実験に先立って、被験者が最終段の増幅器の利得を調整し、音のおおきさを快適レベル(MCL)にセットするよう指示した。

結果は、音節明瞭度(被験者3人の平均値、最大値及び最小値)として、Fig.2-12(b)にまとめられている。図に基づいて次の推論が得られた。

- (1) (A)の結果より、圧縮条件が厳しくなっても圧縮歪は少なく、良好な音声を提示できる。
- (2) (B)の結果より、補聴器を用いなければ、②の聴力特性での聴き取りは不十分であり、音声レベルの減少③では殆ど聞き取れない状態になる。
- (3) (C)の結果により、本提案の補聴器を適用することで、②の狭い可聴領域の特性や③の入力音声レベルの減少に対してもほぼ十分に聴取可能な音声を与えることができる。



condition ③ : -20[dB] input under the condition ②
 (a) Conditions of the experiments



(b) Results of hearing experiments

Fig. 2-12 Effects of amplitude compression in SRA system.
 ---Listening test using /VCV/ syllables---

2-3-3.3 合成音声の母音カテゴリー判断に対する振幅圧縮の効果

前節の/VCV/による実験では、母音、子音ともに圧縮による改善が示された。ここでは F_1 - F_2 図上に一様に分布する合成音声を用いて、模擬難聴状態が母音カテゴリー判断[25]にどのように影響し、提案する単共振分解型振幅圧縮処理によってどの程度回復できるかを探った。母音カテゴリーの情報は通常使われる異聴表と密接な関連を持っておりかつ聴力特性の影響を異聴表より詳細に分析することができる。

(1) 合成母音の作成

ここでは、 F_1 - F_2 図上での母音カテゴリーを調べるのに通常用いられる方法を応用することとした。合成母音は次のような条件で、ホルマント合成方式によって作成した。信号は、ピッチ、振幅とも、400 [ms] の定常部の前後に直線的に変化する各 50 [ms] の立ち上がり、立下り部を設け、時間長 500 [ms] とした。 F_1 , F_2 は、 F_1 - F_2 図上で格子状に配置した合計 60 試料の音声として作成された。また、 F_3 は、あらかじめ求められている F_1 , F_2 を変数とする 2 次の回帰式[22]から定めた。上記以外のパラメータはすべて固定とし、次のように与えられている。すなわち、 $F_0=130$ [Hz], $F_4=3500$ [Hz], $F_5=4500$ [Hz], 帯域幅は、 $B_1=50$ [Hz], $B_2=80$ [Hz], $B_3=115$ [Hz], $B_4=150$ [Hz], $B_5=200$ [Hz] とし、高次極補正は、6 [kHz] (標本化周波数 12 [kHz]) の折り返しのみを利用する。音源として、ピッチ以外のパラメータを一定に保った Rosenberg 波を用いた。

(2) 実験用試料の作成

上述の合成母音の各試料を原音声 (A) とすると、それに 2 種類の加工を施した音

声試料を作成する。すなわち、ひとつは、模擬難聴状態とするためにマスクノイズを加算したもの(B)、もうひとつは、圧縮処理をした上でマスクノイズを加えたもの(C)である。難聴特性と圧縮条件を決める HTL は、2-3-3.2 と同じ、すなわち、(3.1)とし、UCL は、2-3-3.2 の場合より 10 [dB] 低くして 80 [dB] とした。したがって、可聴領域は、300 [Hz] で 40 [dB]、3000 [Hz] で 20 [dB] となっている。このようにして得られた (A) , (B) , (C) の 3 種の試料、それぞれについてランダムに並べ直した実験用音声試料を 5 種類作成した。

(3) 実験方法

被験者は成人男性（健聴者、大学生）1人で、(A) , (B) , (C)の順にそれぞれ5種の音声試料を聞き、5 母音のいずれかを強制選択するよう求められた。音声は片耳にヘッドホンで提示し、被験者は判定実験に先立ち、5 種ある実験試料のうち、直接関係しない別の試料の一部を用いて、出力レベルを快適レベル (MCL) に調整した。

(4) 実験結果

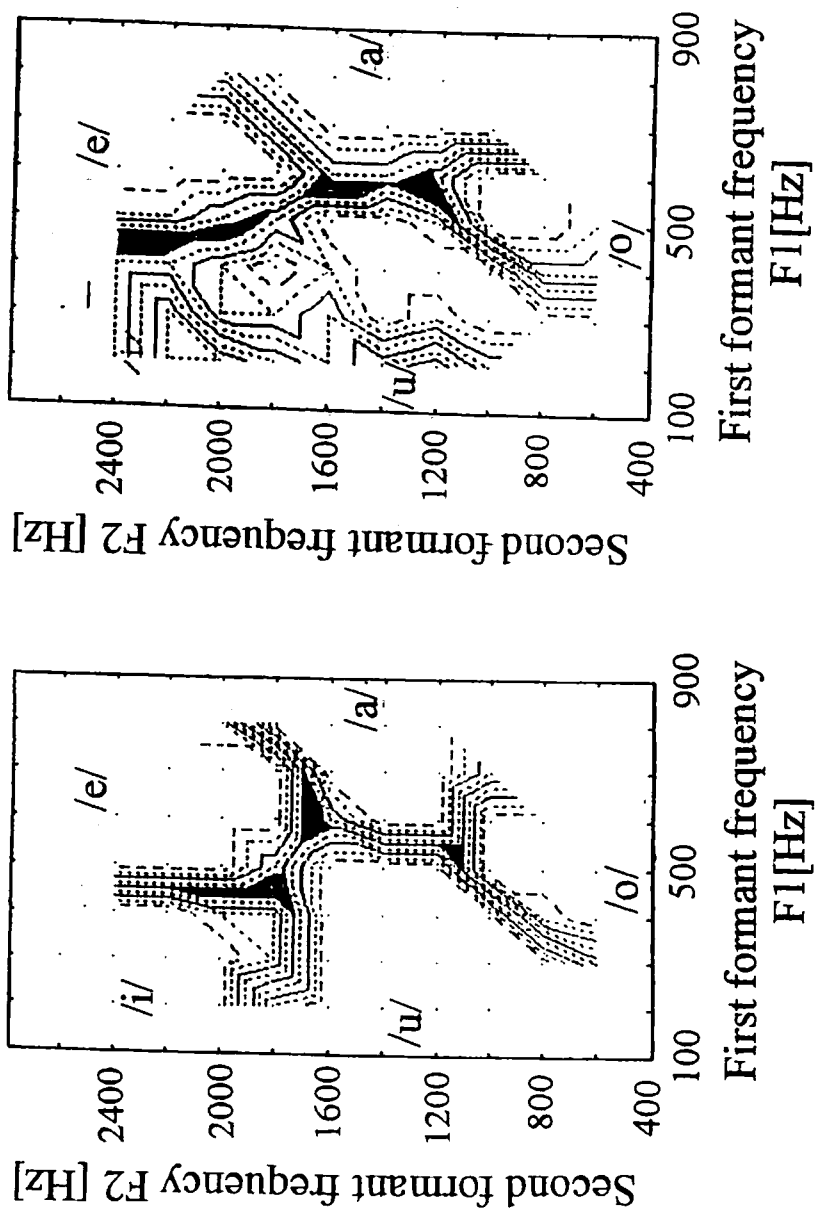
Fig.2-13 は、判定結果を示している。図中、カテゴリー境界を示す線は、 F_1 - F_2 図の各点の母音同定率をもとに点間の同定率を内挿、算出して得た等高線である。一番内側の等高線が 2 母音を 50[%]に分ける母音境界として示され、以下、順次 10[%]刻みで 90 [%] まで描かれている。(線のない部分は 100 [%] その母音と判定されたことを示す。)

原音声 (A) の判定結果を基準として、他の条件の判定は次のようにまとめられる。

(i) 圧縮処理 (補聴器) を使用しない模擬難聴状態 (B) では、/ i / と / u / の

カテゴリーが乱れていて、 F_1 が低い場合の F_2 の判別が困難であることを示している。これは、 F_1 の低下が高い F_2 の振幅を相対的に減少させ、高域のマスク量増加と相俟って判定が困難となったと解釈できる。(A)と(B)の判定結果が一致する各母音カテゴリーの類似度[25]は77 [%]にとどまる。

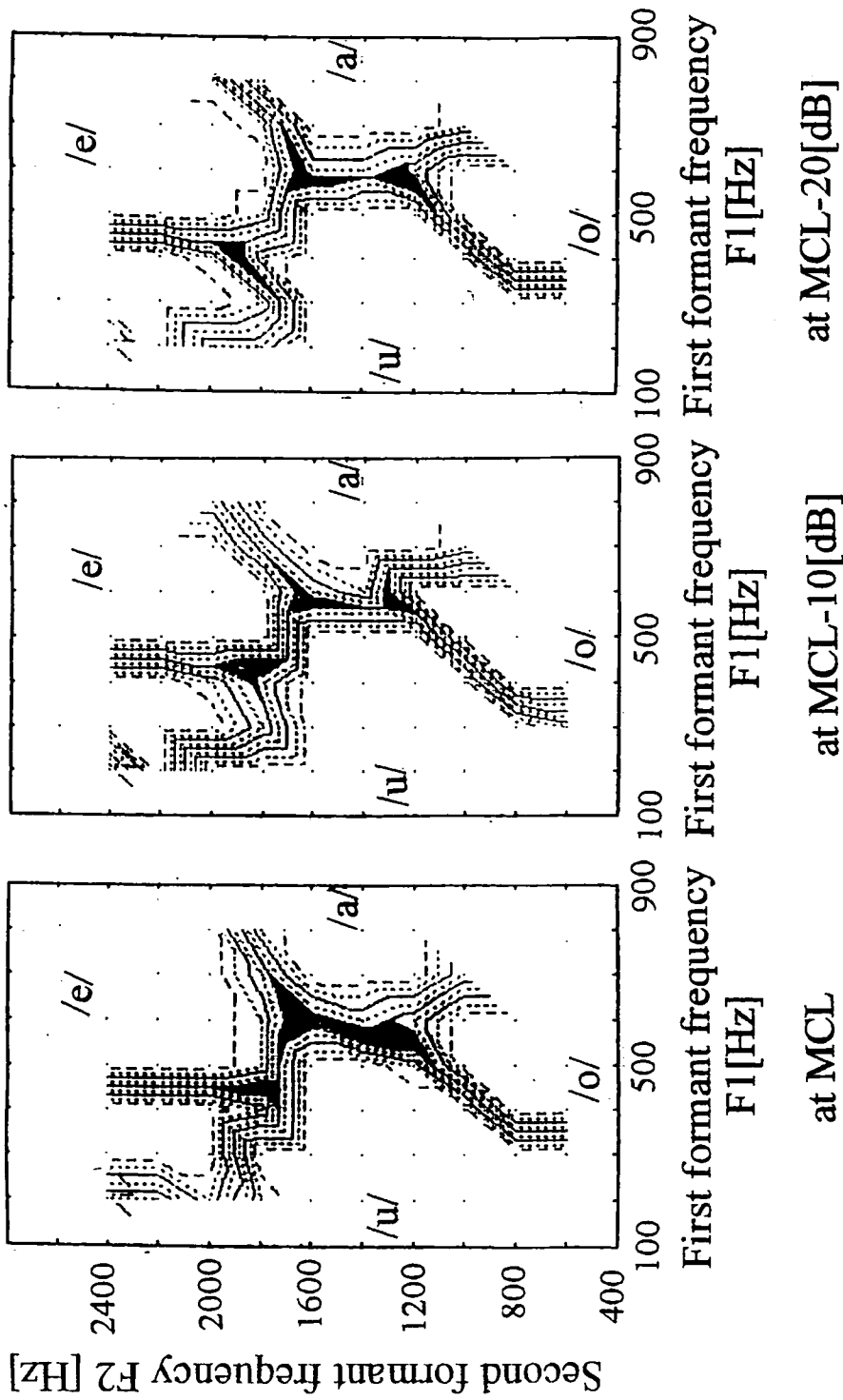
- (ii) (C)に示されるように模擬難聴状態で圧縮処理(補聴器)を用いると、母音カテゴリーが原音声(A)に近くなり、母音カテゴリーの類似度[25]は、92 [%]に達する。このことより、明らかに、本提案方式の振幅圧縮が有効であるといえる。



(A) Speech without masking noise. (B) Addition of masking noise.

Fig. 2-13 Effects of amplitude compression in SRA system.

---Change of vowel categories (A), (B) ---



(C) Restoration by the amplitude compression.

Fig. 2-14 Effects of amplitude compression in SRA system.
 ---Change of vowel categories (C)

2-4.まとめ

本章では、感音性難聴者のため高品質音声を伝達する補聴方式として、単共振分解型振幅圧縮補聴方式を提案し、補聴器としての基本的性能について検証した。本方式は、逆フィルタ制御法により抽出したホルマント周波数に基づいて分離抽出した単共振スペクトルのホルマントピークが可聴領域に入るように振幅圧縮することにより、通常用いられるスペクトル分解による振幅圧縮法に比べてスペクトルの谷に相当する部分に歪が生じにくい。また、単共振分解された信号の実効値、ホルマント周波数での圧縮上限・下限から圧縮係数を算定するので、振幅の瞬時圧縮方式で生ずる非線形変換による高調波歪を避けることが可能になる。具体的に電話帯域音声を対象とする補聴器をシミュレーションで実現し、信号処理としての有効性と模擬難聴状態にした健聴者の音声認知への効果を調べた。

まず、感音性難聴者のオーディオグラムを模擬した振幅圧縮条件下で実音声を圧縮し、スペクトル分析によって期待した圧縮効果が得られるかを検証した。その結果、母音部、子音部において圧縮前に可聴領域に入っていないホルマントピークが、圧縮後は可聴領域内へ入っていることを確認した。次に実音声の VCV 試料を同様の圧縮条件で圧縮し、マスクノイズにより模擬難聴状態にした健聴者による聴取実験を行ない圧縮処理によって、母音部、子音部ともに聴取実験で音声の認知が改善されることを示した。またマスクノイズを重畳しないで圧縮条件を変化させて聴取しても音声の認知に低下傾向はなく圧縮品質も良好であると推定できた。更に合成音声による母音カテゴリの判定実験で、模擬難聴下で圧縮しない音声を聴取する場合に比べ、模擬難聴状態で圧縮音声を聴取する場合の音声の聴取正答率から判定された母音カテゴリは、圧縮により大幅に改善した。これらの結果から本章で提案した単共振分解型圧縮方式の有効性を示した。

第3章 音声画像で強化する単共振分解型補聴器による聴能訓練とその効果

感音性難聴者に適合した補聴方式として第2章で提案した、単共振分解型補聴器では、ホルマントピークレベル(単共振信号のスペクトルピークレベル)が可聴領域に入るように音声信号を圧縮する[19]。更に明確な補聴器の設計思想を得るためには、感音性難聴者の聴覚に与える音声情報と認知能力との間の定量的関係を明らかにする必要がある。その場合、感音性難聴者の認知能力は、聴能訓練によって改善される可能性があるので、十分に改善された上限値を推定した後にそれらの関係を明らかにしなければならない。本章ではまず音声画像を併用して、提案補聴器による聴能訓練を効率化する方法とその効果の分析について述べる[20]。

3-1. 音声パラメータの画像表現とその特徴

ここで使用する音声画像は何ら認識手法を用いずに音声信号から視覚情報への変換を行うもので、音声から抽出したホルマント周波数、ピッチ周波数、ニューラルネットの処理出力などで音声の持つ音響的な諸情報をカラー画像と白黒パターンの融合として表現する。Fig.3-1に音声画像の生成のブロック図を示す。音声信号から抽出した各特徴パラメータを入力として色彩とパターンが決定される。RGB(3原色)信号は低次の3個のホルマント周波数の巡回比によって決定され、各信号の係数は中性母音で無色化されかつ日本語の5母音の分離が明瞭になるよう考慮されている[26]。入力パラメータの応答としてニューラルネットワークより出力される各子音性の強度により輝度変調された子音ごとの特徴パターンと色彩とが融合、表示される[20]。出力要素は調音様式、調音位置の出力の積をとって16カテゴリー(母音、/m/, /n/, /ng/, /N/, buzz, /b/, /d/, /g/, /p/, /t/, /k/, 有

声摩擦音,無声摩擦音,弾音,無音)出力が設定されており,ニューラルネットの学習が充分になされれば健聴者が聴覚を通じて音声を聴き取り,理解するのと同じように被験者が表示画像から音素系列を読み取ることが可能になると期待できる。音声の時間情報は,画面の上下方向に,ピッチ周波数の対数は画面の水平方向に対応し色対比効果による調音結合の補償がなされて直観的に音素を読み取りやすい表現として提示される[26]。

日本語の5母音を画像化した例を Fig.3-2 に示す。図において各母音が明瞭に色彩により直観的に理解できるように表現されており,男声(左列)女声(右列)におけるピッチ周波数の相違が水平方向の色彩の長さ(対数表示)の違いになり,各母音の時間長は上下方向の色彩の幅に変換されている。色彩画像上に重畳された子音のパターンはニューラルネットワーク出力による輝度制御を受けるので,音声に含まれる各音素情報の内優勢なものが白い明瞭なパターンとして画像上に重ねられるから,被験者は総合的に音声画像を読み取り言葉を理解できる。これは,ちょうど,聴覚を用いて音声を聴き取ると同様に,視覚によって読み取ることで達成していると考えられ,まったく認識手法を用いずに,音声情報を音響信号から視覚情報に変換しているといえる。

音声画像での色彩生成と子音情報の重畳の原理を Fig.3-3 に示す。また単語の音声画像の例を Fig.3-4 に示す。RGB 信号で合成された有声音の色彩の上に子音パターンが重畳している様子が分かる。被験者はまず,5母音の色彩を理解した後に,子音パターンの情報を学習すると容易に,直観的理解ができるようになる。

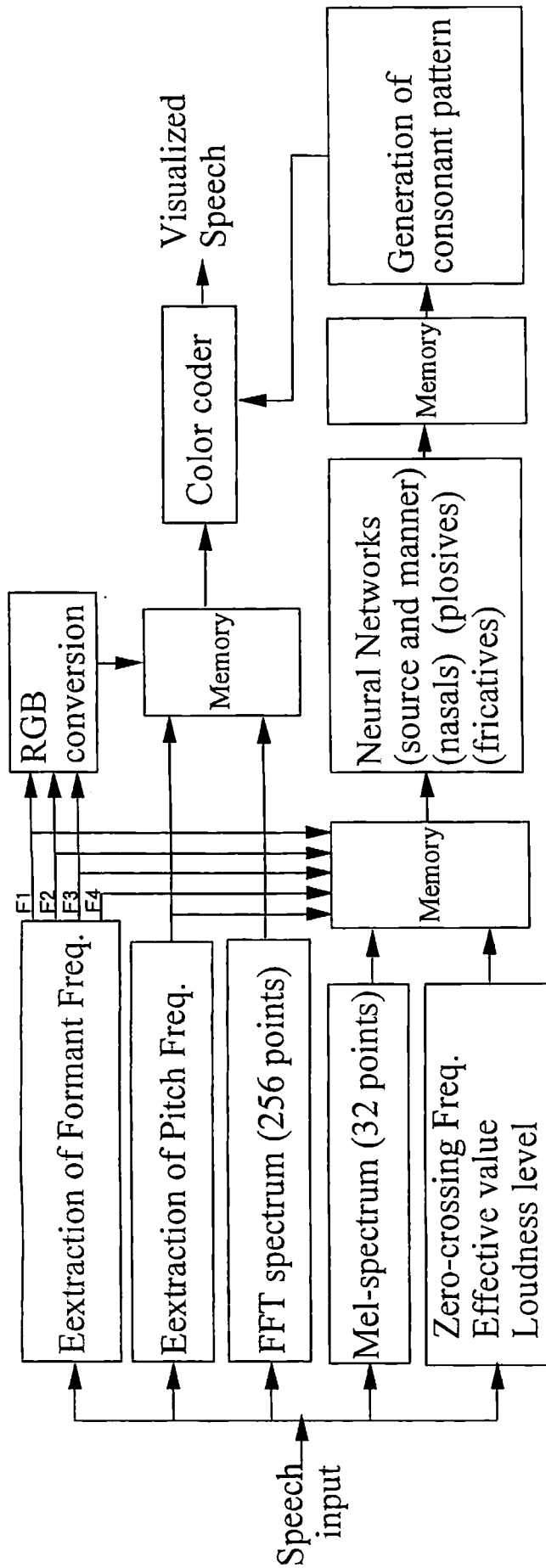
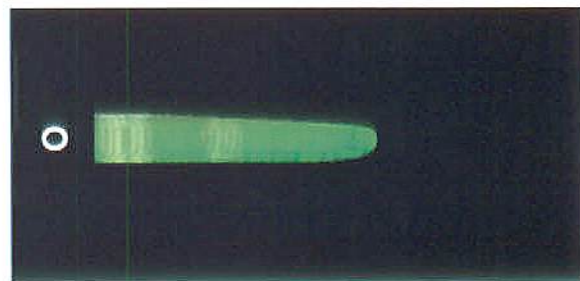
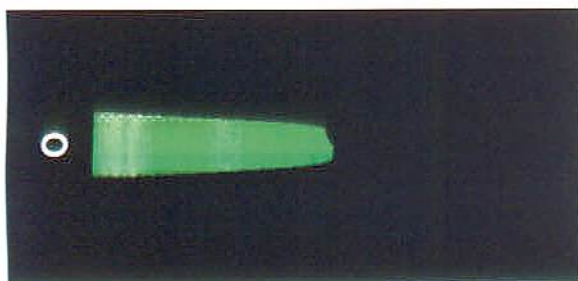
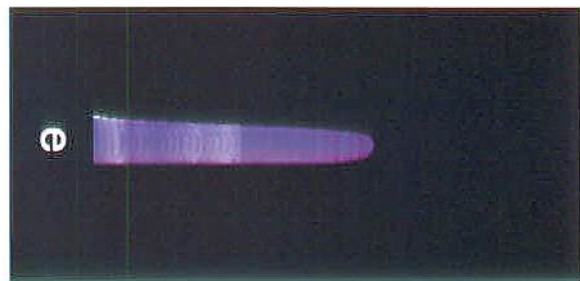
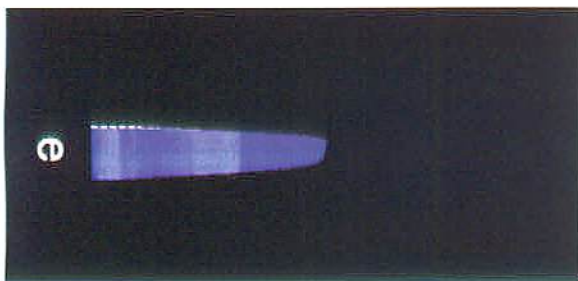
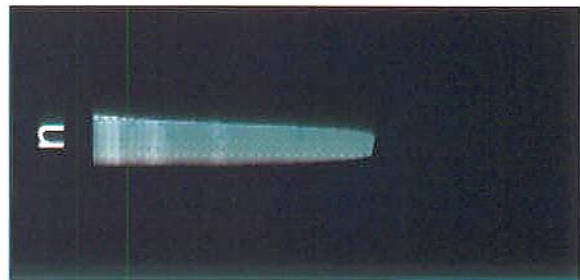
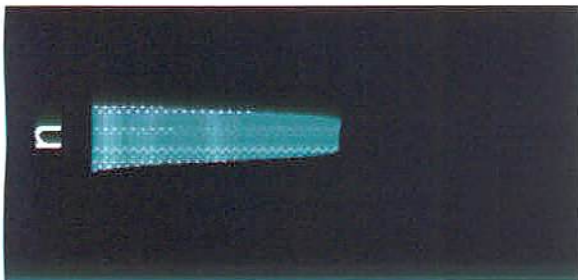
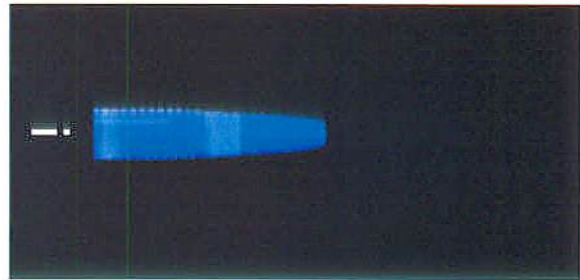
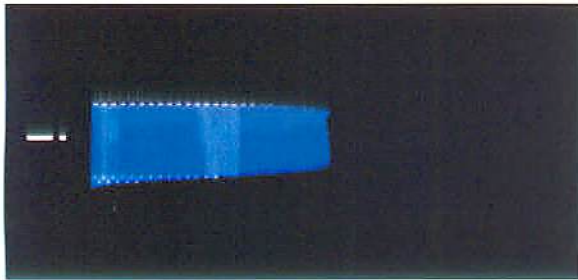
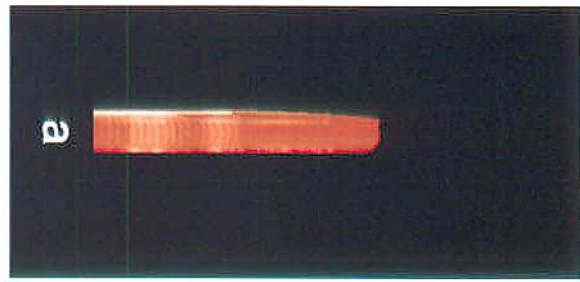
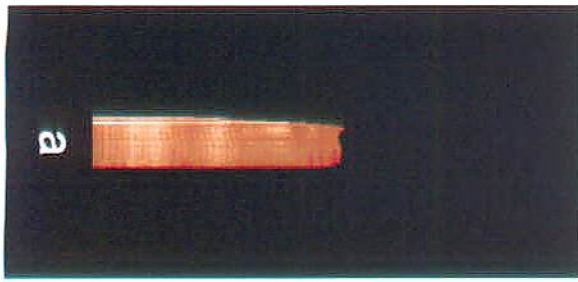


Fig. 3-1 Block diagram of the system to generate visualized speech.



Male

Female

Fig. 3-2 Display examples of visualized speech for vowels.

R, G, B Signals

$$E_R = \frac{5F_1}{F_3}$$

$$E_G = \frac{3F_3}{5F_2}$$

$$E_B = \frac{F_2}{3F_1}$$

where F_1, F_2, F_3 : Formant frequency

Consonantal Information
(no recognition but adding each pattern)

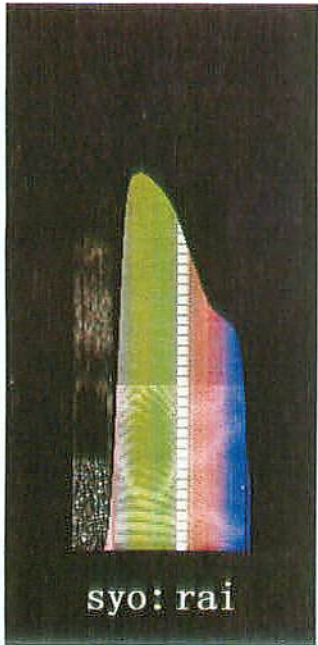
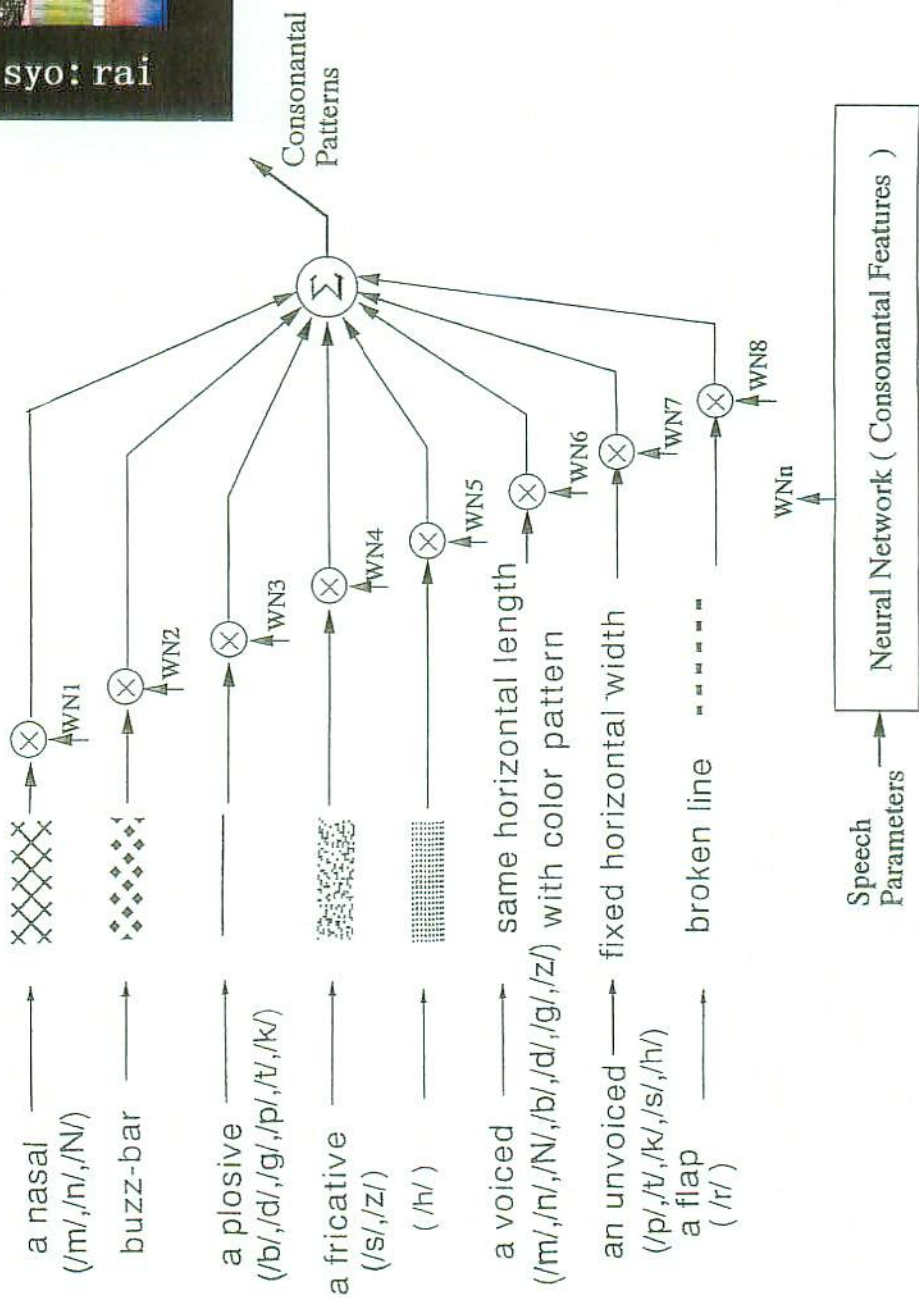


Fig. 3-3 A simplified model to explain the consonantal information.

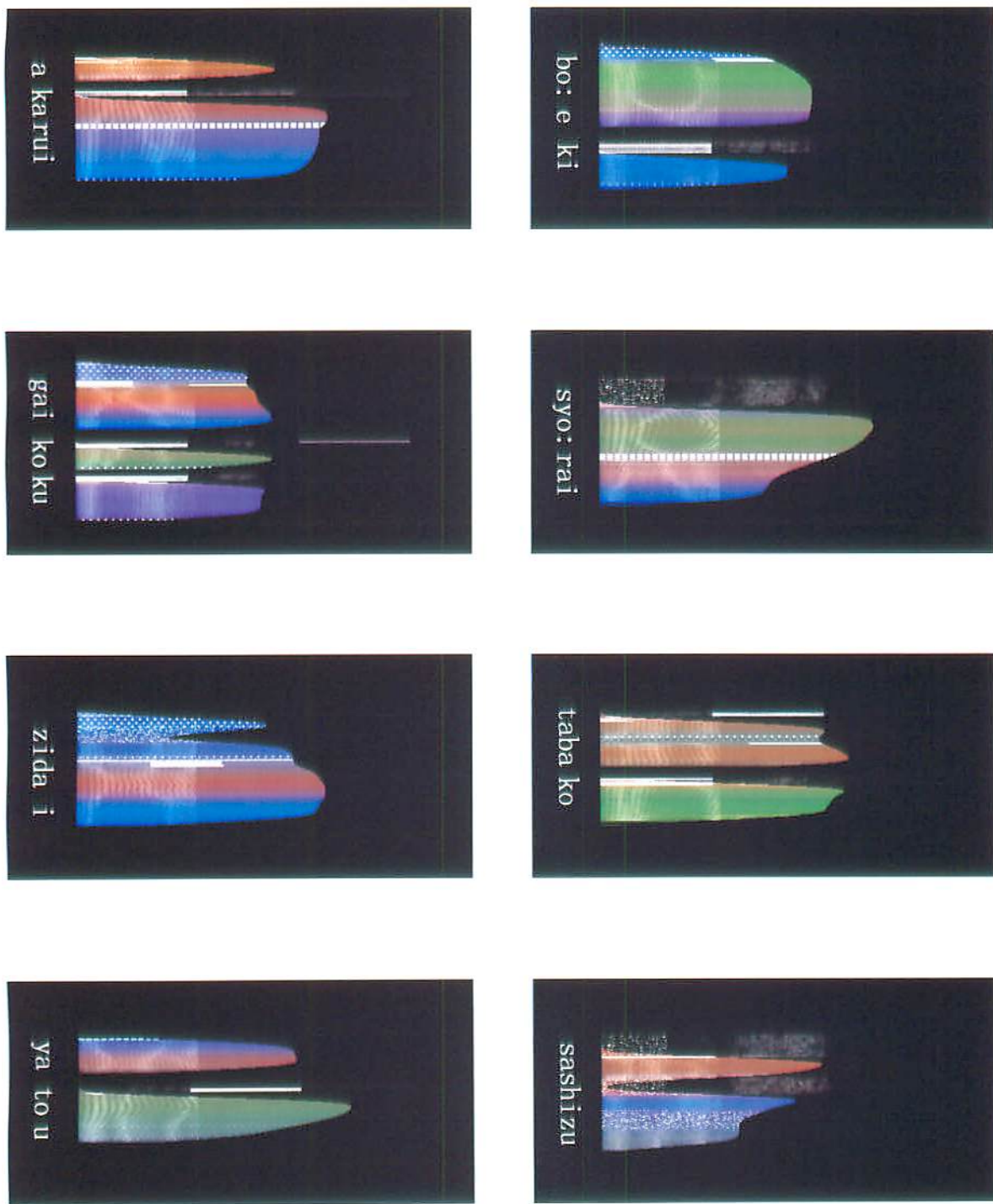


Fig. 3-4 Display examples of visualized speech for words.

3-2. 聴能訓練システムの構成と機能

システムの構成を Fig.3-5 に示す。パソコンとオーディオシステムを連結し、音声と音声画像のデータベース及びそれらの提示ソフトウェア、訓練後の聴取テスト試料を組み合わせたシステムである。音声信号から音声画像を生成するためには Fig.3-1 に示したように音声から第1～第4ホルマント周波数、ピッチ周波数、零交差周波数、実効値、ラウドネスレベルなどの諸パラメータの抽出、計算及びカラー画像の合成など、信号処理が複雑なので現時点ではリアルタイム処理はまだ難しい。そこで、音声から予め音声画像を生成しデータベース化しておく。提示する音声は被訓練者の聴力特性に基づき圧縮上限と下限を定め、データファイルとしてパソコン内に保持しておく。これらの音声画像、圧縮音声試料をテキストファイルに収められた提示順に従って被験者に提示する。本研究の実験では上述のデータファイルを再生提示するが、被訓練者が自己の補聴器を使用する場合や人工内耳などの場合にはスピーカから音声提示可能である。

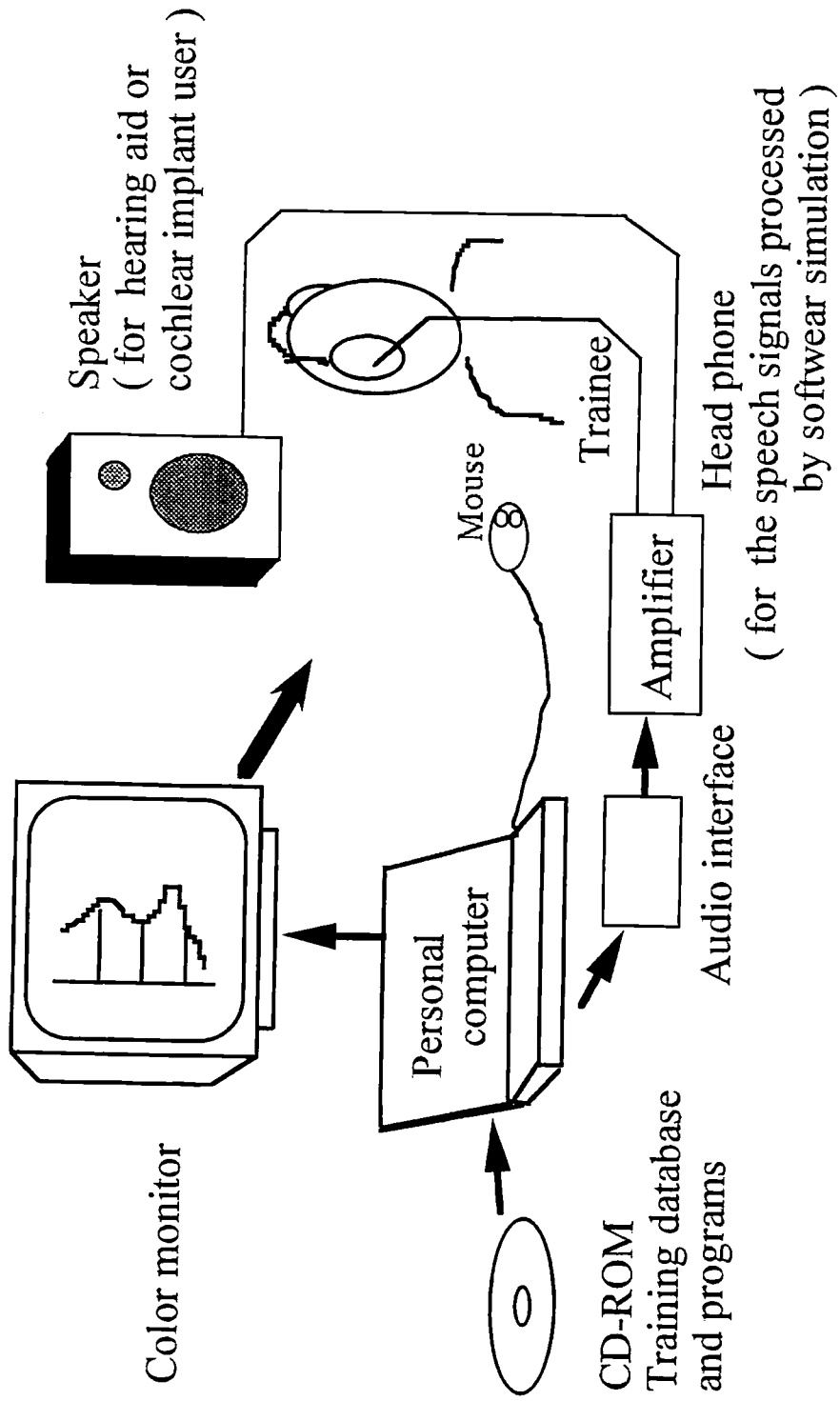


Fig. 3-5 The hearing training system for the hearing impaired with hearing aid and visualized speech.

3-3. 訓練方法

音声画像を併用する聴能訓練では、簡易テキストにより音声画像の基本的な読み取り方（母音、ピッチ、子音パターンの基本表現例など）を約2時間程度解説した後、音声画像を用いた訓練を行う。単共振分解圧縮処理により可聴領域内のホルマントピーク分布を高めた音声を聴取しながら、視覚からの情報によって、聴覚から聴き取る音素のかすかな手がかりを強化し音素認知の改善を試みる。

Fig.3-5 に示した聴能訓練システムで、視覚情報として音声画像と文字(または文字のみ)を表示しつつ、次のような手順で聴能訓練を行う。被験者は防音室内にヘッドホンをつけ着席し、あらかじめ用意された単語の振幅圧縮音声(被験者毎に定められた圧縮の上下限の特性を与えて単共振分解型補聴器のシミュレーションで生成されたもの)を聴きながら音量をMCL(最適レベル)に調節する。その後音声画像や文字を画面に、補聴器の出力音声をヘッドホンから被験耳(片耳)に提示する。一連の提示の流れはFig.3-6 のようになる。まず訓練用単語群から抽出した単語音声を1回提示し約1.5秒の間をおいて、音声画像(または、文字)と同時に更に1回、画像(または、文字)は静止表示させたまま約1.8秒の間をおいて最後に文字も提示しながらもう1回音声を提示する、計3回の音声提示を1組として1単語を終了し、50単語について繰り返し1つの学習セットとした。Fig.3-7に文字と同時提示した音声画像例を示す。被験者の認知能力を正しく判定するためにこれらの学習セットが終了する度に別話者(男女各20名の試験単語群(ATR重要単語データベース520語)からランダムに抽出した50個の単語により聴取テストを行った。単語群にはATR重要語(520語)データベースを用い、訓練を実施した[23]。訓練回数は被験者により異なり14-40回に及ぶが、いずれの被験者も聴取テストの最後の数回では正答率が飽和していることを確認して終了した。

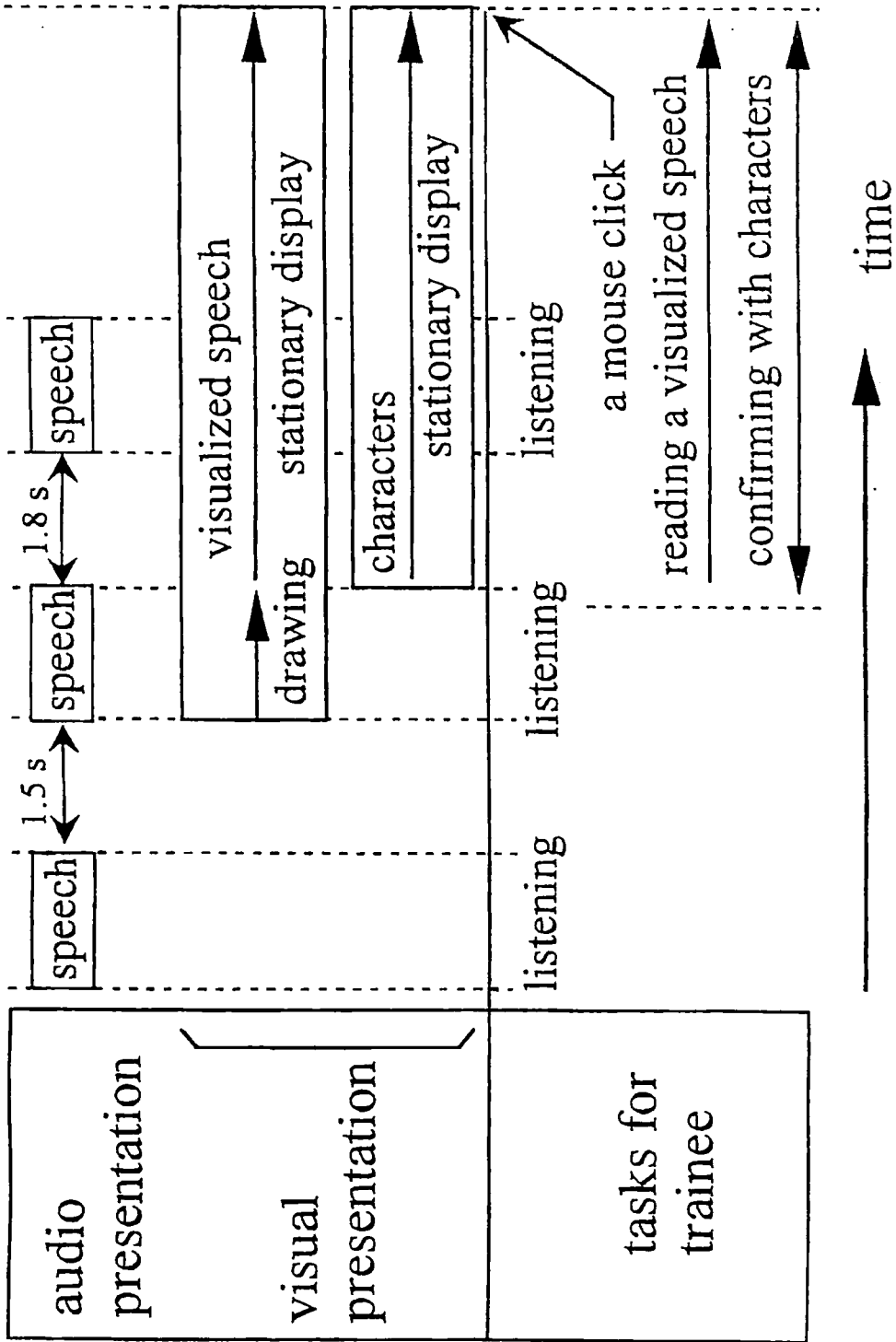
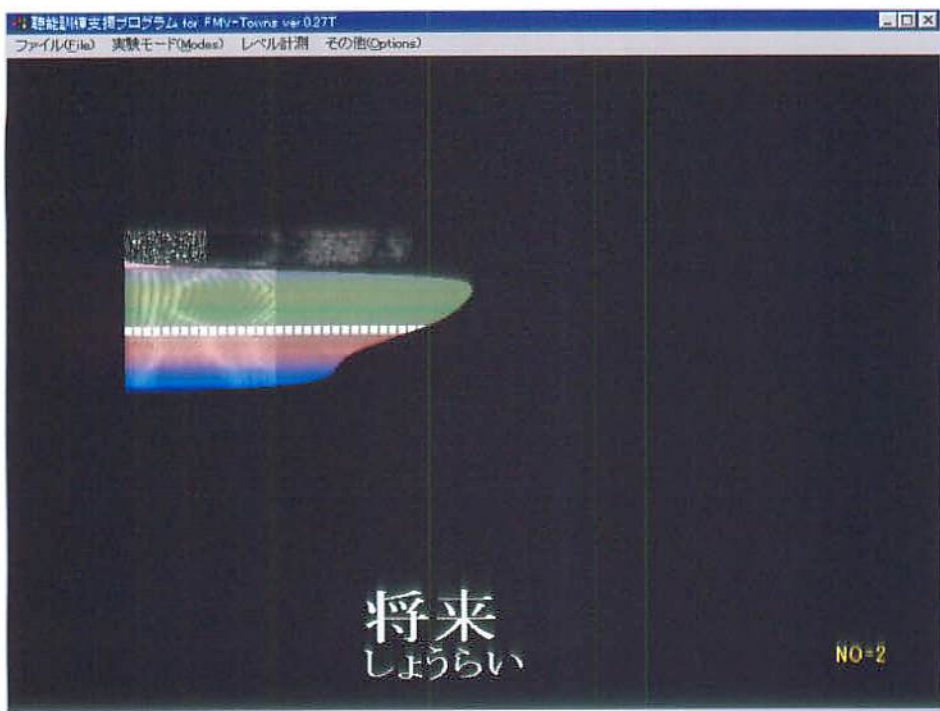
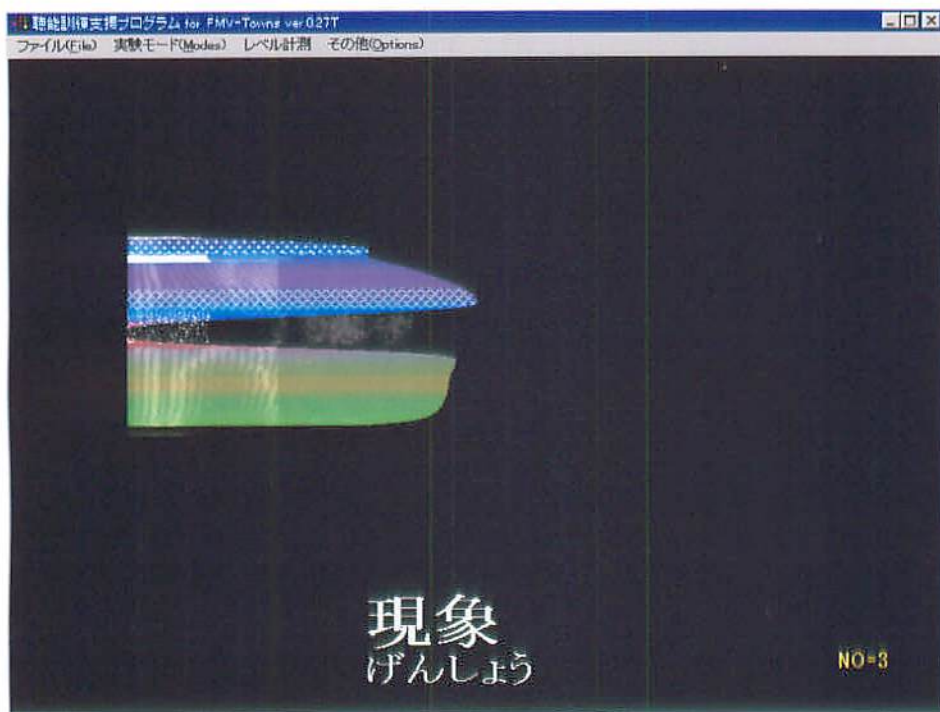


Fig. 3-6 The timing chart for presenting a word and a visualized speech in the training.



(1)



(2)

Fig. 3-7 The examples of the displayed image in the training.

3-4. 訓練実験に参加した感音性難聴者の聴力特性と補聴器の振幅圧縮特性

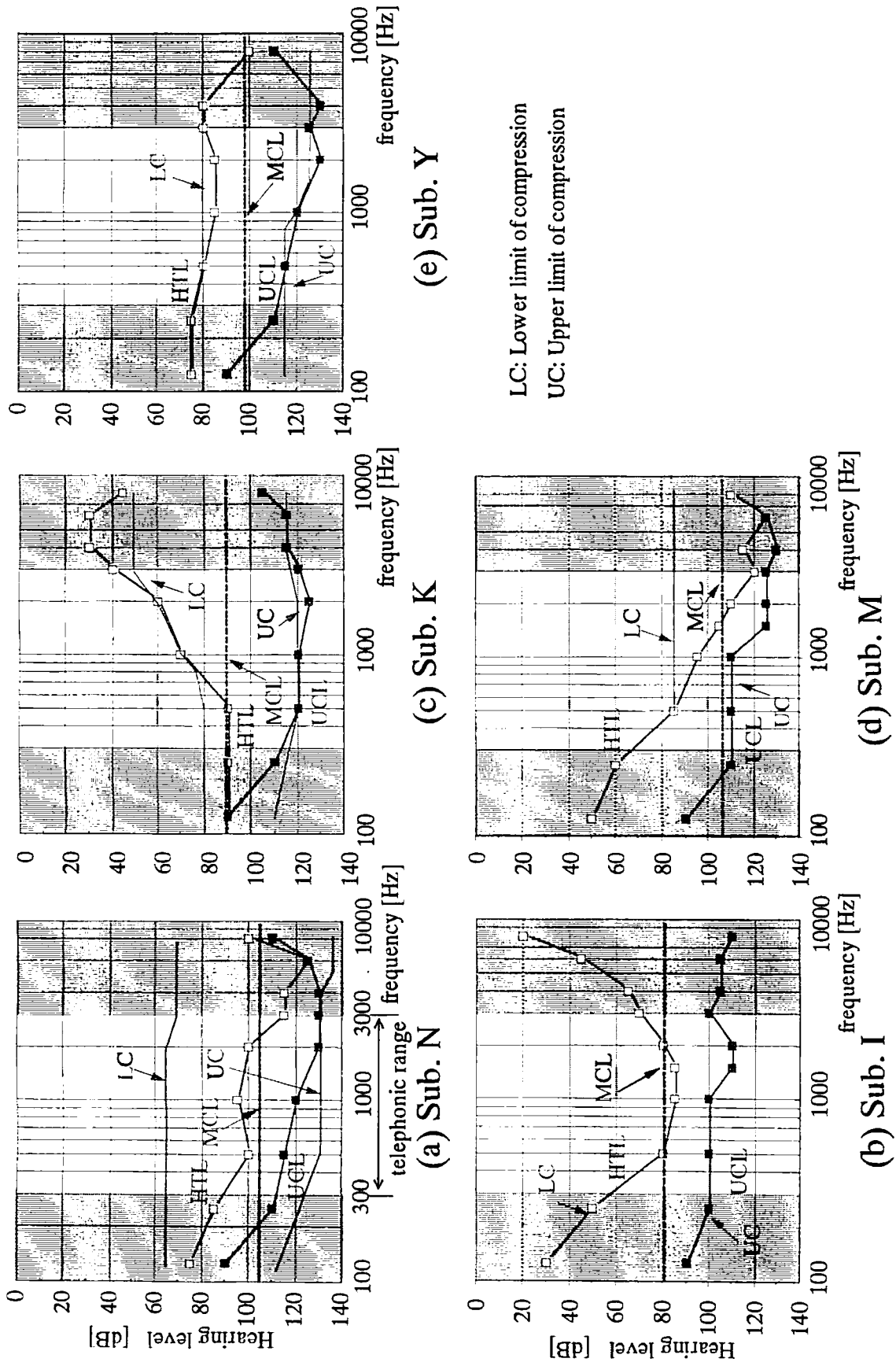
聴能訓練には5人の被験者(M, Y, K, N, I, いずれも成人女性, 23才-39才, 感音性難聴)が参加した。各被験耳のオーディオグラムと補聴器に用いた圧縮上・下限及び音声聴取時のMCLの平均がFig.3-8に示されている。Sub. NとMが高音漸傾型, Sub. Yは平坦型, Sub. Iは中域のダイナミックレンジが狭く, Sub. Kは高域に行く程ダイナミックレンジが広がる傾向を示している。Table.3-1に各被験者の補聴器使用開始時期などを示す。HTLから計算された四分法による平均聴力レベルは71-104[dB]である。1名の被験者(I)を除き全員が日常的にそれぞれの補聴器を使用している。

各被験者の振幅圧縮特性は, 次のように設定した。まず HTL と UCL, の近傍に LC(Lower limit of compression)及び UC(Upper limit of compression)を一致させたものから, 5[dB]-10[dB]程度ずつ変化させた圧縮条件ごとに 10 個の単語試料を圧縮処理しておく。次に各圧縮条件の群を 2 群抽出し, 各群から 1 個ずつ単語を選んで, 提示順をランダムにして被験者にヘッドホンの片耳に提示し, 前, 後どちらの音声が聴き良いかを強制選択させた。その結果, 聴き良いと選択された圧縮条件を被験者の聴能訓練用圧縮条件とした。

Table. 3-1 The observations on the subjects who participated in the experiment.

sub.	average hearing level [dB]	type of hearing impairment	congenital/postnatal	sex	age	age of start to use hearing aid
M	104[L]*	sensory-neural	postnatal	F**	39	5
Y	84[R]	sensory-neural	postnatal	F	23	12
K	71[R]	sensory-neural	postnatal	F	36	6
N	98[R]	sensory-neural	postnatal	F	38	20
I	83[L]	mixed	postnatal	F	29	no hearing aid

*L(left), R(right) ; tested ear. **F : female



LC: Lower limit of compression
 UC: Upper limit of compression

Fig. 3-8 Audiogram of the subjects and their compression conditions.

3-5. 訓練の効果と聴力特性の関係

3-5-1. 最小可聴閾値 (HTL) による平均聴力レベルとの関係

このシステムは文字と音声による訓練あるいは音声画像, 文字及び音声を提示する場合を任意に選択して訓練が可能であり, 被験者 5 人のうち 3 人 (N, I, K) は文字と音声画像, 1 人 (Y) は文字のみ, 残る 1 人 (M) は, 文字のみの訓練が殆ど効果がないことを確認した後文字と音声画像を併用して訓練を行った。この割り当ては平均聴力レベルから同等の被験者を異なるグループに配置するようにし M, Y を 1 グループ, I, K, N を他のグループとし, 各グループを文字のみと文字と音声画像の訓練に割り当てて学習効果を比較する意図があった。しかし後述するように, 平均聴力レベルは音声の認知能力とはあまり関係なく学習効果の比較を示すには至らなかった。毎回の聴能訓練後に行った単語聴取テスト結果と MCL の推移を Fig.3-9 ~Fig.3-13 に示す。各訓練時経過の観察では, 被験者が MCL をどの程度の変化範囲に設定しているかを調べるため, パソコンから一定振幅の校正用正弦波信号を出力し, ヘッドホン接続状態で, ヘッドホンの端子電圧実効値を測定した。この電圧実効値からヘッドホンの音圧を知るため予め, オーディオメータのヘッドホンでの音圧を人工耳に装着したマイクロフォンで電気信号に変換しその実効値を測定する。更に実験に使用したヘッドホンでの校正用正弦波信号に対する音圧レベルを同様に実効値電圧として測定しておく。オーディオメータのヘッドホン及び実験時の実効値電圧から実験時ヘッドホンでの提示音圧レベルに変換する変換式を準備する。訓練経過の各図中には訓練中の音圧レベル変化として示した。各被験者毎に, MCL 設定レベルは異なるが訓練期間中のレベル変動は小さいことが分かる。

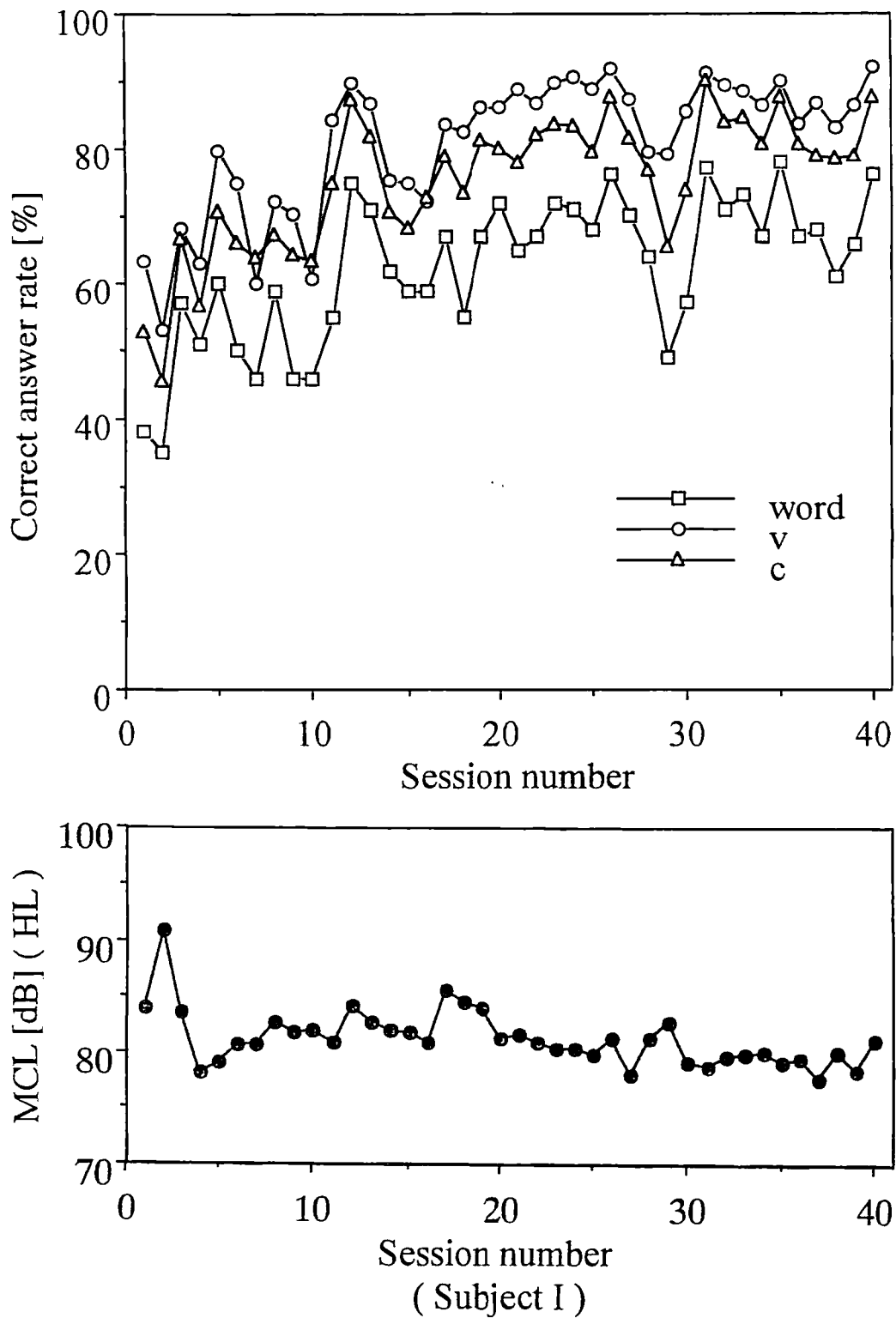


Fig. 3-9 The changes of correct answer rate and MCL in listening tests of words after the audio-visual trainings.(Sub. I)

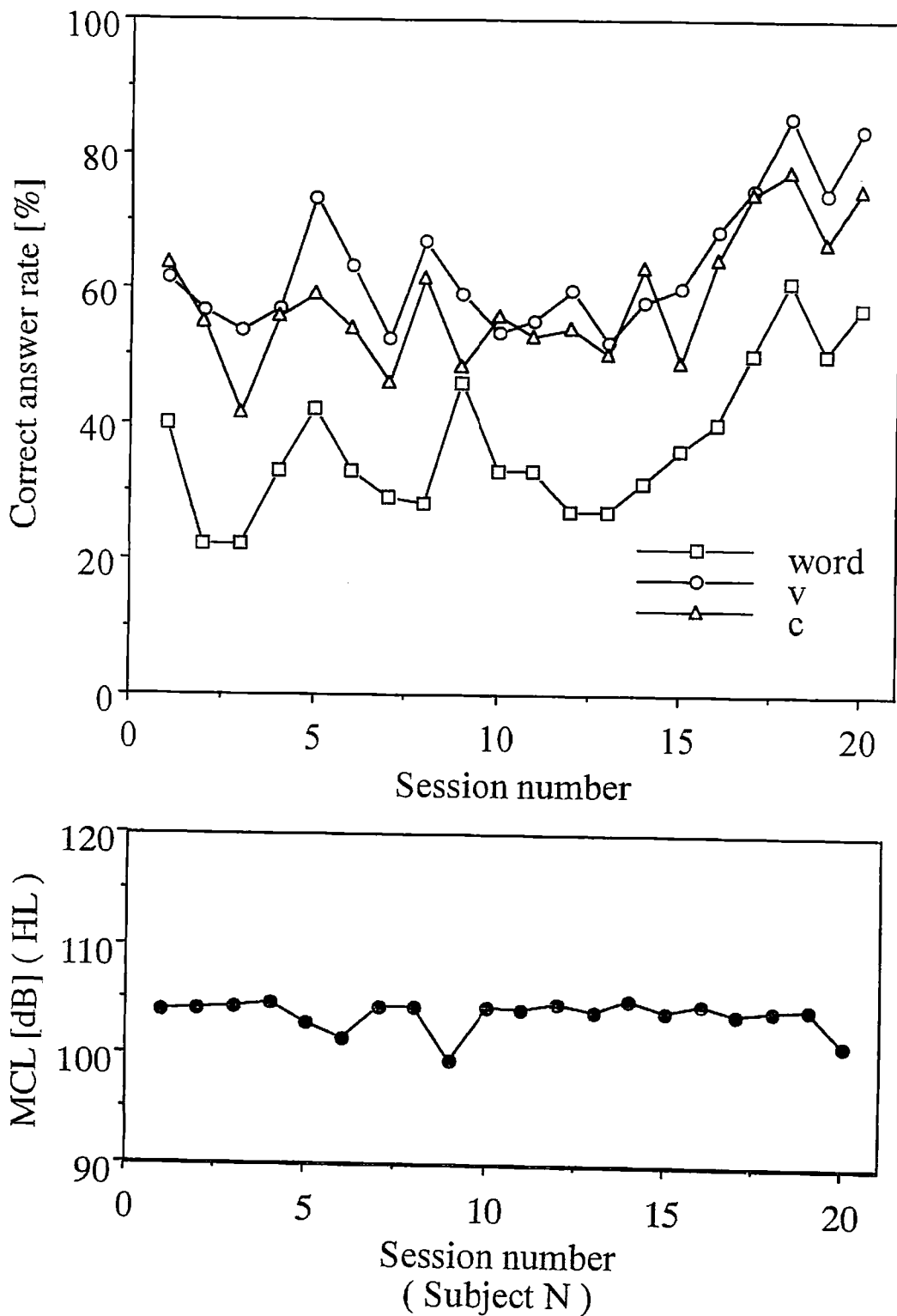


Fig. 3-10 The changes of correct answer rate and MCL in listening tests of words after the audio-visual trainings.(Sub. N)

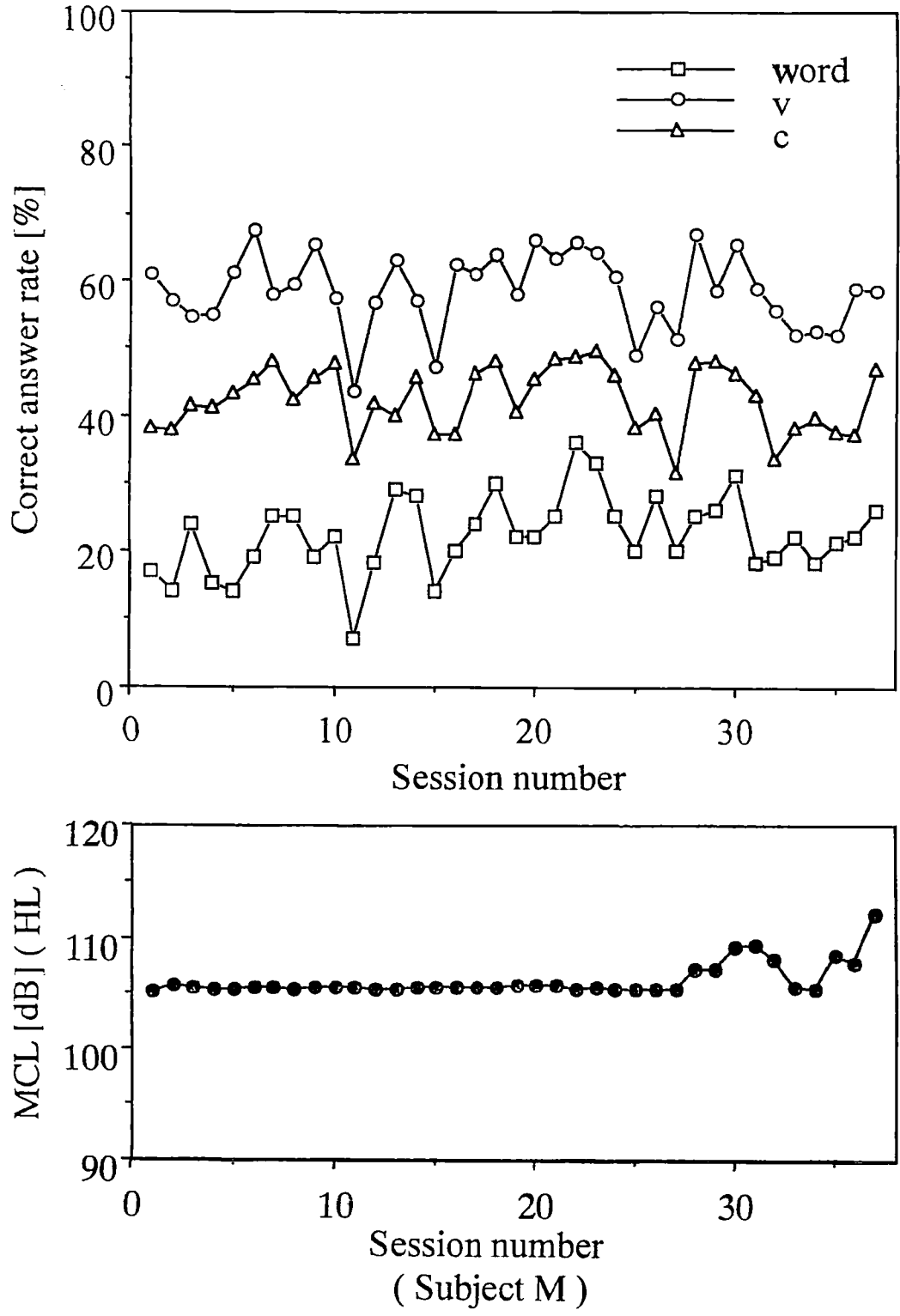


Fig. 3-11 The changes of correct answer rate and MCL in listening tests of words after the audio-visual trainings.(Sub. M)

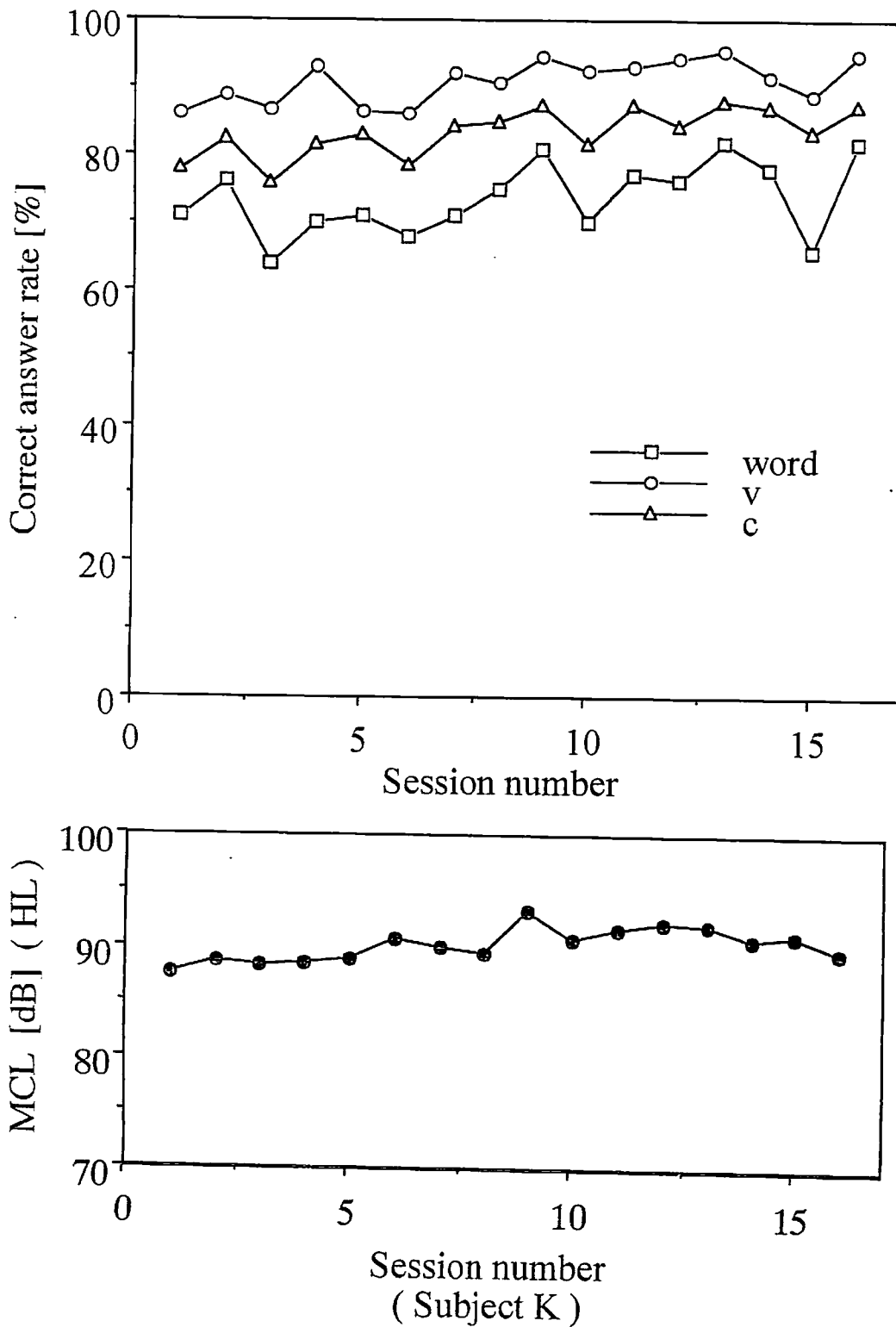


Fig. 3-12 The changes of correct answer rate and MCL in listening tests of words after the audio-visual trainings.(Sub. K)

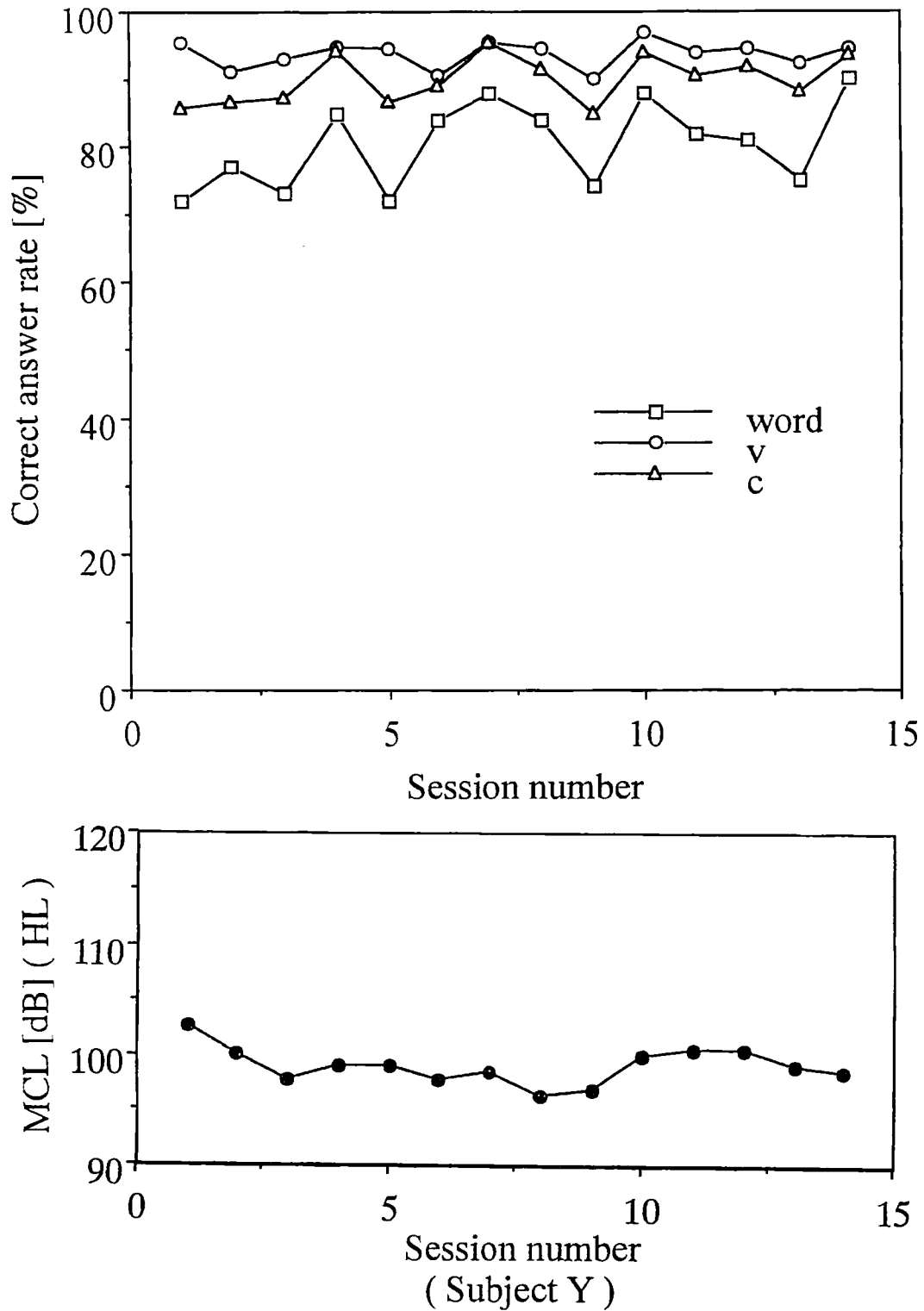


Fig. 3-13 The changes of correct answer rate and MCL in listening tests of words after the audio-visual trainings.(Sub. Y)

各々の被験者で訓練回数が異なるので被験者毎に訓練の開始直後,中間時点,訓練最後の各時点において4回ずつのテスト結果から平均正答率をとって Fig.3-14 に示した。図中に平均に用いたテストの番号を記してある。被験者のうち Y と K は当初から聴取正答率が高かったので訓練時の音声及び聴取テストの音声にはホワイトノイズを重畳し,聴取を難しくして訓練を行った。図中(K)及び(Y)はマスクノイズを加算しない場合の聴取正答率を示している。各被験者の個々のデータから訓練前,後のスコアを有意性検定 (t 検定) して統計的に有意な結果 (5%有意水準) に*記号を付した。その結果訓練効果が有意と認められるのは I, N の2名のみであることが分かった。有意性検定の結果より訓練効果の傾向は3グループに分け得る。即ち,初めからかなり高い正答率を維持し,訓練により更に高い正答率を望めない被験者(Y,K),訓練効果が顕著に認められる被験者(I,N),更に低い正答率で認知能力の改善が望めない被験者(M)に分けられる。

一般に,最小可聴閾値(HTL)による平均聴力レベルが,聴力損失の程度を示す指標として平均聴力レベルが用いられるので,各被験者の単語聴取正答率との対応を調べ Fig.3-15 に示す。この図から平均聴力レベルが 90[dB]を越える領域で聴取正答率の低下が観察されるが,両者の間に単調な傾向は認められないのが分かる。

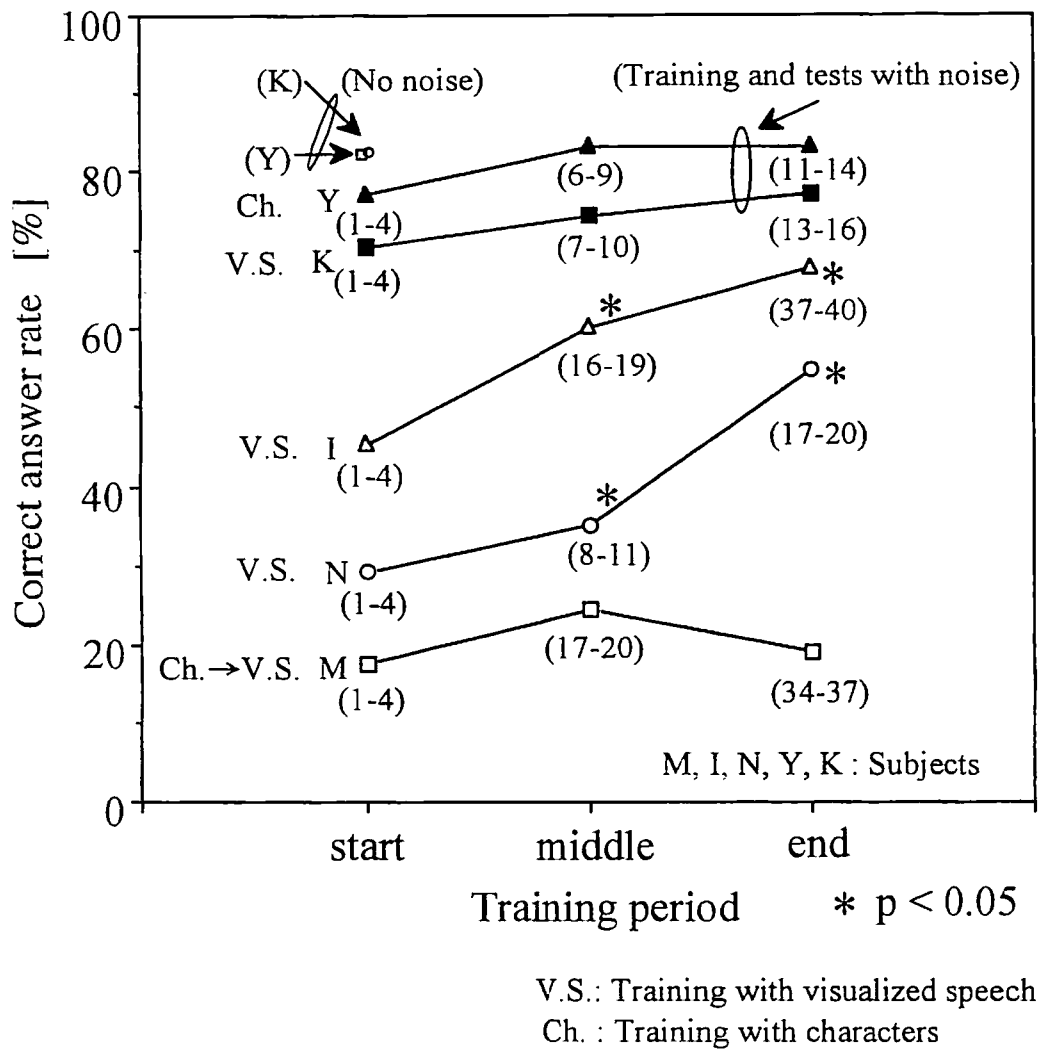


Fig. 3-14 The results in listening test of words after the trainings.

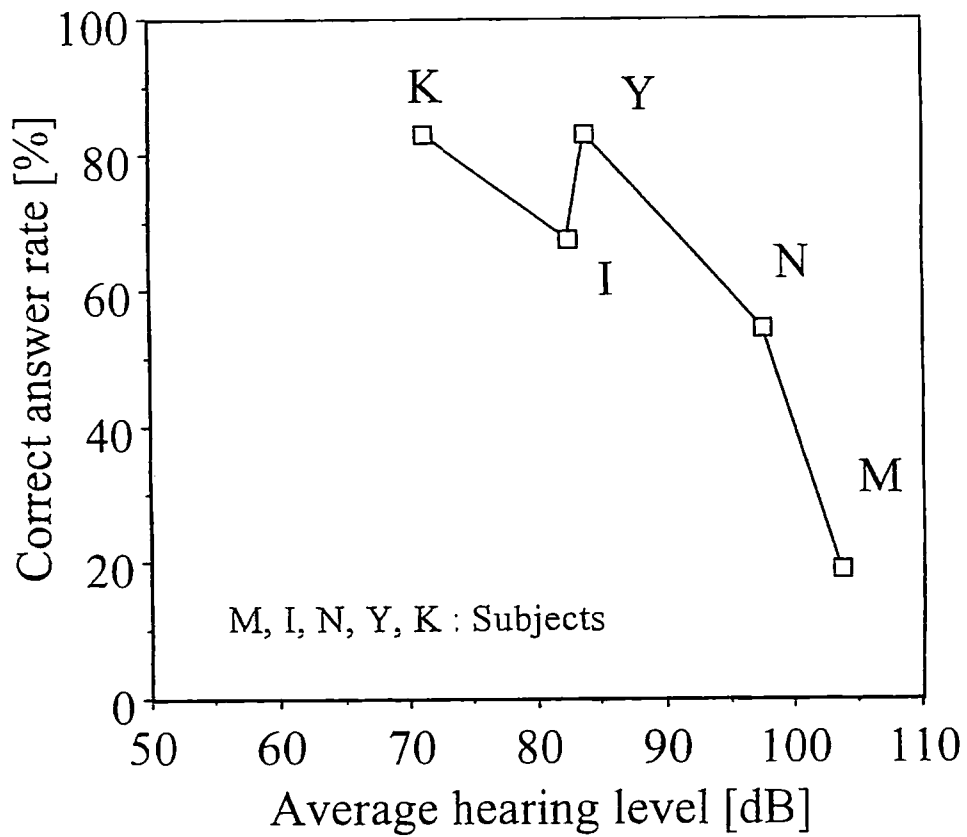


Fig. 3-15 The relation between the correct answer rate and average hearing level.

3-5-2. 提案する平均ダイナミックレンジの定義による訓練効果の解釈

前節では, Fig. 3-15 に示したように訓練効果の有効性と平均聴力レベルの間に説明のつく関連がみとめられなかった。そこで訓練効果の有無と関係すると考えられる聴覚的特性の表現として, 四分法により平均ダイナミックレンジ H_r [dB] を定義した。

$$H_r = (H(f_1) + 2H(f_2) + H(f_3)) / 4 \quad (3.1)$$

$$H(f_k) = UCL(f_k) - HTL(f_k) : k = 1, 2, 3$$

$$f_1=500[\text{Hz}], f_2=1000[\text{Hz}], f_3=2000[\text{Hz}]$$

Table 3-2 に, 各被験者の平均ダイナミックレンジと通常聴力の指標として用いられる平均聴力レベルを示す。前述した画像の併用効果が音声認知能力を改善したと見られる被験者 I と N は, この平均ダイナミックレンジが 20-30[dB] の被験者である。それに対して 40-50[dB] の平均ダイナミックレンジを持つ Y, K は当初より 80% 程度の正答率で改善の余地が少なく, 被験者 M は余りにも可聴領域が小さく (8.7[dB]), 聴能訓練の効果が上がらないと解釈できる。

次に, 平均ダイナミックレンジに対し最終的な聴取正答率を整理すると Fig. 3-16 のようになる。図において, 平均ダイナミックレンジ 22.5[dB] の被験者 N を除けば平均ダイナミックレンジの増大によって, 正答率が大になる傾向がある。被験者 M, N 以外の被験者は Fig. 3-8(b) (c) 及び(e) に例示したように各自の HTL と UCL に近い圧縮の下限, 上限の特性で振幅圧縮した音声を好んだが, 被験者 M, N は, 圧縮特性調整時に自己の聴力特性よりかなり緩やかな圧縮特性を選択した (Fig. 3-8(a)(d) 参照)。被験者 N の場合オージオグラムに比べて, 振幅圧縮が緩やかなために可聴領域内の情報が減少する。従って, Fig. 3-16 に示すように, N と I が同程度の平均ダイナミックレンジを有するにも拘わらず正答率が被験者 N が I

よりも 15%程度低くなると説明できる。Fig.3-17 に提案する平均ダイナミックレンジと平均聴力レベルの関係を示す。この図から両者には一定の関係が薄いことが分かる。

Table. 3-2 The average hearing level and average dynamic range of the subjects.

sub.	average hearing level [dB]	average dynamic range [dB]
M	104[L]*	8.7
Y	84[R]	37.5
K	71[R]	48.8
N	98[R]	23.8
I	83[L]	20.0

*L(left), R(right) ; tested ear.

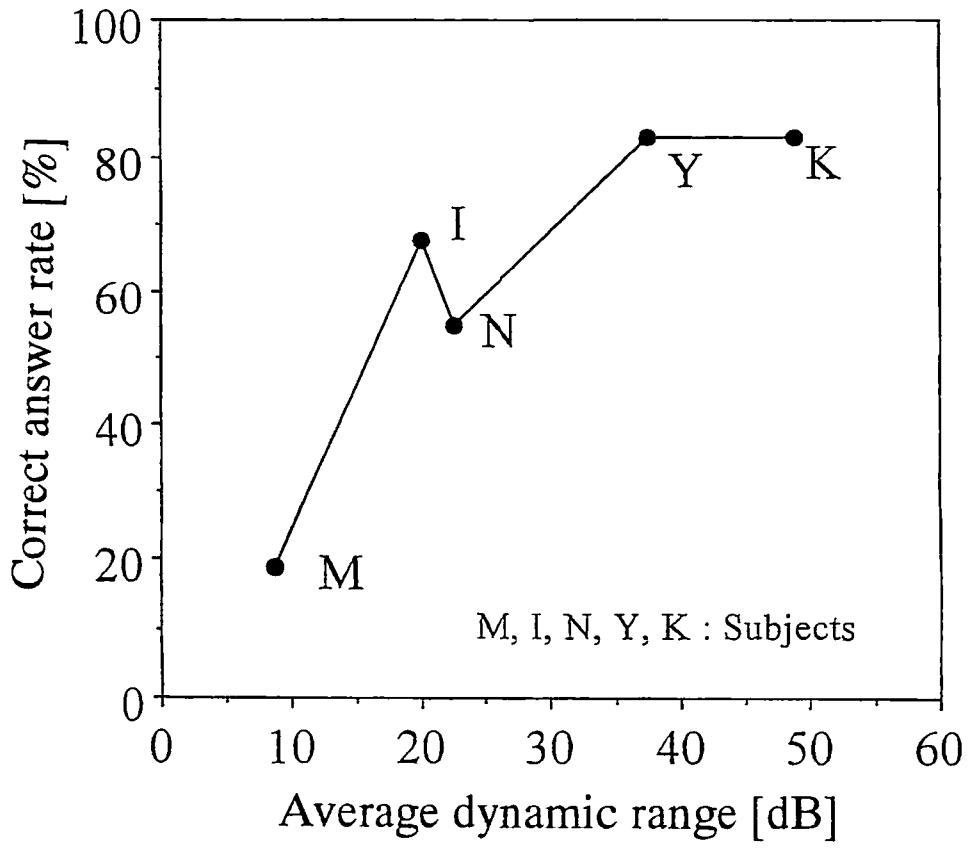


Fig. 3-16. The relation between the correct answer rate and average dynamic range.

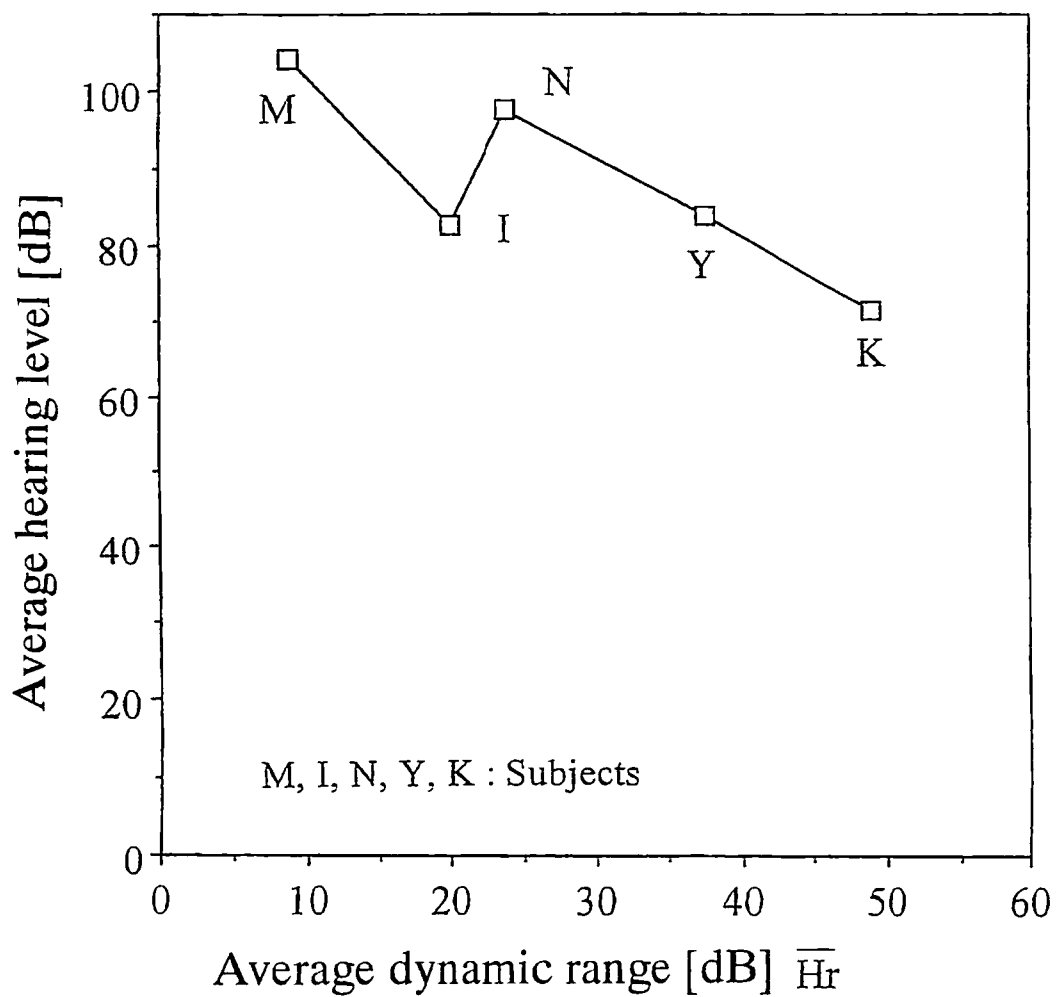


Fig. 3-17 The relation between average hearing level and average dynamic range.

3-6.まとめ

本章では感音性難聴者の聴覚に与える音声情報と音声認知能力との定量的関係を明かにするため、まず聴能訓練によって認知能力を改善した上限値を求め、両者の関係を検討した。聴能訓練には音声画像を併用したシステムを提案した。音声データベース音声を被験者毎の圧縮特性に従い圧縮処理した振幅圧縮音声と音声画像を提示し補聴処理により可聴領域内へ圧縮されたホルマントピーク情報のキューを音素と対応させる手助けとして、音声認知能力の改善を図る。訓練では文字と音声画像を一定タイミングで画面に提示しつつ音声をヘッドホンに片耳提示して一定語数（50語）の聴能訓練の後聴取テストを1回行う。5人の難聴者による訓練を実施し、毎回の訓練後に行った単語聴取試験の結果を分析したところ、被験者に対する訓練効果は次のように整理できる。

訓練の開始直後、中間時点、訓練最後の各時点において4回ずつのテスト結果から平均正答率をとって有意検定（t検定）した結果、統計的に有意な結果を示したのは2名の被験者であった。訓練効果の傾向は、初めからかなり高い正答率を維持し、訓練により更に正答率の向上が望めない被験者、訓練効果が顕著に認められる被験者、更に低い正答率で訓練による向上が認められない被験者の3グループに分けられる。聴覚障害者の聴力損失は一般にHTLの四分法に基づく平均聴力レベルが用いられるが一連の訓練結果を説明する尺度として適さなかったため感音性難聴者の音声認知を反映するものとして平均ダイナミックレンジを定義した。各被験者について結果を検討した結果、平均ダイナミックレンジが20-30[dB]の被験者で、聴能訓練の効果が顕著であることが明かになった。また振幅圧縮の条件にHTL,UCLかそれらに近い特性を用いた場合には平均ダイナミックレンジは音声認知と関連が強く、緩やかな圧縮特性では被験者の音声認知が低下する傾向が認められた。

第4章 難聴者の可聴領域内ホルマント情報と語音認知能力の関係

4-1. 難聴者の平均ダイナミックレンジと振幅圧縮音声による語音認知能力の関係

可聴領域内にホルマントピークが入る確率は圧縮時の圧縮下限, 上限と密接に関係する。被験者の聴力特性に一致するような圧縮条件で振幅圧縮した場合には, 平均ダイナミックレンジの広さがホルマント情報の伝達量に直接の関連を持つといえる。しかし被験者の可聴領域と一致した条件で圧縮する場合に比べ緩やかな圧縮上限, 下限で圧縮した場合には, 可聴領域内に入るホルマントピークの割合が減少する。実際に聴取する難聴者が圧縮音声の音質をややきついつと感じたりする場合には圧縮上限と下限を緩やかに設定する場合がある。このことが前章において圧縮特性が聴力特性に一致しているときには平均ダイナミックレンジから聴き取りの正答率を良く推定できるのに対し, 被験者 N の場合のように聴力特性に比べ緩やかな振幅圧縮をする場合には, 推定値より正答率が小さくなる原因になると考えられる。このことは, 圧縮の程度に応じ, 可聴領域内に入るホルマント情報の量が変化するためと推定される。そこで, 聴取正答率を聴力特性から算定した平均ダイナミックレンジで説明できないような圧縮条件の場合について, 可聴領域内のホルマントピーク分布に着目し, 音声認知との関連を検討する。

4-2. 各難聴者の可聴領域内ホルマントピーク分布の推定法

可聴領域内にあるホルマントピークの分布割合を F_1, F_2 ごとに算出し, 難聴者による圧縮音声の聴取結果とその関係を検討する。逆フィルタ制御法により抽出されたホルマント周波数に基づき, 第2章において述べた(2.3)(2.4)式の導出と同様にして, ホルマントの共振を打ち消す逆フィルを用いて単共振成分に分解しこれに

512 点の FFT 処理を行い,ホルマント周波数の上下 1 ピッチ周波数の範囲でピークレベルを探索しホルマントピークとする。試料音声の聴取には F_1 , F_2 が相互に関連して貢献していることから,両者を関連づけるパラメータが音声の聴取し易さを表すと思われる。Table.4-1 に各被験者の聴取した音声試料での F_1 , F_2 それぞれのホルマントピークの可聴領域内分布割合とその加算平均,幾何平均及び聴取時の正答率を示す。また聴取正答率と上述の各パラメータとの相関係数を最下段に示した。

Table. 4-1 Correlation between formant peak distribution in the hearing areas and correct answer rates.

Subject (Hard of hearing)	F_1 [%]	F_2 [%]	$\frac{(F_1[\%]+F_2[\%])}{2}$	$\sqrt{F_1[\%]\times F_2[\%]}$	Correct ans. rate [%]
I	71.5	47.8	59.7	58.5	67
N	57.5	36.7	47.1	45.9	54
Y	85.8	81.5	83.7	83.6	83
M	62.6	0.7	31.7	6.6	19
K	90.1	93.5	91.8	91.8	83
Correlation coefficient	0.79	0.97	0.94	0.99	

4-3. 難聴者における可聴領域内ホルマントピーク分布と語音認知能力の関係

Table.4-1で F_1 , F_2 ピーク分布幾何平均と聴取正答率が最も高い相関(0.99)を示していることから, 難聴者の音声認知結果と試料のホルマントピーク分布の幾何平均値との関係を Fig.4-1 に示す。この結果から難聴者の音声明瞭度は F_1 , F_2 ピーク分布の幾何平均に対し単調増加の傾向を持つことが分かる。被験者 N, I の正答率も共にこの単調な関係を満足し可聴領域内にあるホルマント情報に見合った音声明瞭度を示していると考えられる。

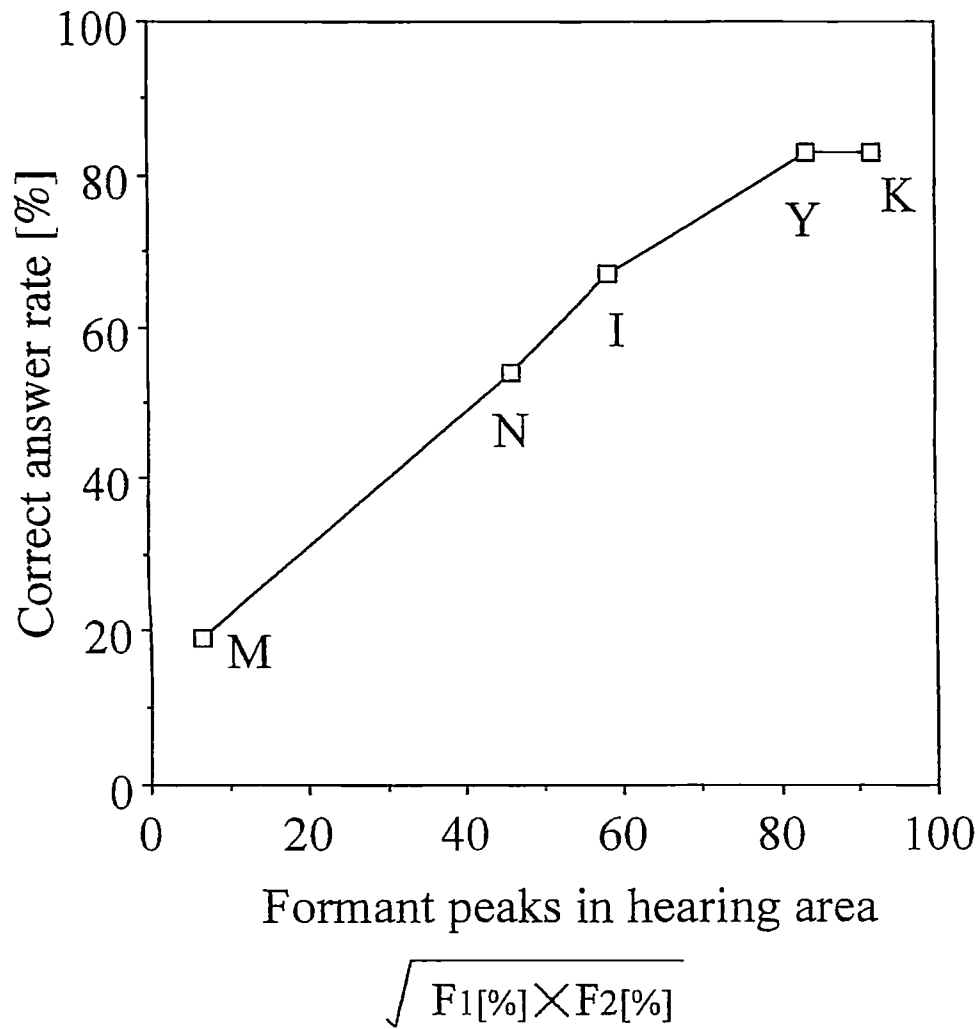


Fig. 4-1 The distributions of formant peak levels in hearing area and correct answer rate (Hearing impaired).

4-4. 難聴者の語音認知能力に関する一仮説

前節までの結果から、難聴者が聴取する際の音声明瞭度を可聴領域内のホルマント情報から定量的に推定できるという仮説を得た。即ち被験者の HTL, UCL と LC, UC がそれぞれ対応しているときには可聴領域内に入るホルマント情報量と関連する、平均ダイナミックレンジの広さが、また、被験者が自己の聴力特性より緩やか LC, UC を好む場合には、可聴領域内に存在するホルマントピークの全フレーム数に対する分布割合の幾何平均値が音声の聴取正答率と密接な関係にあり、振幅圧縮条件の変化によりこれらのホルマントピークが、可聴領域内にどの程度分布するかを知る事により、難聴者の音声認知正答率の推定が可能である。

4-5.まとめ

前章の結果では、難聴者の音声認知で、振幅圧縮特性が聴力特性に一致している場合、聴力レベルにより算定した平均ダイナミックレンジが大きい程音声認知正答率が高くなる傾向が認められた。本章では、振幅圧縮音声の圧縮条件が、聴力の測定結果に一致しない緩やかな圧縮を好む場合にも、音声認知正答率を説明するパラメータについて検討した。圧縮が緩やかになると圧縮の程度に応じ、可聴領域内に入るホルマント情報の量が増加すると推定される。そこで、可聴領域内のホルマントピーク分布に着目し、音声認知との関連を検討した。

難聴者の可聴領域内に分布する第1,第2ホルマントのピーク双方が音声認知に関係が深いことから、第1,第2ホルマントピークのそれぞれの可聴領域内ピーク分布を求め、両者の加算平均、幾何平均と音声の聴取正答率の相関を検討したところ、幾何平均値が高い相関係数(0.99)を示した。更に、難聴者に提示した音声試料の第1,第2ホルマントピークの可聴領域内分布割合の幾何平均値と聴取正答率は、単調増加の関係を示した。その結果、前章で測定された聴力特性より緩やかな圧縮特性を好む被験者の正答率も、緩やかな圧縮特性に見合ったものであることが明らかになった。

以上の結果から難聴者の音声認知は、可聴領域内のホルマントピーク分布と、難聴者の音声認知正答率の間には一定の関係が存在し、音声明瞭度を推定することが可能になるといふ仮説を得ることができた。これは新たな補聴器の設計指針や補聴効果の予測に、有用性が高いと考えられる。

第5章 健聴者の模擬難聴実験による難聴者の語音認知能力に関する仮説の検証

本章ではまず第4章で得た難聴者の改善された音声認知能力と可聴領域内の音声情報の定量的関係から、補聴器を通して伝達すべき音声情報についての仮説を健聴者の種々の模擬難聴状態において可聴領域内の音声情報の量を制御する実験で検証し、かつ模擬難聴が実際に感音性難聴の状態を正しく模擬していることを証明する。このためマスクノイズで可聴領域を模擬した健聴者に対して、提示する音声の圧縮割合を制御し、その可聴領域内のホルマントピーク分布を変えたときの音声明瞭度の変化を調べる。

5-1. 仮定された難聴の聴力特性と振幅圧縮の制御

3人の感音性難聴者(N, I, Y)のHTLの形状に近い周波数特性のマスクノイズを用いて難聴状態をシミュレートした。振幅圧縮のために補聴器に与える特性として圧縮の上限をHLの90[dB]に設定し、下限の方はマスクノイズのレベルに一致するところからFig.5-1のように3~4段階、20[dB]ずつ緩やかになるように決めた。また、各難聴者の圧縮音声以外に非圧縮音声も実験対象とした。

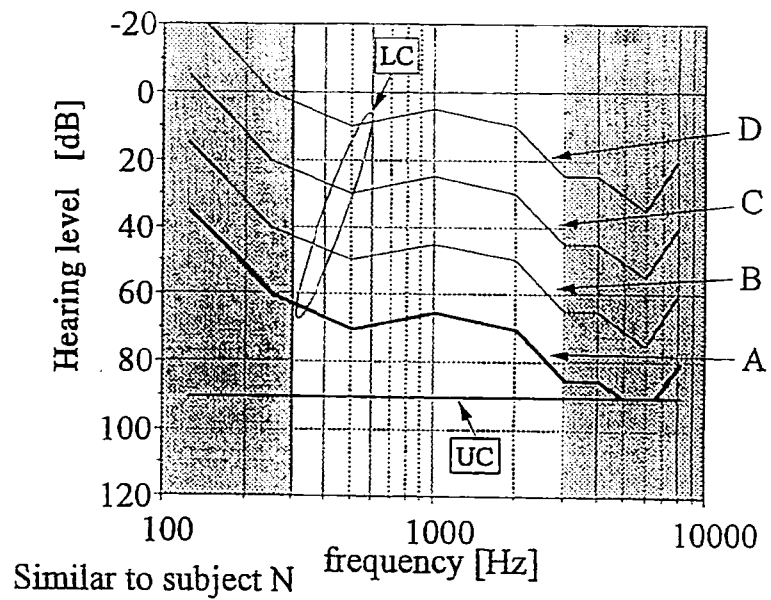
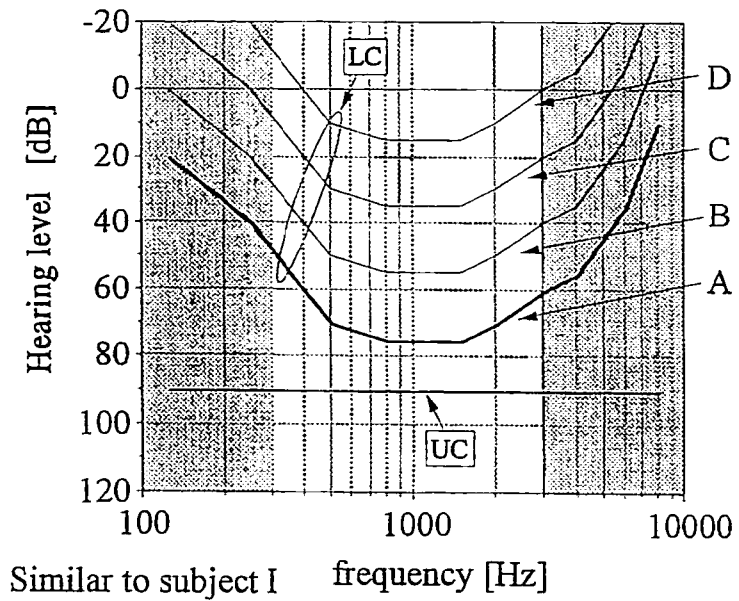
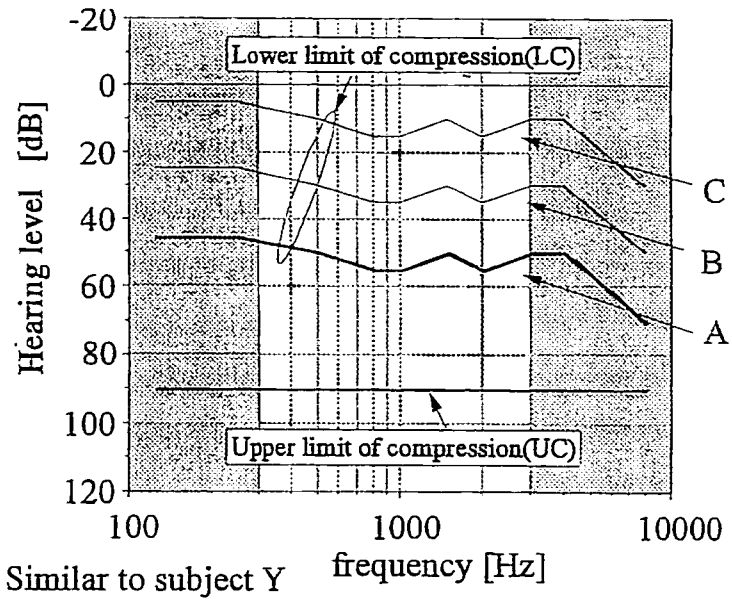


Fig. 5-1 Compression conditions for the simulated hard-of-hearing.

5-2. 模擬難聴状態の生成方法

難聴状態を模擬するマスクノイズは実効値の周波数スペクトルが Fig.5-1 に与えられた特性 A に一致するものとした。即ち, マスクノイズは各難聴者の特性 A に一致するようホワイトノイズを FIR フィルタに通して作成し, 難聴者毎の音声試料に加算して聴取実験に使用した。FIR フィルタは 125-6000[Hz] の帯域で HTL の振幅特性を持つフィルタのインパルス応答とホワイトノイズの畳み込みを行う。フィルタは帯域内の測定点間は直線内挿、帯域外は 0 とする特性を与えた。300-3000[Hz] の BPF に通してマスク信号とし各音声試料に加算して実験に使用した。各難聴者の聴力特性これにより圧縮条件の違いにより一定のマスクノイズをこえて可聴領域内に入るホルマントピークの分布割合を変化させながら音声を提示した。

5-3. 実験方法

音声試料は男女話者各 10 名の 400 単語であり, テストの度にランダムに抽出した 80 単語でピーク分布を算出すると共にそのうちの 50 単語をランダムに抽出し聴取試験に用いた。更に男女各 2 人が発話した拗音を除く/VCV/音節試料 (62 個) から全音節 62 個を抽出してその聴取試験をも行った。被験者は健聴者 4 名 (大学生) である。被験者は防音室内に着席し, 聴取試験に用いない単語群で MCL に音量を調整した後, 音声試料をヘッドホンで片耳聴取した。上記の方法により非圧縮音声と各圧縮条件による圧縮音声を用いた聴取テストを行い, 音声の正答率 (明瞭度) と可聴領域内のホルマントピーク分布割合の関係を調べた。

5-4. 実験結果と仮説に関する検討

Fig.5-2 に実験結果を示す。図は F_1 , F_2 の可聴領域内ホルマントピーク分布割合の幾何平均値に対する次の3種の聴取正答率を示している。

- (I) 模擬難聴による単語音声の正答率
- (II) 模擬難聴による/VCV/音声の正答率
- (III) 感音性難聴者の単語音声の正答率

模擬難聴の場合の正答率は健聴者4人の被験者の平均として示されている。また図中の曲線は、いずれも2次関数の最小2乗近似による表現である。この結果より次の事実を指摘することが出来る。

(1) ホルマントのピーク分布の幾何平均値と音声明瞭度の関係は明らかに単調増加の傾向を示す。この場合、非圧縮音声（図中に N. と表示）も圧縮音声の正答率と同一の関数で推定可能である。

(2) 可聴領域内のホルマントピーク分布が60%以上の場合健聴者の/VCV/音節聴取結果は難聴者の単語聴取結果に近い正答率を示している。これは難聴者が単語の持つ冗長性を音声認知に活用することの不得手さを示しているという可能性があるが、その他に次節で述べるように難聴者の場合、UCL を越える成分の存在が MCL を下げる要因となりホルマントピークの分布割合がこの評価よりも実際には少なくなっているのに対し、健聴者の UCL は仮想的に与えられているので、MCL によって、ホルマントピーク分布の割合は変化しないという原因も考えられる。実際には難聴者と健聴者の単語音声の聴取におけるスコアの差はこれらの2つの原因が混在した結果によると考えている。それに対してホルマントピーク分布が80[%]をこえると、模擬難聴状態の健聴者の単語聴取試験結果に難聴者のそれが近づく傾向が認められる。これは可聴領域の広い被験者は健聴者同様、単語の冗長性を利用しながら判

断できると共にホルマントピークレベルが UCL を越える頻度が少なくなることを示していると思われる。

以上の結果より、難聴者の語音の聴取において、可聴領域内のホルマントピーク分布割合が増加するにつれて音声の明瞭度が増加するという仮説が、模擬難聴状態の健聴者による実験結果で裏付けられた。このことはマスクノイズで模擬する難聴状態による健聴者の聴取判断は感音性難聴者のそれとかなり良く一致すること即ち、模擬難聴によるテストが信頼性の高いものであることをも示している。

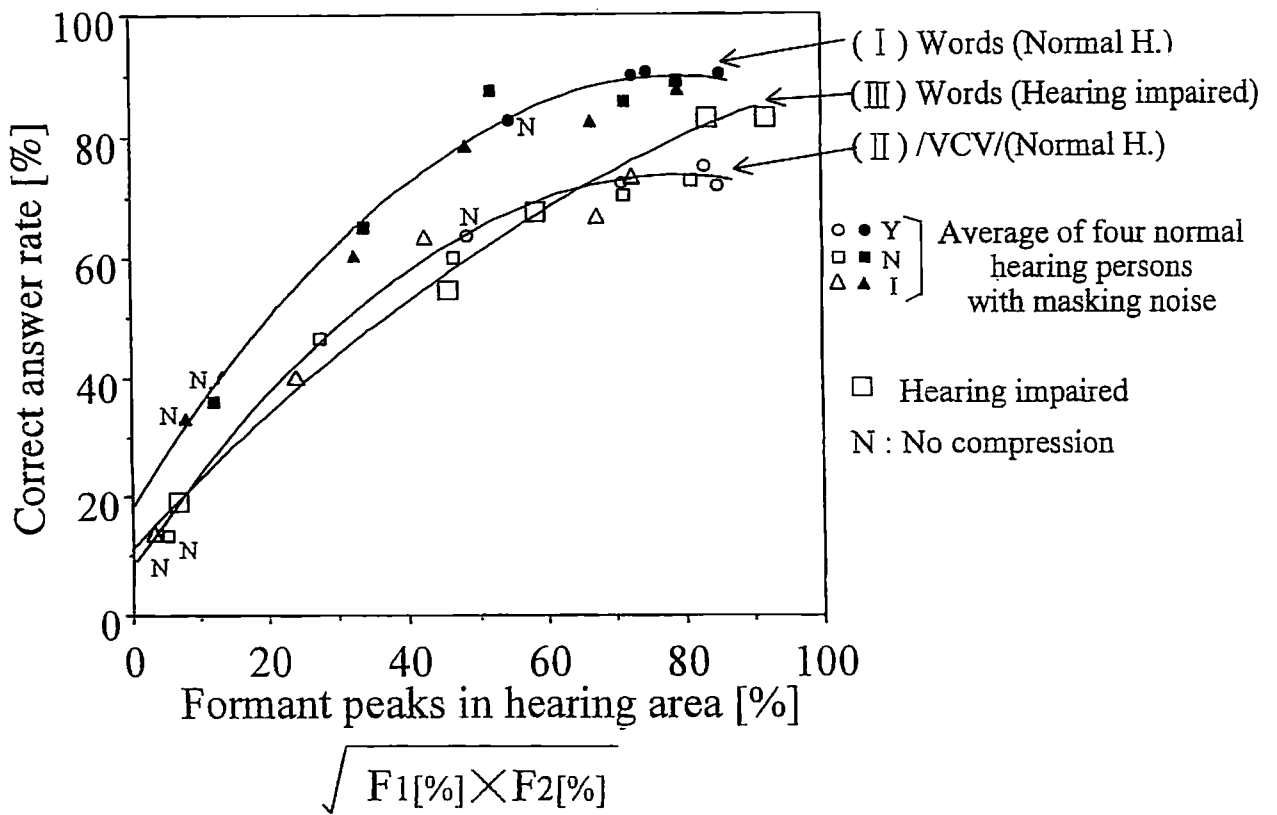


Fig. 5-2 The relations among distributions of formant peak levels in hearing area and correct answer rate.

5-5.まとめ

本章では音声認知能力と可聴領域内の音声情報の定量的関係から、補聴器を通して伝達すべき音声情報についての仮説を健聴者の種々の模擬難聴状態において可聴領域内の音声情報の量を制御する実験で検証を試みた。難聴者の HTL 特性に近い 3 種類のマスクノイズを準備し健聴者を模擬難聴状態にした。そこに音声の振幅圧縮条件を変化させ可聴領域内に存在するホルマントピークの割合を変化させた圧縮音声試料を提示した結果、音声明瞭度（正答率）は、先の感音性難聴者の場と同様の傾向、即ち、第 1 第 2 ホルマントピークの可聴領域内分布割合の幾何平均に対し単調増加傾向を示した。また非圧縮音声の聴取結果も同一関数の延長上近く圧縮音声とともに一定の関数で推定が可能ながことが明らかになった。

以上の結果、難聴者の語音の聴取において、可聴領域内のホルマントピーク分布割合が増加するにつれて音声の明瞭度が増加するという仮説が、模擬難聴状態の健聴者による実験結果で裏付けられた。またこの実験結果は、マスクノイズで模擬する難聴状態による健聴者の聴取判断は感音性難聴者のそれとかなり良く一致し、模擬難聴によるテストが信頼性の高いものであることをも示している。

このように、補聴器の設計、フィティングにおいて可聴領域内のホルマントピーク分布と音声明瞭度（正答率）の関連を予め調べておけば、使用時の音声明瞭度推定を可能にすることを示唆ものである。これは聴能訓練での目標値としてとらえることも可能と考えられる。

第6章 単共振分解型補聴器の振幅圧縮性能の改善に関する検討

本章では、可聴領域に音声情報を伝達する単共振分解型補聴器の改良法とその効果の推定について述べる。

6-1. 振幅圧縮特性が真の難聴と模擬難聴に与える影響の違い

前節までの実験結果によれば感音性難聴者の認知能力を改善するには常に可聴領域内に音声のホルマントピークがより多く入るように補聴器で処理制御することが望ましい。ここで使用した電話音声用単共振分解型補聴方式では、単共振信号の実効値からホルマントピークレベルを推定して、圧縮特性を定めている[19]から、その推定に誤差がある場合には感音性難聴の被験者は MCL の調整によってそれを補正しようとすると思われる。即ち、圧縮上限(UC)を UCL に一致させた場合 UCL を越えるホルマントピークの頻度がある程度多ければ意図的に MCL を下げることで耳を保護しようとする可能性がある。このため、模擬難聴の健聴者に比べ可聴領域内に入るホルマントピーク分布割合が低下している可能性がある。そこでこのような可能性を軽減するため振幅圧縮特性の改善を図る。

6-2. 振幅圧縮特性の改善法

先に提案し本実験に使用している圧縮方式で UCL を越える成分の存在を明確にするために第2ホルマント周波数のダイナミックレンジに対する圧縮後の F_2 ピークと UCL のレベル差を分布図として Fig.6-1 に示した。(第1ホルマントは UCL を越えることが殆どない。) 図の x 軸はホルマント周波数における可聴ダイナミックレンジの広さを現す。図は男女各2名の話者の発声した200単語を被験者 I の聴

力特性で圧縮した音声の分析結果であり縦軸（ホルマントピークレベルと UCL の差）の値が 0[dB]を越えれば圧縮信号が UCL を越えていることを意味する。この図からダイナミックレンジが 25[dB]程度以下になると UCL を越えるホルマントピークレベルの発生が顕著になる。この現象は、圧縮係数の計算の元になっている Fig.2-6 に示した単共振波の実効値とホルマントピークレベルの関係から推定すると、同一の実効値に対し F_2 のピークレベルの分布が 20-40[dB]程度の広がりを持つため、ホルマントピークの存在する周波数での UC と LC の間のダイナミックレンジの減少につれて、圧縮音声のホルマントピークもばらつきを生じていると考えられる。このような現象を防ぐために、一律に UC を下げるのではなくホルマント周波数での可聴領域の広さが 25[dB]以下の場合に圧縮上限(UC)の値を次のように修正して UCL を超えないようにする改良を試みる。

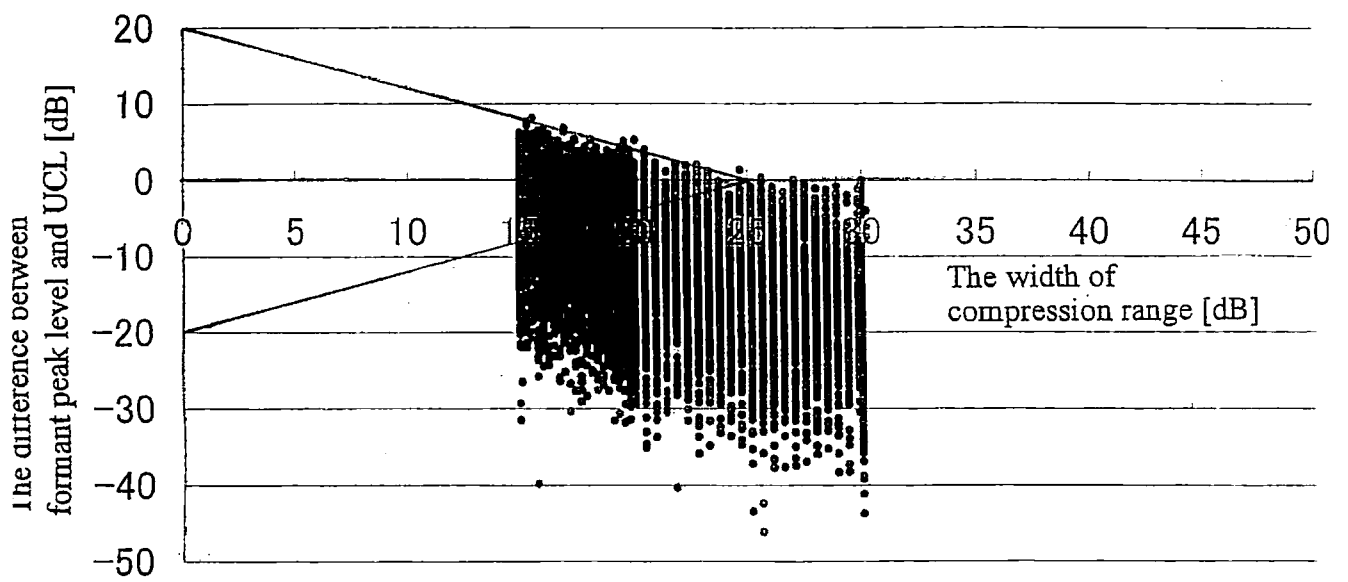


Fig. 6-1 The difference between formant peak level and UCL .

Fig.5-4

即ち次式の圧縮上限 UL_c をこれまでの UC (UCL に等しい)の代わりに, UC と LC の間のダイナミックレンジが 25[dB]程度から減少しても UCL を越えるホルマントピークが増加するのを防ぐような振幅圧縮処理を用いることとした。

$$H_r(f) = UCL(f) - HTL(f)$$

$$UL_c(f) = UCL(f) + 0.8H_r(f) - 20 \quad ;(H_r(f) \leq 25[\text{dB}]) \quad (6.1)$$

$$UL_c(f) = UCL(f) \quad ;(H_r(f) > 25[\text{dB}])$$

ここで

f :ホルマント周波数[Hz]

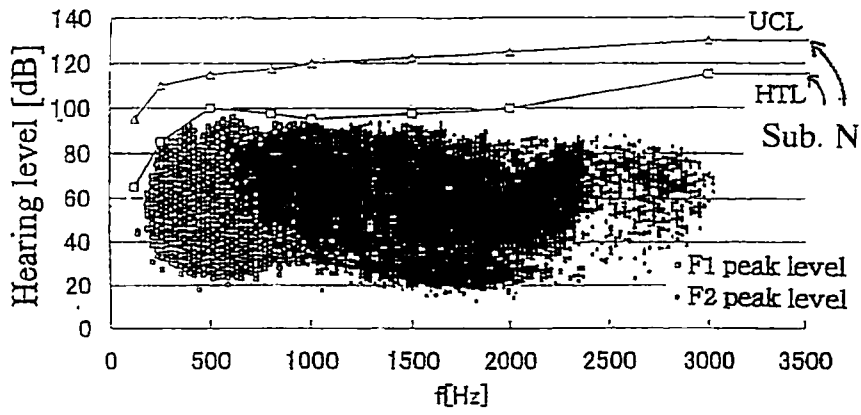
$H_r(f)$:ホルマント周波数 f [Hz]での可聴範囲[dB]

$UL_c(f)$: f [Hz]での圧縮上限[dB]

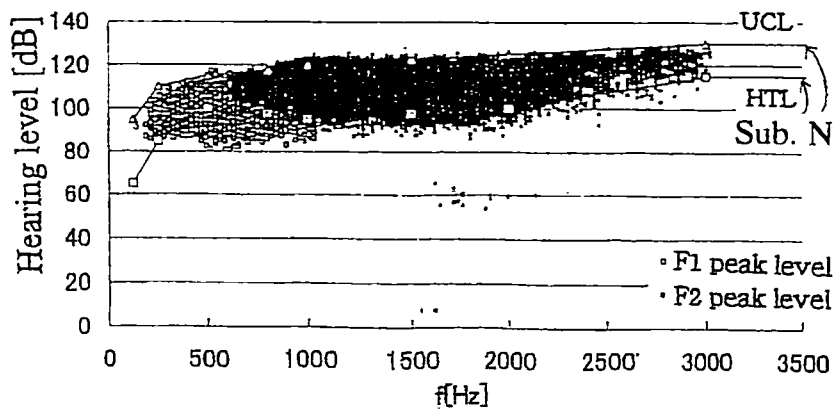
$UCL(f)$: 被験者の f [Hz]での UCL[dB]

Fig.3-8 に示した難聴者の聴覚特性 (HTL と UCL) について上記の圧縮上限 UL_c と HTL に一致した圧縮下限による圧縮効果をホルマントピークの測定によって調べる。音声試料は成人男女各 2 名の発話による 50 単語 (合計 200 単語) のすべてのフレームを対象とし前節と同様の分析を行う。

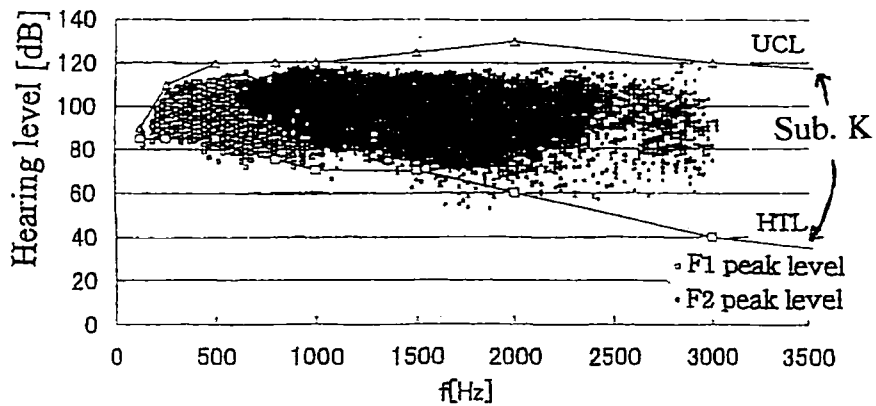
Fig.6-2 は振幅圧縮前後の音声の F_1, F_2 のホルマントピーク分布の例を示す。(a) の原音声では例示される難聴者(N)の可聴領域より下に 70[dB]以上のダイナミックレンジでホルマントピークが広がっている。(b)は同難聴者の UCL から(2)式を用いて UL_c に補正して振幅圧縮を行った後 F_1, F_2 のホルマントピーク分布を示したものである。同様に(c)には難聴者(K)の場合が示されている。Fig.6-3(d)~(f)はそれぞれ難聴者 M, I, Y の聴力特性に対応する圧縮結果である。これら 5 名の感音性難聴者何れの場合においても圧縮によって UCL を越えるホルマントピークは極く少数になり MCL が適正に設定できると推定される。



(a) Non-compressed speech; $\sqrt{F1[\%] \times F2[\%]} = 0 [\%]$

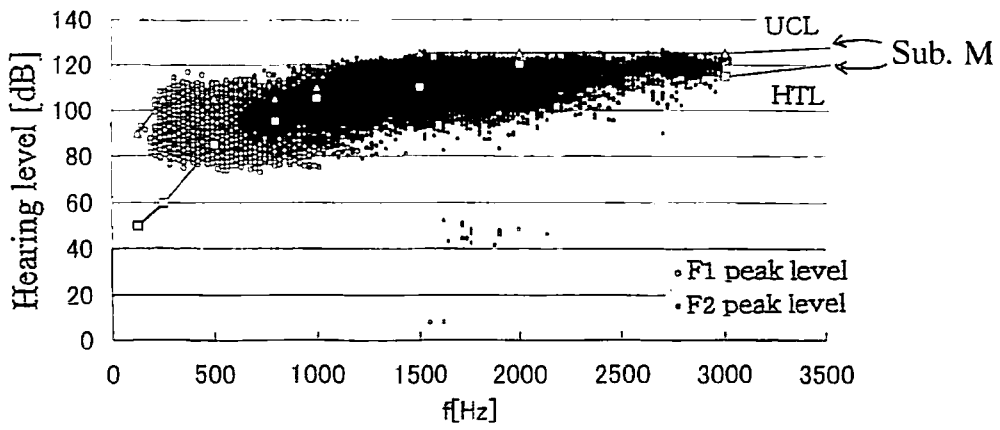


(b) Compressed for sub. N; $\sqrt{F1[\%] \times F2[\%]} = 84 [\%]$

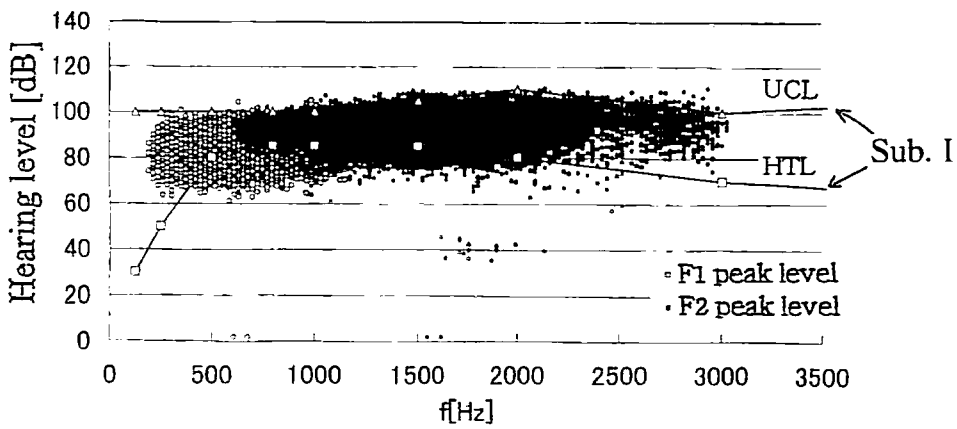


(c) Compressed for sub. K; $\sqrt{F1[\%] \times F2[\%]} = 99 [\%]$

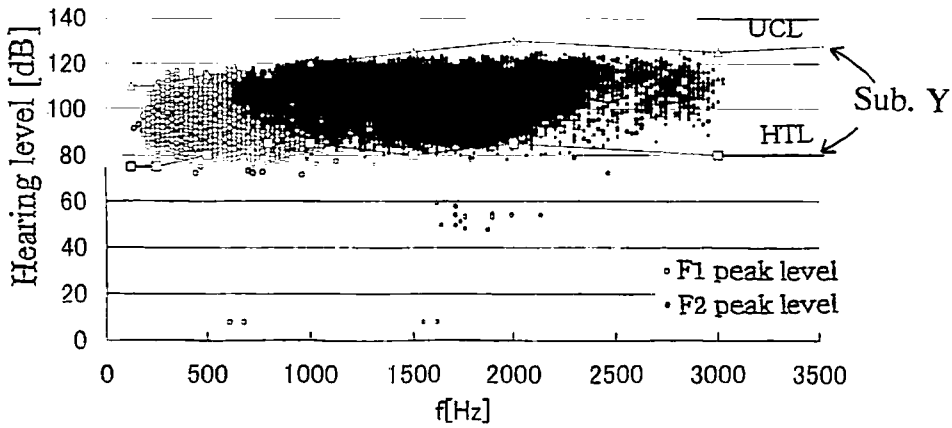
Fig. 6-2 Distributions of formant peak levels before and after the amplitude compression (Sub. N, K).



(d) Compressed for sub. M ; $\sqrt{F1[\%] \times F2[\%]} = 53 [\%]$



(e) Compressed for sub. I ; $\sqrt{F1[\%] \times F2[\%]} = 83 [\%]$



(f) Compressed for sub. Y ; $\sqrt{F1[\%] \times F2[\%]} = 98 [\%]$

Fig. 6-3 Distributions of formant peak levels after the amplitude compression (Sub. M, I, Y).

6-3. 特性の改善による認知能力の向上の推定

圧縮方式の改良による音声明瞭度の改善を推定すれば Fig.6-4 のようになり,各被験者とも聴取スコアの大きな改善が期待される。更に Fig.6-4 に,これら 5 名の改善後の推定値と前節の健聴者による聴取実験の結果(I, Y, N の特性を近似して得られるもの)を記入すれば Fig.6-5 を得る。この場合圧縮特性の下限はマスクノイズの特性(Fig.5-6 の特性 A)に等しく,平均ダイナミックレンジは Fig.5-6 より計算されている。図から,圧縮特性の改善により難聴者の聴取正答率が上昇する結果,模擬難聴状態の健聴者の単語の正答率に接近すると推定される。なお残っている正答率の差が単語の冗長性の利用の巧拙に起因するものと考えられる。上記の結果より平均ダイナミックレンジが狭くとも可聴領域内のホルマントピーク分布割合が増大すれば音声の聴取改善が見込めることが分る。

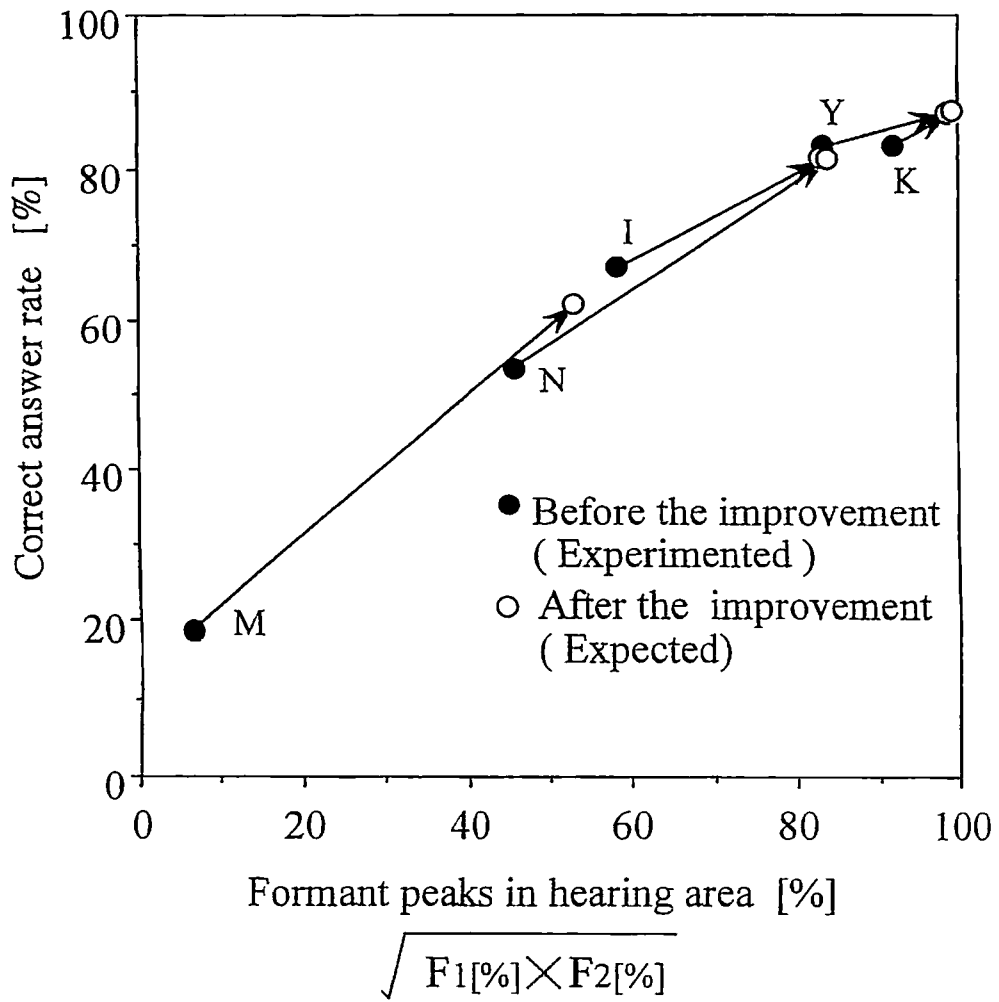


Fig. 6-4 The expected changes in correct answer rates after the improvement.

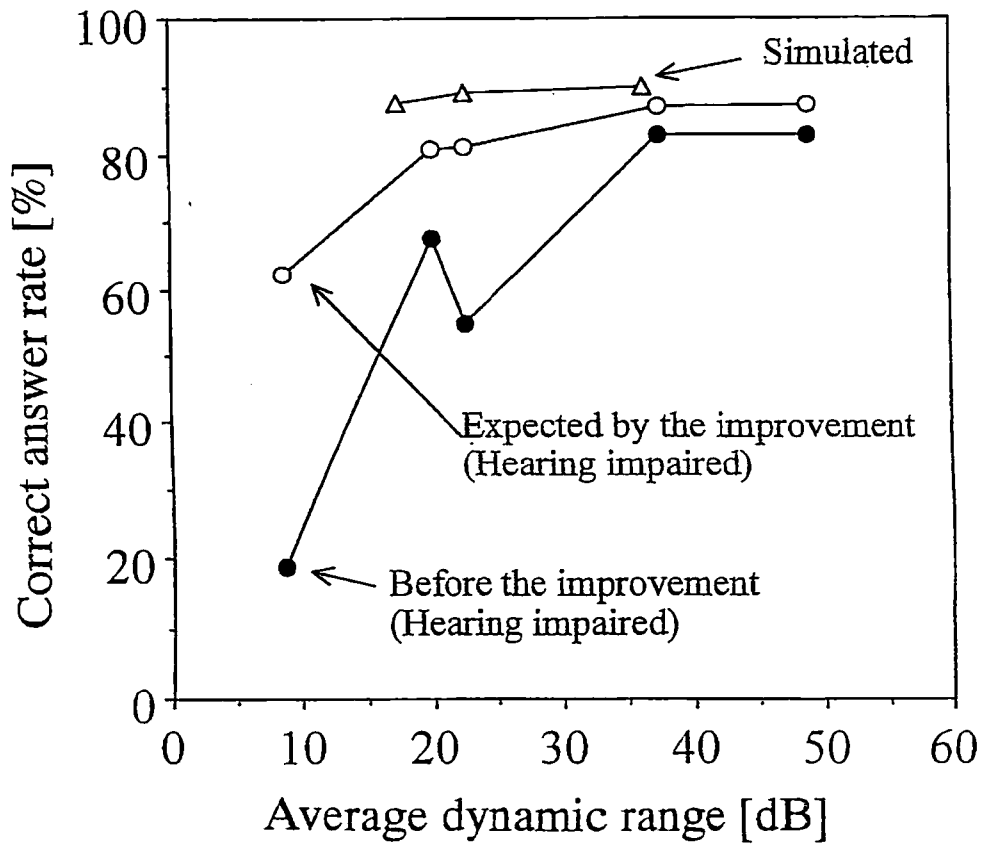


Fig. 6-5 The relations between average dynamic range and correct answer rate after the improvement.

6-4.まとめ

単共振分解型補聴方式での、単共振信号の実効値からホルマントピークレベルを推定し圧縮特性を定めている過程での推定誤差がある場合の感音性難聴の被験者はMCLの調整の際に低めにレベルを設定する可能性があるので、これを軽減するため振幅圧縮特性の改善を図った。

音声の圧縮処理によりホルマントピークがUCLを越えないよう、UCを一様に引き下げるのではなく、ホルマント周波数におけるダイナミックレンジに応じて補正する方式を試みた。その結果、この改良により可聴領域内のホルマントピーク分布率を上げることが可能で、難聴者による音声認知の更に大きな改善が期待できることを示し、本章で用いた補聴処理方式の具体的な改良により達し得る認知能力を算定した。今後、難聴者による補聴器音声の認知実験により実証を行いたい。

第7章 結 言

本研究は、ホルマント情報を伝達する聴覚障害補償システムとして、新しい補聴方式である単共振分解型補聴器を提案し、この補聴信号処理音声を用いて伝送されるホルマント情報と感音性難聴者の音声認知の関連を明かにすることを目的に行われた。

まず第2章では、感音性難聴者のため高品質音声を伝達する補聴方式として、単共振分解型振幅圧縮補聴方式を提案し、具体例として電話帯域音声を対象とした補聴器のシミュレーションを実現しその基本的性能について検証した。本方式は、逆フィルタ制御法により抽出したホルマント周波数に基づいて分離抽出した単共振スペクトルのホルマントピークが可聴領域に入るように振幅圧縮することにより、通常いられるスペクトル分解による振幅圧縮法に比べてスペクトルの谷に相当する部分に歪が生じにくい。また、単共振分解された信号の実効値、ホルマント周波数での圧縮上限・下限から圧縮係数を算定するので、振幅の瞬時圧縮方式で生ずる非線形変換による高調波歪を避けることが可能になる特徴を持つ。信号処理としての有効性と模擬難聴状態の健聴者による音声認知への効果を調べるためまず、感音性難聴者のオーディオグラムを模擬した振幅圧縮条件下で実音声有感音性難聴者の聴力特性を想定して圧縮し、スペクトル分析によって母音部、子音部共に圧縮前に可聴領域に入っていないホルマントピークが、圧縮後は圧縮の上限・下限の間へ入っていることを確認した。次に実音声（VCV 試料）を、マスクノイズにより模擬難聴状態にした健聴者に提示し圧縮処理によって、母音部、子音部ともに聴取実験で音声の認知が改善されることを確かめた。またマスクノイズを重畳しないで圧縮条件を変化させて聴取しても音声の認知に低下傾向はなく圧縮品質も良好であると推定で

きた。更に合成音声による母音カテゴリの判定実験で、模擬難聴下で圧縮しない音声を聴取すると乱れていた母音カテゴリが圧縮処理により回復することを示した。感音性難聴者による実験に先だってこれらの予備実験で単共振分解型圧縮方式の有効性を示した。

次に第3章では、感音性難聴者の音声認知とホルマント情報との関連を明かにする前に、難聴者の音声認知能力は聴能訓練によって改善される可能性を持つことから、音声の画像を併用する聴能訓練の効果と認知能力の上限を求める実験を行った。音声画像により効率化した聴能訓練システムを提案し単共振分解型補聴器を用いて聴取能力の改善を図り、音声認知とホルマント情報の関連について検証した。その結果、音声認知能力の向上が認められる難聴の状態は本研究で新たに定義した平均ダイナミックレンジで 20-30[dB]程度の難聴でありそれより著しく広い可聴領域でも、著しく狭い可聴領域でも効果が認められなかった。

次に、感音性難聴者での実験で推定された音声認知能力と平均ダイナミックレンジには単調増加の関係が存在するが、通常、聴力の基準として用いられる HTL(最小可聴閾値)との関係は明確でないことが分かった。

更に第4章では、このような可聴領域と認知能力との関係は、可聴領域に存在する音声情報との関係に置き換えることが出来ることから、ホルマントピーク分布と音声認知能力の関係を調べた。その結果、難聴者の聴こえについて可聴領域内の第1, 第2ホルマントピーク分布率の幾何平均値により音声認知能力を推定できるという仮説を得た。この仮説を証明するために、第5章で、マスキングノイズで模擬難聴状態とした健聴者により、圧縮状態を変化させた音声の聴取実験を行い感音性難聴者の結果と全く同様な関係の存在を確認した。このことは、可聴領域内ホルマントピーク分布と音声認知能力との定量的な関係のみならず、模擬難聴という実験

形態の正当性をも示す結果である。

最後に、第6章において、本研究の結果に基づいて振幅圧縮処理の改良により可聴領域内のホルマントピーク分布率を上げれば、難聴者による音声認知の更に大きな改善が期待できることを示し、本研究で用いた補聴処理方式の具体的な改良により達し得る認知能力を算定した結果、難聴者の音声認知を更に高める圧縮処理の改善方法を示した。

本研究によって、音声に含まれるホルマント情報は難聴者及び模擬難聴状態の健聴者の音声認知に密接に関連しており、可聴領域内の第1、第2ホルマントピークの分布割合の幾何平均と音声明瞭度には単調増加の傾向が存在することが明らかになった。このことは、今後の補聴器設計、可聴領域内に存在するホルマントピーク分布から聴能訓練の目標値を設定するなど難聴者の音声認知の向上にも活用できる成果である。

今後この知見を具体的に応用した補聴器フィッティングやホルマントピークレベルに対する聴力測定などの分野への展開が期待できる。また、一般の音声情報機器の評価、騒音環境下での音声認知などへの応用も可能と考える。更にマスキングノイズを用いた健聴者による模擬難聴が実際の難聴者の結果と一致することが示された結果今後の実験を効率良く進めることも可能となった。

謝 辞

音声,とりわけ聴覚障害の補償,補助に関する研究とめぐり合う機会を与えていただき,本研究を進めるにあたって終始熱心に情熱を傾けてご指導頂いた熊本大学渡邊亮教授に心からの感謝の意を表します。

本論文をまとめるにあたり懇切なご指導とご助言を頂いた熊本大学園田頼信教授,内村圭一教授に深く感謝致します。

多くの研究討論や御助言と懇切なご指導をいただいた熊本大学上田裕市助教授に厚く御礼申し上げます。研究討論や研究遂行にご協力を頂いた熊本大学坂田聡助手に厚く御礼申し上げます。研究をすすめるにあたりご協力とご支援を頂いた,熊本大学岩田一樹技官,伊元友明技官に心から御礼申し上げます。

被験者のご紹介をはじめ実験にご協力頂いた,熊本県聴覚障害者情報提供センター小野康二氏,熊本県福祉総合相談所黒川一也氏,熊本県難聴者中途失聴者協会会長宮本隆安氏,同事務局宮本せつ子氏,ひばり園園長丸内春美氏,また音声認知実験にご協力頂いた熊本県内在住の聴覚障害者の方々をはじめ多くの皆様に厚く御礼申し上げます。

また,本研究の開始以来,日々ご協力いただいた熊本大学渡邊研究室,上田研究室の卒業研究生諸氏に厚く感謝致します。とりわけ宗研一郎君, Niko Frasherri 君, 田崎晃二君, 宮浦武士君, 藤井雅和君, 山副公成君, 日野光章君, 高山和彦君, 作村剛君, 池田博君, 野下正裕君, 坂本英之君の大学院生諸氏には, 研究の遂行にあたり, 機器の開発, プログラム作成, 各種の音声認知実験及びデータまとめなどで多大のご協力を頂いたことをここに記し厚く感謝致します。

最後に, 研究への理解と協力を惜しまず援助してくれた妻孝美に感謝致します。

参 考 文 献

- [1] 切替一郎, 藤村靖監修, 神山五郎, 戸塚元吉共訳, “話し言葉の科学”, p.4, 東京大学出版会(1966)
- [2] 高橋信雄監修, “音遊びの聴覚学習”, 学苑社, p.2, (1992)
- [3] 神崎仁編集, “聴覚情報処理とその異常”, メジカルビュー社, pp.213-214, (1996)
- [4] 坂部長正, 須藤貢明, 出口利定, “難聴児と聴能およびその補聴”, pp.170-172, 学苑社(1985)
- [5] 今井秀雄, “聴覚活用ハンドブック”, pp.64-67, 財団法人 心身障害児教育財団(平成元年)
- [6] Volker Hohmann and Birger Kollmeier, “The effect of multichannel dynamic compression on speech intelligibility”, J. Acoust. Soc. Am. 97, 1191-1195 (1995)
- [7] H.Levitt, “聴覚障害者の語音弁別能力: スペクトルの考察,” 中西訳: 「補聴器適合の基礎」, 第2章, pp23-27, 医学書院(1986)”
- [8] 浅野 太, 鈴木陽一, 曾根敏夫, 林 哲也, 佐竹充章, 大山健二, 小林俊光, 高坂知節, “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 47, 373-379 (1991)
- [9] Mekata, T., Suzuki, R. and Yoshizumi, Y., “Consonant enhancement algorithm for the hearing-impaired with reduced temporal resolution”, Int. Symp. Speech and Hearing Science, pp.75-85(1991)
- [10] R. W. Guelke, “Consonant burst enhancement : A possible means to improve intelligibility for the hard of hearing”, J. Rehabili. Res. Rev., 24,

pp.217-220 (1987)

[11] P. J. Blamey, G. J. Dooley, J. I. Alcantara, E. S. Gerin and P. M. Seligman, "Formant-based processing for hearing aids", *Speech Communication* 13, pp.453-461, North-Holland(1993)

[12] H.Levitt, "Digital master hearing aid" , *J. Rehabili. Res. Dev*, 23(1), pp.79-87, (1986)

[13] 伊福部達, "音の福祉工学" , pp144, コロナ社(1997)

[14] 小野溢夫, "デジタル補聴器の現状と将来" , *音響学会誌* 47, pp778-784 (1991)

[15] G. Fant, "Acoustic theory of speech production, " ,Mouton & Co. (Hague), pp.48-62 (1970)

[16] 粕谷英樹他, "年令、性別による日本語5母音のピッチ周波数とホルマント周波数の関係," *音響学会誌* vol 24, 6, pp.373-379 (1968)

[17] A. Watanabe, "Formant estimation method using inverse-filter control," *IEEE Trans. Speech and Audio Processing*, vol.9, No.4, pp317-326(2001)

[18]池田 隆, 渡邊 亮, "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).

[19]池田隆,上田裕市, 渡邊亮, "高品質音声を伝達する単共振分解型補聴器", *音響学会誌* 57, pp326-335 (2001)

[20]A. Watanabe, S. Tomishige and M. Nakatake, "Speech visualization by integrating features for the hearing impaired", *IEEE. Trans. Speech and Audio Processing*, vol.8, No. 4, pp.454-466, (2000).

[21] A. WATANABE, "Speech visualization system as a basis for speech training

and communication aids, "Profound Deafness and Speech Communication", (Edit by G. Plant and K. Spens), Whurr Publishers Ltd, pp.528-550, (1995)

[22]渡邊 亮, 濱崎良介, 上田裕市, "電話音声の画像化を目的としたパラメータ抽出・推定法," テレビジョン学会誌 45, 233-243 (1991).

[23]桑原尚夫, 匂坂芳典, 武田一哉, 阿部匡伸, "研究用 ATR 日本語音声データベースの作成," ATR 自動翻訳電話研究所テクニカルレポート TR-I-0086(1986).

[24] D. W. Robinson and R. S. Dadson, " Threshold of hearing and equal-loudness relation for pure tones and loudness function" , J. Acoust. Soc. Am., 29, 1284(1957)

[25] Yuichi Ueda and Akira Watanabe, "Visible/tactile vowel information to be transmitted to the hearing impaired", J. Acoust. Soc. Jpn. (E) 8, 99-108 (1986)

[26] A. Watanabe, Y. Ueda and A. Shigenaga,

"Color display system for connected speech to be used for the hearing impaired" , IEEE Trans. Audio, Speech, Signal Processing, vol. ASSP-33, pp.164-173(1985)