

電子科学研究科

GD

0

0002515500

R

58

静岡大学附属図書館

静岡大学 博士論文

深々度潜水における音声通信と潜水病予防のための  
デジタル信号処理に関する研究



平成6年8月

中井孝芳

# 内容梗概

潜水活動は人類の新たなフロンティアと言ってよい。それを発展させるには安全な潜水活動を支える技術が必要であり、そのために電子技術の一層の役割が期待されている。

集積回路技術の進展に伴い、高速で高機能なデジタル IC が開発されるようになった。本論文ではこれらの IC を用いて、潜水活動を行うのに最も重要と考えられる情報通信と潜水病対策に用いられる二つの支援システムを開発した。前者は音声通信に必要なヘリウム音声の修復システムであり、後者は潜水浮上時に潜水土の血液中に発生する気泡の自動検出システムである。

まず、ヘリウム音声の修復について述べる。深度数十メートル以深の潜水では潜水病を避けるため、ヘリウムを多量に含んだヘリウム空気を用いる。この環境で発声された音声は「ヘリウム音声」と呼ばれ、了解度が低く、そのままでは音声による通信は殆どできない。ヘリウム音声の了解度が低い原因は大別して二つあり、それはマイクロホン等の音響機器と音声自身である。

本研究では、まず、数種類のマイクロホンを高圧タンクで実測した結果から B&K 製の 1/2 インチコンデンサマイクロホン 4133 と松下通工製のエレクトレット型マイクロホン MW063 の周波数特性がヘリウム空気中で 10kHz まで平坦であることを示した。

これらを使用してヘリウム音声を収録した。

この音声を分析し、ヘリウム音声は通常空気中の音声に比べ、ホルマント周波数が非線形に上昇する、母音に比べ子音の音圧レベルが低化する、ピッチは高くなるが高々30%である、ことを示した。

この結果より、実時間向きのホルマント周波数を非線形に変換できる線形予測法 (LP 法) を用いた新しいアルゴリズムによる修復法を提案した。これは、LP 法により声道の極周波数を抽出し、この特性を LP 法の一つである LSP 係数を用いて非線形に変換し、この変換された係数と残差波形を用いて再合成する方式である。開発当時 (1979 年)、使用可能な高速で高機能の IC は高速積和器 (TRW 製 1010J) 等であり、これらを用いて実時間ヘリウム音声修復装置を試作した。本装置では演算速度を能力限界まで高くするため固定小数点方式を採用した。演算誤差等の評価を行うため浮動小数点演算と比較および評価を行い、装置の各部の演算語長を決定した。各部はパイプライン方式で結合され、音声が入力されてから変換され出力されるまでに 50ms の遅れがある。IC は約 800 個を要し、幅 51cm、奥行き 40cm、高さ 35cm の箱に収納されている。

この装置について市販の修復装置 (アナログ電子回路によるピッチ同期型波形伸長方式) とともに実際のヘリウム音声を用いて性能の評価をした。また、修復に用いる変換方式と実時間動作のデジタル装置としての設計製作上の問題点、およびヘリウム音声の物

理的性質などについて検討した。その結果、深度 300 メートルの音声は、原音の明瞭度が 10%に対し、修復音では 70%に向上し、了解度、声質とも実用上十分に修復できることが示された。

次に、この結果を踏まえ、固定小数点演算の高速の DSP(富士通製 MB8764、第 2 世代 DSP) を用いて改良システムを試作した。すなわち、エラー処理を強化して突然の雑音による発振等はなく実用機として十分であることが示された。システムは全体として 5 つの処理ユニットで構成され、パイプライン方式で処理される。DSP は 5 個、IC 総数約 200 個で、A3 の大きさ 1 枚の基板に納められている。消費電力は約 40W、試作 1 号機の約 1/4 の大きさである。

本論文で取り組んだ二つの課題のうち第二の課題は潜水浮上時における血液中の気泡検出である。潜水土は急速に浮上すると体内に気泡が発生し潜水病の原因となる。従って、血液中の気泡が少ない段階で検出することは安全な潜水活動のためには極めて重要である。身体を循環した血液は必ず肺動脈を通るので、従来法は肺動脈に超音波を照射しドップラの原理による変調信号(ドップラ信号)を聴覚的に判断して気泡音を検出する。この信号には鼓動音等の気泡音以外の雑音が混入しているため専門家でも判定は容易ではない。本研究では、ドップラ信号中の気泡音の音響的性質を明らかにし、それに基づいて気泡音を検出するアルゴリズムを提案した。まず、鼓動音を軽減するため線形予測法により残差信号を求める(線形予測法によって周期的な鼓動音は予測できるので除去し、突発的に発生する気泡音はほぼそのまま残る)。次に、この残差信号より振幅と零交差数、周波数分析を求め気泡音を検出した。この結果、Spencer の 5 つの段階区分のうち気泡のない段階 0 から段階 3 までの区分を正しく判定できた。気泡が少ない場合の判定は専門家でも難しく、本装置の性能は潜水病の予防という観点からは十分であるといえる。また、海洋科学技術センターでの有人潜水模擬実験における潜水土のドップラ信号では専門家と同一の段階区分で、正しく判定できた。

さらに、鼓動音区間の設定等を自動化して DSP を用いた準実時間の気泡検出装置を試作した。設定を自動化しても、結果は上述のプログラムの判定と一致した。

また、アルゴリズムを改良し、鼓動区間を含めて分析した方がよいことを実際のデータで示した。段階 3 と 4 の判定は気泡音の周波数 1500 から 2000Hz にある個数で判定すればよいことをデモテープのデータで示した。

本論文では安全な潜水に不可欠な二つの支援システムの研究の成果を述べた。どちらも、デジタル信号処理の標準的な技法の一つである線形予測法を用いているが、ヘリウム音声の修復では、声道の伝達特性の抽出とその変換に、気泡検出では、気泡の検出には雑音である鼓動音の軽減に用いた。また、実時間あるいは準実時間で動作するデジタル電子装置を開発した。

# 目次

<b>1</b>	<b>序論</b>	<b>1</b>
1.1	ヘリウム音声の修復	3
1.2	潜水浮上時の潜水士の血液中に発生する気泡検出	10
<b>2</b>	<b>ヘリウム音声のデータの収集</b>	<b>12</b>
2.1	はじめに	12
2.2	ヘリウム音声のデータ資料	12
2.3	マイクロホンの高圧環境における特性	14
2.3.1	測定方法について	14
2.3.2	測定システム	18
2.3.3	測定結果と検討	18
2.3.4	録音に使用したマイクロホンの特性	22
2.4	まとめ	24
<b>3</b>	<b>ヘリウム音声の性質</b>	<b>29</b>
3.1	はじめに	29
3.2	ピッチ周波数	29
3.3	発話速度	31
3.4	ホルマント周波数	31
3.5	母音と子音の相対的な音圧強度について	33
3.6	まとめ	40
<b>4</b>	<b>ヘリウム音声修復装置 (1号機)</b>	<b>41</b>
4.1	はじめに	41
4.2	修復のアルゴリズムと誤差	42
4.2.1	アルゴリズム	42
4.2.2	変換誤差	42
4.3	システムの設計	46
4.4	前処理部	46
4.5	分析部	47
4.5.1	アルゴリズムとシミュレーション	47
4.5.2	分析部のハードウェア	49



4.6	LSP 係数変換部 . . . . .	50
4.7	ダウンサンプリング部と LSP 合成部 . . . . .	55
4.8	回路構成とヘリウム音声修復装置の外観 . . . . .	55
4.9	まとめ . . . . .	55
<b>5</b>	<b>ヘリウム音声修復装置 (1 号機) の評価</b>	<b>59</b>
5.1	まえがき . . . . .	59
5.2	聴取試験 . . . . .	59
5.2.1	音声資料 . . . . .	59
5.2.2	聴取試験の結果の概要 . . . . .	60
5.3	聴取試験による修復装置の評価 . . . . .	63
5.3.1	方式による明瞭度の違い . . . . .	63
5.3.2	ヘリウム音声の子音部の脱落 . . . . .	63
5.4	装置化上の問題点 . . . . .	67
5.4.1	非線形変換の実行上の問題点 (声道最低共振周波数 $F'_{wa}$ について) . . . . .	67
5.4.2	修復音の音質 (個人性の再現性) . . . . .	69
5.5	まとめ . . . . .	71
<b>6</b>	<b>DSP を用いたヘリウム音声修復装置 (2 号機)</b>	<b>72</b>
6.1	まえがき . . . . .	72
6.2	LSP 変換方法の改良 . . . . .	72
6.3	その他の改良点 . . . . .	74
6.3.1	前処理部 . . . . .	77
6.3.2	分析部 . . . . .	77
6.3.3	逆フィルタリングとダウンサンプリング部 . . . . .	79
6.3.4	LSP 係数変換部 . . . . .	79
6.3.5	LSP 合成部 . . . . .	82
6.4	回路構成 . . . . .	82
6.5	ヘリウム音声の修復の結果と考察 . . . . .	82
6.6	まとめ . . . . .	88
<b>7</b>	<b>線形予測分析法を用いた潜水浮上時における血液中の気泡検出</b>	<b>90</b>
7.1	はじめに . . . . .	90
7.2	超音波ドップラ法 . . . . .	91
7.3	ドップラ信号の資料 . . . . .	91
7.4	ドップラ信号の特徴 . . . . .	91
7.5	気泡音検出システム . . . . .	92
7.5.1	アルゴリズムの概要 . . . . .	92
7.5.2	線形予測による鼓動音成分の影響の軽減 . . . . .	92
7.5.3	鼓動区間の除去 . . . . .	97
7.5.4	振幅、零交差数と線形予測による気泡音検出 . . . . .	100
7.5.5	重み係数と閾値の決定方法 . . . . .	101

7.5.6	アルゴリズムの使用方法	102
7.6	結果と考察	102
7.6.1	デモテープによる気泡検出の実験	102
7.6.2	有人潜水模擬実験における気泡検出実験	103
7.6.3	気泡数の増大とドップラ信号のスペクトルピーク周波数の低下に関する考察	103
7.7	まとめ	105
<b>8</b>	<b>DSP による気泡検出とアルゴリズムの改良</b>	<b>107</b>
8.1	はじめに	107
8.2	DSP を用いた気泡検出システム	107
8.2.1	DSP	107
8.2.2	システム構成	107
8.3	気泡検出アルゴリズムと DSP での計算法	108
8.4	処理時間と結果	108
8.5	アルゴリズムの改良	111
8.6	ドップラ音の資料	111
8.7	改良アルゴリズムによる結果	116
8.8	まとめ	116
<b>9</b>	<b>結論</b>	<b>120</b>
<b>A</b>	<b>ヘリウム音声収録用テキスト</b>	<b>126</b>
<b>B</b>	<b>音速と密度の求め方および声道の最低共振周波数の推定</b>	<b>132</b>
<b>C</b>	<b>ヘリウム音声修復装置のアルゴリズムとハードウェアの詳細</b>	<b>134</b>
<b>D</b>	<b>日本語ヘリウム音声データベース</b>	<b>146</b>
	<b>参考文献</b>	<b>155</b>

# 第 1 章

## 序論

海洋は地球の約 3 分の 2 という広大な面積を占める。しかし、長い間、その表面を利用する航海や、ごく浅い範囲を利用する漁業などに限られていた。近年、深度 300 メートルまでの大陸棚には石油などの資源があり、また、貝の採取等の漁業や娯楽として手軽に潜水できるスクーバ・ダイビングが開発され、潜水が盛んに行われるようになった。

潜水には大きく分けて二つのタイプがある。

一つは、「しんかい 2000」や「しんかい 6500」のように潜水艇に人間が乗り、外気圧に関係なく人間は通常の 1 気圧のまま潜る方法 (大気圧潜水) である。このようにすれば深海の観察やマニピュレータによる採取などが可能であるが、外界とは隔離されるから制約も大きい。

もう一つは人間が外界にさらされて潜る方法 (環境圧潜水) である。古来から行われてきた海女に代表される素潜りはこの方法の一つである。素潜りでは通常深度は 5~10 メートルまでで、息こらえは約 1 分であるから作業できる範囲は非常に狭い。海洋開発の場合にはこれでは制約が大き過ぎることから、空気を圧搾したポンペを背負って潜る方法や海上から圧搾空気をホースなどで送る方法がとられた。この方法はすでに 18 世紀にはヨーロッパで行われているが、窒素過多による窒素麻酔、高酸素による中毒や潜水病 (潜水作業が終了した後地上に戻るまでの間に起こる減圧障害) という生死にかかわる問題がある。窒素麻痺と酸素中毒により圧搾空気のポンペを背負う自給気式 (スクーバ) 潜水器ではその限界は 70 メートルといわれている [1]。また、潜水病を起こさないで潜水できる時間は数十分程度である。

これ以上の深さまたは長時間の潜水では飽和潜水と呼ばれる方法がとられている。飽和潜水は血液に環境気体を完全にとけ込ませた状態で潜水する方法である。特に、深々度では非常に大がかりなシステム潜水と呼ばれる方法がとられている。これは 1965 年ころアメリカ海軍が行った大がかりな潜水装備をほどこした加圧式の水中エレベータにより本格化し、潜水時間と潜水深度は増大した。また、生理学的障害を避けるためヘリウムを主体とした人工空気 (以下、「ヘリウム空気」という) が用いられるようになり、潜水深度は 300 メートル以深となった。日本では海洋科学技術センター (神奈川県横須賀市夏島) により昭和 46 年 (1971 年) から開始された。これが深度 100 メートルまでの飽和潜水技術を用いたシートピア実験計画で、昭和 50 年度 (1975 年) まで続いた。続いて昭和 51 年

より大陸棚全域を対象とする深度 300 メートル海域の開発に必要な潜水作業技術の研究開発が進められた。現在までに実際に深度 300 メートルで海中作業実験も行われている [2]。

このように 300 メートルという深々度でも潜水実験が行われるようになったが、安全に潜水するためには問題点もいくつか残されている。これには、高圧ヘリウムによる生理学的な影響、音声通信、潜水後の時間当たりの減圧量、住居環境としての高圧チャンバー内の騒音等が含まれる。この論文では潜水活動を支援するために最も重要と思われる二つの装置の開発について述べる。

第 1 の課題は音声通信に必要なヘリウム音声の修復である。潜水して作業を行うには潜水土同士や海上(地上)との命令伝達や作業確認等の通信が必須である。音声は人間にとって簡便で、手足や目の自由を妨げることなく作業をし続けることができる。また、深々度では太陽光は到達せず、ライトをつけても海底では人間の移動により堆積物が舞い上がり極端に視界が悪くなるのでテレビ等での監視も不十分である。よって、音声による通信が作業の安全性を高めるとともに効率の上からも望ましい。しかし、深度数十メートル以深では上述したように生命維持のためヘリウム空気を用いるので、気体の音速と密度が通常空気と異なり音声がひずむ。この音声は「ヘリウム音声」と呼ばれ、了解性が悪く、通信は困難である。したがって、この音声を了解できる音声に実時間で変換(この変換を「修復」という)できれば、音声による通信は可能になる。また、長時間の潜水の場合、ヘリウム音声では潜水土同士の会話ができないので、潜水土にとって精神的な大きなストレスになる。このストレスを軽減する意味でもヘリウム音声の修復は重要である。この修復装置の開発を第 1 の目標とする。

第 2 の課題は潜水土の血液中に発生する気泡検出である。潜水が深度 300 メートルまで可能になったとは言え、減圧過程では潜水病 [3] の起こる可能性があり、完全には解決されていない。実際の潜水においては、減圧過程で、潜水土の自覚症状および潜水土の血液中の気泡検出によって時間当たりどの程度の減圧ができるかを実験的および経験的に決めている。海洋科学技術センターでの 300 メートル有人潜水模擬実験では深度にもよるがおよそ 1.8m/h で減圧する。就寝時(8 時間)と昼の 2 時間は圧力を変えないので、減圧過程だけで約 12 日かかることになる。したがって、安全でしかも効率よく潜水するには、如何に潜水病を起こさずに減圧するかにかかっている。潜水病を未然に防ぐ方法として後述するように減圧過程で潜水土の血液中の気泡を音としてモニターする方法がある。現在は人間がこの音を聞いて判断している。しかし、専門家でも個人差があり、また、潜水病を起こさない段階での気泡のモニターは集中力がいて長時間は大変である。そこで、この気泡音を自動的に検出し、それにより診断するシステムの開発が必要である。これを第 2 の目標とする。潜水病は素潜り以外、つまりスクーバ・ダイビング等でも発生するのでこの装置の利用範囲は広い。

これら二つの課題に共通することは上述したように潜水活動のための支援システムである。ヘリウム音声では情報通信手段として実時間で修復する必要がある、気泡検出では生命に関係するのでできるだけ早い診断が必要で即時性が要求される。また、両者とも人間が聞くことができる「音」である。つまり、ヘリウム音声の修復では声という「音」であり、血液中の気泡検出では後述するように超音波を照射し、血液中の移動している気泡からの反射または散乱によるドップラシフトしたものをヘテロダイン検波することにより

得られる「音」である。よって、一次元のデジタル信号処理による技法で解決可能なものといえる。

ヘリウム音声では通常音声と同様に線形予測法が適応できるので、如何に実時間で声道の特性を非線形に、安定で、誤差が少なく変換できるかが鍵となる。本論文では新しい実時間向きの変換アルゴリズムを提案し、その後にシステムの設計と製作およびその評価について述べる。

一方、気泡音検出では、鼓動音が雑音となる。これを如何に軽減するかが最大の問題点であり、この軽減に線形予測法を適応するとよいことを示した後にその検出法について述べる。

以下に、現在までに行われているヘリウム音声の修復法と超音波ドップラ法による気泡検出について述べる。

## 1.1 ヘリウム音声の修復

まず、音声の生成とモデルについて述べる。音声の発声器官の正中断面の概略図を図 1.1(a) に示す。声帯の周期的な振動によりできる空気のかたまりの列または声道の一部をせばめることによって発生する乱流雑音が音源であり、声道は音響管で音源の共鳴器となる。声道の形と音源の位置により、種々の音が発声できる。母音は /a/、/i/ などのように通常声帯を振動させ声道は一定の形に保ったまま発声される。有声子音は /ga/、/ba/、/za/ の /g/、/d/、/z/ のように声帯振動のほか声道の一部のせばめや閉鎖を伴い、通常最初にバズバー (閉鎖やせばめがあるときに声帯振動により声道壁等から放射される 200Hz 程度の低い音) があり破裂や摩擦の後母音を伴う。無声子音は /ka/、/pa/、/sa/ の /k/、/p/、/s/ のように声帯を振動させず声道の一部をせばめたり閉鎖させて、摩擦性または息を瞬時に流出させて破裂性の音を出す。鼻子音は /ma/、/na/ の /m/、/n/ のように声道の一部を閉鎖させ、軟口蓋をあげ声道が鼻腔に枝分かれした状態で声帯を振動させて発声する。ただし、通常の会話のように連続的に音声を発声するときは前後の母音や子音による相互作用等でその音自身に変形される。

図 1.1(b) は音声の簡単な生成モデルであり、音源 (ピッチまたは雑音) を時間的に変化するフィルタ (声道が変化することに対応) に入力することに相当する。音源と時変フィルタとの間には実際には相互作用があるが、簡単なモデルではこの相互作用がないとして取り扱う。

ヘリウム音声では発声する環境は空気ではなくヘリウム空気となる。ヘリウム音声のひずむ理由は大別して二つある。一つはマイクロホン等の音響機器にあり、もう一つは音声自身である。どちらも音速が速くなることと媒質の密度が高くなることによる。マイクロホンについては第 2 章で述べるのでここでは取り上げない。ヘリウム音声のひずみは、主に声道の伝達特性による。これは、図 1.1(b) で時変フィルタが変化したことに対応する。この変化は、音速が速くなり声道が見かけ上短くなったことによる線形的なホルマント周波数の上昇と、密度の上昇により声道壁が剛体とはみなされず、特に低いホルマントでの周波数の上昇である [5]。

このヘリウム音声を修復する試みは多数ある。

1964年 Holywellらのテープの低速再生による方法 [6] が始まりである。これは実時間処理が不可能なので、以後実時間で可能な種々の方法が開発された。これらの手法については文献 [7] に詳述されているので、その表に以後公表されたものを併せて一つの表にしたものが表 1.1である。以下に主な修復法についてその特徴を述べる。

1. 低速再生方式 [6] テープの再生時、録音時よりも低速で再生することによって得られる。ホルマント周波数は線形的に下降するが、ピッチ周波数も低下する。音声はわかりやすくなるが、実時間では不可能である。
2. ピッチ同期型の波形伸長法 [8] 有声音のようにピッチがある場合はピッチに同期して、ない場合は適当な間隔で波形の一部を引き延ばして切り張りする。その方法の一例を図 1.2に示す。ホルマントは線形的に下降する。ピッチ周波数は変えられるが、通常は変更しない。欠点はホルマントが線形的にしか降下しないことと、無声子音の破裂部が、ピッチがないので適当な間隔で間引かれ、脱落する可能性があることである。この脱落は降下させる割合が大きいほど大きくなる。また、ピッチ抽出が必要なので、抽出誤りを起こすと音声が悪化する。
3. SPAC-H[7] 波形ではなく自己相関波形を伸長する方式である。雑音に強いのが特徴である。図 1.3に方法を示す。ホルマントは線形的にしか降下できないが、ピッチは変更可能である。深度 96m のヘリウム音声を修復し明瞭度 44%である。ただし、報告でも指摘しているが、マイクロホンの特性に問題がある。
4. 線形予測後に系のインパルス応答を求め時間軸でそれを伸長する方式 [9] 図 1.4に示すように線形予測法で声道の伝達特性を推定の後、インパルス応答に変換する。このインパルス応答を直線内そうし、インパルス応答を変換する。この変換されたインパルス応答から伝達特性を求め、残差信号を入力として合成する。ホルマントは線形にのみ降下できる。ピッチは変更できない。

以上は線形変換のみが可能である。

以後の方法は原理的に周波数領域で非線形変換可能な方法である。これらの手法は図 1.1(b) に示したように音声を音源と声道とに分け、それらを分離し、主に声道特性を変換し、再合成する方法である。

5. チャンネル・ボコーダ (Channel Vocoder) を応用した方法 [10],[11] 図 1.5に示すように多数の帯域フィルタ (声道特性の推定) を用意し、その振幅を検出する。この値を周波数を降下させたい周波数でそれぞれチョップし帯域フィルタで整形しそれを足し合わせる。チョップする周波数は入力をスペクトル平坦化した帯域フィルタで得る。アナログ技術でもデジタル技術でも装置化は可能であり、ホルマント周波数の非線形変換が可能である。ピッチ周波数は変わらない。ただし、帯域フィルタの通過帯域での遅れがいずれも等しくないとはっきりしなくなり、音質の劣化をまねく。しかし、通過帯域での遅れを等しくするのは難しい。
6. FFT を用いてその包絡から声道特性を求める方法 [12] 声道の周波数特性を短時間 FFT (高速フーリエ変換) 包絡で求め、この包絡を 50 の折れ線で近似する。FFT

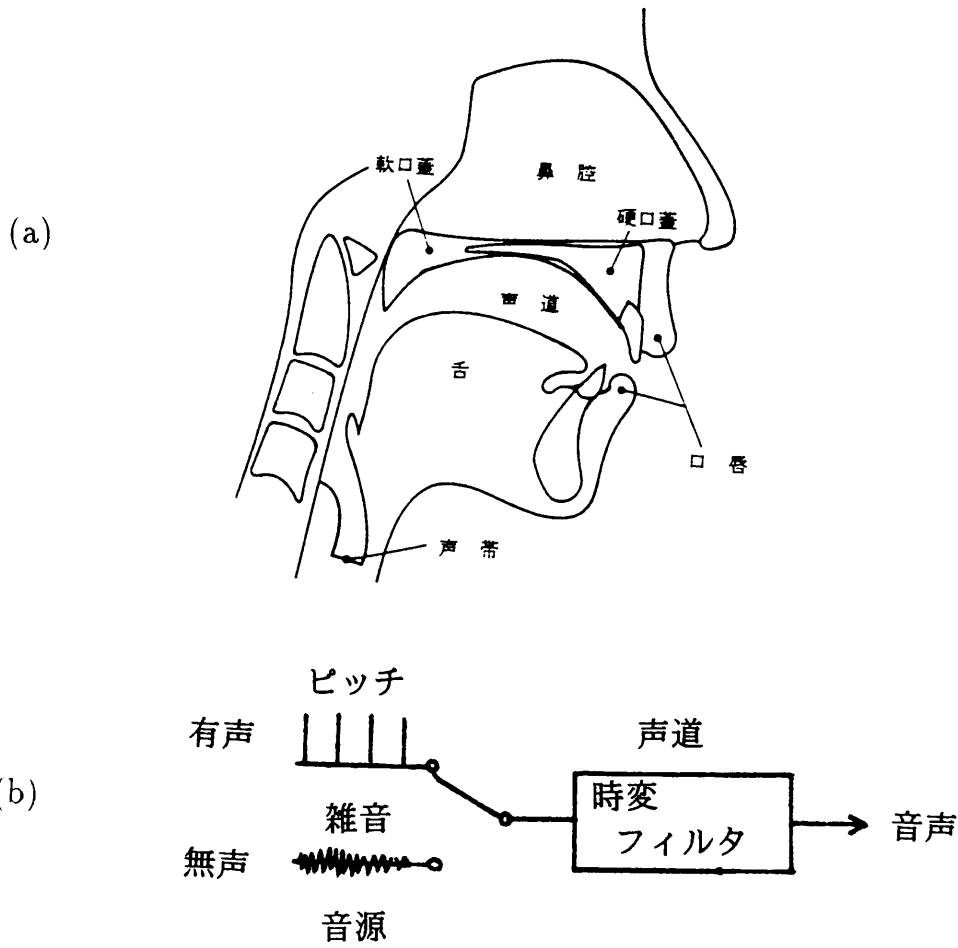


図 1.1: (a) 発声気管の概略図 (文献 [4] より引用) (b) 音声の最も簡単な生成モデル

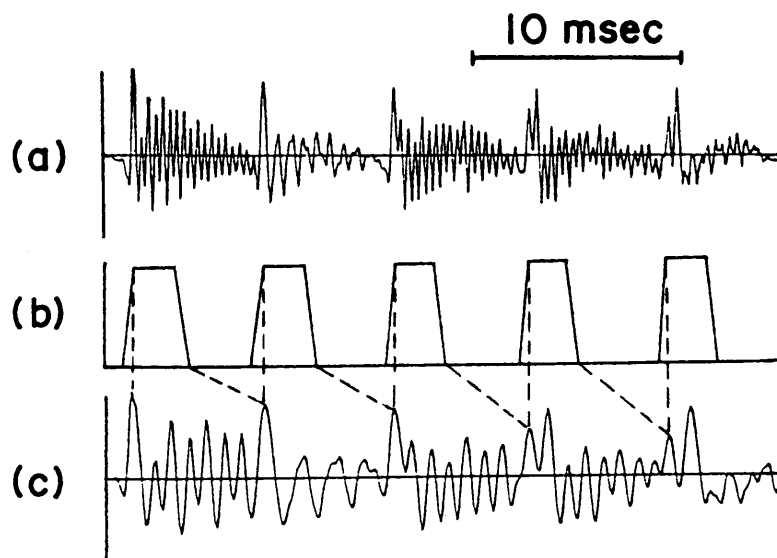
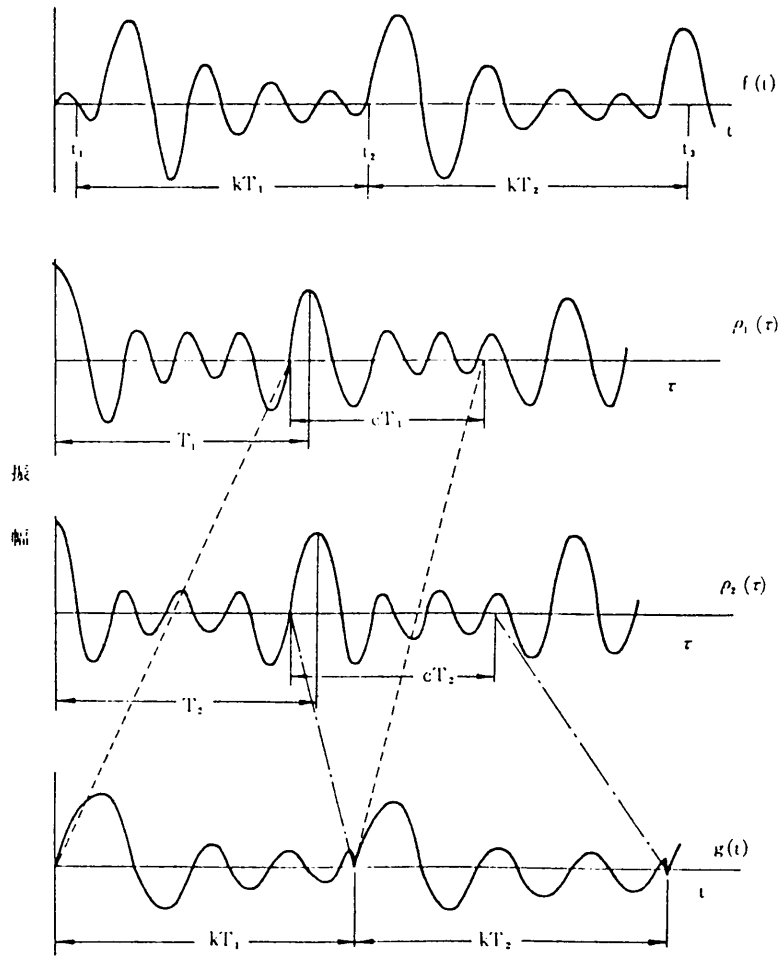


図 1.2: ピッチ同期波形伸長方式の一例 上から (a) 深度 300m で発声された原音/a/, (b) 波形を抽出する窓、(c) 修復音

表 1.1: ヘリウム音声の了解性改善に関する諸報告のあらまし (文献 [7] より引用し、以後の報告を追加)

報告者 (年)	方式	実験資料	圧縮比、移動周波数	了解度の改善
Holywell ら [6] (1964)	低速再生	実験室	0.625, 減速再生	87.5 → 92.5
		潜水中	0.625, 減速再生	3.6 → 36.0
Copel[14] (1966)	ヘテロダインシフト	Sealab I	可変	20 → 90
Golden[10] (1966)	チャンネルボコーダ	Sealab II	1/1.5, 1/1.83	良好
Stover[8] (1967)	ピッチ同期波形伸長	実験室	後部 1/4 を除き伸長	大気中母音と差がない
Flower[15] (1969)	ピッチ同期波形伸長	高圧タンク (27atm)	後半を除き伸長	音素の正当率 88%に
Stewart[16] (1969)	ヘテロダインシフト		可変 (約 750Hz)	40 → 60
Roworth[11] (1970)	チャンネルボコーダ	高圧タンク (11atm, 93%He)	1/2.2	43 → 86 文中 Key Word
Quick,Jr.[17] (1970)	波形領域分析合成	高圧タンク (27atm, 96%He)	音速比及び、非直線的	比較的的良好
Helmut[18] (1971)	ピッチ同期波形伸長		1/2	自然性回復
鈴木誠史ら [7](1975)	SPAC-H	高圧タンク (10.6atm, 91.3%He)	1/2.31	CV 音節明瞭度 26.1 → 44.2
鈴木久喜ら [9] (1975)	線形予測インパルス応答	深度 450feet	1/2	明瞭度試験なし
鈴木久喜ら [13] (1978)	線形予測声道形状変換			理論のみ
Richards[12] (1982)	FFT 包絡変換合成	深度 560feet, 1000feet	線形、非線形	了解度試験 70%





$f(t), g(t)$  : 入力, 出力  
 $\rho_1(\tau), \rho_2(\tau)$  : 短時間自己相関関数

図 1.3: SPAC-H の原理 文献 [7] より引用

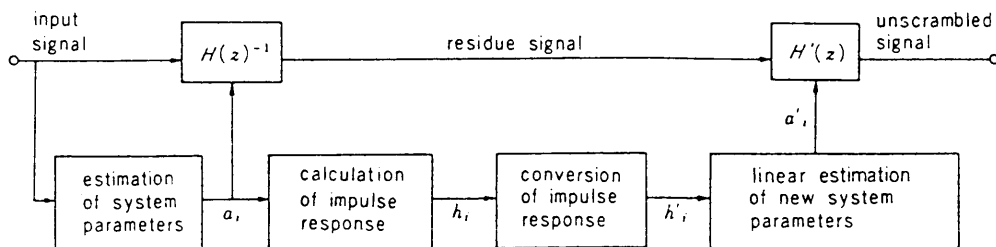


図 1.4: 線形予測後に系のインパルス応答を求め時間軸でそれを伸長する方式のブロック図 文献 [9] より引用

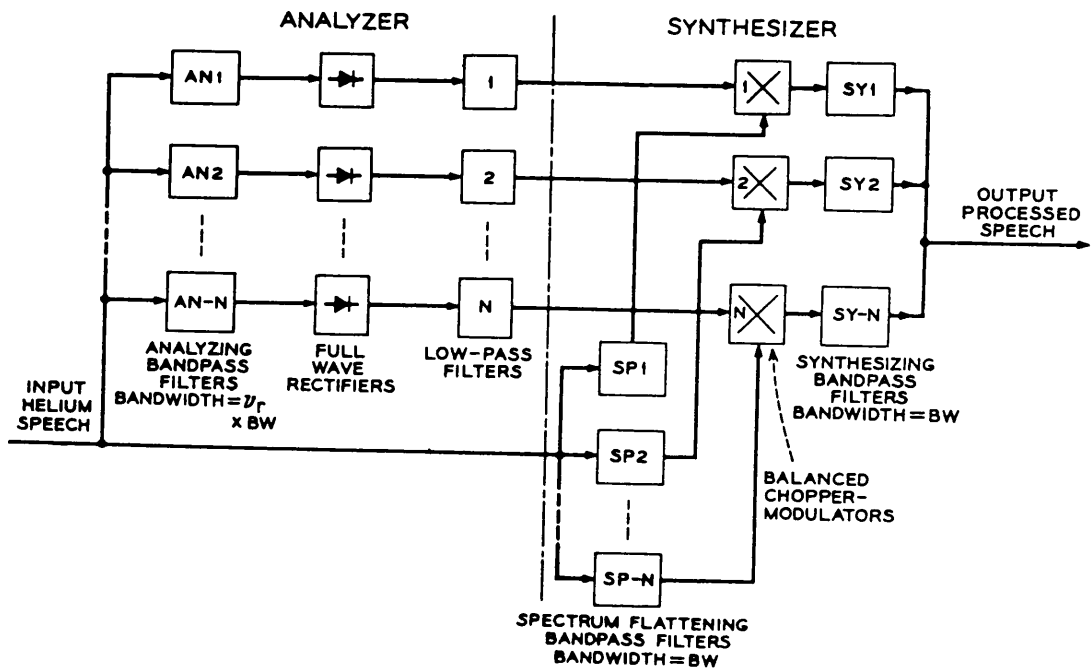


図 1.5: チャンネル・ボコーダを応用したヘリウム修復のブロック図 文献 [10] より引用

の各々の値を対応する 50 の折れ線で近似した包絡の値で割る (音源の平坦化に当たる)。50 の折れ線で近似した包絡を線形または非線形に変換する。周波数ごとにこれら二つを掛け合わせ、その IDFT(逆離散的フーリエ変換)により波形を得る。周波数変換は線形、非線形とも行える。ピッチ周波数は変更できない。ヘリウム音声を修復した結果では、四つの単語 (例えば "sud"、"sum"、"sub"、"sun" のように一部の発音だけが異なる) から選択する試験を行い、深度 1000 フィート (約 300m) における修復音の了解度は 70%程度である。

7. LPC 係数から推定した声道形状を変換する方式 [13] LPC で求めた PARCOR 係数から声道形を求め、これを変形してホルマント周波数を変換する。音源は残差を用い、PARCOR 係数で合成する。線形変換できる。非線形変換は声道形により変換式が同じにならず、非線形性が大きいと困難である。

以上、ヘリウム音声の修復についての現在までの手法について述べた。文献 [12] 以外はヘリウム音声の発声された深度は浅い。また、これらの報告で現在実時間で動作可能な装置となっているものはピッチ同期型の波形伸長法、チャンネル・ボコーダ方式によるものである。これらの方式はアナログ技術を使用しており、デジタル技術を使用するにはいたっていない。

1980 年以降デジタル集積回路技術は格段に進歩し高速に演算可能な乗算器や積和器、ビットスライス・マイクロプロセッサ、DSP(デジタル信号用プロセッサ)等が次々に開発された。そこで、これまで装置が大規模で開発が困難であると考えられていた全デジタル方式のものを開発することとし、また、深々度のヘリウム音声のひずみを忠実に変換して修復できるような声道の周波数特性を非線形変換できるアルゴリズムの開発に目標をおいた。声道の周波数特性を非線形変換するためには声道の特性を知らなければならない。これには線形予測法を用いた。また、音源としては、声道の特性を抽出した後の残差波形を用いた。これは、ピッチ周波数は変更できないが、ピッチや雑音源の情報は残差波形に残されているからである。また、ピッチ抽出はそれだけでも一つの研究テーマであり、雑音の多いヘリウム音声では音圧レベルの低いときにピッチ抽出を誤る可能性が大きく、複雑な処理の割合には効果が期待されないからである。よって、本方式では音源と声道を分離して、声道特性を抽出した後、声道特性を非線形に変換し、ピッチを変更しないで再合成する方式を採用した。特に、ここで提案する方式は、声道特性を非線形に変換するアルゴリズムが今までの方式と異なる。

本論文では、まず、修復の評価に使用するヘリウム音声の資料とそれを収録するためのマイクロホンの特性について第 2 章で述べる。次に、この資料を用いて通常空気中で発声した音声とヘリウム音声を比較、検討した。この結果を第 3 章で述べる。これらの特性を踏まえて、開発当時入手可能な高速積和器とビットスライス・マイクロプロセッサを用いたヘリウム音声修復装置修復装置 (1 号機) の原理とその設計方針と製作について第 4 章で述べる。これは、ビット数の制約を受けた固定小数点演算で、如何に浮動小数点演算の結果に近づけるかという計算機シミュレーションによる検討を含む。この実時間システムの評価を第 5 章で述べる。その後、16 ビット固定小数点演算型の DSP が開発されたので、

修復アルゴリズムをさらに改良し、2号機を試作した。第6章ではこの改良と試作について述べる。

## 1.2 潜水浮上時の潜水士の血液中に発生する気泡検出

大気圧潜水と素潜り以外では、ダイバーは圧搾空気あるいは  $\text{He-O}_2$ ,  $\text{He-N}_2\text{-O}_2$  などの高圧混合ガスを呼吸する。吸入した気体のうち、窒素やヘリウムのように生体に不活性なガスは徐々に体内組織に溶け込んで行く。浮上によって環境圧力が下がると溶解していた気体は溶けなくなるので、いくつかの経路で体外に排泄される。この排泄速度が速ければ問題はないが、急激な浮上をすると、排泄が間に合わなくなり気泡が発生する。体内に残った気泡は小血管の血栓を起こしたり神経を圧迫したりして、さまざまな機能障害を引き起こす。これらを総称して減圧症、一般には潜水病と呼ぶ。減圧症の予防および治療には、浮上時の気泡発生を即刻知って浮上を一時停止するとか、浮上後の残存気泡を検出して適当な治療を施さなければならない。それには即時的で高感度の気泡検出技術が必要である。

ただし、急性の生命に関わるような潜水病は非常に短時間で起き、また、昏睡状態に陥るので、気泡検出を行うまでもなく再加圧を行う必要がある。ここで取り上げる気泡検出は、それよりも緩慢ではあるが浮上した後1日程度で進行する亜急性の潜水病や、短期間では比較的症状が軽いが長期間の気泡の存在により発生する骨壊死 [3] の予防や診断に用いるものである。

気泡を検出するには一般に超音波ドップラ法と呼ばれる方法が用いられる。これはドップラ・フロー・メータと呼ばれる超音波装置を使用して体外から超音波を照射し、血流中の気泡に当たって反射してくる超音波のドップラ変調音を聴覚的に判断して気泡を検出する。超音波照射の目標位置は肺動脈である。肺動脈は肺でガス交換される直前の血液が全て通り、体表に比較的近いからである。超音波の周波数としては  $5\sim 10\text{MHz}$  が用いられる。ドップラ・フロー・メータは元来は流体の流速測定用の装置だが、Gillis や Spencer が血流中の気泡検出に応用できることを示した [19],[20],[21]。肺動脈に超音波が照射されると、血液や気泡に当たり反射または散乱する。肺動脈中では血液も気泡も速度を持っているので超音波が反射または散乱するときドップラ・シフトする。この音をヘテロダイン検波すると、聴覚で聞くことができる周波数となる(以下、「ドップラ信号」という)。血液に比べ気泡からは反射または散乱が多く、また速度が早いので、鞭をふったような音または小鳥が鳴くような音がする。これを人間が聞いて検出する。ただし、血液からの反射または散乱があり、鼓動によって大きく変化するので、これが気泡検出の雑音となり、特に気泡が少ないときは判定が難しくなる。

Spencer は気泡の検出頻度によって次の5つの段階に区分している [21]。

段階 0：気泡が存在しない状態

段階 1：鼓動 4~5 回に対し 1~2 の割合で孤立した気泡が検出される状態

段階 2：鼓動 1~2 回に対し気泡が検出され、それらの半分以下には気泡の集団が検出される状態

段階 3：各鼓動間に孤立した気泡、あるいは集団の気泡が検出されるが、  
鼓動音はまだ聞きとれるような状態

段階 4：鼓動音をおおい隠すような多数の気泡が検知される状態

段階 2 以上では一般に自覚症状があり、減圧症を予防するには段階 1 で検出する必要がある。しかしこのドップラ信号は鼓動音が大きく、段階 1 では気泡は小さく数も少ないので気泡音を聞き分けるのは熟練者でも容易でなく、判定は最低でも 1 分間以上 (血液が全身をめぐる戻って来るまでの時間) 聞かなければならないので神経を集中させるのが大変である。また、しばしは熟練者同士でも判定が異なる。そこで客観的でしかも段階の低いときにも判定可能な自動診断装置の開発が必要である。気泡検出を自動化する試みは現在まであまりない。Belcher は 1980 年にコンピュータを応用してこの信号を分析し、ミニコン PDP-11/40 による気泡検出を試みた [22]。そのシステムではアナログ回路の帯域フィルタバンクでドップラ信号をスペクトル分析し、鼓動音を軽減することを行っているが、ドップラ信号はやぎとひつじのデータのみである。

本論文では、減圧症を予防するためにデジタル信号処理 (主な処理は線形予測法) により自覚症状の現れない段階 1 でも気泡を検出できる高感度で高速のシステムを開発することを目的とし、その音響分析とそれに基づいて提案したアルゴリズムの検討について第 7 章で述べる。また、第 8 章ではハードウェアとして実現する部分とその後のアルゴリズムの改良点について述べる。

## 第 2 章

# ヘリウム音声のデータの収集

### 2.1 はじめに

深度 300 メートルにおよぶ高圧環境下での本格的な潜水実験は、海洋科学技術センターで 1980 年から行われるようになった。ヘリウム音声の収録はすべて海洋科学技術センター内の有人潜水シミュレータ用のチャンバで行われた。図 2.1 に示すようにこのチャンバは直径 2.3 メートル、長さが 7 メートルの円筒形で、両端に潜水用のプールのあるウエット・チャンバと球形のサブ・チャンバにつながっている。ダイバーは 4 名で、300 メートルの実験ではおよそ一ヶ月このチャンバ内で暮らすことになる。ここでは、このチャンバで録音したヘリウム音声のデータ資料と、その録音の際に使用したマイクロホンの測定方法と特性について述べる。

### 2.2 ヘリウム音声のデータ資料

音声データは聴取試験のできるように、「明瞭度試験用平等率音節表」、「2 音節ワード・リスト」および「3 音節のワード・リスト」、「文章了解度試験用単文表」を文献 [23] からのものを一部変更をして使用した。また、のちに「101 音節表」、「明瞭度試験用平等率音節表」にあらたに作成したものと「まぎらわしい音節および単語」を追加した。また、発声できるピッチの高さを調べるため、音階を発声することも行った。表 2.1 に、発声したテキストの種類を示す。なお、詳しい発声用テキストの内容は、付録を参照されたい。

次に各年における収録方法を述べる。収録年度は 1980 年、1982 年、1984 年、1987 年、1991 年である。

1980 年は初めての収録であり、ダイナミック型マイクロホンを用いた。テープレコーダはソニー製 TC-D15M、メタルテープを使用した。音声資料は 50 音節表、単語及びアクセント、文章である。録音状態はクリップしているものが多く使用できるものは少なく、ピッチの測定や発話速度の測定のみ取り上げた。また、クリップしていないものでも周波数特性はその後の測定と比べるとあるべきホルマントがないなどマイクロホンの特性に問題があるので、修復等には用いていない。

ここで、1980 年以外のダイバーについて出身地とともに表 2.2 に示す。各年とも成人

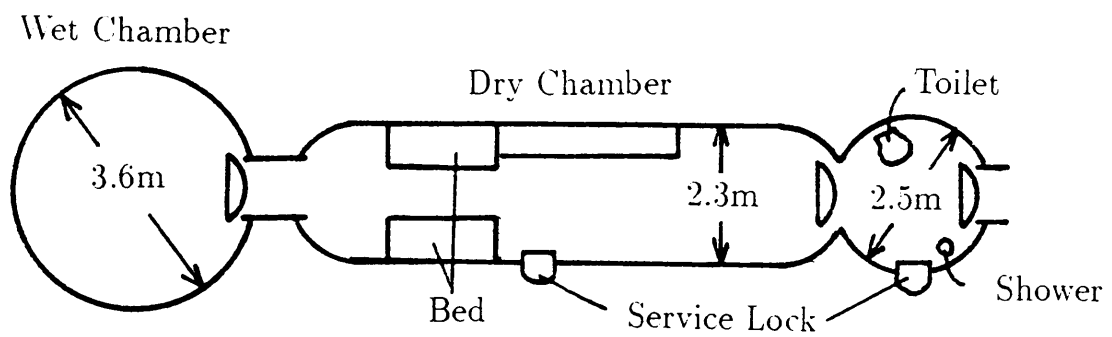


図 2.1: 有人潜水シミュレータのチャンバの概略図 (使用部分のみ記述)

男子各 4 名である。

1982 年は 11 月 22 日から 12 月 12 日まで収録した。テープレコーダはソニー製 TC-D15M、メタルテープを使用し、録音はチャンバ内で、VU メータがおよそ -10dB になるように発声するように指示した。また、ドライチャンバーの両側には寝具用の毛布をつるし反響を防いだ。また、発声位置も反響をできるだけ避けるため、ベッドのある位置で行った。マイクロホンは 3 種類で表 2.3 に示す通りである。ヘリウム音声の修復装置の評価に用いたデータは主にこの年のデータである。録音テープは周波数特性のよいメタルテープを使用した。

次に、1984 年の 9 月 29 日から 10 月 25 日までの内 10 日間収録したデータについて述べる。マイクロホンは 1982 年で使用したのと同じエレクトレット型コンデンサマイクロホンとコンデンサマイクロホンを使用した。コンデンサマイクロホンの収録の時ベッドに三脚を置き、それにマイクロホンを固定したこと以外の収録条件は同じである。また、ヘリウム空気中で発声できるピッチ周波数の範囲が変化するか調べた。マイクロホンは MW063 と B&K4133 を使用した。録音テープはクロムテープを用いた。

1987 年にはマイクロホンはアコー 7017 を用いたが、周波数特性は 10kHz 以上まで平坦であったが、高圧では収録途中で突然収録不能となることがあり、高圧では使用できないことがわかった。録音テープはメタルテープを用いた。

1991 年は 10 月 10 日から 10 月 30 日まで収録した。マイクロホンは MW063 を使用した。録音はデジタルオーディオテープデッキ (DAT デッキ) を用いた。

## 2.3 マイクロホンの高圧環境における特性

### 2.3.1 測定方法について

この節での目的は高圧環境でマイクロホンの感度と周波数特性を測定することであるが、このような特殊な測定は既知の確立された方法がないので、いくつかの異なる測定方法を用いてそれぞれの結果をよく吟味する必要がある。常識的に可能性のある方法として、次の三つを考えてみる。

- (i) 無響室内で、感度既知のマイクロホンと比較して特性を求める。(置換法)
- (ii) 音響カプラを用いる相互校正法。
- (iii) 静電加振器 (Electrostatic Actuator) を用いる方法

(i) は、常識的で融通性のある方法だが、既設の無響室を高圧ガスで充満することは不可能であり、高圧環境において特性の既知なマイクロホンがまだ存在しないという理由から現在は実施できない。

(ii) は、相互校正法であるから絶対感度を測定でき、またカプラとマイクロホンはそれほど大きな物でないのでこれらを高圧タンクに入れて測定できる点で現実性がある。ところが、通常のカプラでは、3 個のマイクロホン A, B, C に対して AB, AC, BC の 3 対について電圧 - 音響 - 電圧の伝送量を測定する必要があるため、同一環境下でマイクロ



表 2.1: 各年度における発声テキストの種類

- 1982 年度
  - ・ 2 音節ワードリスト
  - ・ 3 音節ワードリスト 3-A
  - ・ 明瞭度試験用平等率音節表 B-3
  - ・ 文章了解度試験用短文表 1、2
  
- 1984 年度
  - ・ ピッチの測定
  - ・ 長母音
  - ・ サ行音・ガ行音・ガ行
  - ・ 3 音節ワードリスト 3-A
  - ・ 明瞭度試験用音節表
  - ・ 文章了解度試験用短文表
  
- 1987 年度
  - ・ 101 音節表
  - ・ まぎらわしい音節
  - ・ 2 音節ワードリスト
  - ・ 3 音節ワードリスト 3-A
  - ・ 明瞭度試験用音節表 N-1, N-1b, B-3a, B-3b
  - ・ 文章了解度試験用短文表 1、2
  
- 1991 年度
  - ・ 101 音節表
  - ・ まぎらわしい音節
  - ・ 2 音節ワードリスト
  - ・ 3 音節ワードリスト 3-A
  - ・ 明瞭度試験用音節表 N-3b
  - ・ 文章了解度試験用単文表 1

表 2.2: 各年度の実験におけるダイバーと出身地

ただし、話者 C と話者 E、話者 K と話者 P、話者 L と話者 O は同一話者

実験年	ダイバーと出身地
1982	A : 神奈川県 B : 愛媛県 C* : 滋賀県 D : 山口県
1984	E* : 滋賀県 F : 香川県 G : 岩手県 H : 大阪府
1987	I : 和歌山県 J : 岩手県 K* : 岩手県 L* : 岩手県
1991	M : 岩手県 N : 岩手県 O* : 岩手県 P* : 岩手県

表 2.3: 各年度に実験で使用したマイクロホン

実験年	マイクロホンの型番
1982	エレクトレット型コンデンサマイクロホン (松下通信工業 WM063) ダイナミック型マイクロホン (ソニー F560) 1/2 インチ・コンデンサマイクロホン (ブリュエル・ケア 4133)
1984	エレクトレット型コンデンサマイクロホン (松下通信工業 WM063) 1/2 インチ・コンデンサマイクロホン (ブリュエル・ケア 4133)
1987	エレクトレット型コンデンサマイクロホン (松下通信工業 WM063) 1/4 インチ・コンデンサマイクロホン (ブリュエル・ケア 4133)
1991	エレクトレット型コンデンサマイクロホン (松下通信工業 WM063)

ホン対を交換しなければならず、その際必ず高圧タンクを開閉するのでガスを損失する。従って経済性、環境の一定性、および時間がかかり過ぎる問題がある。もしもカブラ内に音源を組み込んでしまえば、その音源とマイクロホン A, B の間の感度比、感度積を測定することにより 2 つのマイクロホン A, B の感度を求めることができる。このようにカブラ内に音源を組み込んだものを使えば、高圧タンクの開閉は設置する時だけで良いのでガスの損失の問題はなくなる。そこで、電子総研で試作したカブラ (アクティブカブラ) を使用した。

アクティブカブラを用いる相互校正法ではマイクロホン A と B の感度  $M_A$  と  $M_B$  は、

$P_A$  : カブラ内音源 (チタバリ振動子) からマイクロホン A への電圧減衰量

$P_B$  : カブラ内音源からマイクロホン B への電圧減衰量

$P_{AB}$  : マイクロホン A からマイクロホン B への電圧減衰量

とすれば、

$$M_A = \frac{1}{2}(-P_A + P_B - P_{AB}) - 10 \log_{10} \frac{C_A \gamma P}{V} \quad (2.1)$$

$$M_B = \frac{1}{2}(P_A - P_B - P_{AB}) - 10 \log_{10} \frac{C_A \gamma P}{V} \quad (2.2)$$

$C_A$  : マイクロホン A の静電容量

$\gamma$  : 環境気体の比熱比

$P$  : 環境気体の全圧

$V$  : アクティブカブラの全容積

で与えられる。各伝送量は精密抵抗減衰器を用いた置換法によりそれぞれのマイクロホンの出力電圧から求める。ただし、測定された感度には一定の補正 [24](カブラ内波動補正と振動膜インピーダンス補正) を加える。

(iii) は、コンデンサマイクロホンの振動板を静電的に振動させるもので [25]、感度の絶対値を求めることはできず、B&K 製のマイクロホン用のものしか製造されていないので測定対象のマイクロホンが限定されるが、構造が簡単であり、高圧タンク内での測定が可能なので、(ii) の方法と合わせて実施した。静電加振器による測定の原理は次の通りである。加振器のグリッドへの印加電圧を  $e_g$  とすればコンデンサマイクロホンの振動膜に加わる圧力  $P_d$  は、

$$P_d = \epsilon e_g^2 / (2d^2) \quad (2.3)$$

$\epsilon$  : 環境気体の誘電率

$d$  : グリッドと振動膜との等価距離

となる。従って、 $e_g = E_0 + e_0 \cos \omega t$  ( $E_0 \gg e_0$ ) のとき、

$$P_d \simeq \frac{\epsilon}{2d^2} \left[ \left( E_0^2 + \frac{e_0^2}{2} \right) + 2e_0 E_0 \cos \omega t \right] \quad (2.4)$$

となり、振動膜を微小信号  $e_0 \cos \omega t$  で振動させることができる。従って、角周波数  $\omega$  の信号入力に対するマイクロホン出力を測定すれば、その角周波数に対するマイクロホンの感度が求められる。

### 2.3.2 測定システム

図 2.2 に測定システムのシステム図を示す。高圧チャンバは、海洋科学技術センター (横須賀) の動物実験用潜水シミュレータを用いた。測定は、アクティブカプラと静電加振器にマイクロホンを組み合わせた 2 系統の装置を高圧チャンバに入れ、プログラマブル発振器から信号を入力して、マイクロホン出力をトラッキングフィルタ (NF 社、TF425) を通してデジタル電圧計で読む。このフィルタは雑音除去のため、帯域幅 5Hz で中心周波数は測定周波数に連動する。マイクロホンの組み合わせの切り替え、発振周波数の設定、電圧読み取り、記録、繰り返し測定、監視などはマイコン制御の下に行なった。2 系統の切り替えは手動で行なった。コンデンサマイクロホンの成極電圧は、アクティブカプラ法の場合は安全のため通常の 2 分の 1 の 100V とし静電加振器法の場合は 100V と 200V の二通りとした。また、静電加振器へは直流バイアス 500V に測定すべき周波数の微小信号を重ねた電圧を印加した。成極電圧は、2 分の 1 に下れば周波数によらず感度が 6dB 下がるだけであり、静電加振器への直流バイアスについても微小振幅励振では同様である [25]。

環境条件は潜水シミュレータにより厳密に調節した。温度は測定時  $23.5 \pm 0.5 \text{ }^\circ\text{C}$ 、加減圧速度は 10atm/時だが、その時も温度変化は  $\pm 3 \text{ }^\circ\text{C}$  に抑えた。湿度は 2% 以下である。測定は 10 気圧の圧力変化ごとに定圧に保って、定常状態になってから行ない、100Hz ~ 30kHz で 30 ~ 40 点の周波数で、各 3 ~ 5 回繰り返した。測定時間は 1 つの圧力で 40 ~ 60 分であった。

### 2.3.3 測定結果と検討

アクティブカプラ法と静電加振器法によって得られた高圧環境下でのコンデンサマイクロホンの周波数特性を、図 2.3 と図 2.4 に示す。気体の組成は表 2.4 に示したものである。表 2.4(b) の組成の時、組成はほぼヘリウムなので、以下「ヘリウム」という。アクティブカプラ法による測定値にはカプラ内波動補正量及びマイクロホンの振動膜インピーダンス補正量 [24] を加算してある。静電加振器法による測定値には回折効果による感度上昇分を加算する必要があるが、「ヘリウム」中では 10kHz まで回折効果による影響は少ない (10kHz で 2dB 程度)[28]。この特性は要約すると以下のようになる。

- (a) 圧力上昇に伴い全体的な感度の低下が見られる。
- (b) 空気中でのアクティブカプラ法による測定では、13 ~ 30kHz にカプラ内の空洞の共振によるピークが現われている。「ヘリウム」気体中では、音速が空気より速いのでこれに対応するピークは本測定の帯域外となる。

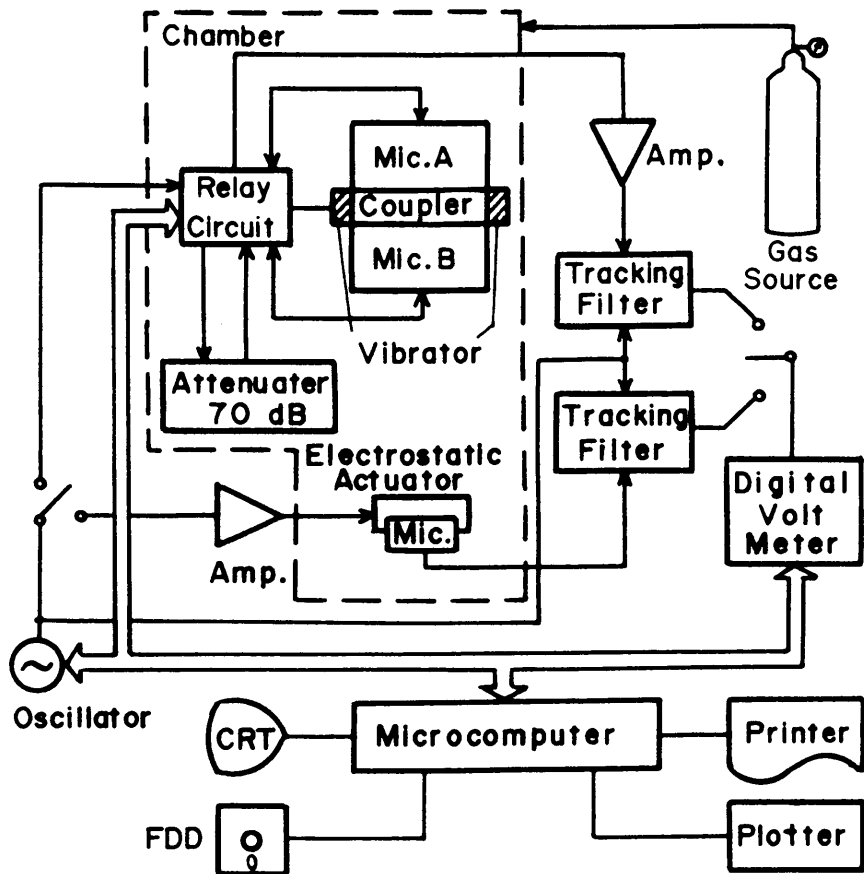


図 2.2: コンデンサマイクロホンの測定システムのブロック図

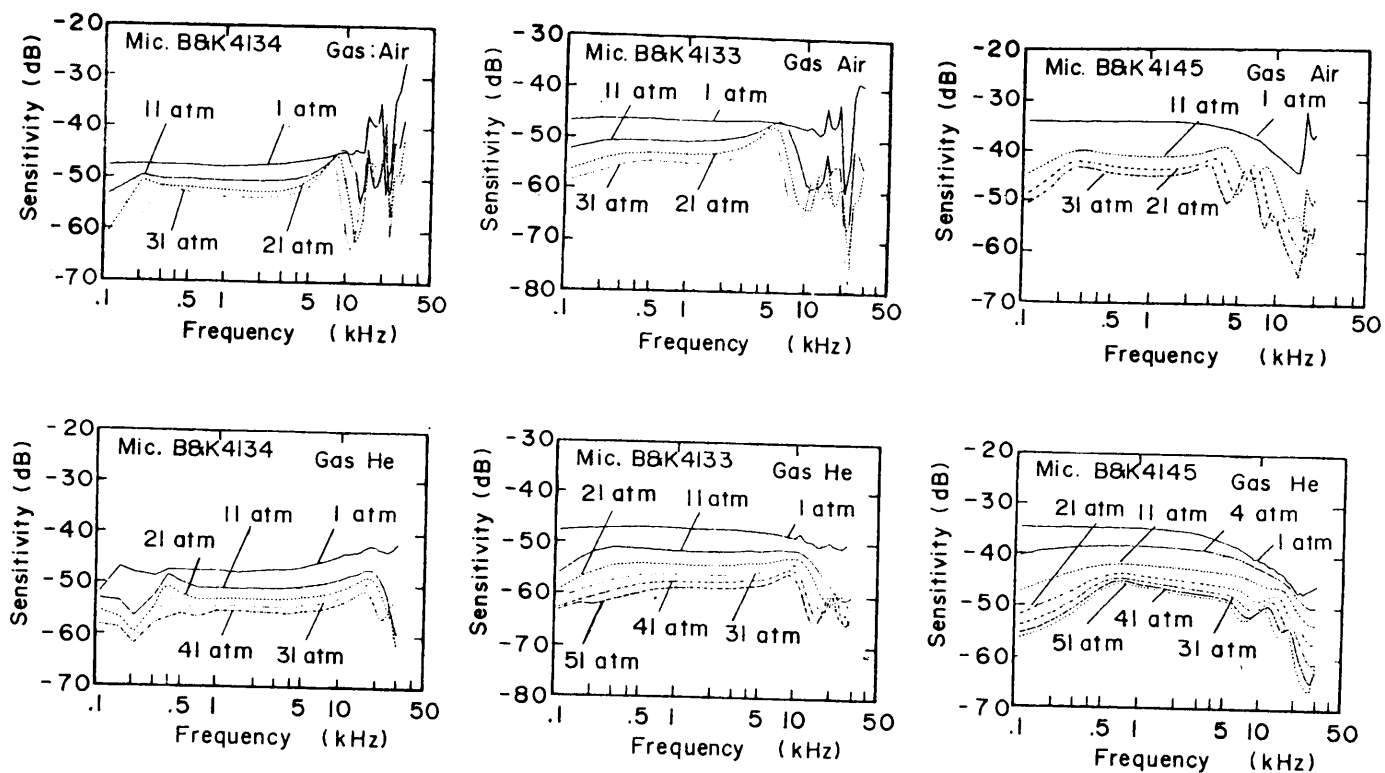


図 2.3: 種々の環境圧力でのアクティブキャパ法によるマイクロホンの感度

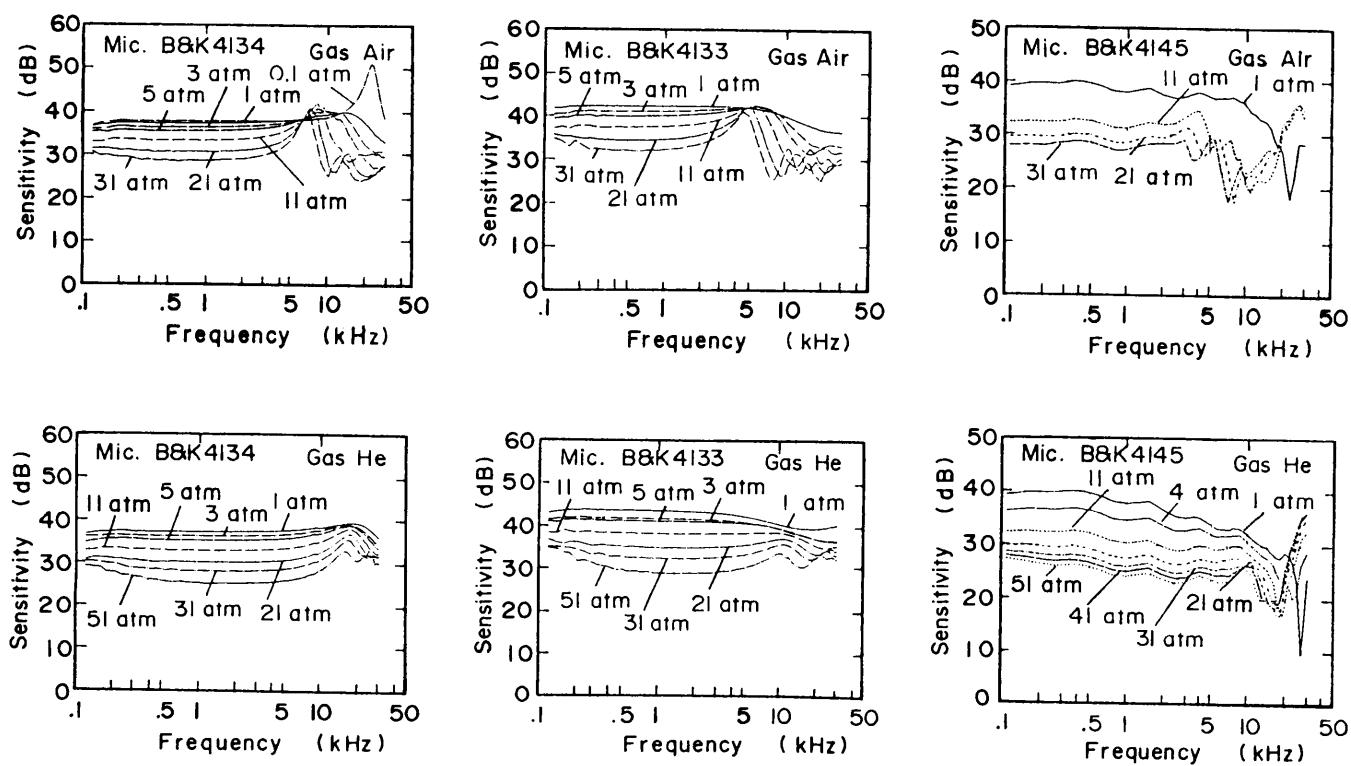


図 2.4: 種々の環境圧力での静電加振器法によるマイクロホンの感度

表 2.4: 各圧力での物理定数 (a) 空気 (N<sub>2</sub>:78.0%, O<sub>2</sub>:21.0%, Ar:1.0%)、 (b) ヘリウム  
 空気混合気体 (N<sub>2</sub>:1.5%, O<sub>2</sub>:0.4%, He:98.1%) ただし、1 atm = 1013 hPa

(a)

圧力 (atm)	密度 (kg/m <sup>3</sup> )	音速 (m/s)	粘性係数 (kg/ms)
1	1.19	346	$1.80 \times 10^{-5}$
11	13.1	346	$1.82 \times 10^{-5}$
21	24.9	346	$1.83 \times 10^{-5}$
31	36.8	346	$1.85 \times 10^{-5}$

(b)

圧力 (atm)	密度 (kg/m <sup>3</sup> )	音速 (m/s)	粘性係数 (kg/ms)
1	0.183	956	$1.98 \times 10^{-5}$
4	0.734	956	$1.98 \times 10^{-5}$
11	2.02	956	$1.98 \times 10^{-5}$
21	3.85	956	$1.98 \times 10^{-5}$
31	5.69	956	$1.98 \times 10^{-5}$
41	7.52	956	$1.98 \times 10^{-5}$
51	9.36	956	$1.98 \times 10^{-5}$

- (c) 高圧になると広域に複数のピークが生じ、圧力上昇に伴い低域に移動する。この減少は「ヘリウム」気体中よりも空気中の場合の方が顕著である。
- (d) アクティブカプラ法による測定では、約 500Hz 以下の低域において感度の低下が見られる。

この論文での主旨はヘリウム音声の修復であるので、このような特性になる理由については光ヘテロダイン干渉法によるマイクロホンの膜の振動の観測 [29] や文献 [30] と [31] の計算機シミュレーションの結果にゆずるが、(a) の感度が低下する特徴は全ての場合に見られ、(c) の原因はマイクロホンの背極にある穴の影響で膜が複雑な振動を起こすためである。また、(d) の現象は静電加振器法では現われないので、マイクロホン自体ではなくキャピラリチューブの影響等も含む測定法に原因があると考えられる。

### 2.3.4 録音に使用したマイクロホンの特性

ヘリウム音声の収録時に主に使用した2本のマイクロホンの特性を示しておく。図 2.5 のエレクトレット型コンデンサマイクロホン WM063 の特性は空気 1 気圧にヘリウム 50 気圧を加えた後、減圧中に測定したものである (表 2.4(b) と同じ組成)。簡易無響箱で既知のコンデンサマイクロホン (B&K4134) を基準として求めたものである ([27] より引用)。また、図 2.6 のコンデンサマイクロホンの特性は空気 1 気圧にヘリウムを加えた場合、静電加振器法 [32] で求めた。この環境は潜水時の環境に非常に近い。どちらも、100Hz から 10kHz までほぼ平坦な特性をしていてヘリウム音声の特性を論ずるときにマイクロホンの周波数特性を特別考慮する必要はないと思われる。以下、コンデンサマイクロホンとエレクトレット型コンデンサマイクロホンのどちらで収録したヘリウム音声も周波数特性に補正は加えていない。

上記で述べた環境はコンデンサマイクロホンのみが実際の潜水環境に近いのでその他のマイクロホンについて、実際音声を収録して検討した。1982 年に行われた模擬実験の際に使用した3種類のマイクロホンで収録した母音のスペクトルや単語の平均スペクトルを求めたのが、図 2.7 と図 2.8 である。窓長 30ms、ハミング窓を掛けた後、高速フーリエ変換 (FFT) により分析したものである。単語の平均スペクトルを求める場合は分析フレームを 10ms とし、おのおのの周波数毎に絶対値の平均を求めた。発音は毎回微妙に変わるが、マイクロホンの大まかな特性はこの方法でも知ることができる。図 2.7 は深度 300 メートルで「しれる」と発声した時の /i/ であるが、(a) のエレクトレット型コンデンサマイクロホンと (c) のコンデンサマイクロホンは高域を除きほぼ同じような特性をしている。(a) の 11kHz 付近にみられるピークは図 2.5 で示した特性のピークに対応するものである。(a) と (c) に比べると (b) の特性は大きく変わっている。まず、200Hz 付近に大きなピークがある。また、3.5kHz、6.2kHz、7.2kHz、10.8kHz 付近に谷がある。ただし、このデータは符号+11 ビットで A-D 変換しているからダイナミックレンジは約 66dB であり、この谷はもっと深い可能性がある。よって、このダイナミックマイクロホンは高圧では特性に大きな谷ができ、音声収録用には不向きであることがわかる。また、図 2.9 に



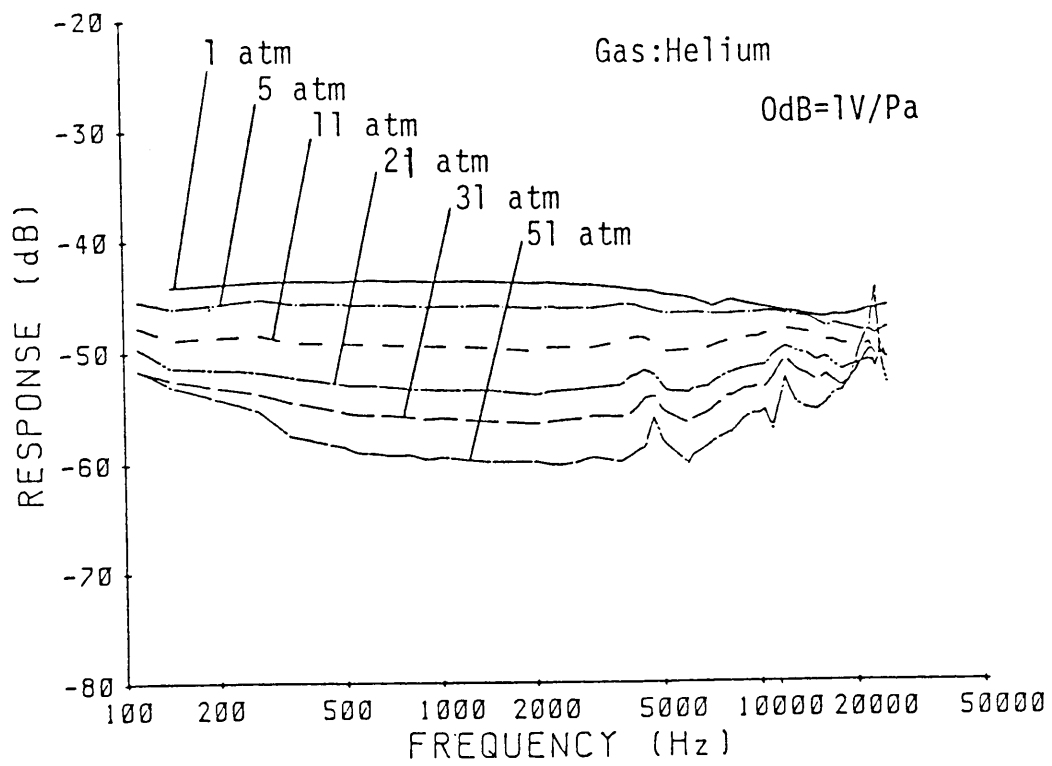


図 2.5: エレクトレット型コンデンサマイクロホン WM063 のヘリウムにおける周波数特性 1atm 増加すると深度は 10m 深くなる

示すように深度 145 メートルでも、ダイナミック型マイクロホンで収録したものには深度 300 メートルほどではないが 200Hz 付近にピークがあり、3kHz 付近に谷があるので音声収録には向かないことがわかる。

## 2.4 まとめ

以後のヘリウム音声の修復の際に用いる音声リストについてその収録環境とあわせて説明した。また、マイクロホンの高圧環境における特性を測定し、エレクトレット型コンデンサーマイクロホンまたは、コンデンサーマイクロホンが音声収録用には適していることを示した。これらのマイクロホンでヘリウム音声を収録したが、特性の良いマイクロホンで深度 300 メートルまでの日本語のヘリウム音声を収録したものは、世界的に見てもこれらデータのみである。

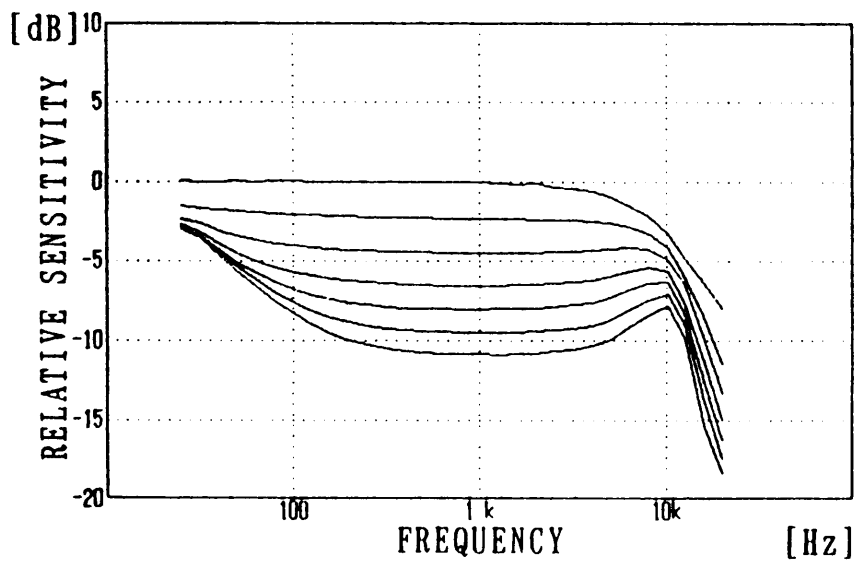


図 2.6: コンデンサマイクロホン B&K4133 のヘリウム空気における周波数特性  
上から通常空気、深度 40m、90m、140m、190m、240m、290m 相当

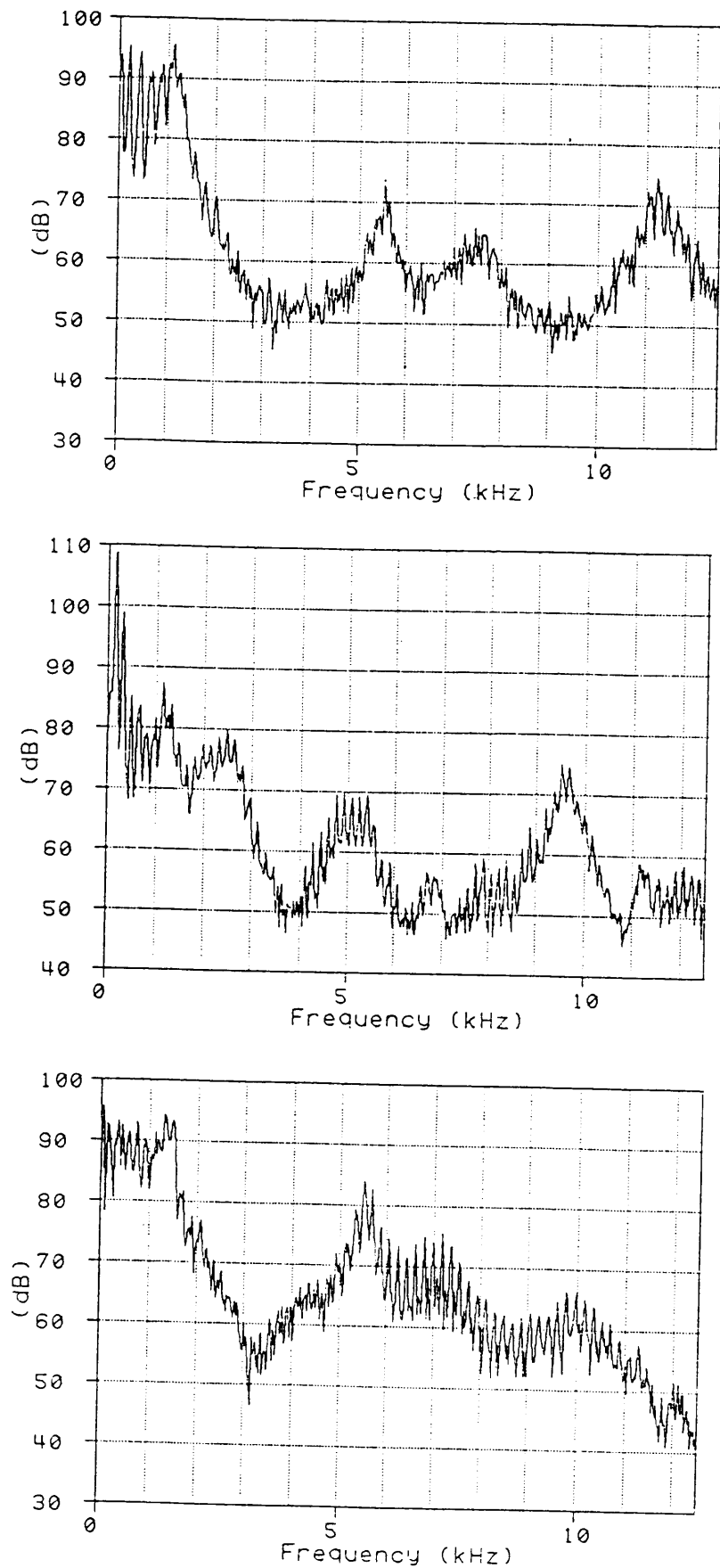


図 2.7: 深度 300m で「しれる」と発声した時の /i/ (/fireru/) のスペクトル  
 上からマイクロホン MW063、F560、4133

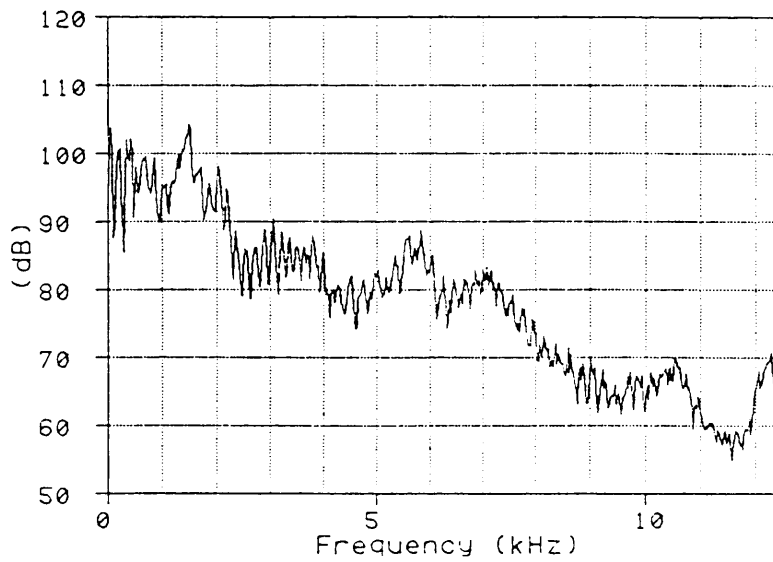
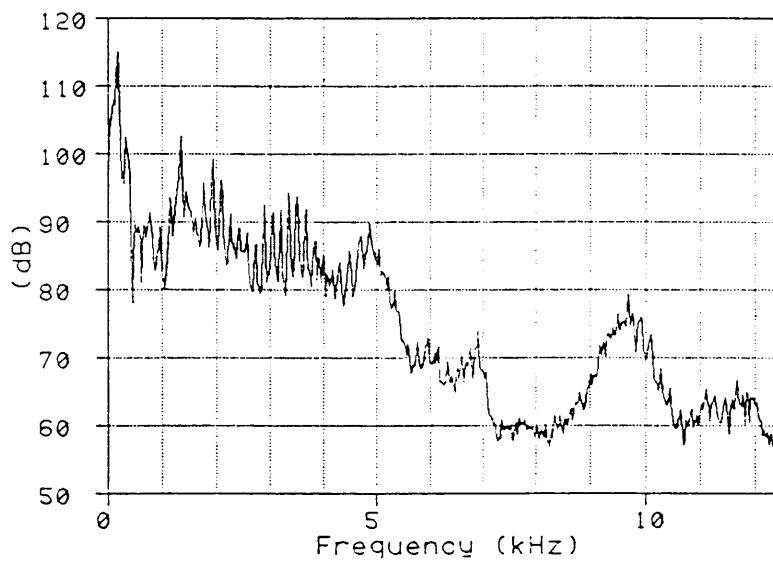
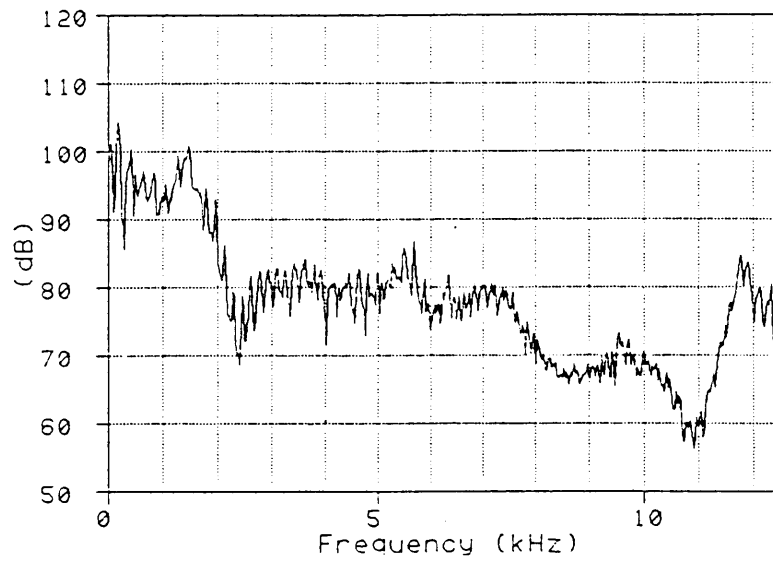


図 2.8: 深度 300m で「さめる」と発声した時の平均スペクトル  
上からマイクロホンは MW063、F560、4133

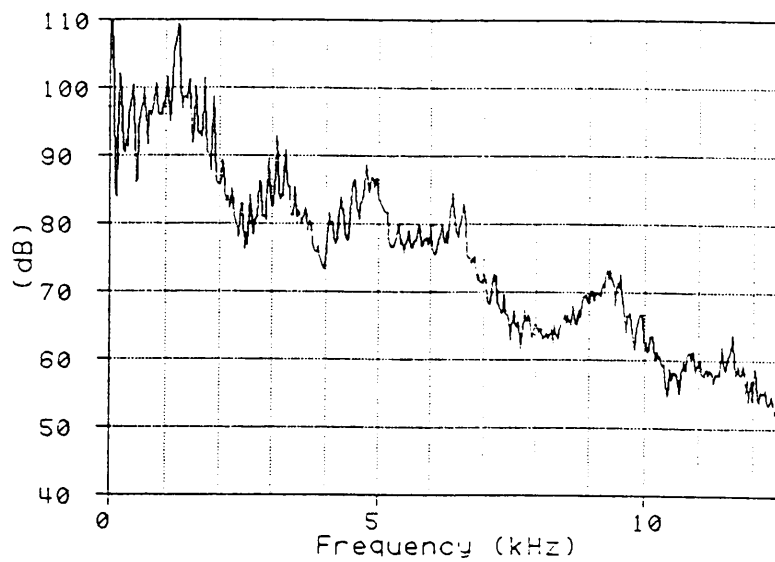
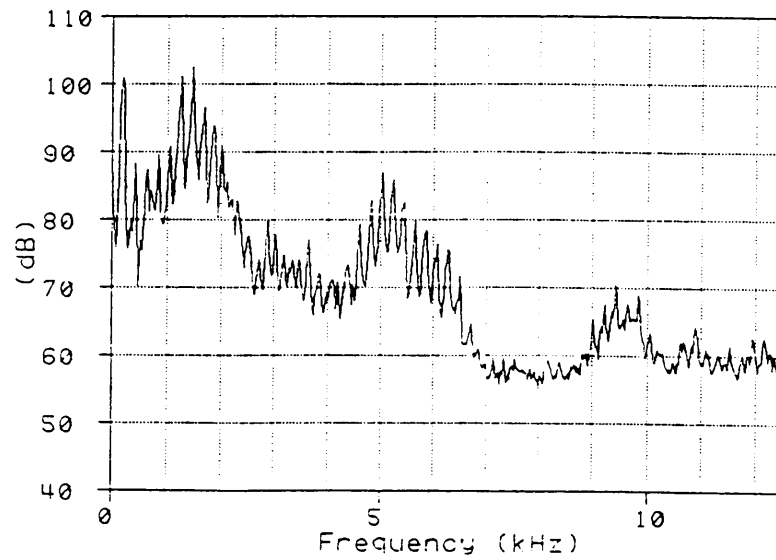


図 2.9: 深度 145m で「さめる」と発声した時の平均スペクトル  
上からマイクロホンは F560、4133

## 第 3 章

# ヘリウム音声の性質

### 3.1 はじめに

ヘリウム音声は特殊な音声であり通常の音声で知られている特性とは異なるので、通常音声と比較しながらヘリウム音声について分析した結果を以下に述べる。ただし、修復に関わる詳しい分析は第 5 章で述べる。

### 3.2 ピッチ周波数

音声における「ピッチ周波数」とは声帯が振動する基本周波数のことである。ヘリウム音声ではピッチ周波数が上昇するという報告があるが深度は 100m 程度である。深度 300m という深深度までのデータでピッチ周波数を求めたものがないので、以下に分析方法とその結果を示す。使用したデータは 1980 年の話者 1 名 (以後、「話者 $\alpha$ 」とする) と 1984 年のデータである。

話者 $\alpha$ の発声した 50 音節の母音部分を分析した。深度は 3m、178m、300m である。ピッチ周波数の推定には自己相関法を用いた [33]。まず、音声信号は 10kHz で標本化する。次に遮断周波数 800Hz の低域通過フィルタをかけ、このデータの自己相関からピッチ周期を求める。分析窓長 30ms、分析周期 10ms、おのこの音節に対して分析数は連続する 10 個で各深度ごとに合計 500 個である。図 3.1 にその結果を示す。また、その平均と分散と深度 3m に対するピッチ周波数上昇率を表 3.1 に示す。深度 3m と深度 178m のピッチ周期は平均値はほぼ同じであるが分散が深度 178m のほうが小さい。また、深度 300m では他の二つと比べピッチ周期は短く、分散も小さくなっている。

1984 年の録音の際に 1 本のマイクロホンでの収録 (方法 i) のほかに、箱を使用して口とそれ以外の放射音を分離して収録 (方法 ii) することを行った。その箱を利用した測定の概略を図 3.2 に示す。箱を使用したことで周波数特性に与える影響があるが、ピッチについては補正せずに使用できる。1 本のマイクロホンを使用したときと箱を使用して測定した結果を表 3.2 に示す。音声資料は 75 音節で、母音部の定常部各 10 ピッチを用いた。方法 i では 2 名の話者ともにピッチ周波数は深度 0 メートルに比べ 300 メートルで上昇することがわかるが、分散は一方はおおきくなり他方は小さくなっている。方法 ii では話者

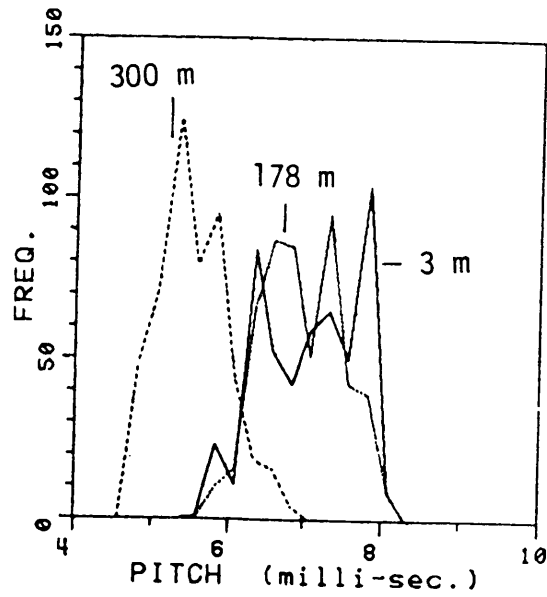


図 3.1: ピッチ周期の各深度における頻度分布 (話者 $\alpha$ )

表 3.1: ピッチ周期の平均と分散とピッチ周波数の上昇率 (話者 $\alpha$ )

深度 (m)	ピッチ 周 期 (ms)	分散 (ms)	ピッチ周波数	3m に対するピッ チ周波数上昇率
3	7.12	0.40	140Hz	—
178	7.01	0.28	143Hz	1.02
300	5.59	0.18	179Hz	1.27



Eでは深度0メートル、300メートルともに方式*i*の0メートルよりもピッチが高く、分散が大きい。話者Gでは方式*i*の0メートルより多少高い程度で分散のそれほど大きくなっていない。これは箱で収録すると口から出ている声がほとんど聞こえないという状況が話者Eでは心理的に影響を与えているが、話者Gではそれほどではないことを示している。また、話者Eでは方法*i*の0メートル以外のピッチ周波数が同じということは耳からのフィードバックを無視した結果と思われるのに対し、話者Gではピッチの上昇のしかたが異なるのは耳からのフィードバックがないほうが影響が少なく、あるほうが影響が大きいことを示唆している。

また、声帯を二質量モデル、声道を音響管モデルとしてシミュレーションした結果でも深度0メートル、300メートルでピッチ周波数は大差なかった [34] ので、ピッチ周波数の上昇は心理的な影響と考えるのが妥当である。

### 3.3 発話速度

発話速度については遅くなるというデータがいくつか報告されているが、深度は100m程度である。

これについても深度300mにおよぶデータはないので分析しておく。結果を表3.3に示す。分析データはピッチ周期を求めたものと同じ話者 $\alpha$ の短文である。深度178mでは深度3mに比べて発話速度が遅くなっているが、深度300mでは深度3mに比べて大きな差はない。

遅くなる理由を文献 [7] では、明瞭度が低くなるためわかりやすくしようとする心理的なものであると述べている。高圧ヘリウム空気を肺に急激に吸い込むことおよび吐き出すことは実際は大変であるとのダイバーの話もあり、高圧空気による影響も考えられるので、すべて心理的なものといえるかは今後の課題として残されている。

### 3.4 ホルマント周波数

話者G、Hの長母音について線形予測を用いて分析した。分析次数は20、発声時の安定した区間を目視で選び連続する6フレームの平均を求めた。深度は0mと300mである。この5母音の第1と第2ホルマント周波数の関係を図3.3に示す。ホルマント周波数は深度300mでは深度0mに比べ高い方に移動しているのがわかる。

次に、同じデータについて、深度0mと300mの同一長母音のホルマントごとに対応させて表したのが図3.4と図3.5である。なお、ここでは音声は発声ごとにホルマント周波数は多少異なるが、同一長母音ではほぼ同じようになると仮定している。また、これらの図で曲線で示したのがFantらが理論的に求めたものである。その式を以下に示す。

$$F_h = k \sqrt{F_a^2 + (r-1)F_{ua}^2} \quad (3.1)$$

ここで、 $F_h$  : ヘリウム音声のホルマント周波数

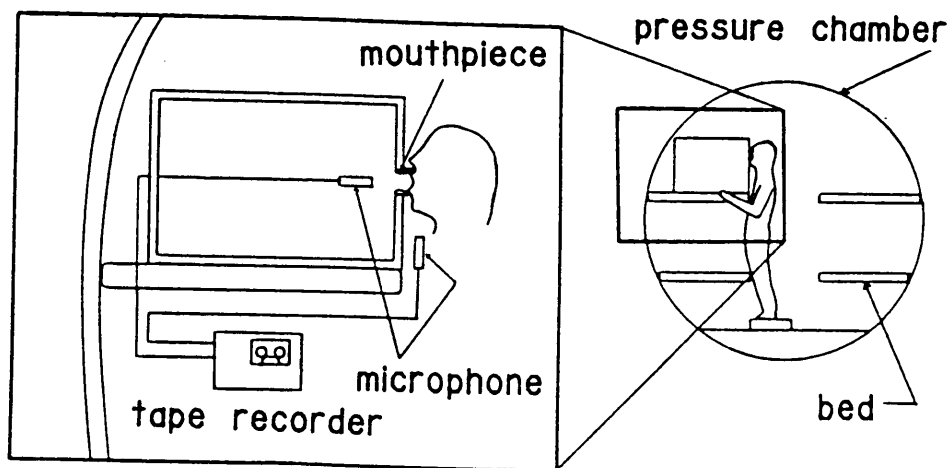


図 3.2: 箱を使用して口とそれ以外の放射音を分離して収録する概略図

表 3.2: ピッチ周期の平均と分散とピッチ周波数の上昇率

(a) 通常の録音方法 (1本のマイクロホンを使用), (b) 箱を使用して口からの放射を分離して録音

録音方法	話者	深度 (m)	ピッチ周期 (ms)	分散 (ms)	ピッチ 周波数	ピッチ周波数 上昇率
(a)	E	0	7.96	0.31	126Hz	—
(a)	E	300	6.07	0.26	165Hz	1.31
(a)	G	0	7.12	0.35	140Hz	—
(a)	G	300	5.53	0.48	181Hz	1.29
(b)	E	0	6.20	0.45	161Hz	—
(b)	E	300	5.86	0.49	171Hz	1.06
(b)	G	0	6.42	0.35	156Hz	—
(b)	G	300	6.30	0.41	159Hz	1.02

$F_a$  : ヘリウム音声と同一の声道形状で発声したときの  
 通常空気中におけるホルマント周波数  
 $F_{wa}$  : 通常空気中における声道の最低共振周波数  
 $k$  : ヘリウム空気の通常空気に対する音速比  
 $r$  : ヘリウム空気の通常空気に対する密度比

である。 $F_{wa}$ は主に声道の長さや断面積、および声道壁の柔らかさによるが、ここでは200Hzとして図の曲線は示した。話者Gの場合、 $F_{wa}$ は200Hzより低く、話者Hの場合は200Hzより高いことがわかる。話者Gの場合、第1ホルマント周波数のみを用いて最小二乗法で $F_{wa}$ を求めると142Hz、話者Hの場合は230Hzとなり、文献[36]で示された150~200Hzにほぼ近い。また、被験者は異なるが文献[35]で測定した頬の単位当たりの質量が1.17~1.71g/cm<sup>2</sup>であり、この値から推定される $F_{wa}$ は169~204Hz(付録B参照)である。よって、Fantらの式が成り立つとともに $F_{wa}$ の範囲は140~230Hz程度である。

同じ話者の無声子音の破裂部について調べたのが図3.6、図3.7である。分析方法は破裂部を目視で切り出し、線形予測の中で比較的短い時間でも分析可能な共分散法で、分析次数は12とした。ピーク周波数を深度0mと300mで求め、同じ音節を対応させて表した。話者はGとHである。母音ほど曲線上にはないがFantらの式でほぼ表されていることがわかる。

Fantらの式は一様な管についての理論であるが、現実の音声のように複雑な管の形をしている場合にも深度300mで発声された音声で成り立つことがここで示された。

### 3.5 母音と子音の相対的な音圧強度について

話者Bについて分析した。音声資料は明瞭度試験用音節表の100音節で、深度0mの通常空気中と深度300mのヘリウム空気中のデータで母音と子音の相対強度を求めた。音声の発声強度はいつも同じとは限らず、ヘリウム音声を修復するという目的には個々の母音の強度とそれに対する子音の相対的な強度を求めることは意義がある。

通常空気中の音声は4.8kHzの低域ろ波後10kHzの標本化周波数で、深度300mのヘリウム音声は90Hzから12kHzの帯域ろ波後25kHzの標本化周波数で標本化した。なお、ヘリウム音声ではファンの影響で50Hzの電源周波数の雑音が大きく、この影響をできるだけさげ、しかも音声にできるだけ影響がないように、90Hz以下の低域を遮断した。

分析は、破裂音などの発声時間の短いものを考慮して、以下の式に示すように5msの区間の短時間エネルギーをとり、重複させずに5msごとのエネルギーをまず求めた。短時間エネルギーEを求める式を以下に示す。

$$E = (1/N) \sum_{n=1}^N x(n)^2 \quad (3.2)$$

ただし、Nは深度0mで50、深度300mで125である。

表 3.3: 短文朗読時間の変化 (深度 3m に対する比)

文章番号	深度 3m での持続時間 (秒)	深度 178m	深度 300m
1	2.05	1.22	0.98
2	2.77	1.34	1.36
3	3.81	1.63	1.10
4	1.66	1.51	1.10
5	1.57	1.18	1.02
6	4.12	1.38	1.03
7	1.83	1.15	0.87
8	3.46	1.35	1.07
9	3.85	1.14	0.95
10	5.11	1.12	0.98
11	4.57	1.15	1.05

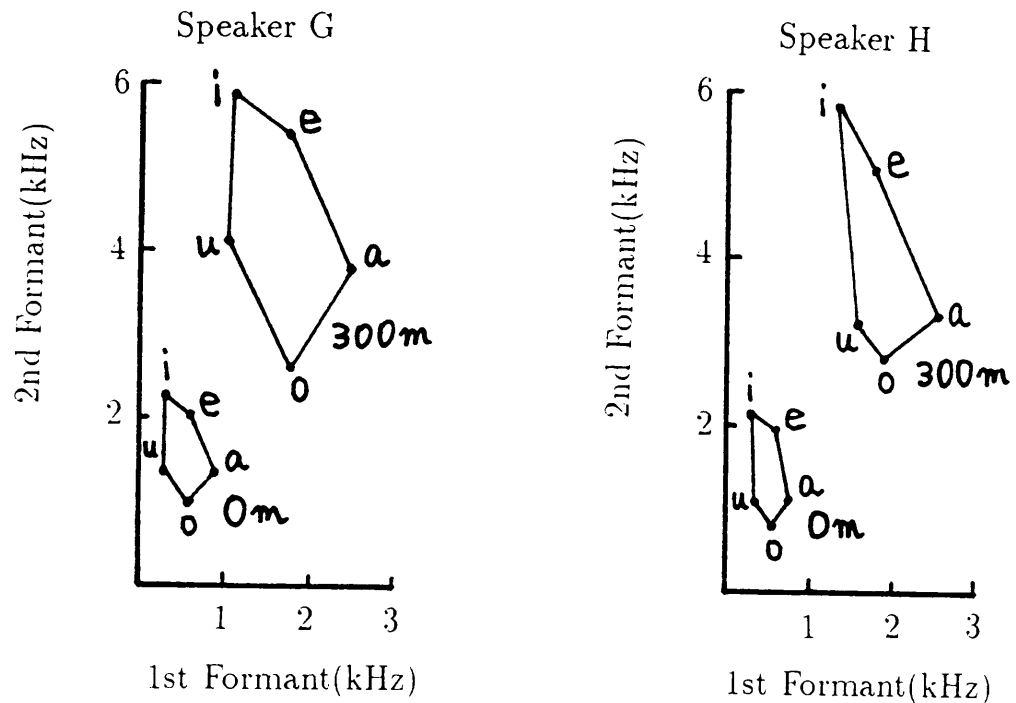


図 3.3: 通常空気 (深度 0m) と深度 300m の 5 母音の第 1 と第 2 ホルマント周波数の関係

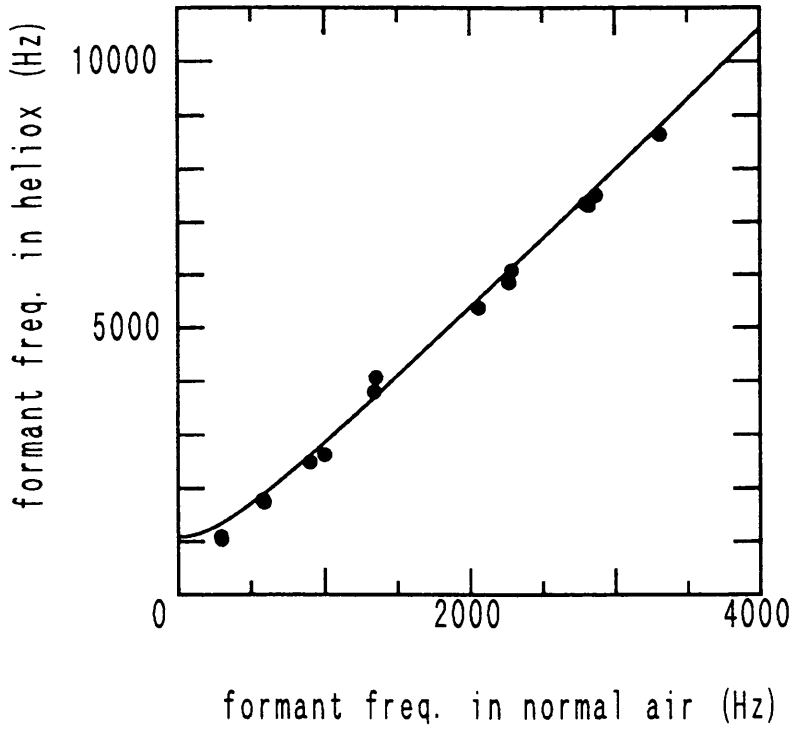


図 3.4: 通常空気と深度 300m の母音のホルマント周波数の関係,  
 ・母音のホルマント周波数、曲線は Fant らの式

話者 G

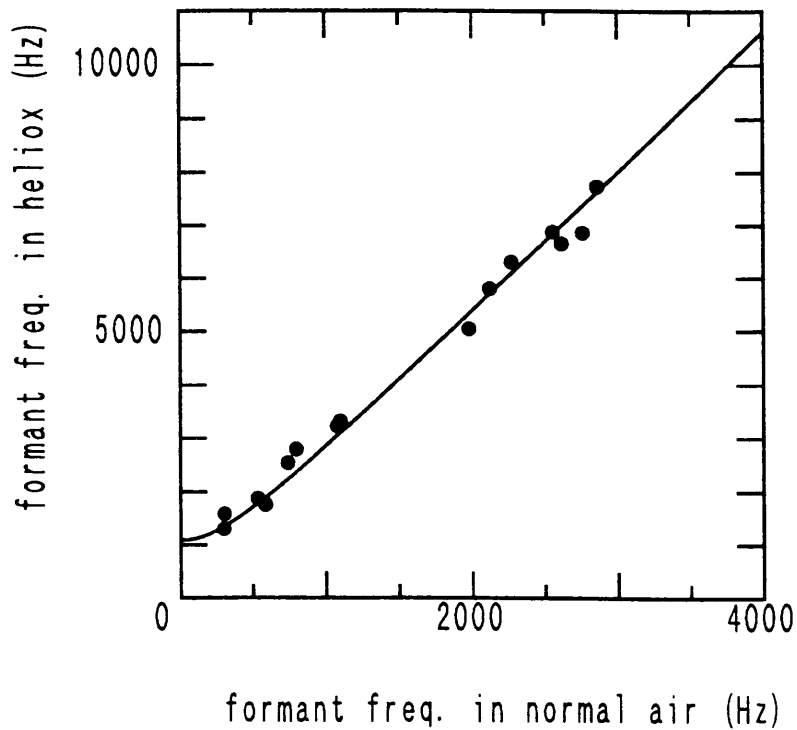


図 3.5: 通常空気と深度 300m の母音のホルマント周波数の関係,  
 ・母音のホルマント周波数、曲線は Fant らの式

話者 H

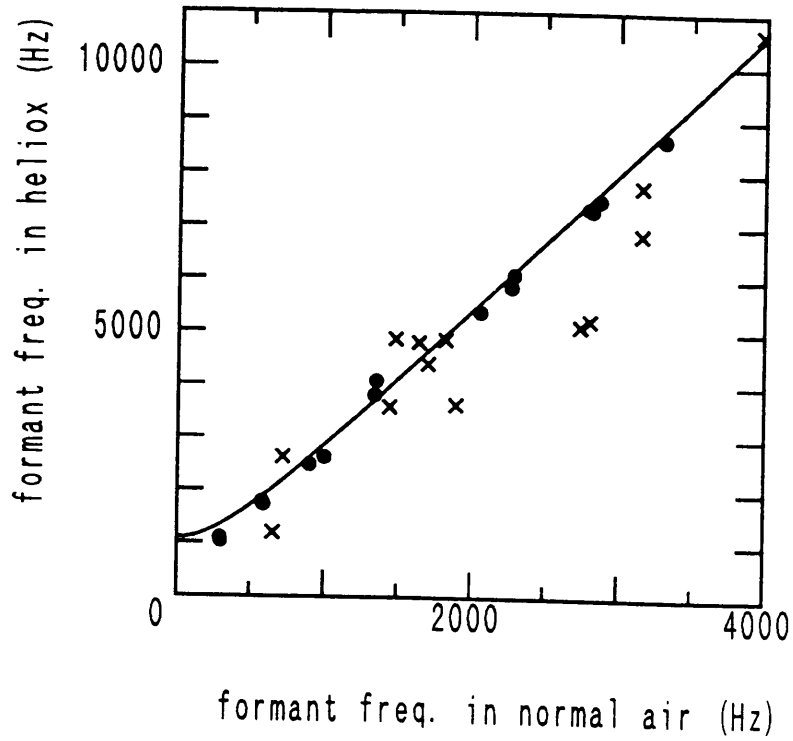


図 3.6: 通常空気と深度 300m の無声子音のピーク周波数の関係, 話者 G  
 ・母音のホルマント周波数、× 無声子音のピーク周波数、曲線は Fant らの式

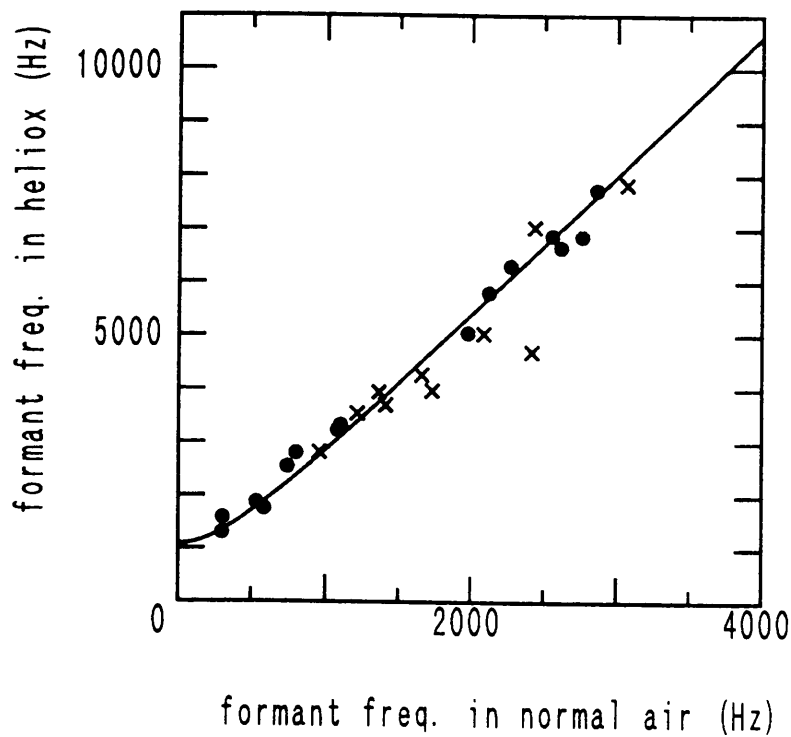


図 3.7: 通常空気と深度 300m の無声子音のピーク周波数の関係, 話者 H  
 ・母音のホルマント周波数、× 無声子音のピーク周波数、曲線は Fant らの式

次に、波形から目視により母音部、子音部を求める。特に有声破裂子音ではバズバーも求める。子音部では破裂または摩擦、鼻子音部分の開始から終了まで、バズバーでは開始から終了までの連続した数フレームを、母音部では、その区間の最大となる短時間エネルギー  $E$  を示すフレームの 2 つ前のフレームから 5 つ後のフレームの、連続する 8 フレーム (40ms) を抽出し、各フレームのエネルギー  $E$  を平均して、それぞれの音声部の代表値とし、dB 値を求めた。

また、エネルギーの値自体は、発声時の声の大きさや録音時のボリュームの大きさなどによるので、子音のエネルギーは、その発声時の後続母音を基準とした相対エネルギーで表すようにした。

雑音の大きさは明瞭度試験音節表を読み上げているうちの数カ所の雑音部分の平均値である。通常空気中で 40.5dB(母音 56.2dB から 74.7dB、子音 46.1dB から 63.8dB)、深度 300m で 41.5dB(母音 63.2dB から 75.7dB、子音 41.4dB から 62.7dB) であった。深度 300m での雑音は 90Hz 以下を遮断しても子音によっては雑音とかわらないレベルのものがある。それは、「ソ」の /s/ の 41.4dB、「ツ」の /ts/ の 42.0dB、「チュ」の /ch/ の 43.6dB である。以下に示すデータはこのような雑音のある環境で発声された音声の結果である。

まず、図 3.8 に通常音声と 300m のヘリウム音声の母音 /i/ に対する他の母音のエネルギーレベルを示す。この値は各音節における母音部の 5ms ごとの分析における最大エネルギー部を同じ母音ごとに平均した結果である。通常音声とヘリウム音声では、/a/、/e/、/o/ の大小関係は多少変化するが、他の /i/、/u/ に比べて大きいことは変わっていない。大きさの最大の母音と最小の母音との差はほぼ 6dB でこれも変化はない。

次に、母音の大きさは母音ごとに異なるので、後続母音ごとに子音部 (破裂または摩擦、鼻子音) の後続母音に対する大きさを調べ、その通常空気中の子音に対する深度 300m の子音の大きさを示したのが図 3.9、図 3.10 である。

全体的にマイナスの値を示していて、ヘリウム音声では通常空気中の音声に比べ子音が母音に比べエネルギーが低下することを示している。

後続母音 /i/ では /k/ のように深度 300m では破裂部が確認できないまでに低下しているものもあった。他の子音の低下量は 10dB 以内である。後続母音 /u/ では、低下量が 10dB 以上の子音はなく、/p/ のようにむしろ 300m の方が大きいものもあった。後続母音 /a/ では、有声子音でのレベル低下が大きく /g/ では通常空気中に比べ 10dB 以上の低下である。後続母音 /o/ では有声子音 /z/ の低下量が 10dB 以上で、/k/、/t/、/b/ が 10dB 近い低下量を示した。後続母音 /e/ では、無声子音 /k/、/p/、有声子音 /r/ で 10dB 以上の低下である。

この結果より、後続母音で異なるが鼻子音 /m/、/n/ での低下量が少なく、特に有声破裂音 /g/、/d/、/b/ および /k/ で低下量が大きいものがあることがわかる。また、バズバーについて比較したのが図の 3.11 である。例外はあるが、子音と同様に深度 300m のヘリウム音声のほうが低下している。ただし、/g/、/z/ の子音を含む音節と /bo/ では通常音声で観測されたバズバーが観測されず、有声子音でも口腔の比較的奥で破裂を伴うものでバズバーがなくなることがわかる。

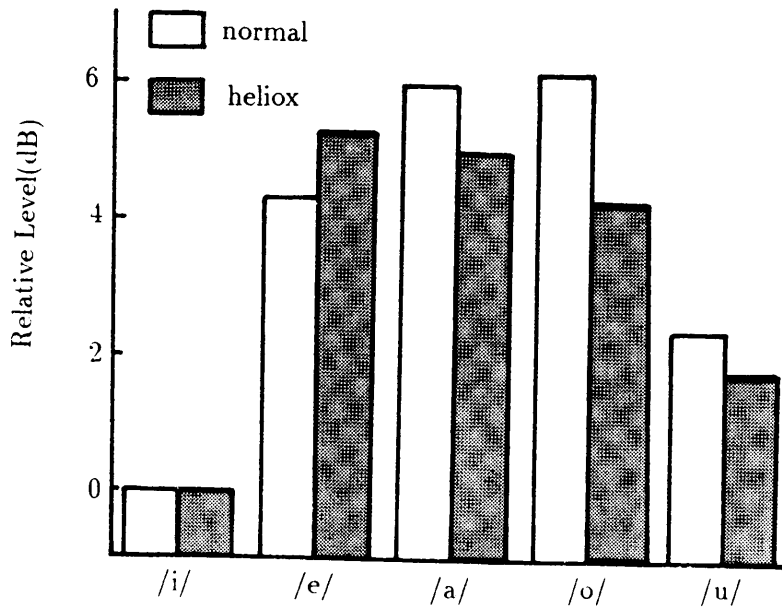


図 3.8: 通常音声と 300m のヘリウム音声の母音 /i/ に対する他の母音の相対エネルギーレベル

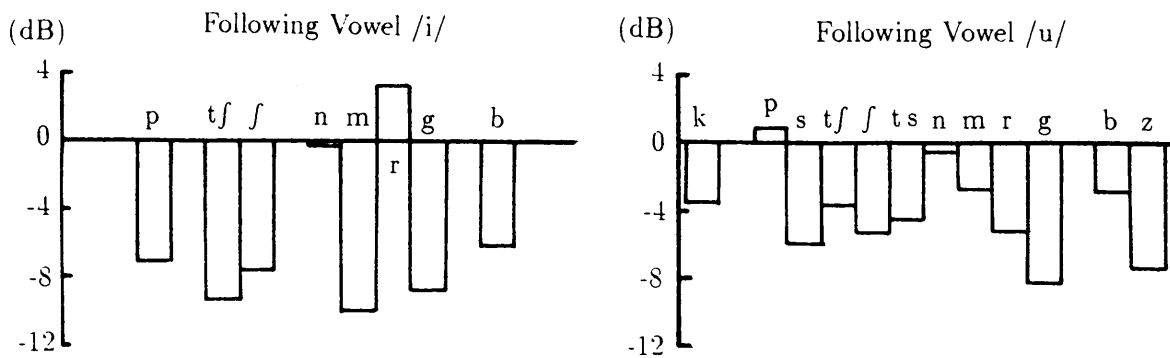


図 3.9: 通常音声に対する深度 300m における子音の大きさ (後続母音で正規化)



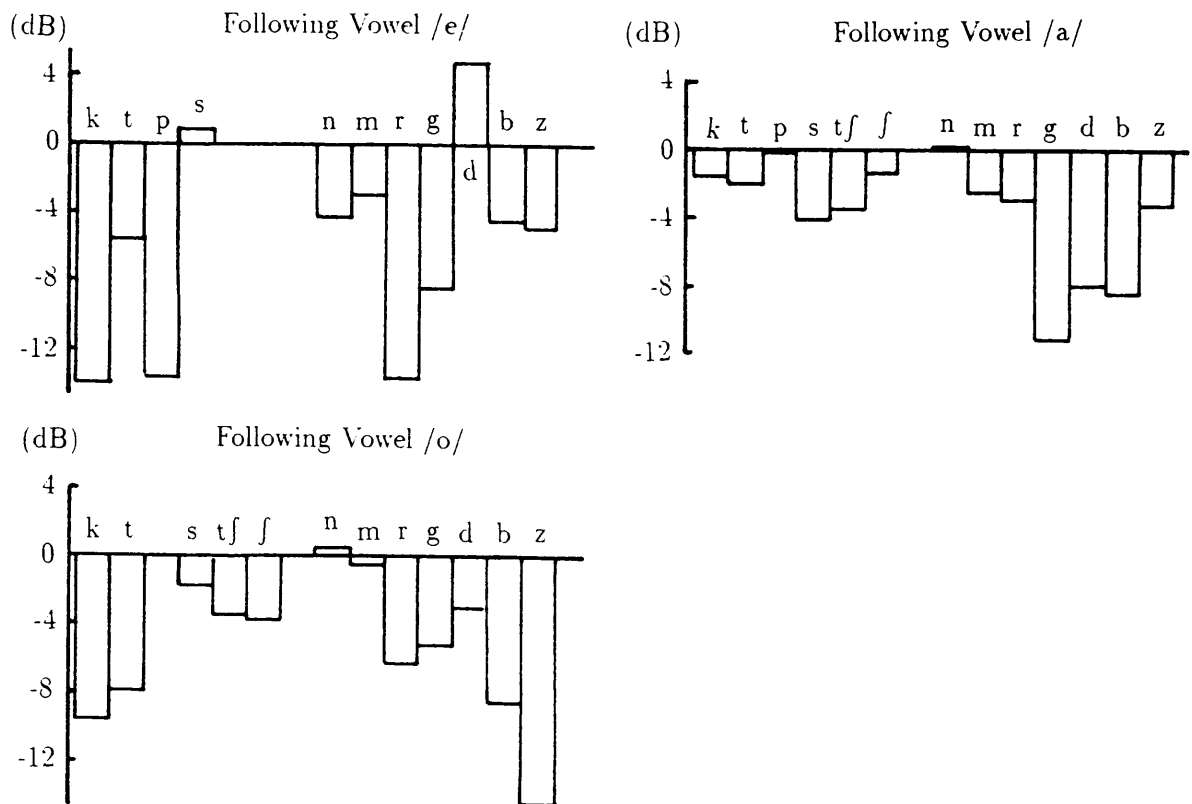


図 3.10: 通常音声に対する深度 300m における子音の大きさ (後続母音で正規化)

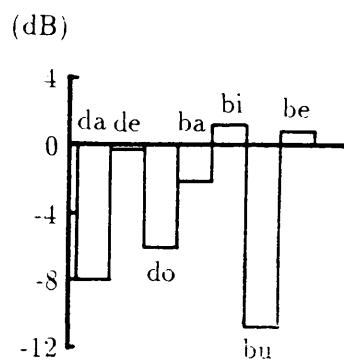


図 3.11: 通常音声に対する深度 300m におけるバズバーの大きさ (後続母音で正規化)

## 3.6 まとめ

以上より、ヘリウム音声の性質は以下のようにまとめられる。

1. ヘリウムという音速の速い媒質のため声道が空気中より短く見えるためホルマントが高い周波数に音速の速度に比例して移動する。
2. 高圧では密度も高くなるので声道壁が剛体とはみなせなくなり低いホルマントが相対的により高い周波数に非線形に移動する。つまり、Fantらの式が成り立つ。
3. Fantらの式は母音だけでなく無声子音の破裂部でもほぼ成り立つ。
4. ピッチ周波数が高くなるが、深度 300m でも高々30%である。
5. 子音の大きさが母音に比べて通常空気より小さくなる。一部に欠落もおこる。
6. 発話速度が遅くなることがある。

その他として、以下のことに注意が必要である。

- 騒音が大きい。
- マイクロホン自体も高圧環境でひずむ。

ヘリウム音声の修復にあたってはこれらの点に注意を払いアルゴリズムを考案する必要がある。

# 第 4 章

## ヘリウム音声修復装置 (1 号機)

### 4.1 はじめに

前章でヘリウム音声の性質の概略を述べた。ここで開発するヘリウム音声修復装置は深度 300m 以深のヘリウム空気中で発声した音声の修復を目標とした。深度が深くなると前章で述べたようにホルマント周波数が非線形に上昇するので、非線形に変換可能な方法がよい。現在までに提案されているヘリウム音声の修復方法は 1 章で述べたように線形に変換するものが大部分であり、実時間で可能なものはアナログ技術によるもののみである。そこで、デジタル技術を用いたホルマントの非線形変換が可能な方式を提案する。ピッチ周波数は前章で述べたように高々 30% 高くなるだけなので、ピッチの変換はここでは考えないことにする。声道のホルマント周波数を抽出する方法で最も良く使用されている方法は線形予測法 (LPC) である。しかも、ヘリウム音声ではピッチ周波数と第 1 ホルマント周波数が通常空気中の音声と比べて離れているので精度よく抽出できることが期待される。ここで提案するアルゴリズムは実時間処理を念頭におき、解が安定で収束も早く、その安定性も容易に判断できるものを用いた。それは、LPC の一種であり、その係数自身が周波数の関数である LSP 係数である。修復方法は、LPC 分析で抽出した声道の極周波数を LSP 係数自身を非線形変換し、再合成はこの LSP 係数と残差波形を用いる。原理的な評価は計算機シミュレーションで可能であるが、潜水実験の現場に持ち込んで実際のヘリウム音声を修復して評価するには実時間でできるシステムが必要である。以下に、この装置の修復の原理とそれに伴う誤差、使用できるデジタル IC の制約の中でのシステムの設計方針について述べ (具体的なハードウェアの設計については付録 C 参照)、次章でその評価について述べる。

## 4.2 修復のアルゴリズムと誤差

### 4.2.1 アルゴリズム

ヘリウム音声と通常の大気中の音の違いは、声道伝送特性に基づく成分と音源特性に基づく成分とがあるが、前者の方が顕著である。声道伝達特性はアンチホルマント (零点) もあるがおおよそホルマント周波数と帯域幅によって規定され、前章で示した Fant らの式で近似される [5]。ここでは後の説明上式 3.1 を書き直したものを示す。

$$F_a = \sqrt{(F_h/k)^2 - (r-1)F_{wa}^2} \quad (4.1)$$

ここで、 $F_h$  : ヘリウム音声のホルマント周波数

$F_a$  : ヘリウム音声と同一の声道形状で発声したときの  
通常空気中におけるホルマント周波数

$F_{wa}$  : 通常空気中における声道の最低共振周波数

$k$  : ヘリウム空気の通常空気に対する音速比

$r$  : ヘリウム空気の通常空気に対する密度比

ヘリウム音声の修復の主要な機能は入力音声のホルマントを式 4.1 に従って低下させ、出力音声のホルマントを  $F_a$  になるように変換することである。一方、ヘリウム音声の音源となる声帯振動の周波数は前章で述べたように深度 300m でも空気中の発声と同じか、高々 30% 高くなるだけなので基本周波数は原音のままとする。従って、修復のためには残差駆動型のボコーダを用い、ヘリウム音声进行分析してホルマント周波数を抽出し、それを式 4.1 に代入して  $F_a$  を求め、残差信号を音源にして再合成すれば良い。しかし、 $F_a$  を  $\alpha$  係数から直接求める演算は複素数演算であって計算量が多く、ままた解が求まらないことがあり安定度に問題があり実時間で行なうには不向きである。

ここでは式 4.1 のホルマント周波数を LSP 係数 [38] の角周波数で置き換えることによって非線形変換を行なう新しい方式を提案する。LSP 係数は周波数に関するパラメータだからホルマント周波数を簡単な式で非線形に変換できる。また、解の安定性の判別が容易で実時間向きである。この方式によるヘリウム音声修復装置のブロック図を図 4.1 に示す。入力された音声は前処理を施した後、声道の伝達特性を推定する。この伝達特性の変換には LSP 係数の変換を用いる。また、音源は声道の伝達特性の逆フィルタリングによってもとめ、これらを再合成することにより修復音を得られる。

つぎにこの方式によるホルマント周波数変換の誤差について述べる。

### 4.2.2 変換誤差

表 4.1 に各深度における通常空気に対する音速比、密度比 (算出方法は付録参照)、表 4.2 に 5 母音と中性母音のホルマント周波数を示す。

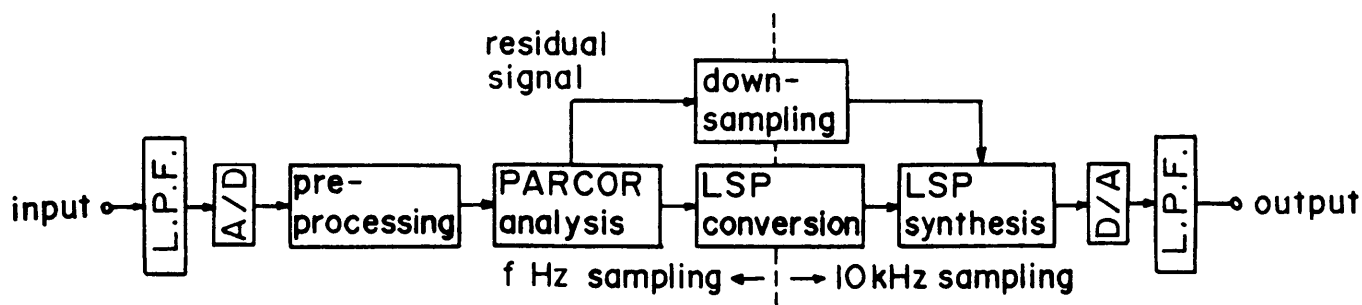


図 4.1: ヘリウム音声修復装置 (1号機) のブロック図

表 4.1: 各深度におけるヘリウム空気の深度 0m の通常空気に対する音速比、密度比

深度 (m)	音速比 ( $k$ )	密度比 ( $r$ )
100	2.27	2.47
200	2.53	3.85
300	2.64	5.23
400	2.71	6.62
500	2.75	8.01

表 4.2: 深度 0m の通常空気における 5 母音と中性母音のホルマント周波数 (Hz)  
ただし、 $F_4 = 3500\text{Hz}$ ,  $F_5 = 4500\text{Hz}$

	/i/	/e/	/a/	/o/	/u/	/ə/
$F_1$	250	450	800	500	330	500
$F_2$	2100	1900	1200	700	1000	1500
$F_3$	3000	2500	3000	3000	2200	2500

これらのデータを用いるとヘリウム空気中と通常空気中で発声された音声のホルマント周波数の関係は式 4.2によって与えられるので、LSP を使用して変換する場合の誤差はヘリウム空気中のホルマント周波数をここで提案する式で変換し、対応する通常空気中のホルマント周波数と比較することによって評価できる。この評価の具体的方法を以下に述べる。

1. ヘリウム音声の第  $j$  ホルマント周波数を  $F_{hj}$ 、帯域幅  $B_{hj}$  とし、空気中での第  $j$  ホルマント周波数を  $F_{aj}$  とするとき、表 4.1 の音速比、密度比、表 4.2 のホルマント周波数を用いて次の式 4.2 を計算する (単位は Hz)。帯域幅については詳しいデータがないので、ここでは通常空気の場合と同様な関係式 4.3 が成り立つものとして取り扱う。

$$F_{hj} = k \sqrt{F_{aj}^2 + (r - 1)F_{wa}^2} \quad (4.2)$$

$$B_{hj} = 0.03F_{hj} + 35 \quad (4.3)$$

2.  $F_{hj}$ 、 $B_{hj}$  より  $\alpha$  係数を計算し、第  $i$  LSP 係数の角周波数  $\omega_{hi}$  を求める。

$$\omega_{ai} = (f/f') \sqrt{(\omega_{hi}/k)^2 - 4\pi^2(r - 1)(F'_{wa}/f)^2} \quad (4.4)$$

ここで、 $f$  は分析時の標本化周波数 (Hz)、 $f'$  は合成時の標本化周波数 (Hz)、 $F'_{wa}$  は変換時に用いる声道の最低共振周波数 (Hz) である。

3. 第  $i$  LSP 係数  $\omega_{ai}$  とすると、これより  $\alpha$  係数を計算し、ホルマント周波数  $F'_{aj}$  を求める。
4. この変換による第  $j$  ホルマント周波数の誤差  $\Delta F_j$  を次のように定義する。

$$\Delta F_j = F'_{aj} - F_{aj} \quad (4.5)$$

ここで示される誤差は LSP 係数を媒介としてホルマント周波数変換を行なう本方式における原理的なものである。この変換の模式図を図 4.2 に示す。立て線が LSP 係数の角周波数であり、スペクトル包絡は LPC スペクトルである。

深度は 200、300、400、500m を想定した。分析時の標本化周波数は深度に応じて変えられるが、ここでは 25kHz と一定にしてある。合成時の標本化周波数は 10kHz である。変換誤差はシミュレーションの結果、低域で大きくなることがわかった。そこで変換誤差の大きい第 1 ホルマントについてまとめたのが図 4.3 である。第 1 ホルマント周波数が低いものほど誤差が大きい深度 500m で高々 20Hz である。

次に、誤差の大きい /i/ の深度 300m における第 1 ホルマントの帯域幅のみを式 4.3 によらず種々変えた時の変換後のホルマント周波数と帯域幅を表 4.3 に示す。これより、 $F'_{wa}$  が 0Hz、つまり線形変換ではヘリウム音声の帯域幅  $B_h$  に関係なく変換後のホルマント周波数はほぼ同じで、誤差は非常に小さい。また、変換後の  $Q (= F/B)$  についてもヘリウム音声の  $Q$  と変わらないことがわかる。 $F'_{wa}$  が 0Hz でない時は帯域幅  $B_h$  の広いものほど

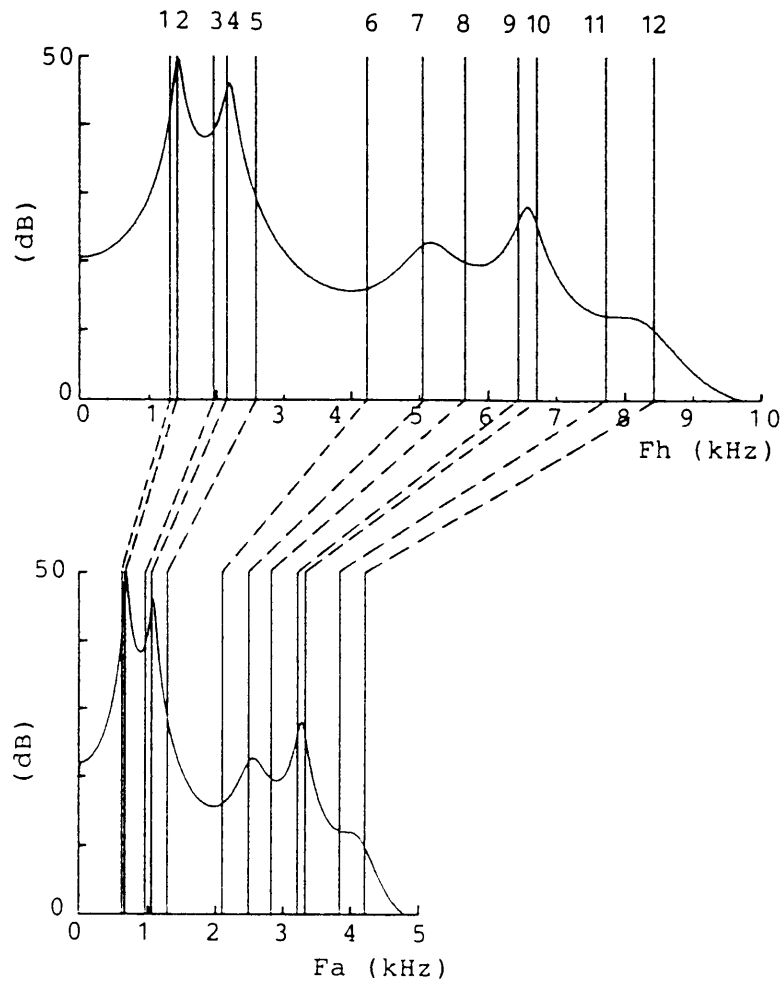


図 4.2: LSP 係数による変換方法の模式図

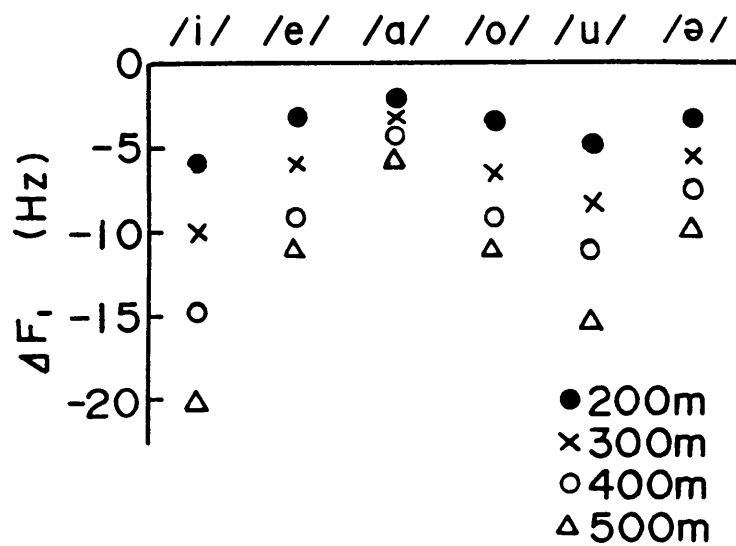


図 4.3: 種々の深度における第 1 ホルマント周波数の本アルゴリズムによる誤差  $\Delta F_1$

変換後のホルマント周波数が低くなり、変換後のバンド幅は  $B_h/k$  に近く  $Q$  が低下する傾向がある。しかし、式 4.4 の変換時に  $F'_{wa}$  を多少小さくすれば第 1 ホルマント周波数の誤差は高々  $\pm 25\text{Hz}$  におさまることがわかる。

### 4.3 システムの設計

ヘリウム音声を修復する 1 号機の開発は 1979 年に始まる。1979 年当時はデジタル信号用のプロセッサという考え自体がまだなく、インテル社の i8086 の発売が 1978 年、モトローラ社の 6800 の発売が 1980 年であり、汎用のマイクロプロセッサ自身もまだ 8 ビットが主流であった。そこで、高速性を要求されるヘリウム音声修復装置では汎用のマイクロプロセッサでは演算が実時間でできないので、TRW 社の高速乗算器 1010J と AMD 社のビットスライス Am2901 を、他は市販の IC を使用することにして設計を進めることにした。

システムの概略は図 4.1 にすでに示した。ここでは、システムの設計に当たって考慮したことを列記する。

1. 音声波形のダイナミックレンジを考慮して入力に使用する A-D 変換器は 12 ビット (符号+11 ビット) とする。
2. 分析の標本化周波数は修復後の帯域が 5kHz となるように深度にあわせて可変にできるようにする。分析次数は修復後の帯域が 5kHz であり、この帯域にあるホルマント数がおよそ 5 であるので 12 とする。標本化周波数の最大値は 30kHz となる。また、ピッチ抽出は実時間ではむずかしく、装置がさらに大きくなるので行なわない。この場合、分析の標本化周波数と再合成時の標本化周波数が異なるため再合成時の入力は LPC 係数より逆フィルタリングしたものをダウンサンプリングする必要がある。これをデジタルフィルタで行なうには、分析の標本化周波数が可変であるためすべての周波数において実時間で実現するのは困難である。よって、D-A 変換、4.8kHz 低域ろ波後、10kHz で A-D 変換する方式とする。
3. プロトタイプなので市販の IC を使用し特殊なものは使用しない。
4. 発話速度自体は大きく変化しないのでフレーム周期は 10ms、フレーム長は 30ms とする。
5. 浮動小数点演算では今のところ本装置に使える程に高速の演算器がなくシステムも大きくなるので固定小数点演算とする。

以上を踏まえて、次節から前処理部、分析部、変換部、合成部とダウンサンプリング部に分けて述べる。

### 4.4 前処理部

分析する前に行う処理としては、低域ろ波、A-D 変換、プリアンファシス、最大値検出、左シフト、窓掛け、パワーの計算がある。



低域ろ波は、アナログ信号のデジタル処理に伴う重複歪み (alias) を避けるために必要なもので、通常分析の標本化周波数の 1/2 よりやや低めに設定する。本装置ではスイッチド・キャパシタ・フィルタを用いて標本化周波数に応じて変化できるものにした。A-D 変換器は、符号+11 ビットのもので、標本化周波数 30kHz まで可能なものを用いた。左シフトは最大値検出によって得られた値により、この最大値が最下位ビット (LSB) から 12 ビット目になるようにシフトする。この理由については分析部の項で述べる。パワー  $U_0$  は窓掛けされたデータの二乗和である。このアルゴリズムとハードウェアの詳細は付録 C に示す。

演算時間は、1 データあたり、プリアンファシスに 600ns、データシフトに最大  $1\mu\text{s}$ 、窓掛けに  $1\mu\text{s}$ 、 $U_0$  の計算に 600ns かかる。よって標本化周波数 30kHz のとき最大約 2.9ms となり 1 フレームが 10ms なので仕様は満たされる。

## 4.5 分析部

### 4.5.1 アルゴリズムとシミュレーション

分析には線形予測法を用いるがこれにもいくつかの方法があるので、代表的な 3 つについて考察する。抽出できる情報、演算回数、メモリ数等について表 4.4 に示す。ここで  $N = 900$  としたのはフレーム長が 30ms、最大標本化周波数が 30kHz のためである。また、分析次数  $p=12$  としたのは修復時の帯域が 5kHz になるようにしたためである。この表より以下のことが言える。DLI 法 [39] は PARCOR 係数、 $\alpha$  係数が一度に求まるが、 $\alpha$  係数の値は 3 以上になることがあり同じビット数では精度が落ちる。Le Roux 法 [40] は、DLI 法より多少演算数が多いが、PARCOR 係数だけなので演算精度はよい。しかし、残差波形を得るには別の計算が必要になりアルゴリズムが多少複雑である。単純化した変形格子法 [41] は演算数が多いが、アルゴリズムは単純な繰り返しが多く、残差波形も得られる。繰り返しが多いということは回路が簡単になることなので、今回はこの方法を用いることにした。

単純化した変形格子法のアルゴリズムの概略を以下に示す。N 点からなる 1 フレームの音声信号の両端に  $p$  個のゼロを付加し、 $U_i$  の漸化式を式 4.8 で置き代えたものを使用する。この式を用いれば式 4.11 を用いる場合の約 3/5 の演算回数でできることが示されている [41]。以下に式を示す。

$$k_i = W_i / U_i \quad (4.6)$$

$$W_i = \sum_{n=1}^N \varepsilon_{i-1}^f(n) \varepsilon_{i-1}^b(n) \quad (4.7)$$

$$U_i = U_{i-1} (1 - k_{i-1}^2) \quad (4.8)$$

ただし、 $U_0 = \sum_{n=1}^N x^2(n)$ ,  $k_0 = 0$

表 4.3: 深度 300m における /i/ の第 1 ホルマント周波数の帯域幅  $B_h$  (Hz) と LSP 変換における通常空気中での最低共振周波数  $F'_{wa}$  (Hz) のみを種々変えたときの本方式による変換後の第 1 ホルマント周波数と帯域幅の値 (Hz)

$B_h$	$F'_{wa}$	200	195	190	124	0
50		243( 22)	260( 22)	276( 22)	407( 20)	481( 19)
73		241( 32)	257( 32)	273( 32)	406( 29)	481( 27)
100		236( 44)	254( 44)	270( 44)	405( 40)	481( 37)
150		229( 67)	247( 66)	264( 65)	404( 60)	481( 56)
200		221( 89)	241( 88)	258( 87)	402( 80)	481( 75)
250		212(111)	233(110)	252(109)	400(100)	481( 94)
300		203(134)	226(133)	245(132)	399(120)	481(112)

表 4.4: 3 種類の線形予測分析の 1 フレーム当たりの演算回数等の比較  
(分析次数  $P = 12$ 、窓長  $N = 900$ )

\* Durbin-Levinson-Itakura 法

	DLI 法*	Le Roux 法	簡単化した変形 格子法
抽出できる情報	PARCOR 係数, $\alpha$ 係数, 残差パ ワー	PARCOR 係数, 残差パワー	PARCOR 係数, 残差波形
除算	12(p)	12(p)	12(p)
積和	$168(p^2 + 2p)$	$144(p^2)$	$33300(3Np + N)$
相関 (積和)	$11700(Np + N)$	$11700(Np + N)$	
残差波形 (積和)	$10800(Np)$	$10800(Np)$	
メモリ語数	$953(N + 4p + 5)$	$949(N + 4p + 1)$	$1812(2N + p)$

$$\varepsilon_i^f(n) = \varepsilon_{i-1}^f(n) - k_i \varepsilon_{i-1}^b(n) \quad (4.9)$$

$$\varepsilon_i^b(n) = \varepsilon_{i-1}^b(n-1) - k_i \varepsilon_{i-1}^f(n-1) \quad (4.10)$$

ただし、 $i = 1, 2, \dots, p$   
 $\varepsilon_0^f(n) = x(n)$ ,  $\varepsilon_0^b(n) = x(n-1)$   
 $x(n)$  : PARCOR 分析部への入力信号

一般に、

$$U_i = \frac{1}{2} \sum_{n=1}^N \{ \varepsilon_{i-1}^f(n)^2 + \varepsilon_{i-1}^b(n)^2 \} \quad (4.11)$$

予備実験の結果、固定小数点演算では  $\varepsilon_i^f(n)$ 、 $\varepsilon_i^b(n)$ 、 $k_i$  を 16 ビット、 $U_i$  と  $W_i$  を 32 ビットとすれば単精度浮動小数点演算と大差ないことがわかった。

ただし、入力音声小さくなると、そのまま計算したのでは式 4.9、4.10 で切り捨て誤差が大きくなることが予想される。この問題は固定小数点演算と単精度浮動小数点演算により推定されたホルマント周波数を比較した。その結果、入力を符号プラス 11 ビットに正規化すれば良いことがわかった。

この固定小数点による分析をさらに厳密に評価するために浮動小数点演算で求めた同方法の PARCOR 分析による LPC 包絡  $T_1(z)$  と、この固定小数点による PARCOR 分析による LPC 包絡  $T_2(z)$  から以下のようなスペクトル歪  $DS$ [42] を求めた。

$$DS = \left\langle \left( 10 \log_{10} \left| \frac{T_2(z)}{T_1(z)} \right|^2 \right)^2 \right\rangle_{\omega, t} \text{ dB}^2 \quad (4.12)$$

ただし、 $z = \exp(j\omega T)$ 、 $T$  : サンプル周期  $\langle \quad \rangle$  は  $\omega$ 、 $t$  にわたる長時間平均を表わす。

潜水深度 300m におけるヘリウム音声データ”二千六百四十八と書け”(約 2 秒) を用いてスペクトル歪を求めると  $0.075 \text{ dB}^2$ [37] である。よって、分析部では、聴覚的な検知限  $1 \text{ dB}^2$  [41],[43] 以下に押さえることができる。

#### 4.5.2 分析部のハードウェア

この仕様を満たすハードウェアの構成を考える。分析部では  $U_0$  を除く式 4.6~4.10 を行う。従って、分析部は表 4.4 から最大積和回数がパワー計算を除いた 32,400 回となり、1 フレームを 10ms とすれば 1 積和演算当たり 300ns 以内でなくてはならない。積和器としては TRW1010J を 2 個用いた並列演算方式として上述の条件を満たすようにし、式 4.6 と式 4.8 は Z80 と乗除算器 1855C を用いることにすると、ブロック図は前処理部を含めて図 4.4 のようになる。

図 4.4の動作を信号の流れ図によって示したのが図 4.5である。はじめに式 4.7により  $W_1$ を計算する。 $U_0$ は前処理部で求めているので  $U&k_i$ 部に  $W_1$ を転送すると  $k_1$ が計算される。 $k_1$ の計算の後、 $W&\varepsilon_i$ 部では  $\varepsilon_i^f(n)$ 、 $\varepsilon_i^b(n)$ を式 4.9、4.10で計算する。一方、 $U&k_i$ 部では式 4.8により  $U_1$ を求める。このように式 4.9、4.10と式 4.8は並列して演算されるので両者が終わって初めて次のステップである  $W_2$ を求める。以下同様にして 12 次まで求める。

この計算では式 4.7は 1 演算当たり 150ns、式 4.9、4.10では 1 演算当たり 600ns であり、式 4.6、4.8は  $W_i$ の符号で異なるが、正で  $129.5\mu s$ 、負で  $258\mu s$  である。 $\varepsilon_i^f$ 、 $\varepsilon_i^b$ の計算と、 $U_i$ の計算が並列に行われているため、標本化周波数が 14.3kHz 以下では  $U_i$ が、14.3kHz 以上では、 $\varepsilon_i^f$ 、 $\varepsilon_i^b$ が分析部の演算処理時間に関係してくる。したがって、標本化周波数 30kHz では式 4.6、4.7、4.9、4.10の演算時間の和が処理時間となり、窓長 30ms のとき、約 9.7ms となり仕様は満たされる。

## 4.6 LSP 係数変換部

LSP 係数変換部では以下の処理を行う。

1. 分析部で求めた PARCOR(偏自己相関) 係数より  $\alpha$  係数を計算する。 $\alpha$  係数を求める漸化式を以下に示す。

$$\begin{aligned}\alpha_n^0 &= 1, & \alpha_n^{n+1} &= 0 \\ \alpha_{n+1}^i &= \alpha_n^i - k_{n+1}\alpha_n^{N+1-i} & (i = 1, 2, \dots, n+1) \\ n &= 0, 1, 2, \dots, 11\end{aligned}\tag{4.13}$$

2.  $\alpha$  係数より LSP 係数を求める。これは、LSP 係数の角周波数を  $\omega_h$  とし、 $x = \cos \omega_h$  とすると以下に示す 2 つの 6 次の多項式の根を Newton-Raphson 法で解くことにより求められる。

$$\begin{aligned}x^6 + \frac{a}{2}x^5 + \frac{(b-6)}{4}x^4 + \frac{(c-5a)}{8}x^3 + \frac{(9-4b+d)}{16}x^2 \\ + \frac{(5a-3c+e)}{32}x + \frac{(f-2+2b-2d)}{64} = 0\end{aligned}\tag{4.14}$$

ここで

$$\begin{aligned}a &= \alpha_{12}^1 \mp \alpha_{12}^{12} \pm 1, & b &= \alpha_{12}^2 \mp \alpha_{12}^{11} \pm a, & c &= \alpha_{12}^3 \mp \alpha_{12}^{10} \pm b \\ d &= \alpha_{12}^4 \mp \alpha_{12}^9 \pm c, & e &= \alpha_{12}^5 \mp \alpha_{12}^8 \pm d, & f &= \alpha_{12}^6 \mp \alpha_{12}^{17} \pm e\end{aligned}$$

3. 求めた LSP 係数の角周波数  $\omega_j$  を次式で変換する。ここでは式 4.4を一般化して次式のように書く。

$$\omega_j' = \sqrt{C_1(\omega_j^2 - C_2)}\tag{4.15}$$

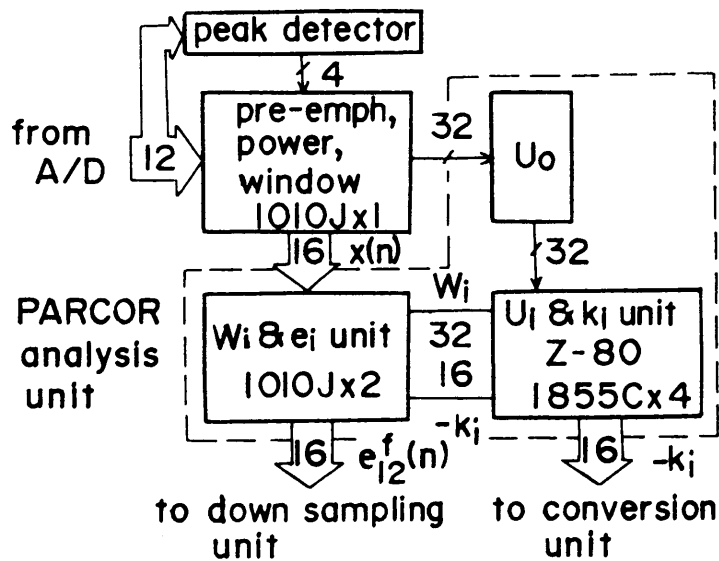


図 4.4: 前処理部と分析部のブロック図

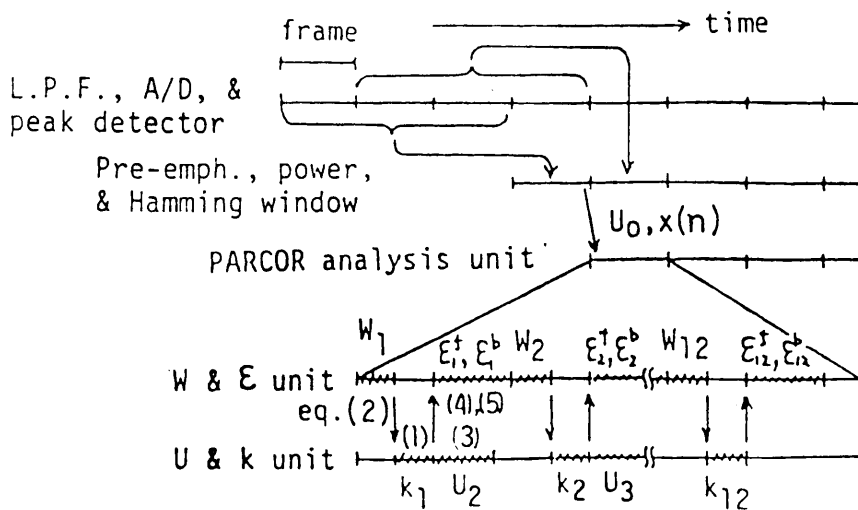


図 4.5: 分析のタイミング図 (図中の式 (1)~(5) は本文中の式 (4.6)~(4.10) にあたる)

ここで、

$\omega'_j$  : 変換後の LSP 係数の角周波数

$C_1$  : 線形変換の係数

$C_2$  : 非線形変換の係数

である。

ハードウェアの語長と小数点の位置を決めるため、PARCOR 係数から  $\alpha$  係数を求める部分と  $\alpha$  係数から LSP 係数を求める部分について 3 つの候補を作り計算機シミュレーションを行った。ただし、LSP 係数を変換する部分は逆余弦と平方根を表引きすることにしたので、シミュレーションからはずした。使用するデータは表 4.5 に示すようなホルマント周波数で、標本化周波数 20kHz で合成された五母音と疑似乱数である。3 つの候補は、

1. 語長 32bit(符号 + 31bit、小数点は bit24 と bit23 の間)
2. 語長 28bit(符号 + 27bit、小数点は bit20 と bit19 の間)
3. 語長 24bit(符号 + 23bit、小数点は bit18 と bit17 の間)

である。

ただし、bit0 が LSB である。Newton-Raphson 法での収束条件は  $|x(n+1) - x(n)| < 2^{-10}$  (ただし、 $x(n)$  は  $n$  回目の解) とした。

この結果の一例を表 4.6 に示す。語長 24bit においても 11bit まで単精度浮動小数点で行ったものと一致している。これは他の合成母音、疑似乱数でもいえ、今回はこの語長を採用することにした。次に、Newton-Raphson 法での収束回数について示したのが表 4.7 である。ここで、+1 と書いてあるのはおののの 10 フレームを分析した結果、その収束回数が 1 回多いものがあることを示している。この結果より、初期値を 1 とすると、第 1 ホルマント周波数の低いものほど第 1 LSP 係数  $\omega_1$  の収束回数が早く、また疑似乱数のようにほぼ LSP 係数が等間隔のものは収束回数がほぼ一定である。以上より平均的な収束回数は 5 回と考えればよいことがわかる。

演算回数を上述のように Newton-Raphson 法の各次数での反復回数を 5 回とすると 1 フレームでの積和が約 350 回、除算が 52 回となり、このほかに  $\alpha$  係数の計算や LSP 係数を変換する計算等がある。よって、積和は約  $10\mu s$  で行う必要がある。これらを満たすためビットスライス型の AMD 社の Am2900 シリーズを用いた。タイミング、プログラム等の詳細は付録 C に示した。ここでは概略を示すためブロック図を図 4.6 に示す。Am2901 を 6 個用いて 24 ビットの語長としている。スタートとシフトの制御を行う Am2904 と、プログラムを制御する Am2910 をおき、全体が通常のマイクロプログラムのレベルで動作するようにした。また、乗算の速度を上げるためマルチプレクサを設けた。式 4.15 の計算を行うときに必要な逆余弦、平方根および余弦は表引きとし、ROM2716(8 ビット  $\times$  1K) を 4 個用いた。プログラムを格納するメモリにはアクセスタイム 85ns の CMOS-RAM、HM6148(4 ビット  $\times$  1K) を 14 個使用し、電池でバックアップしている。CMOS-RAM を使用した理由は、研究室のミニコンピュータ M-70 でプログラム開発を容易に行うためである。

なお、実行時間は 1 フレーム当たり約 9ms であることを確認している。

表 4.5: 深度 100m 相当の合成ヘリウム音声のホルマント周波数と帯域幅 (Hz)  
 ただし  $F_4 = 7820\text{Hz}$ 、 $B_4 = 311\text{Hz}$

	F	B		F	B
/i/	783	66	/o/	1204	97
	4698	217		1648	117
	6665	275		6887	282
/e/	1149	96	/u/	937	81
	4251	203		2239	138
	5245	233		4935	224
/a/	1834	125	/ə/	1246	100
	2564	150		3389	176
	6620	275		5599	244

表 4.6: 浮動小数演算および 24 ビット固定小数点演算における LSP 係数の余弦の値 (左: 10 進数表示, 右: 16 進数表示)、深度 100m 相当の合成ヘリウム音声 /a/, 標本化周波数 20kHz, 分析次数 12

	浮動小数点演算		固定小数点演算	
$\cos \omega_1$	0.873549	6FD0	0.873840	6FDA
$\cos \omega_2$	0.839111	6B67	0.839172	6B6A
$\cos \omega_3$	0.789114	6501	0.788635	64F2
$\cos \omega_4$	0.697727	594F	0.697662	594D
$\cos \omega_5$	0.663670	54F3	0.663879	54FA
$\cos \omega_6$	0.298971	2644	0.298981	2645
$\cos \omega_7$	-0.231891	E252	-0.232056	E24C
$\cos \omega_8$	-0.465638	C466	-0.465820	C460
$\cos \omega_9$	-0.527016	BC8B	-0.526733	BC94
$\cos \omega_{10}$	-0.715631	A467	-0.715179	A475
$\cos \omega_{11}$	-0.780824	9C0E	-0.780945	9C0A
$\cos \omega_{12}$	-0.874466	9012	-0.874725	9009

表 4.7: 母音と疑似乱数の LSP 係数を Newton-Raphson 法で求める場合の収束までの反復回数 (+1 は 10 フレームを分析した結果、分析フレームによっては反復回数が+1 になることを示す)

第 nLSP 係数	/i/	/e/	/a/	/o/	/u/	/ɔ/	疑似乱数
1	3	4	6	5	4	5	5
2	2	2+1	4	3	3	2	4+1
3	4+1	4+1	4	4	5+1	4+1	4+1
4	7	7	4	3	4	6	5
5	5	5	3	3+1	3	4	5
6	4	4	5+1	5+1	6	5	5+1
7	5+1	5	6	6+1	5	6	5
8	5+1	4	5+1	6	3+1	4	5
9	3	4	4	3+1	4+1	4	5
10	5	5+1	5	4	6	5	5
11	1	1	1	1	1	1	1
12	1	1	1	1	1	1	1

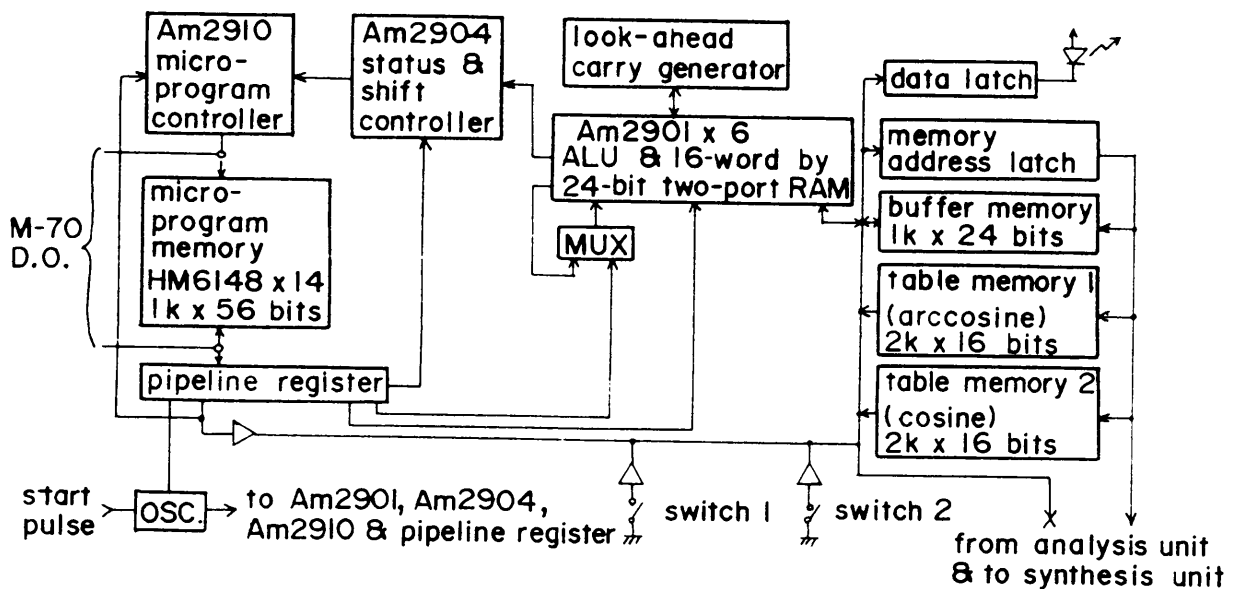


図 4.6: LSP 変換部のブロック図



## 4.7 ダウンサンプリング部と LSP 合成部

合成部は LSP 合成を行なう。その信号の流れ図を図 4.7 に示す。次数は分析部が 12 次となっているので、合成部も 12 次である。標本化周波数は 10kHz である。このように分析部の標本化周波数によらず一定としたのは、標本化周波数を分析部と同一にすると式 4.15 により変換後の LSP 係数の周波数は 5kHz 以下となり、それより高域にいくつかの LSP 係数を導入しないと 4~5kHz では利得が下がり過ぎる。そこで、4~5kHz での利得を補償する LSP 係数が必要となり計算量が増大すること、このようにすると合成出力を 12 ビットとしても内部での数値が大きくなり語長が 20 ビット以上必要なこと、また標本化周波数が高いことにより低域における係数の影響が大きくなることなどが問題となるためである。ただし、標本化周波数を 10kHz とすると残差波形のダウンサンプリングが必要となる。今回は、前述したように分析時の標本化周波数が可変であるため D-A 変換、4.8kHz の低域ろ波後、A-D 変換することとした。合成部の標本化周波数 10kHz は通常用いられる周波数であり、語長は 16 ビットでよい。また、LSP 係数については標本点ごとに係数を補間すると歪みが減るので係数補間回路をつけた。

## 4.8 回路構成とヘリウム音声修復装置の外観

図 4.8 にシステム全体のタイミングの流れ図を示す。計算量が多いので大きく五つの分部に分け、パイプライン方式とした。このため各部にメモリが必要になるが演算速度自体は抑えられる。また、出力される音声は入力より 50ms 遅れることになる。各部での処理、語長、計算量、使用メモリ数、使用 IC の数を表 4.8 に示す。装置は電源やスピーカも含めて一つの箱 (幅 51cm、奥行き 40cm、高さ 35cm) に収められ、35cm × 21cm のガラスエポキシのボード 7 枚に約 800 個の IC で構成されている。ヘリウム空気中の音声の周波数帯域は空気中の音声よりも音速比に比例して広がる。ヘリウムと空気の混合比は深度が深いとヘリウムの比率が高くなるが、ヘリウム 100% でも音速比は高々 3 以下である。本装置では修復後の帯域を空気中の音声に対応する 5kHz とし、その入力ヘリウム音声の帯域をその 3 倍の 15kHz 以下と想定した。そのため標本化周波数は最大 30kHz まで可変とし、深度に応じて、入力の低域ろ波器の遮断周波数、標本化周期、音速比、密度比、声道の最低共振周波数などの変換パラメータの値を装置前面のテンキーボードから 16 進数で入力する。

## 4.9 まとめ

ホルマント周波数を非線形に変換できる新しいアルゴリズムに基づく実時間ヘリウム音声修復装置の修復の原理と誤差について、また、そのアルゴリズムをどのように実行するかを各部にわけて述べた。変換に伴う誤差は低域で大きくなるが、変換に使用する最低共振周波数を多少低くすれば高々 25Hz であることを示した。修復装置の演算は固定小数点演算であり、入力音声は符号+11 ビット、前処理と分析部での演算語長は基本的には 16 ビットで一部に 32 ビットを用いれば十分であることを示した。変換部は 24 ビット固

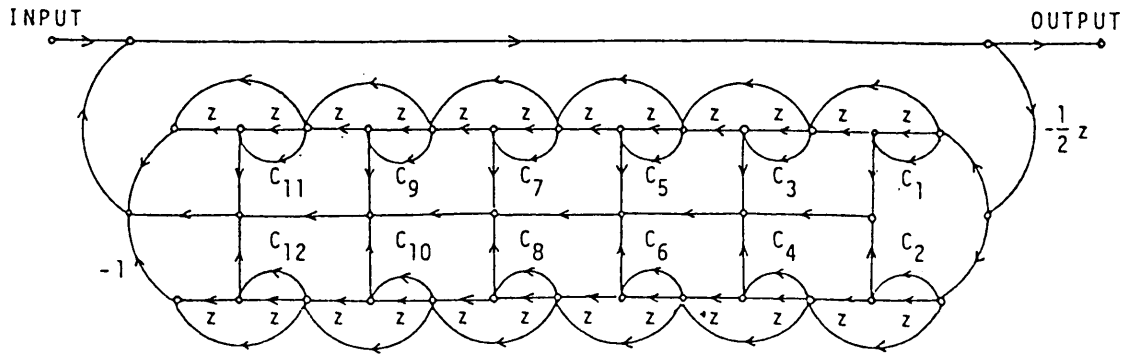


図 4.7: LSP 合成部の信号の流れ図 (次数  $p = 12$ )

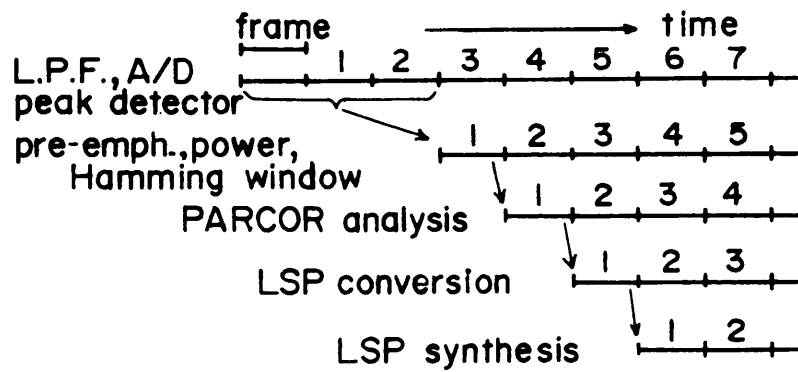


図 4.8: システム全体のタイミングの流れ図

表 4.8: システムの仕様 (最大標本化周波数 30kHz、窓長 30ms、計算量は最大値)

部分	処理	語長 (ビット)	フレーム当 りの計算量	メモリ数 (Kビット)	IC の数
前処理	低域ろ波, A-D 変換, プリアンファシス, パワー, 窓掛け, 最大値検出	16	加算 900 積和 1800	15(RAM) 32(ROM)	TDC1010 1 その他 129
分析	簡単化した変形格 子法による分析	16 16	積和 32400 除算 12	50(RAM) 16(ROM)	TDC1010 2 Z80 1 CDP1855 4 その他 227
変換	PARCOR $\rightarrow \alpha$ $\alpha \rightarrow$ LSP LSP $\rightarrow$ LSP'	24	積和 450 除算 52	54(RAM) 64(ROM)	Am2901 6 Am2904 1 Am2910 1 その他 120
合成	LSP 合成, LSP 係数補間	16	加算 1800 積和 1200	1(RAM)	TDC1010 1 その他 215
ダウン サンプル リング	D-A 変換 $\rightarrow$ 低域ろ波 $\rightarrow$ A-D 変換	12		15(RAM)	その他 20
その他	キー入力等				その他 60

定小数点演算で十分であること、Newton-Raphson 法での収束回数が各係数当たり 5 回であることを示した。合成部は 16 ビット固定小数点でよいことを示した。実時間で演算処理するため、各部をおのおの 10ms で処理し、パイプライン方式とした。入力された音声に変換されて出力されるまでに 50ms の時間の遅れがある。装置は IC が約 800 個で、幅 51cm、奥行き 40cm、高さ 35cm の箱に納まっている。この装置は実時間で動作することを確認した。次章でこの修復装置の評価を行う。

# 第 5 章

## ヘリウム音声修復装置 (1 号機) の評価

### 5.1 まえがき

本章では、前章の試作機による処理結果に基づいて、修復原理・修復方式および改良の指針など、実用化に不可欠な項目について検討する。

検討に当っては、

- 修復法の原理的な性能とその限界を知るために汎用のミニコンピュータによる修復結果を参照データとする
- 音声資料としては実際の 300m までの潜水模擬実験において収録したダイバーの音声を用いる
- 従来商品化されている装置と比較する
- ホルマント周波数の変化率を各周波数とも一定にする方式（線形変換）と周波数によって変化させる方式（非線形変換）のそれぞれの特徴と問題点を明らかにする。特に LSP パラメータを用いる方法に付随する問題を明確にする
- 明瞭度試験や了解度試験の結果から、修復の良好なものとそうでないものについて波形やスペクトルにより検討し、改良の指針を得る

という方針で行なった。

### 5.2 聴取試験

#### 5.2.1 音声資料

音声資料は 2 章で述べた内、1982 年 11 月 22 日～12 月 13 日に海洋科学技術センターで行なわれた 300m 有人潜水模擬実験の際に収録したものを使用した。4 人のダイバーのうち録音が比較的良好な話者 B による明瞭度試験用平等率音節表、単語了解度試験用 3 音節ワードリスト、文章了解度試験用リストを用いた。録音時の潜水深度は 0～300m ま

で10段階であるが、今回の評価実験には0m、87m、200m、300mの4段階での資料を使った。その通常空気に対する音速比と密度比を表5.1に示す。

図5.1に一例として単語（「ラジオ」、/rajiɔ/）について、深度300mのヘリウム音声と本装置による（非線形変換）修復結果、および同一発声者の通常の音声のスペクトルの時間変化を示した。

## 5.2.2 聴取試験の結果の概要

聴取試験は、聴取者12名（教官2名と学生10名）が無響室内でヘリウム音声の原音と修復音を聴取して書き取り、結果を計算機で集計した。使用スピーカはヤマハNS-450で、音量は平均70dB（騒音計A特性）に設定した。

表5.2は、音節明瞭度試験の結果を集計したものでa:原音声、b:試作した装置、研究室の汎用音声処理用ミニコンピュータMelcom-70（以下M70）によるシミュレーション、およびMalconi社製の修復器の3種の処理による各深度における12人の聴取結果の平均値である。各欄とも音節としての正答率と、その音節の母音部と子音部を分けた正答率を表示してある。図5.2は、この表から原音の聴取結果と試作した修復装置による結果をグラフにしたものである。

意味のある単語の了解度試験の結果の例として図5.3に3音節ワードリスト（単語数50個）による了解度を示す。図5.2と図5.3を比べると修復音は音節明瞭度よりも了解度が20%程度高くなっていることがわかる。また、300m潜水時の原音の単語了解度は20.5%で通話はほとんど不可能だが、修復音のそれは83%に向上し、この程度になれば通話可能である。この単語了解度試験は本実験の初期にヘリウム音声のような特殊な音に対しても単音明瞭度の改善が了解度の改善に寄与するかを確認しておくために行ったものである。この試験では87m、200mでは線形変換としたが、300mでは非線形変換とした。それは300mの音に対する非線形変換は5.4.2節で述べるように線形変換よりも明瞭度は低い音色が好ましいからである。結果からみれば、300mの音声を線形変換すれば了解度はこの値よりはいく分高くなることが推察される。

これらの聴取試験の数値には個人差があり、同一条件における最高の正答率を示した聴取と最低のそれとの間には20%を越える場合もあった。ただし、正答率の高い人はどの条件でも高く、低い人は常に低く、安定しているため、12人の聴取による正答率を平均化した数値には、十分意味があると考えられる。

表5.3に深度300mのM70によるシミュレーションの線形変換における子音毎の明瞭度を原音と修復音についてまとめたものを示す。ここで、子音は101音節表に従い、たとえば、/V/は5母音の前に子音がないものを、/t/はタ行の音を示し、/ky/はキャ行の音を表わし、/ky/のように後に/y/のついたものは拗音を表わすものとする。表5.3をみると非常によく修復されているものとあまり修復がうまくいっていないものがある。次節ではこれを、方式による明瞭度の違い、ヘリウム音声自体の子音部の脱落あるいはレベル低下、本装置の装置化上の問題点、非線形変換の実行上の問題点、修復音の音質などにわけて考察する。

表 5.1: 各深度におけるヘリウム空気の深度 0m の通常空気に対する音速比、密度比

深度 (m)	音速比 ( $k$ )	密度比 ( $r$ )
87	2.21	2.29
200	2.53	3.85
300	2.64	5.23

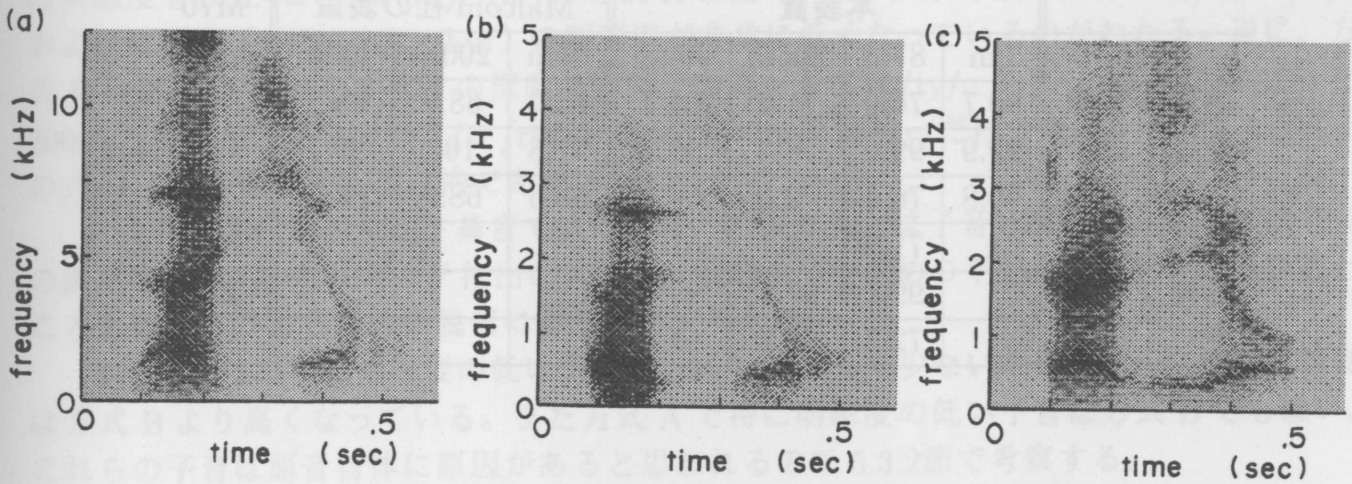


図 5.1: ラジオ (/rajiou/) という単語のサウンドスペクトルグラム

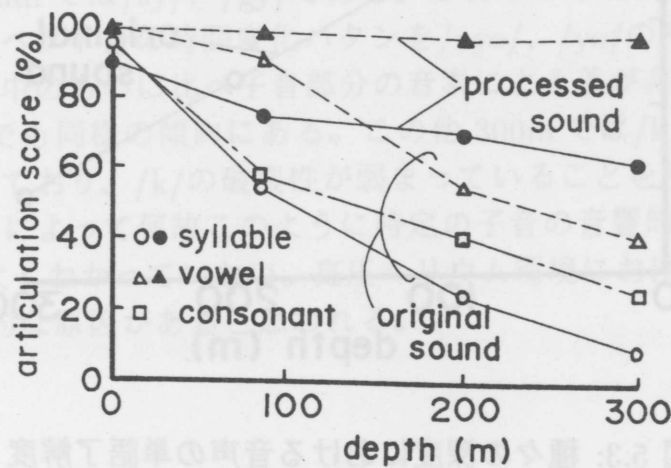


図 5.2: 種々の深度における音声の明瞭度

表 5.2: 各深度における原音の明瞭度と三つの方法による修復音の音節明瞭度

(a) 原音

深度	0m	87m	200m	300m
音節	96.8	54.5	24.6	9.8
母音	99.8	91.7	55.0	41.7
子音	97.0	58.2	41.3	25.2

(b) 修復音

深度		本装置				Malconi 社の装置			M70
		0m	87m	200m	300m	87m	200m	300m	300m
線形 変換	音節	90.7	76.0	71.2	64.2	65.0	68.9	66.2	74.2
	母音	99.9	99.0	98.3	99.2	97.8	100	99.7	99.2
	子音	90.8	76.4	72.0	64.4	65.6	68.9	66.2	74.5
非線 形変 換	音節		71.9	59.1	61.2				73.3
	母音		99.1	97.7	99.0				99.5
	子音		72.3	59.6	61.4				73.6

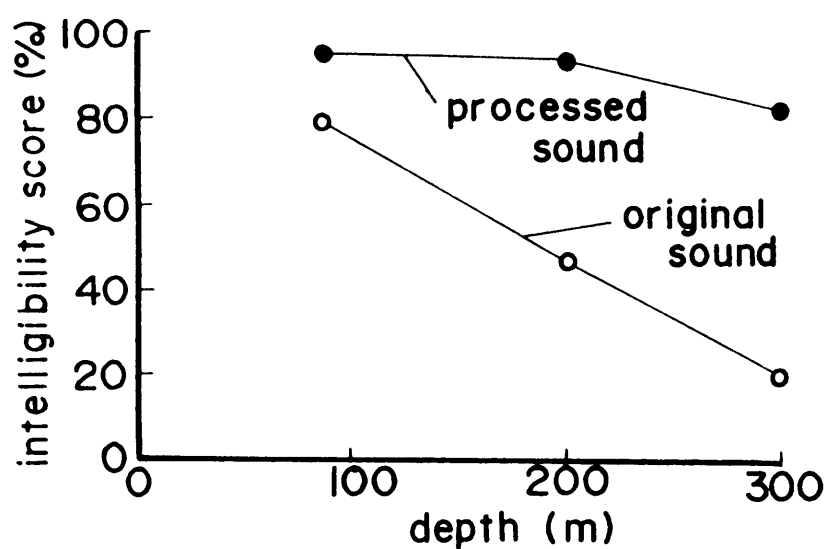


図 5.3: 種々の深度における音声の単語了解度



## 5.3 聴取試験による修復装置の評価

### 5.3.1 方式による明瞭度の違い

ここでは、本修復法（以下「方式 A」）とピッチ同期型波形伸長方式の代表として Malconi 社の修復装置（以下「方式 B」）とを比較する。

方式 A は、前章で述べたようにスペクトル包絡を LPC 分析で求め、それを変換し、残差を利用して再合成するものである。方式 B は図 5.4 に示すようにピッチに同期させて波形の一部を伸長しつなぎあわせるものである。ピッチのない部分では適当な周期で波形の一部を伸長しつなぎあわせる。

方式 A としては深度 87m と 200m では修復装置の結果を使用する。深度 300m では、5.4 節に述べるように装置化上の問題点も顕著になることがわかったのでここでは M70 による線形変換の結果を用いて方式の違いのみを明確にするようにする。表 5.2(b) に示すように、母音についてはどの方式でも 97% 以上であり、大きな違いはない。次に主な子音部の明瞭度を表 5.4 に示す。(a)、(a') の方式 A に比べ、(b) の方式 B では鼻音 (/m/、/n/) および破裂音 (/h/、/p/、/hy/) の明瞭度が非常に低くなっているのがわかる。逆に、方式 B が方式 A よりもすべての深度で明瞭度の高いものは /ty/、/y/、2 つの深度 87m と 200m で明瞭度が高いのは /t/、/ry/、87m と 300m では /ky/ だけであり、その差も 200m の /by/、/ry/ を除き 11% 以内である。

方式 B における誤りは、鼻音では破裂音への誤りまたは子音の脱落、破裂音では子音の脱落であり、これはピッチ抽出のエラー、波形のつなぎめの不連続、波形を間引いたところに破裂部がある場合の脱落によると考えられる。

方式 A では極端に明瞭度の低い子音が方式 B に比して少ないため、子音全体の明瞭度は方式 B より高くなっている。また方式 A で特に明瞭度の低い子音は方式 B でも低い。これらの子音は原音自体に原因があると思われるので 5.3.2 節で考察する。

### 5.3.2 ヘリウム音声の子音部の脱落

どちらの修復法によっても明瞭度が 50% 以下と低いものは深度 87m と 200m では /ry/、/gy/、深度 300m では /sy/、/gy/ である。これらの中で特に明瞭度の悪い /gy/ について /gyo/ のパワースペクトルの時間変化パターンを /nyo/、/yo/ のそれとともに図 5.5 に示す。図 5.6 の通常空気中のものに比べ子音部分の音素による差が非常に不明瞭になっている。これは他の深度でも同様の傾向にある。この他 300m では /k/ の明瞭度も 40% 台で、20% 以上が /h/ に誤っており、/k/ の破裂性が弱まっていることを示している。

環境の気圧や気体によって何故このように特定の子音の音響的な性質が不明瞭になるかについては、まだよくわかっていない。高圧ヘリウム環境における発声機構そのものに別に究明すべき本質的な原因があると思われる。

表 5.3: 深度 300m における音声の子音部の明瞭度マトリックス (原音と線形変換) (V:あ行音)

		修復音						
		10~19	40~49	50~59	60~69	70~79	80~89	90~100
原 音	0~9	gy				z,by,my	ry	w
	10~19				d,ky,hy	g	ty,zy	py
	20~29		k,sy		ny	n,y,r		t
	30~39				m	b	s	
	40~49					h		
	50~59							p
	60~69							
	70~79							V

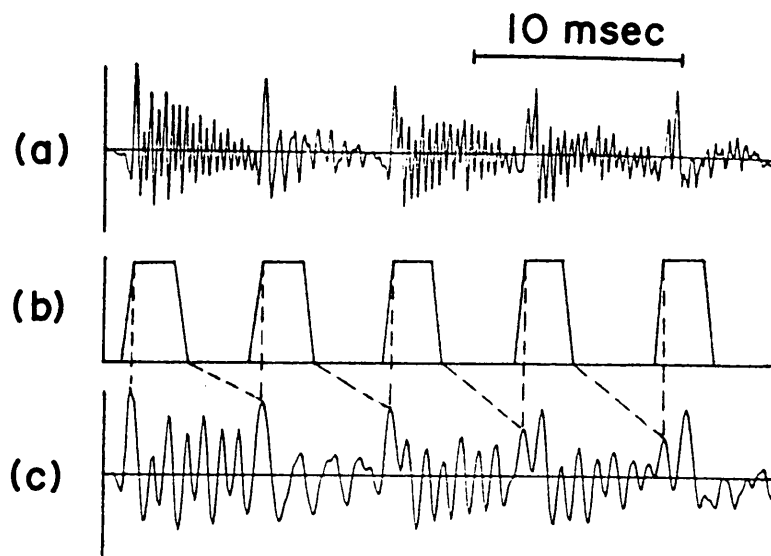


図 5.4: ピッチ同期波形伸長方式の一例 上から (a) 深度 300m で発声された原音/a/, (b) 波形を抽出する窓、(c) 修復音

表 5.4: 各深度における音声の主な子音部の明瞭度 (%)

(a) 本装置、(a')M70 によるシミュレーション、(b)Malconi 社の装置

深度	87m		200m		300m	
	(a)	(b)	(a)	(b)	(a')	(b)
/n/	58	45	77	57	70	43
/m/	82	50	80	50	67	58
/ny/	75	75	81	72	67	58
/my/	92	86	81	89	75	61
/p/	82	57	97	72	92	87
/t/	82	88	77	82	90	85
/k/	90	83	77	80	48	40
/h/	95	52	87	70	77	40
/py/	92	47	83	75	100	97
/ty/	86	94	83	89	86	97
/ky/	83	86	86	83	69	75
/hy/	75	31	58	31	64	36
/by/	56	50	50	89	75	72
/ry/	39	50	19	42	81	72
/y/	75	83	47	56	75	81

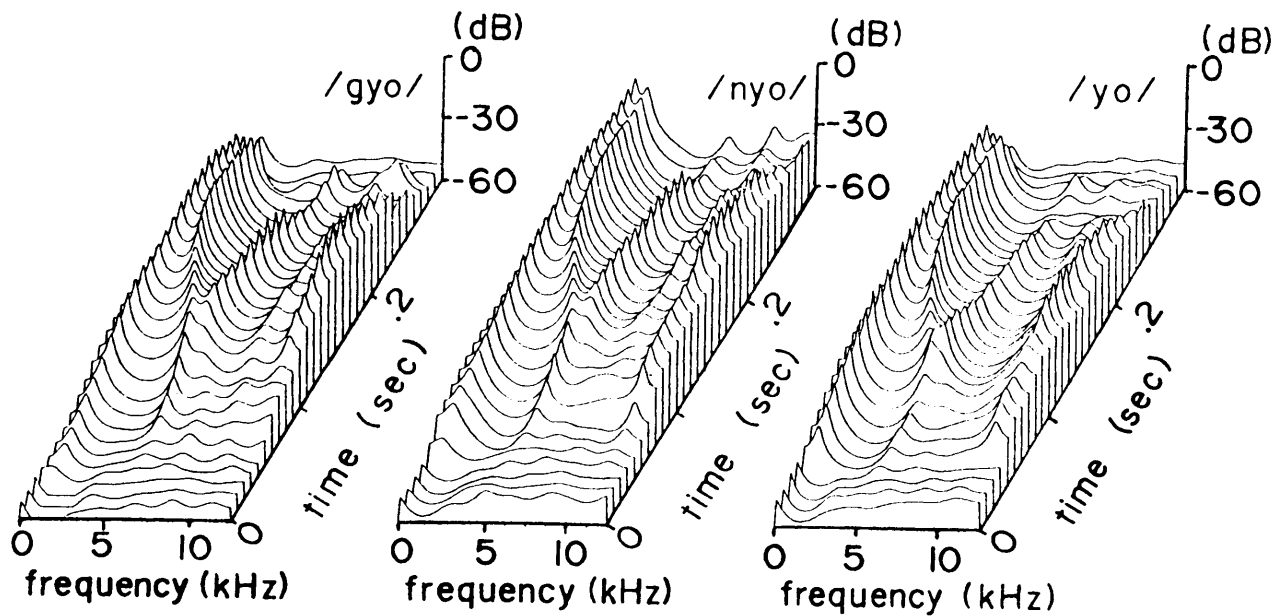


図 5.5: 深度 300m におけるヘリウム音声 /gyo/, /nyo/, /yo/ のスペクトルの比較

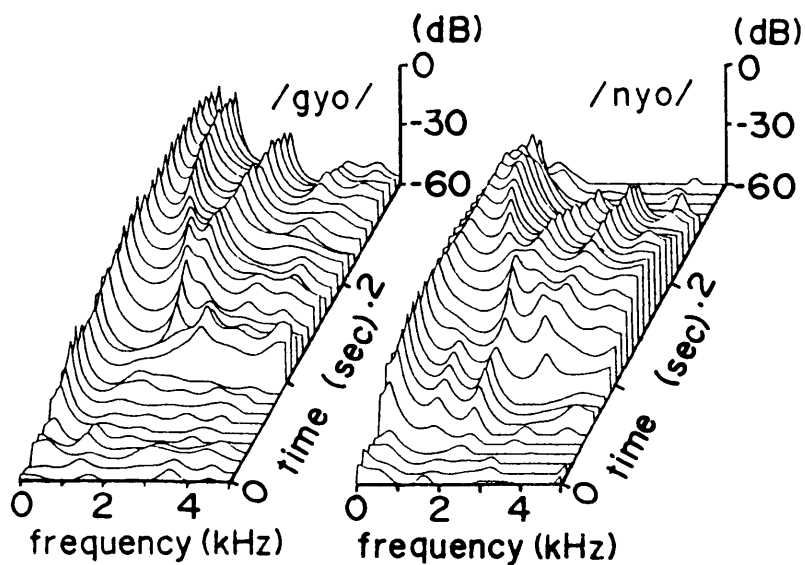


図 5.6: 通常空気中の音声の /gyo/ と /nyo/ のスペクトルの比較

## 5.4 装置化上の問題点

深度 300m では本装置とシミュレーションの両方の修復実験を行なったのでこれを用いて装置化上の問題を考察する。本装置は比較的高速な演算が要求されるため固定小数点演算方式を用いている。PARCOR 係数を求め、式 4.4 で変換する部分は 4.5、4.6 節で述べたようにシミュレーションの結果大きな問題はなかった。ただし、残差部についてはダウンサンプリングをアナログ回路を用いて行なっているため S/N 比劣化の問題がある。

300m の線形変換における結果を比較すると、本装置で特に明瞭度が低いものは鼻音と有声子音である。誤りは鼻音では有声破裂音に、有声子音では他の有声破裂音に誤聴される。入力を小さくしても装置の出力に有色の音が出ており、それは残差信号における信号対雑音比の低下が最大の原因である。また、深度 87m、200m の音声に比べ 300m の音声の修復後の明瞭度が修復装置で悪いのは、原音自体の信号対雑音比が低くなっていることと、更に上述の残差部での雑音の付加があるためである。

### 5.4.1 非線形変換の実行上の問題点 (声道最低共振周波数 $F'_{wa}$ について)

子音セグメントに対する明瞭度は先に表 5.2 で示したように、全体としては線形と非線形の変換方式による明瞭度の差は少ないが (深度 300m の音声の場合、M70 のシミュレーションでは 0.9%、試作装置では 3% 前者が高い)、子音の種類によってはかなり異なる。表 5.5 に 300m の音声に対してシミュレーションによる線形と非線形の方式で修復した場合の明りよう度の差が 5% 以上ある子音を示した。この表の結果から、第 1 ホルマント周波数が低い音である /by/、/m/ のようにうまくいっているものと、/zy/、/g/、/my/ のようにうまくいっていないものがある。非線形変換は原理的にはよいはずであるが、現修復法では十分でない。また、深度 87m、200m の修復装置による修復音において非線形変換の結果がよいものは、87m では /g/、/zy/、/by/、200m では /z/、/by/、/zy/ である。逆に悪いものは、87m では /d/、/ny/、/my/、/gy/、200m では /m/、/n/、/g/、/ny/、/my/ である。やはり、第 1 ホルマント周波数が低いものでもよいものと悪いものがある。これは声道最低共振周波数  $F'_{wa}$  の設定に関する問題と思われるので以下に考察する。

式 4.4 における声道最低共振周波数  $F'_{wa} > 0\text{Hz}$  は変換の非線形の程度を支配するパラメータである。 $F'_{wa} = 0\text{Hz}$  とすれば線形変換に帰着し、 $F'_{wa} > 0\text{Hz}$  で大きくすれば非線形の度合いが強くなる。図 5.7 は 300m ヘリウム音声 /yu/ の原音と、その修復音の波形の一部で、(b)~(f) は  $F'_{wa}$  を 0Hz から 180Hz まで変化させた場合である。原音の波形 (a) と (b) 以下の波形を比べると明らかにピッチ周期は同じでホルマント周波数が低下し、 $F'_{wa}$  が大きくなるとホルマント周波数がいっそう低下していることがわかる。しかし、(f) では波形に大きなうねりと各ピッチ区間内の波形の乱れが生じており、 $F'_{wa} = 180\text{Hz}$  では過剰修正になってしまうことが示されている。これは後で述べるように非線形変換においては LSP 係数の角周波数に下限の値が設定されているため、 $F'_{wa}$  が高いと最低次の LSP 係数の角周波数が 0 または 0 近くになり、強い超低周波成分ができるためである。従って、音色がその発声者の声らしく、且つ明瞭度の高い音を得るには  $F'_{wa}$  を適当に低いある値に設定する必要がある。

さて、変換の式 4.4 の右辺は平方根の演算であるが、ホルマント (ここでは LSP) 周波

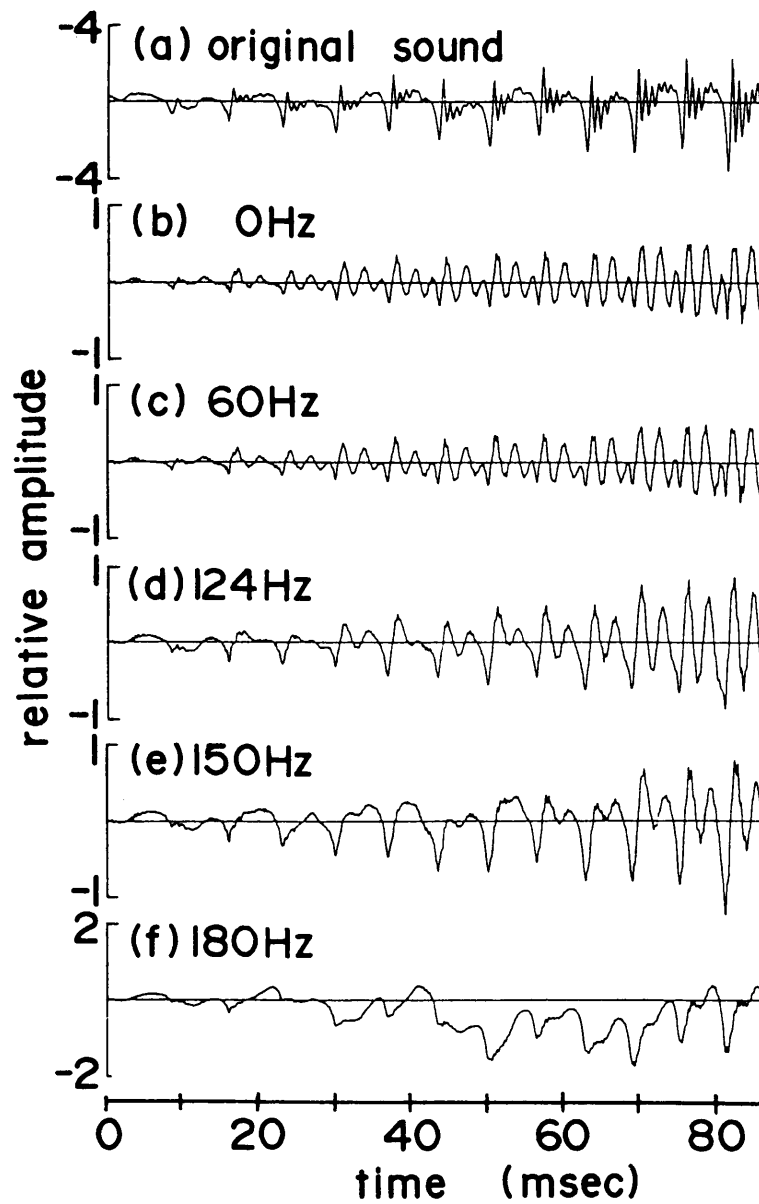


図 5.7: 深度 300m におけるヘリウム音声の音節 /yu/ とその修復音の立上り部分の波形  
 (a) 原波形、(b) 線形変換 ( $F'_{wa} = 0\text{Hz}$  に対応)、(c)~(f)  $F'_{wa} = 60\sim 180\text{Hz}$  の非線形変換

数は非負の実数だから平方根の中が負になる事は許されない。しかし、実際には LSP 係数の値が何らかの原因によって低く出て、平方根の中が負になることはあり得る。この現象は修復時の  $F'_{wa}$  が高い程起こりやすい。また入力として周波数が低くバンド幅の広いスペクトルの山を持つ音声や雑音が入った場合、LSP 係数の値はその山の両側に対になって現われることが多いので、最低次の LSP 係数の角周波数はかなり低くなり、この現象が起り得る。従って、非線形変換の式を使う方式ではこの現象への対策が必要である。この対策として、あるフレームで平方根の内部が負になった場合、(a)：そのフレームの最低 LSP 係数の角周波数を 0 に代替する（他の LSP 係数はそのフレームの値をそのまま用いる）、(b)：そのフレームの LSP 係数全部を直前（10ms 前）のフレームの LSP 係数で代替する、の二つの方法について検討した。図 5.8 は深度 300m のヘリウム音声の /no/ について、両対策の効果を比較したもので、(a)、(b) はそれぞれ対策 (a)、(b) による出力信号、(c) はこの音に対する残差信号である。

この図からわかることは、第 1 に、対策 (b) の方法では、出力波形が遅れて立ち上がる事である。合成部は次数 12 の LSP 合成フィルタなので、フィルタリングの遅れ時間は 1.2ms 程度はあるが、(b) の波形は残差信号に対して 30ms も遅れている。これは、/no/ の立ち上がり部分の LSP 係数の値が低過ぎたために、変換式の平方根の中が負になり、一つ前のフレーム、すなわち無音区間のパラメータで代替されたためにピッチがはっきりしない。フレーム間隔は 10ms なので、音声の立ち上がりで 3 フレーム分だけエラーが生じた事になる。このような対策では、第 1 LSP 係数が低く推定された音は大きく歪むことになる。特に音声の立ち上がりでこの現象が起ると修復音では語頭の子音部が短縮・脱落あるいは歪みが生じて明瞭度が損われる。

一方、対策 (a) に対応する波形 (b) は立ち上がりの遅れはあまり無く、従って、この場合語頭の子音区間の短縮あるいは脱落は少ない。

このような検討から、試作装置では対策 (a) を施した。先に述べた子音明瞭度 61.2% はこの結果である。対策 (b) を施した時の明瞭度は 58.2% であった。

LSP 係数の変換に関して上述した事をまとめると、非線形変換方式には原理的な優位性があるが、実際は種々の原因によって変換式の平方根の内部が負あるいは 0 に近くなるような入力信号に対して、大きな歪や雑音、あるいは語頭区間の脱落などの欠陥が生ずる。従って、 $F'_{wa}$  の値を理論的な声道最低共振周波数よりも低く設定するか、あるいは  $F'_{wa} = 0\text{Hz}$  すなわち線形変換が無難である。300m 以上の深々度における音声の修復は非線形変換が是非とも必要になるが、その場合は上述の問題に対する一層の検討を要する。

## 5.4.2 修復音の音質（個人性の再現性）

明瞭度については上述の通り線形と非線形の変換方式による差は小さいかむしろ非線形の方が悪い。しかし、 $F'_{wa}$  をうまく設定すれば明瞭度は下がらない。一方音質あるいは発声者の声の特徴の再現性という観点から評価すると、明らかに非線形変換の方がすぐれている。これは非線形変換の方が低次ホルマント周波数をより正確に修復できるからで、非線形変換による修復音の方が、声に柔らか味があり、発声者の声質がよく聞きとれ、話者が識別できる。この事は実際の潜水作業では大変重要な事であり、今回の実験とは別にヘリウム音声を聞きなれているダイバーによる評価実験をしたところ特に非線形変換の

表 5.5: 深度 300m の線形変換と非線形変換（いずれもシミュレーション）における子音部の明瞭度（単位%、変換方式による差が5%以上のもの）

子音部	線形変換	非線形変換
/by/	75	86
/k/	48	57
/y/	75	83
/m/	67	75
/p/	92	97
/z/	77	82
/d/	67	61
/my/	75	67
/g/	72	50
/ry/	81	58
/zy/	89	63

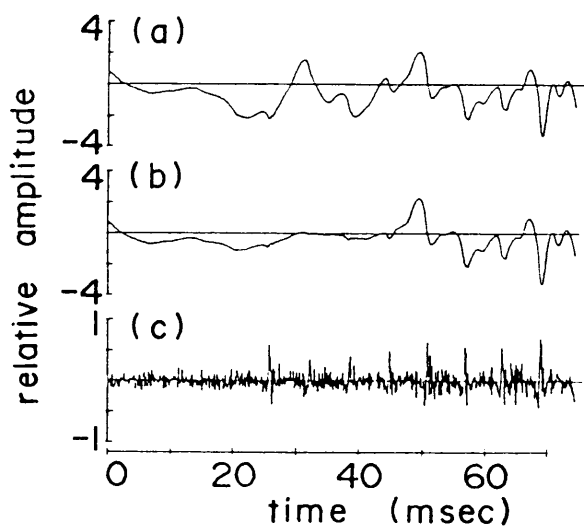


図 5.8: 非線形修復における LSP 係数補正方式の違いによる修復音の比較

(a): そのフレームの最低 LSP 係数の角周波数を 0 に代替する（他の LSP 係数はそのフレームの値をそのまま用いる）、(b): そのフレームの LSP 係数全部を直前（10ms 前）のフレームの LSP 係数で代替する、(c): 残差波形



修復音の方が好感を持たれた。

## 5.5 まとめ

本章では筆試作したヘリウム音声修復装置（1号機）を評価するために、修復に用いる変換方式と実時間動作のデジタル装置としての設計製作上の問題点、およびヘリウム音声の物理的性質などについて検討した。その結果、ここで採用した修復法によって深度 300m 程度までの音声は了解度、声質とも実用上ほぼ十分な程度に修復できることがわかった。ただし、子音の一部にはまだ不十分なものがある。潜水深度 300m 以浅では、周波数軸の線形変換の方が非線形変換よりも明瞭度がやや優れているが音質・個人差の修復には非線形変換の方が優れていることがわかった。潜水深度が 300m よりも更に深くなれば、ヘリウム音声の非線形性が增大するため非線形変換が必要になる。その場合バズバー、鼻音等の第 1ホルマントが低く求まるものでは LSP 係数による方法ではうまく修復できにくいので変換公式の適用のしかたを更に工夫する必要がある。子音の修復効果を高めるには、ヘリウム音声の性質をホルマント周波数以外の面からも更に明らかにしなければならない。

## 第 6 章

# DSP を用いたヘリウム音声修復装置 (2 号機)

### 6.1 まえがき

LSP パラメータの変換を含む分析・合成系によりヘリウム音声を修復する方式の開発と評価を前章までに述べた。ここでは、実際に深々度の潜水作業時の通信の場で使用できる装置の開発に重点をおき、前システムのアルゴリズムの改良と小型化を目的としてデジタル信号用プロセッサ (DSP) を用いた装置の設計と試作について述べ、次に修復上の問題点についてもふれる。

### 6.2 LSP 変換方法の改良

ヘリウム音声と通常の大気中の音声との違いは、声道伝送特性と音源特性の両方にあるが、前者の方が顕著である。声道伝達特性はアンチホルマントもあるがほぼホルマント周波数と帯域幅によって規定され、4.2節で述べたようにホルマント周波数については Fant らによる次の関係式で近似されることが知られている [5]。

$$F_a = \sqrt{(F_h/k)^2 - (r-1)F_{wa}^2} \quad (6.1)$$

ここで、  
 $F_h$  : ヘリウム音声のホルマント周波数  
 $F_a$  : ヘリウム音声と同一の声道形状で発声したときの  
通常空気中におけるホルマント周波数  
 $F_{wa}$  : 通常空気中における声道の最低共振周波数  
 $k$  : ヘリウム空気の通常空気に対する音速比  
 $r$  : ヘリウム空気の通常空気に対する密度比

ヘリウム音声の修復するには入力音声のホルマントを式 6.1 に従って低下させ、出力音声の

ホルマントを  $F_a$  になるように変換することが最も重要である。ヘリウム音声の音源となるピッチ周波数は 3.2 節で述べたように深度 300m においても高々 30% 高いだけなので基本周波数は原音のままかまわない。従って、修復のためには残差駆動型のボコーダを用い、ヘリウム音声を分析してホルマント周波数を抽出し、それを式 6.1 に代入して  $F_a$  を求め、残差信号を音源にして再合成すればよい。

前章までにこの変換を具体的に行う方法として式 6.1 のホルマント周波数を LSP 係数の角周波数で置き換えることによって非線形変換を行なう方式を提案した。しかし、この方法では前章で述べたように低域で帯域幅が広いとホルマント周波数が下がり過ぎる欠点があった。今回はこの点を改良するために以下のように変更することとした。

1. 第  $i$ LSP 係数の角周波数  $\omega_{hi}$  を次式で変換し、 $\omega_{ai}$  を求める ( $i \neq 1$ )。

$$\omega_{ai} = (f/f') \sqrt{(\omega_{hi}/k)^2 - 4\pi^2(r-1)(F'_{wa}/f)^2} \quad (6.2)$$

ここで、 $f$  は分析時の標本化周波数 (Hz)、 $f'$  は合成時の標本化周波数 (Hz)、 $F'_{wa}$  は変換時に用いる声道の最低共振周波数 (Hz) である。

2. ただし、第 1LSP 係数については以下のように変換する。 $\omega_{h1}$  と  $\omega_{h2}$  の平均値：

$$\omega'_h = (\omega_{h1} + \omega_{h2})/2 \quad (6.3)$$

を求め、この  $\omega'_h$  を式 6.2 の  $\omega_{hi}$  に代入し、その変換後の値を  $\omega'_a$  とする。

3. この値と  $\omega_{a2}$  を用いて、

$$\omega_{a1} = 2\omega'_a - \omega_{a2} \quad (6.4)$$

を求める。

線形変換の場合は  $F'_{wa} = 0$  Hz であるから式 6.2 は  $\omega_{ai} = (f/f')(\omega_{hi}/k)$  となり第 1LSP 係数は式 6.2 で直接求めたものと式 6.3、6.4 の変換を行なったものとは同じ値になる。

この変換誤差を次に評価する。表 6.1 にこの評価に用いる各深度における通常空気に対する音速比、密度比、表 6.2 に 5 母音と中性母音のホルマント周波数を示す。

ヘリウム音声の第  $j$  ホルマント周波数を  $F_{hj}$ 、帯域幅  $B_{hj}$  とし、空気中での第  $j$  ホルマント周波数を  $F_{aj}$  とするとき、表 6.1、表 6.2 の音速比、密度比、ホルマント周波数を用いて次の式 6.5 を計算する (単位は Hz)。帯域幅については詳しいデータがないので、ここでは通常空気の場合と同様に以下の関係式 6.6 が成り立つものとして取り扱う。

$$F_{hj} = k \sqrt{F_{aj}^2 + (r-1)F_{wa}^2} \quad (6.5)$$

$$B_{hj} = 0.03F_{hj} + 35 \quad (6.6)$$

$F_{hj}$ 、 $B_{hj}$  より  $\alpha$  係数を計算し、第  $i$ LSP 係数の角周波数  $\omega_{hi}$  を求める。

これらを式 6.2、6.3、6.4に代入し変換後の LSP 係数 $\omega_{ai}$ を求める。  
LSP 係数 $\omega_{ai}$ より $\alpha$ 係数を計算し、ホルマント周波数 $F'_{aj}$ を求める。

この変換による第  $j$  ホルマント周波数の誤差 $\Delta F_j$ を次のように定義する。

$$\Delta F_j = F'_{aj} - F_{aj} \quad (6.7)$$

この $\Delta F_j$ によって誤差を評価する。

深度 300m における 6 母音について改良前と改良後の第 1 ホルマント周波数の誤差 $\Delta F_1$ を求めたのが表 6.3 である。 $\Delta F_1$ が負符号になっているのは目標値よりも低くなり過ぎていることを示す。これより、第 1 ホルマント周波数の低い/i/、/u/での誤差が今回の方式では前方式より小さくなることがわかる。ただし、第 1 ホルマント周波数の高い/a/ではほんの少しだが悪くなっている。

次に、/i/の深度 300m における第 1 ホルマントの帯域幅のみを式 6.6によらず種々変えた時の変換後のホルマント周波数と帯域幅を改良前を表 6.4 (a) に、改良後を表 6.4 (b) に示す。変換時の声道最低共振周波数 $F'_{wa}$ を 200Hz、190Hz、124Hz、0Hz とすると、目標値 $F_{a1}$ はそれぞれ 250Hz、281Hz、408Hz、481Hz となる。一例としてヘリウム音声/i/の第 1 ホルマントの帯域幅を 200Hz とし、 $F'_{wa}$ を 200Hz とすると、改良前ではホルマント周波数が 221Hz、帯域幅が 89Hz に対し、改良後はそれぞれ 242Hz、77Hz となることを示している。よって、ホルマント周波数は改良後の方が目標値 250Hz に近く、帯域幅も改良後の方が $B_{h1}/k$ に近い。これは $F'_{wa}$ が 0Hz でない非線形変換で一般的に言えることである。また、 $F'_{wa}$ が 0Hz である線形変換では改良前の方法とまったく同じ値になるので改良前と同じく帯域幅に関係なく変換後のホルマント周波数は $F_{h1}/k$ であり、帯域幅も $B_{h1}/k$ であり変換誤差はほとんどない。

このように誤差が減る理由は、分析次数を 12 とすると、第 1 ホルマント周波数では第 1LSP 係数と第 2LSP 係数がこのホルマント周波数を挟み、しかも平均値が第 1 ホルマント周波数に近くなるからである。

## 6.3 その他の改良点

上述の他、次に示す点を考慮して改良を行った。

1. 分析時の標本化周波数は先のシステムで予備的に行った試聴実験の結果、深度に応じて 15、20、25、30kHz 程度に大まかに変更しても明瞭度は大差なかったため、今回は 200 から 500m 程度の深々度を想定して標本化周波数は 25kHz に固定した。
2. 合成時に音源として使用する残差は先のシステムのようにアナログ信号で行うと雑音が混入しやすいので LPC 係数より逆フィルタリングした残差をデジタルフィルタでダウンサンプルすることにする。
3. 装置の試作にあたってはシステムがなるべく小型化されるように積和が高速にできる固定小数点演算のデジタル信号用プロセッサ (DSP) を使用することにする。固

表 6.1: 各深度におけるヘリウム空気の深度 0m の通常空気に対する音速比、密度比

深度 (m)	音速比 ( $k$ )	密度比 ( $r$ )
100	2.27	2.47
200	2.53	3.85
300	2.64	5.23

表 6.2: 深度 0m の通常空気における 5 母音と中性母音のホルマント周波数 (Hz)  
 ただし、 $F_4 = 3500\text{Hz}$ ,  $F_5 = 4500\text{Hz}$

	/i/	/e/	/a/	/o/	/u/	/ɔ/
$F_1$	250	450	800	500	330	500
$F_2$	2100	1900	1200	700	1000	1500
$F_3$	3000	2500	3000	3000	2200	2500

表 6.3: 深度 300m におけるヘリウム音声の第 1 ホルマント周波数の変換による誤差  $\Delta F_1$

	/i/	/e/	/a/	/o/	/u/	/ɔ/
改良前	-10.0	-5.8	-3.1	-5.7	-8.0	-5.3
改良後	-5.0	-4.4	-3.2	-4.9	-5.2	-5.2

表 6.4: LSP 係数の変換方式の改良前 (a) と改良後 (b) の深度 300m のヘリウム音声 /i/ の第 1 ホルマント周波数 ( $F_{hi}$ ) の帯域幅  $B_h$ (Hz) と LSP 変換における通常空気中での最低共振周波数  $F'_{wa}$ (Hz) のみを種々変えたときの変換後の第 1 ホルマント周波数と帯域幅 (括弧内) の値 (Hz)

\*  $F_{a1} = \{(F_{h1}/k)^2 - (r - 1)F'_{wa}\}^{1/2}$

\*\*  $B_{h1}/k$  ここで、 $k = 2.64$ 、 $r = 5.23$  である

(a) 改良前

$B_{h1}$	$F'_{wa}$	200	190	124	0
	$F^*_{wa}$	250	281	408	481
50	(19)**	243 (22)	276 (22)	407 (20)	481 (19)
73	(27)**	241 (32)	273 (32)	406 (29)	481 (27)
100	(37)**	236 (44)	270 (44)	405 (40)	481 (37)
150	(56)**	229 (67)	264 (65)	404 (60)	481 (56)
200	(75)**	221 (89)	258 (87)	402 (80)	481 (75)
250	(93)**	212 (111)	252 (109)	400 (100)	481 (94)
300	(112)**	203 (134)	245 (132)	399 (120)	481 (112)

(b) 改良後

$B_{h1}$	$F'_{wa}$	200	190	124	0
	$F^*_{wa}$	250	281	408	481
50	(19)**	246 (20)	278 (20)	407 (19)	481 (19)
73	(27)**	245 (29)	277 (29)	407 (28)	481 (27)
100	(37)**	244 (39)	275 (39)	406 (38)	481 (37)
150	(56)**	243 (58)	274 (58)	405 (57)	481 (56)
200	(75)**	242 (77)	273 (77)	405 (76)	481 (75)
250	(93)**	242 (95)	273 (95)	404 (95)	481 (93)
300	(112)**	241 (114)	272 (114)	403 (114)	481 (112)

定小数点演算のみの DSP では切り捨て等の演算誤差に注意する必要があるが、1号機ですでに検討してある事項もあり多少のアルゴリズムを付加すれば解決できる。また、このシステムでは外部メモリとのやり取りが多くなることから外部メモリとの読み書きが速い富士通製の MB8764 を用いた (1983 年に開発された IC で、開発当時は最高速の DSP)。

MB8764 の主な仕様を表 6.5 に示す。16 ビット × 16 ビットの乗算が 100ns で可能であるがその結果は上位の 26 ビットである。外部メモリとの読み書きも速く 100ns でできる。ただし、割り込み機能がない。

このようにして改良したシステムのブロック図を図 6.1 に示す。

次節以下ではこのシステムについて改良点も含めて述べ、その改良結果をスペクトルひずみを尺度として評価する。

### 6.3.1 前処理部

分析部での切り捨て誤差を減らすため、前処理としてフレームごとに最大値を求めて分析部で用いるデータがいつもほぼ同じ大きさになるようにする。先のシステムではハードウェアの制約のためプリエンファシス後の窓掛け前の値で最大値検出をしていたのでフレームの始めまたは終りのみが大きいような信号に対して分析誤差が大きかった。そこで今回はプリエンファシス後、窓をいったん掛けてから最大値を求めて左にシフトする量を決め、分析部へのデータはその値によって左シフトし窓掛けを行うことで各フレームの最大値がほぼ同じになるように改良した。

### 6.3.2 分析部

分析には線形予測法を用いるがこれにもいくつかの方法がある。そこで、DSP を用いるのに適した方法を求めるため代表的な三つについて考察した。まず、抽出できる情報、演算回数、メモリ数等について表 6.6 に示す。ここで  $N$  を 750 としたのはフレーム長が 30ms、分析時の標準化周波数が 25kHz だからである。また、分析次数  $p$  を 12 としたのは音速比を  $k$  とするとホルマントがほぼ  $k$ kHz ごとに一つ存在することになるからである。DLI 法 [39] は PARCOR 係数と  $\alpha$  係数が一度に求まるが、 $\alpha$  係数の値が 3 以上になることがあり固定小数点ではオーバーフローの問題があるので候補から外した。Le Roux 法 [40] は、DLI 法より多少演算数が多いが、得られるのが PARCOR 係数だけなので 1 以下が保証されており DLI 法より演算精度はよい。しかし、残差波形を得るには別の計算が必要になりアルゴリズムが多少複雑である。ただし、今回は DSP を使用しているのでアルゴリズムは多少複雑でもよい。単純化した変形格子法 [41] はアルゴリズムは単純な繰り返しが多く、残差波形も得られるが、演算数と演算に必要なメモリ数が多いのが欠点である。このように両方法とも一長一短であるので、さらにもう一つの選択基準としてスペクトルひずみを考え、Le Roux 法と単純化した変形格子法のうちスペクトルひずみの少ない方のアルゴリズムを選ぶことにする。

浮動小数点演算で求めた同方法の PARCOR 分析による LPC 包絡  $T_1(z)$  と、この固定小数点による PARCOR 分析による LPC 包絡  $T_2(z)$  から以下のようなスペクトルひずみ

表 6.5: DSP-MB8764 の主な仕様

演算語長	16 ビット
(積和は $16 \times 16 \rightarrow 26$ )	
内部 RAM	$2 \times 128 \times 16$ ビット
外部 RAM	1k ビットまで
積和、加減算	1 マシンサイクル
除算	17 マシンサイクル
1 マシンサイクル = 最小 100ns	

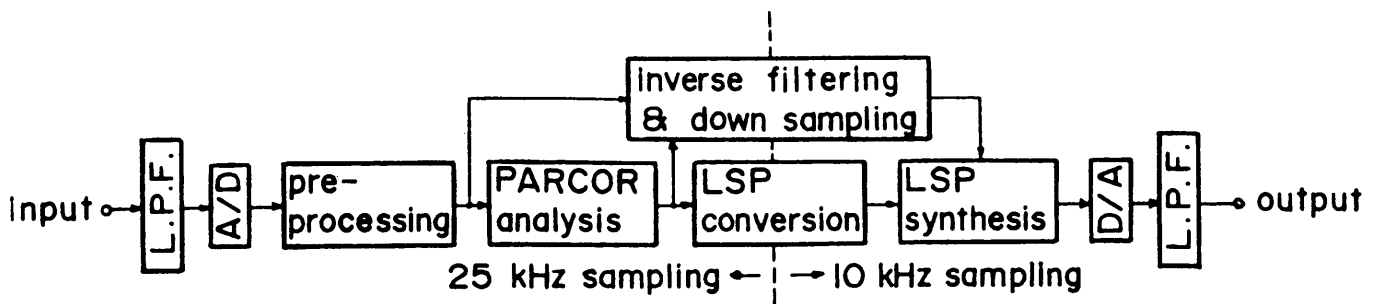


図 6.1: 改良したヘリウム音声修復装置のブロック図



$DS[42]$  を求める。

$$DS = \left\langle \left( 10 \log_{10} \left| \frac{T_2(z)}{T_1(z)} \right|^2 \right)^2 \right\rangle_{\omega, t} \text{ dB}^2 \quad (6.8)$$

ただし、 $z = \exp(j\omega T)$ 、 $T$  : サンプルング周期  $\langle \quad \rangle$  は  $\omega$ 、 $t$  にわたる長時間平均を表わす。

潜水深度 300m におけるヘリウム音声データ「二千六百四十八と書け」(約 2 秒) を用いてスペクトルひずみを求めると、プリエンファシス後の窓掛け前の値で最大値検出の場合 Le Roux 法で  $1.38 \times 10^{-3} \text{dB}^2$  であり、簡単化した変形格子法では  $7.5 \times 10^{-2} \text{dB}^2$  なのでスペクトルひずみの少ない Le Roux 法を用いることにする。また、6.3.1 節で述べたように窓掛け後に最大値検出をし、その値で正規化すると Le Roux 法で  $8.9 \times 10^{-4} \text{dB}^2$  となり、さらにスペクトルひずみは減る。

### 6.3.3 逆フィルタリングとダウンサンプリング部

ここでは、残差を求めるために音声を逆フィルタリングし、次に、分析時の標本化周波数 25kHz に対し合成時の標本化周波数は 10kHz なので、その残差波形をダウンサンプリングする。後者の方法は次の通りである。残差の値と交互に 0 をいれて標本化周波数 50kHz のデータとし、遮断周波数 4.8kHz の FIR の低域フィルタをかけ、そのデータを 5 点につき 1 点にする間引きを行い標本化周波数 10kHz のデータとする。ただし、FIR フィルタの出力は入力に依存しているだけなので合成部に必要な 10kHz の標本化データのみ計算すればよい。また、FIR フィルタの入力は 1 つおきに 0 なので、これも計算しなくてよい。

遮断周波数 4.8kHz の FIR フィルタは文献 [44] のプログラムにより設計した。このフィルタの係数の値を表 6.7 に、また、その周波数特性を図 6.2 に示す。遮断領域で約 40dB の減衰量であり、音声のダイナミックレンジを考慮すれば十分である。タップ数は 32 であるが、入力は 1 つおきに 0 なので実際の積和数は 1 サンプルにつき 16 であり、1 フレーム (10ms) では  $16 \times 100 = 1600$  の積和数となる。

ここでの入力データは前処理部で 1 フレームごとに正規化したデータのままであり、それを逆フィルタリングし、FIR フィルタに通すので、フレームの切れ目で前後のシフト量に応じてフィルタ内のデータをシフトし大きさをあわせる操作をしている。このようにすれば切り捨て誤差は減る。

### 6.3.4 LSP 係数変換部

分析部で PARCOR 係数が求まっている。よって、この部分では LSP 係数を求め、それをさらに 6.2 節で述べたように変換する。この計算は前システムとほとんど変わらないので 4.6 節を参照されたい。

表 6.6: 3 種類の線形予測分析の 1 フレーム当たりの演算回数等の比較  
 (分析次数  $P = 12$ 、窓長  $N = 750(30\text{ms})$ 、フレーム周期 10ms)

\* Durbin-Levinson-Itakura 法

比較項目	DLI 法*	Le Roux 法	簡単化した変形 格子法
抽出できる情報	PARCOR 係数, $\alpha$ 係数, 残差パ ワー	PARCOR 係数, 残差パワー	PARCOR 係数, 残差波形
除算	$12(p)$	$12(p)$	$12(p)$
積和	$168(p^2 + 2p)$	$144(p^2)$	$27762(3Np + N)$
相関 (積和)	$9750(Np + N)$	$9750(Np + N)$	
残差波形 (積和)	$3000(Np/3)$	$3000(Np/3)$	
メモリ語数	$803(N + 4p + 5)$	$799(N + 4p + 1)$	$1512(2N + p)$

表 6.7: ダウンサンプリング用低域フィルタ (FIR フィルタ) の係数

$$\begin{aligned}
 h(0) = h(31) &= -0.006744 & h(8) = h(23) &= -0.029616 \\
 h(1) = h(30) &= -0.000137 & h(9) = h(22) &= -0.041524 \\
 h(2) = h(29) &= 0.005800 & h(10) = h(21) &= -0.033798 \\
 h(3) = h(28) &= 0.014032 & h(11) = h(20) &= -0.000522 \\
 h(4) = h(27) &= 0.020462 & h(12) = h(19) &= 0.055601 \\
 h(5) = h(26) &= 0.020153 & h(13) = h(18) &= 0.122631 \\
 h(6) = h(25) &= 0.009956 & h(14) = h(17) &= 0.182718 \\
 h(7) = h(24) &= -0.008878 & h(15) = h(16) &= 0.218218
 \end{aligned}$$

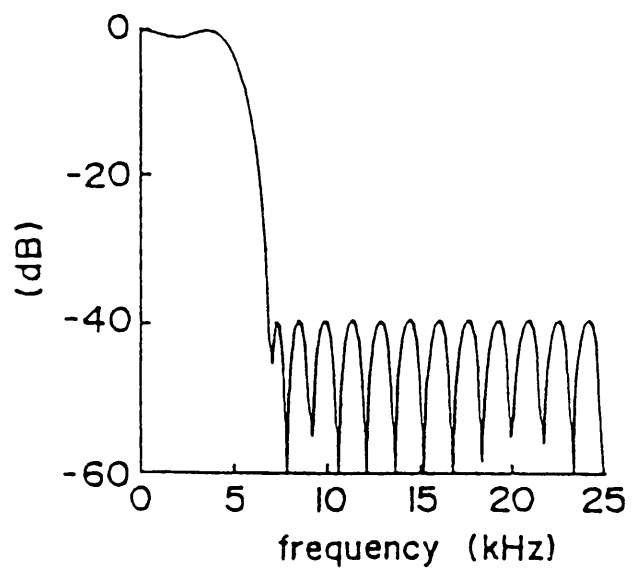


図 6.2: ダウンサンプリング時に使用される FIR フィルタの周波数特性

### 6.3.5 LSP 合成部

ここでは、主な変更点についてのみ述べる。

- (a) 残差信号はシフトしたまま合成し、最後に最初にシフトしたビット数だけ右シフトする。この理由は次の通りである。図 6.3に示すように、合成部の入力を  $n$  ビット左シフト後、合成し  $n$  ビット右シフトすると、 $n = 1, 2, \dots, 8$  と大きくしてゆくにつれて 32 ビット浮動小数点演算と波形が似てくる。ただし、 $n$  をあまり大きくし過ぎるとオーバーフローしてしまう。そこで、適当な  $n$  を選ぶ必要がある。もし、分析部と合成部はともに標本化周波数 10kHz で分析した係数と残差で合成すれば、 $n$  は前処理で決めた値を使用すれば波形はほぼ元の値になりオーバーフローは起こらない。ただし、今回の場合は分析部の標本化周波数が 25kHz でしかも変換をしているので、計算機シミュレーションによって深度 300m のヘリウム音声の 100 音節を修復してみた。その結果オーバーフローは起きず、浮動小数点の結果と同様に音声の立ち上がり部分からピッチが十分に確認されたので、これを採用した。
- (b) 1 フレームを分析して求まる LSP 係数はフレームの平均的な値なので、フレームの中央でこの値であるとして各点の LSP 係数を補間することにした。
- (c) このようにしてもシステムへの入力データによっては合成部でオーバーフローもありうる。LSP 合成は IIR フィルタであるから、いったんオーバーフローが起こるとオーバーフローが起き続ける。そこでオーバーフローのエラーがおきた時は合成フィルタ内のデータを強制的に零クリアすることにした。

## 6.4 回路構成

図 6.4にシステム全体のタイミングの流れ図を示す。

また、図 6.5にシステムのハードウェアの概略図を示す。計算量が多いので、大きく 5 つの部分に分けパイプライン方式とした。このため各部にメモリが必要になるが演算速度自体はこの DSP で十分である。また、出力される音声は入力よりも 55ms 遅れることになる。

表 6.8に各ブロックでの演算量、使用 IC 等の仕様を示す。システム全体で IC の数は 210 個程度、450mm × 320mm の基板 1 枚におさまっている。これは前システムの IC の個数の 1/4 である。この中には開発用のインターフェイスの IC を含んでいるので、プログラムメモリの内容を固定して ROM にすれば IC の数は 160 個程度におさまりさらに小型化可能である。

## 6.5 ヘリウム音声の修復の結果と考察

修復装置の聴取試験による評価として、単語了解度試験を行なった結果について述べる。この試験のリストには 2 章で述べた 3 音節ワードリストを使用した。音声は周波数帯域が 12.5kHz までほぼ平坦であることが確認されているエレクトレット型コンデンサマ

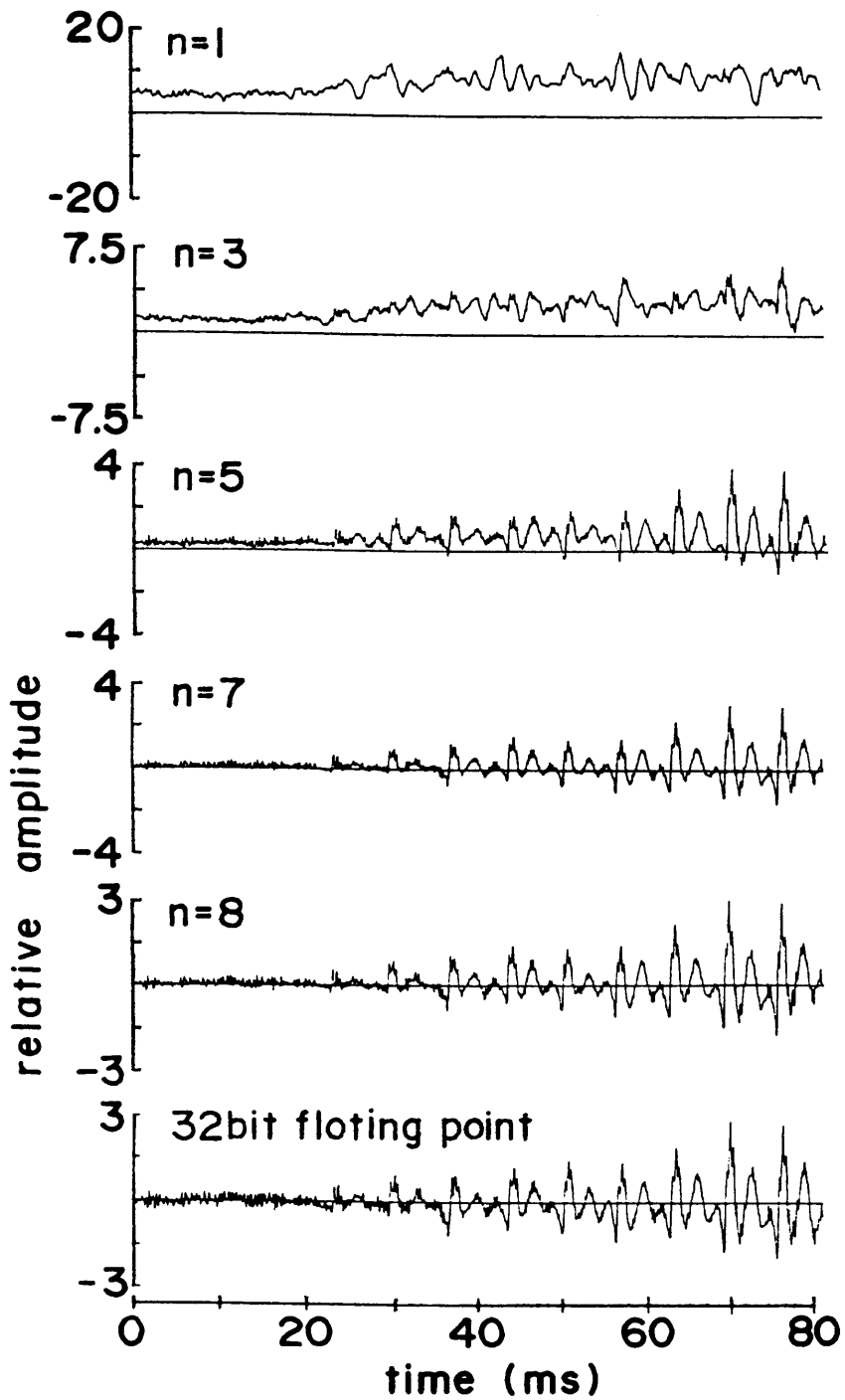
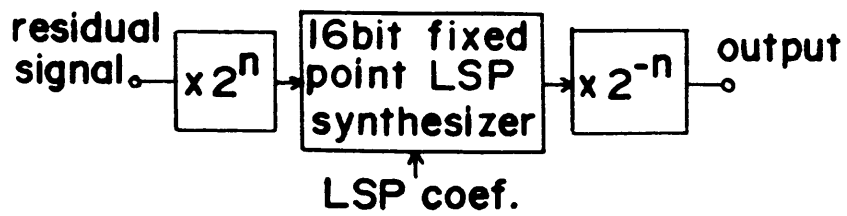


図 6.3: シフト数  $n$  を変えたときの修復音 /ni/ の立ち上がり波形

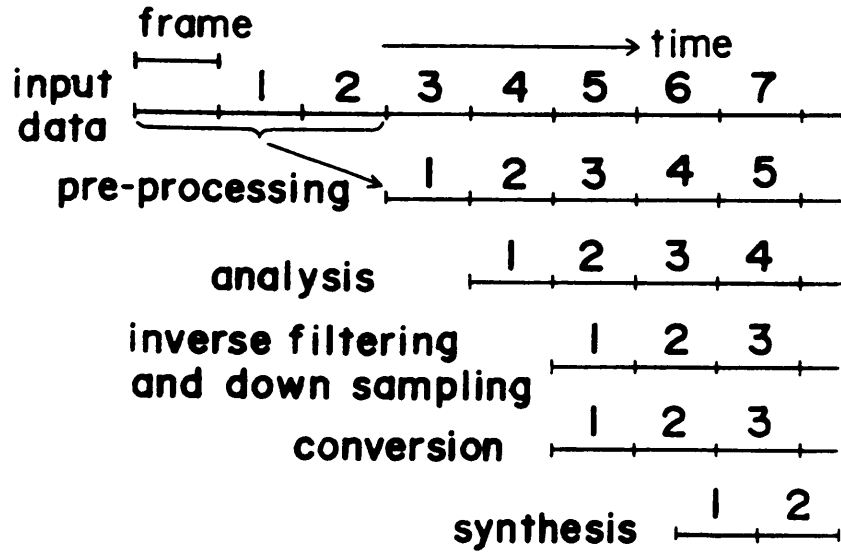


図 6.4: システムのタイミングの流れ図

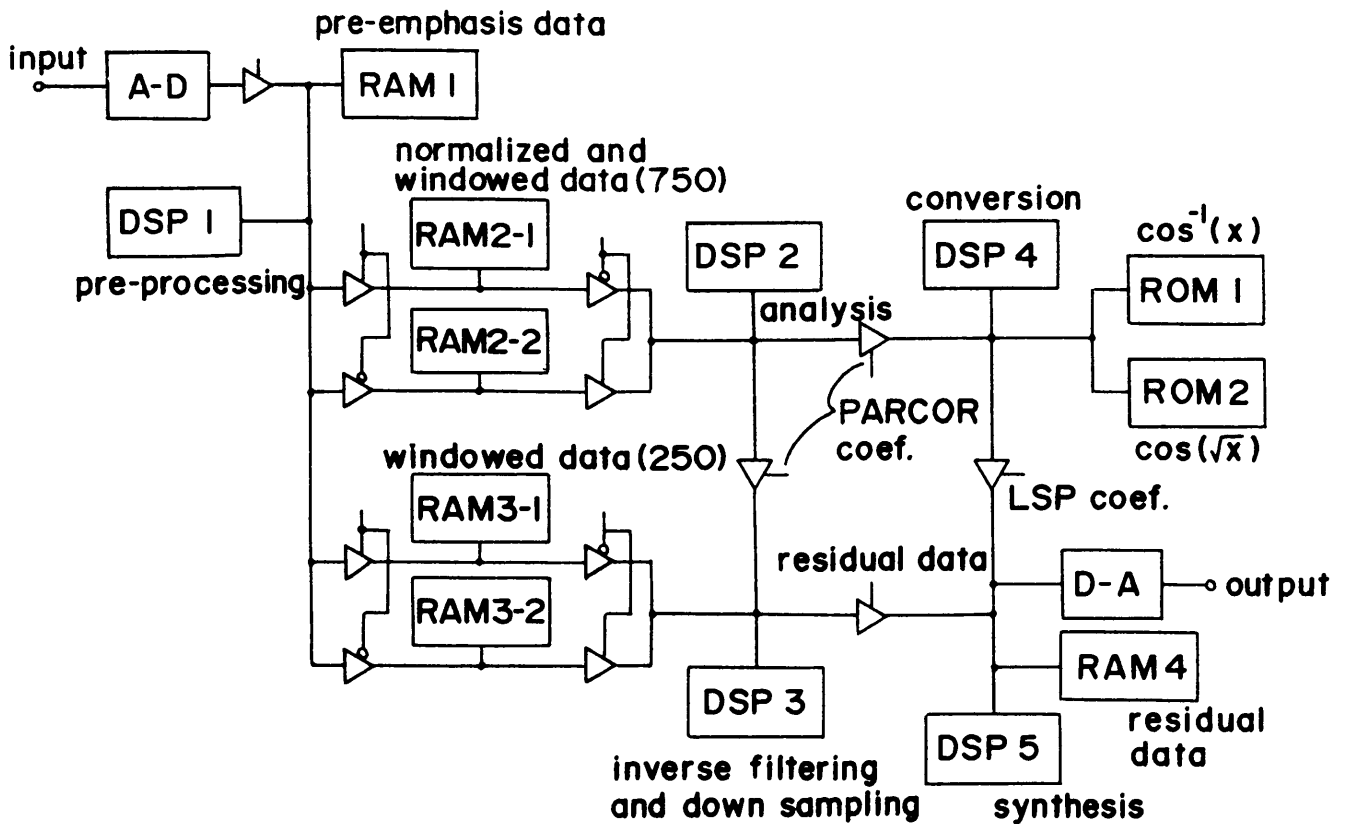


図 6.5: システムのハードウェアの概略図

表 6.8: システムの仕様 (各部の演算量 (最大値))

ユニット	演算量		DSP のプログラム使用メモリ量 (1 語 = 24 ビット)	IC 使用数	
前処理部 (クロック制御部を含む)	積和	1500	893 語 (窓データ 375 語を含む)	A-D	1
	加減算	750		DSP	1
	左シフト	8250		その他	46
分析部	積和	9894	548 語	DSP	1
	除算	12		その他	30
残差及びダウンサンプリング部	積和	4600	182 語	DSP	1
変換部	積和	1800	916 語	その他	34
	除算	52		DSP	1
合成部	積和	1300	325 語	その他	21
	加算	1200		D-A	1
その他インターフェイス等				DSP	1
				その他	50
				合計	207

イクロホンで収録した。話者は男性1名である。被験者はヘリウム音声に慣れていない学生7名であり、無響室内でヘリウム音声の原音および修復音を聴取して書き取った。使用スピーカはヤマハ NS-450 で、音量は平均約 75dB (騒音計 A 特性) に設定した。

3音節ワードリストの聴取試験の結果、深度 300m の原音では了解度 4.8% であり、会話等はほとんどできない。この音声を今回試作した修復装置で線形 ( $k = 2.64$ 、 $r = 5.23$ 、 $F'_{wa} = 0\text{Hz}$ )、非線形変換 ( $k = 2.64$ 、 $r = 5.23$ 、 $F'_{wa} = 120\text{Hz}$ ) を行ったところ、線形、非線形の提示順での被験者 4 名 (第 1 グループ) ではそれぞれ 76%、80%、非線形、線形の提示順の被験者 3 名 (第 2 グループ) では線形で 74%、非線形で 69% であり、一度目の提示による学習効果が大きく線形、非線形変換での了解度の差はないと思われる。これは、音声データが現在入手できるのが最深深度 300m のデータであり非線形性がそれほど大きくないためと思われる。また、第 1 グループの被験者に前修復器により非線形変換を行なった修復音では上記の聴取試験終了後の聴取で 73% であり、今回の試作器の方がわずかながら良い。変換された音声の音質は非線形変換の方が 0m の通常音声に近い。なお、非線形変換で  $F'_{wa}$  を 120Hz としたのはこれ以上  $F'_{wa}$  を大きくすると母音はよく修復されるが、有声子音で第 1LSP 係数の角周波数が 0 以下となるエラーが発生し、修復音の明瞭度が低下するためである。

この原因と思われる一例を図 6.6 に示す。これは深度 300m で「えんぎ」と発声したときの「ん」の部分であるが、1kHz 以下にピークがあり、この値は Fant らの式 6.1 では現れないものである。このピークは「ん」以外の有声子音でも 800Hz から 1kHz に現われる。また、1984 年話者 E の 20 次の LPC 分析 (窓長 30ms、ハミング窓) の結果を図 6.7 に示す。この結果より、母音のみから推定される声道の最低共振周波数から求まる曲線よりヘリウム音声では有声子音の第 1 ピーク周波数が低いものがあることを示している。このピーク周波数は有声音で観測され、母音では母音のレベルが高いので相対的には小さくなるが存在する。これは、声道壁の振動 [35] による放射か、軟口蓋の振動による鼻孔からの放射 [45] と考えられるが特定するまでには至っていない。ヘリウム空気中では気体の密度は高く、通常空気より声道壁や軟口蓋は相対的に質量の少ない壁と考えられるので音声の発声時に振動することが予想される。小型加速度計を使用した頬、鼻、咽頭壁などの振動加速度の測定 [45] をヘリウム空気中で行うとともにモデルによる検討を行いたい。

また、ヘリウム音声データベース (付録 D を参照) のコメントの情報より有声音のバズバーの有無を調べた。話者は 1982 年の話者 A、B、C、1991 年の話者 M、O、P の 6 名である。その結果を表 6.9 と表 6.10 に示す。話者 A は明瞭度試験用音節表を連続して話している部分があるので、単独発声の個数を示してある。話者 C を除きヘリウム空気中ではバズバーがなくなり破裂のみになりやすくしかも、無声音に聞こえてしまうものもでてくる。話者 E ではその傾向が最も強い。バズバーが破裂に変わるのは語中では起きないので、語頭において声帯の振動が起りにくくなっているのが原因と考えられる。

これは次のように説明できる。文献 [45] のモデルを想定すると、バズバーを発声する時には咽頭腔の容積を増大させる。文献 [5] によれば声門での体積速度は  $\rho^{1/2}$  に反比例する。咽頭腔の容積の増大は音響容量の増加であるが、音響容量は  $\rho c^2$  に反比例する。声帯は声門上部と下部とに音圧差がないと振動できない。咽頭腔での音圧の増加は (体積流の時間積分)/(咽頭腔の音響容量) に比例するので、空気中と同様に発声すれば音圧は  $\rho^{1/2} c^2$  だけ大きくなる。よって、同じ圧力差でバズバーを発声できる時間はこの逆数となり深度



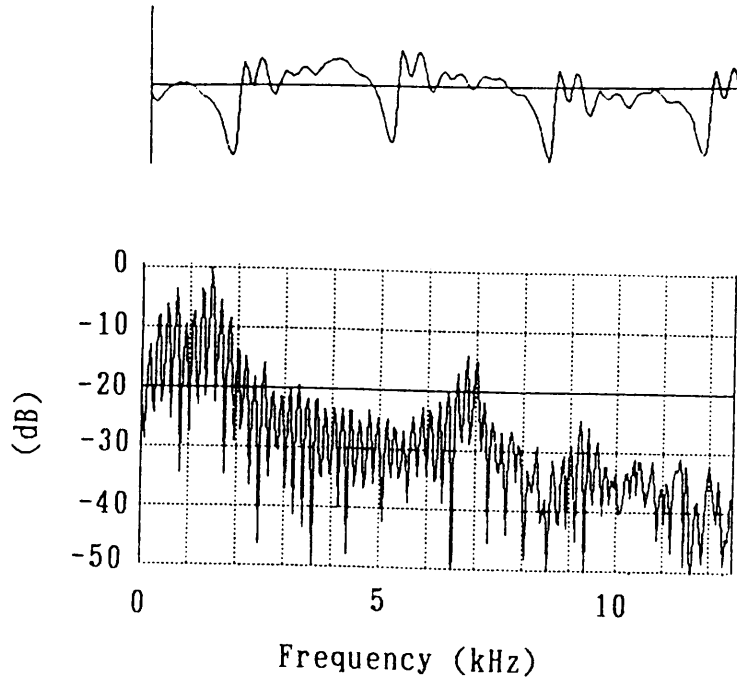


図 6.6: 深度 300m で「えんぎ (演技)」と発声したときの「ん」の波形 (20ms) とそのスペクトル

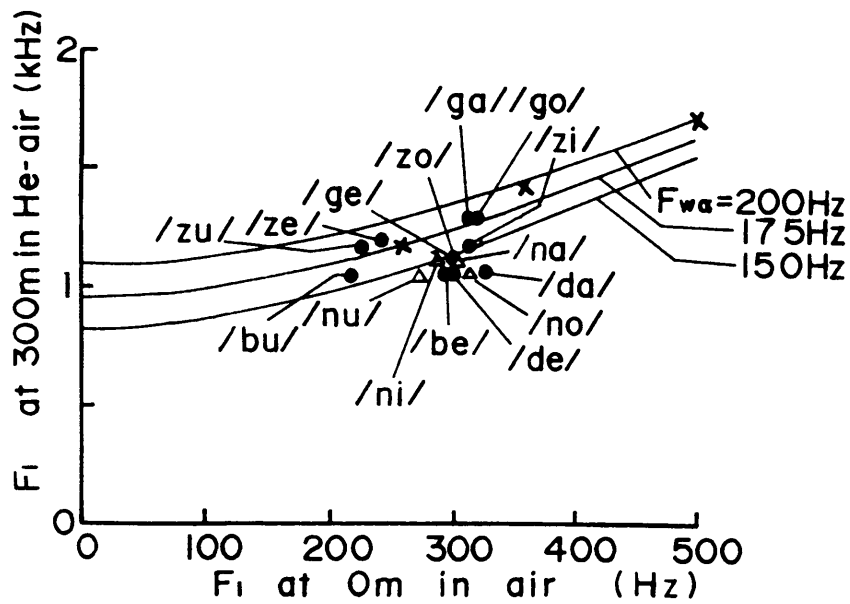


図 6.7: 通常空気と深度 300m における音声の 20 次の LPC 分析時の第 1 ホルマント周波数またはピーク周波数 ×:母音

300メートルではおよそ 1/10 で非常に短い。空気中と同じような長さでバズバーを発声するには咽頭腔をより増大する努力がいることになる。このことは語中でも条件は同じであるが声帯振動はすでに前の母音で起きているのでその慣性で振動は続くと考えられる。また、ヘリウム空気が通常空気に比べ音響インピーダンスが高いので音圧による声帯振動の制御が遅れると考えることもできる。今後、声帯付近の筋電図と音声波形の同時収録等による実証とともに 3次元有限要素法 [46] による検討や音声生成モデルを作り検討したい。

語頭のバズバーに見られるようにヘリウム音声自身に変形し聞こえにくくなる面もある。

## 6.6 まとめ

深々度潜水用の実時間ヘリウム音声修復装置の改良について述べた。LSP 係数の変換方式を今回提案したように改良すれば低域のホルマントの帯域幅が多少広くても変換されたホルマント周波数は Fant らの式に近くなるようになった。前処理部では各フレームで窓かけ後の最大値がほぼ同じになるようにアルゴリズムを改良した。分析部では Le Roux 法を用いることでスペクトルひずみを少なくすることができた。逆フィルタリングとダウンサンプリング部では FIR フィルタを用いて S/N 比の低下を防いだ。先のシステムにおいて過大入力やクリック入力により出力がオーバーフローし続ける欠点については、合成部でのオーバーフロー検出と起きた場合には合成フィルタ内のデータを強制的に零クリアすることで安定に動作させることができた。また、上記のほか前処理部で各フレームで大きさをほぼ同じにする処理をした後は合成まで左シフトしたままでデータを処理し、合成部の D-A 変換直前でデータを元の大きさにすることで修復音を先のシステムより浮動小数点演算の結果に近くできた。

DSP を使用したので IC の数を 207 個と少なくでき小型化された。さらに、DSP のプログラムを ROM にすれば IC は 160 個程度と少なくすることができる。この装置を使用して単語理解度試験を行ったところ深度 300m ではホルマントが非線形に上昇する割合がそれほど小さくなく、また非線形変換で  $F'_{wa}$  を大きくできなかったのも理解度自身は線形と非線形の変換での差はなかった。ただし、音質については非線形変換の方が 0m の通常空気中の音声に近くなった。

非線形性を強くできない原因は有声子音の第 1 ホルマント周波数が Fant の式より低く求まることであるが、その現象をいまのところ特定できていない。また、これを避けて非線形性を強くできる修復方法を求めることが今後の課題として残されている。

また、ヘリウム音声に伴う雑音の問題についても対応できるよう改良していく必要がある。

表 6.9: バズバーの脱落する音素数 (1982 年)

深度 (m)	A				B				C			
	a*		b		a		b		a		b	
0	1	0	/27	2	0	0	0	0	0	0	0	0
87	4	0	/23	3	1	12	2	1	0	0	0	0
145	5	2	/16	2	0	9	0	2	0	0	0	0
200	5	1	/19	3	0	12	2	2	1	0	0	0
260	7	1	/21	2	1	7	2	3	2	0	0	0
300	8	2	/17	1	1	3	1	2	0	1	0	0

- a 明瞭度試験用音節表 (有声子音の個数 27)  
 b 3音節ワードリスト (語頭の有声子音の個数 4)  
     左: バズバーの脱落する音素数  
     右: 無声子音に聞こえる音素数

表 6.10: バズバーの脱落する音素数 (1991 年)

深度 (m)	M				O				P			
	a		c		a		c		a		c	
0	15	0	6	1	20	3	7	2	6	0	5	1
91	7	0	4	0	24	6	7	2	21	3	7	1
200	12	3	8	0	26	7	6	5	12	1	3	0
300	15 <sup>†</sup>	3	4	0	25 <sup>†</sup>	15	7	5	11 <sup>†</sup>	0	9	0

- a 明瞭度試験用音節表 (有声子音の個数 27)  
 c 2音節ワードリスト (語頭の有声子音の個数 9)  
 †, 101音節表 (深度 280m, 有声子音の個数 27)  
     左: バズバーの脱落する音素数  
     右: 無声子音に聞こえる音素数

## 第 7 章

# 線形予測分析法を用いた潜水浮上時における血液中の気泡検出

### 7.1 はじめに

潜水におけるもう一つの課題である潜水浮上時に潜水土の血液中に発生する気泡検出について述べる。序章でも述べたようにこの血液中に気泡が発生すると減圧症(潜水病ともいう)になる可能性が大きい。減圧症の予防および治療には、浮上時の気泡発生を即刻知って浮上を一時停止するとか、浮上後の残存気泡を検出して再加圧する等の適当な治療を施さなければならない。それには即時的で高感度の気泡検出技術が必要である。

気泡を検出するには一般に超音波ドップラ法と呼ばれる方法が用いられる。これはドップラ・フロー・メータと呼ばれる超音波装置を使用して体外から超音波を照射し、血流中の気泡に当たって反射してくる超音波のドップラ変調音を聴覚的に判断して気泡を検出する。超音波の周波数としては5~10MHzが用いられる。ドップラ・フロー・メータは元来は流体の流速測定用の装置だが、GillisやSpencerらが血流中の気泡検出に応用できることを示した[19],[20],[21]。超音波を照射する部位の条件としては二つある。一つは、体表に近く超音波プローブで容易に見つけられることである。もう一つは、血液中のいずれの部位で発生した気泡も通る部位であることである。静脈は雑音が少なく気泡検出には適しているが、これらの条件を満たす部位はない。よって、鼓動音が雑音となるが、これらの条件を満たす肺動脈が選ばれている。

Spencerは気泡の検出頻度によって次の5つの段階に区分している[21]。

**段階 0** : 気泡が存在しない状態

**段階 1** : 鼓動4~5回に対し1~2の割合で孤立した気泡が検出される状態

**段階 2** : 鼓動1~2回に対し気泡が検出され、それらの半分以下には気泡の集団が検出される状態

**段階 3** : 各鼓動間に孤立した気泡、あるいは集団の気泡が検出されるが、鼓動音はまだ聞きとれるような状態

#### 段階 4：鼓動音をおおい隠すような多数の気泡が検出される状態

段階 2 以上では一般に自覚症状があり、減圧症を予防するには段階 1 で検出する必要がある。しかしドップラ信号には雑音が多く、段階 1 では気泡は小さく数も少ないので気泡音を聞き分けるのは熟練者でも容易でない。Belcher は 1980 年にコンピュータを応用してこの信号を分析し、ミニコン PDP-11/40 による気泡検出を試みた [22]。そのシステムではアナログ回路の帯域フィルタバンクでドップラ信号をスペクトル分析し、鼓動音を軽減した信号から気泡検出を試みている。ただし、やぎとひつじのデータである。

本論文では、デジタル信号処理により自覚症状の現れない段階 1 でも気泡を検出できる高感度で高速のシステムを開発することを目的とし、アルゴリズムの検討とハードウェアとして実現する研究をしてきた。この章では研究の前半にあたる気泡音検出のアルゴリズムとその性能評価について述べる。7.2 節～7.4 節で超音波ドップラ法で得られる信号と気泡音の性質を述べ、7.5 節、7.6 節で検出法の詳細を述べる。

## 7.2 超音波ドップラ法

プローブの送信部から投射された超音波の一部は血液中の血球や血小板あるいは気泡に当たって反射し、プローブの受信部で受信される。その反射波は主に赤血球によって生ずるがドップラ効果によって周波数変調を受けている。受信波はヘテロダイン検波器または  $f-v$  変換器で周波数差信号に変換される。この信号を「ドップラ信号」と呼ぶ。ドップラ信号の振幅は反射体の反射率に比例し、周波数は反射体の移動速度に比例する。実際は数 kHz 以下の可聴周波数である。

心臓の収縮時には血液の流量が増加するので、鼓動に同期して振幅が大きくなる。また肺動脈は心臓に近いので太く、血流は乱流になり複雑な周波数分布となる。この期間の信号を以下「鼓動区間」と呼ぶ。

血液中に気泡が発生すると、ドップラ信号には小鳥の鳴き声、鞭を振る音や泡のはじける音のような、短くて周波数の少し高い音が不規則に現われる。これを「気泡音」と呼ぶ。気泡には赤血球の数倍大きいものも多く、反射率は大きい [22]。段階が進むと孤立していた気泡が幾つか結合して大きくなり、流速は低くなる傾向にある [56]。

## 7.3 ドップラ信号の資料

今回用いた資料は Spencer の段階区分に基づいて録音された気泡検出用デモテープと、1984 年 10 月に海洋科学技術センターで行われた 300m 有人潜水模擬実験の際のダイバー 4 名の潜水前と減圧期間におけるドップラ信号を記録したテープである。

## 7.4 ドップラ信号の特徴

図 7.1(a)(b) に気泡検出用デモテープの段階 2 のデータの波形と振幅（絶対値の和）を示す。図 7.1(b) の矢印は気泡の発生時点を示している。これからわかるように、気泡は

反射率が高くても数が少なければ音は大きくない。図 7.2はドップラ信号の周波数 2.5kHz までのパワースペクトルで、図 7.1(a)の気泡音の存在する区間の FFT スペクトルと 4 次の線形予測（自己相関法、以下同じ）で求めたスペクトル包絡を重ねてある。予測次数を 3 から 12 まで変えて分析してみたが、スペクトル包絡は大体 4 次で近似できることがわかった。多くのデータで 200Hz と 1kHz 付近にスペクトルのピークがあり、このうち丸で囲んだ 1kHz 付近のピークが気泡によって生じたものである。

気泡発生の激しさは発生頻度によって段階 0 から段階 4 までに区分されている。各段階の気泡音のスペクトルの様子を見るため、段階 1 以上で気泡のある時点を含む区間の線形予測スペクトルを観察した。図 7.3にその 1 例を 3 次元表示で示す（予測次数 4、1 フレームの窓長 60ms（300 サンプル、ハミング窓）、フレーム周期 20ms、連続 10 フレーム）。丸で囲んだ部分が気泡によるピークである。気泡によって 2 フレーム程度連続して 1kHz 付近にピークが現われ、そのピークは段階が進むに従って鋭くなることがわかる。図 7.4は段階 0 から段階 2 までの各 2 秒間のドップラ信号について、図 7.3と同じ分析条件で求めた極周波数と帯域幅をプロットしたものである。第 1 極周波数は 500Hz 以下に、第 2 極周波数は 1,200Hz 付近に生じている。段階 0 では極周波数が 1kHz 以上で帯域幅が 400Hz 以下のものは発生しておらず、また、このような極は段階 1 より 2 の方が多い。したがって気泡によって生ずる極は周波数が比較的高く帯域幅は狭いことがわかる。以上、気泡音の大略の特徴を述べた。次節では気泡音を正確に検出するアルゴリズムについて述べる。

## 7.5 気泡音検出システム

### 7.5.1 アルゴリズムの概要

図 7.5に気泡検出アルゴリズムの概要を示す。まず、ドップラ信号を低域通過フィルタ（遮断周波数 2.4kHz）に通し、サンプリング周波数 5kHz で A-D 変換する。次に 2 秒ずつを処理単位として鼓動によるドップラ信号の成分を弱めるために線形予測による逆フィルタリングを行う。一方、ドップラ信号に対して心臓の鼓動音を検出し、逆フィルタリングされた信号（予測誤差信号または残差信号）から鼓動区間を捨てる。残った鼓動区間以外の信号について、A および B の 2 段階の処理で気泡音を検出する。気泡音が検出された場合は、気泡音の位置と周波数などを表示する。これらの処理を続く 2 秒の信号毎に繰り返す。以下、おのおのについて詳述する。

### 7.5.2 線形予測による鼓動音成分の影響の軽減

線形予測分析法は一般的性質として周期的な信号に対しては予測が的中して誤差が少なく、逆に不規則な信号に対しては予測誤差が大きくなる。肺動脈からのドップラ信号は心臓の鼓動による血液の流れによって生じているので、ある程度長時間の線形予測を行えばドップラ信号の中の心臓の鼓動による規則的な成分は予測し得るが、気泡は数が少ない場合は不規則で瞬時的にしか発生しないので予測しにくい。従って、ドップラ信号に対して線形予測分析で求めたフィルタ係数により逆フィルタリングして得られる波形（以

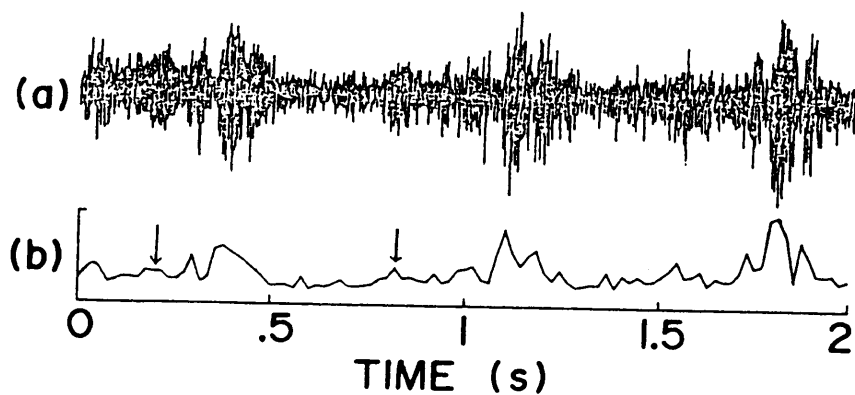


図 7.1: ドップラ音の原波形 (a) とその振幅波形 (b)。矢印は気泡の位置

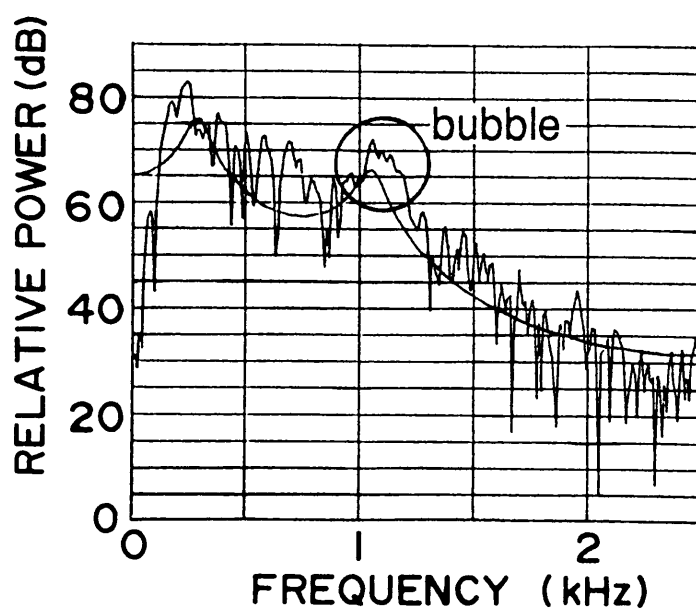


図 7.2: 気泡のある区間のパワースペクトル。FFT と線形予測スペクトルを重ねて表示

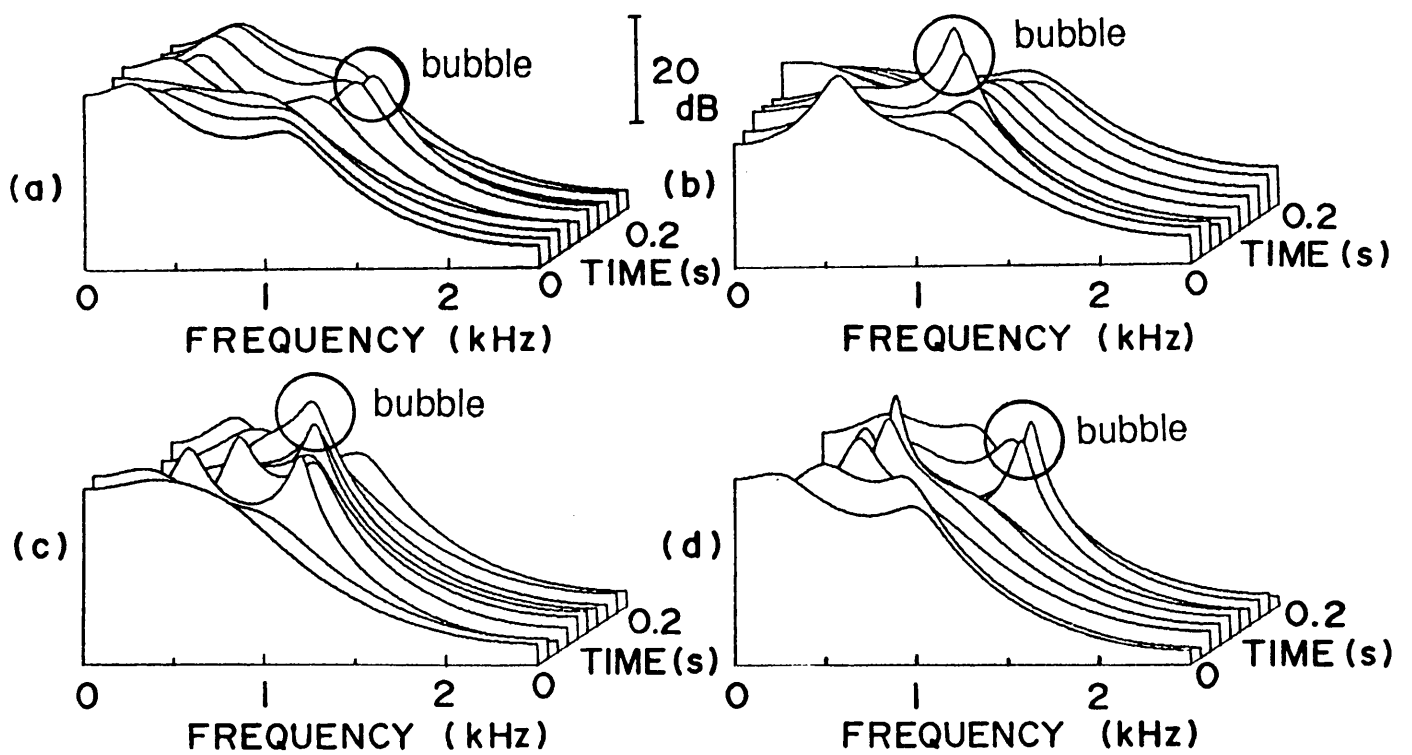


図 7.3: 種々の気泡発生段階における気泡音区間付近の LPC スペクトルの 3 次元表示の例 (a) 段階 1、(b) 段階 2、(c) 段階 3、(d) 段階 4



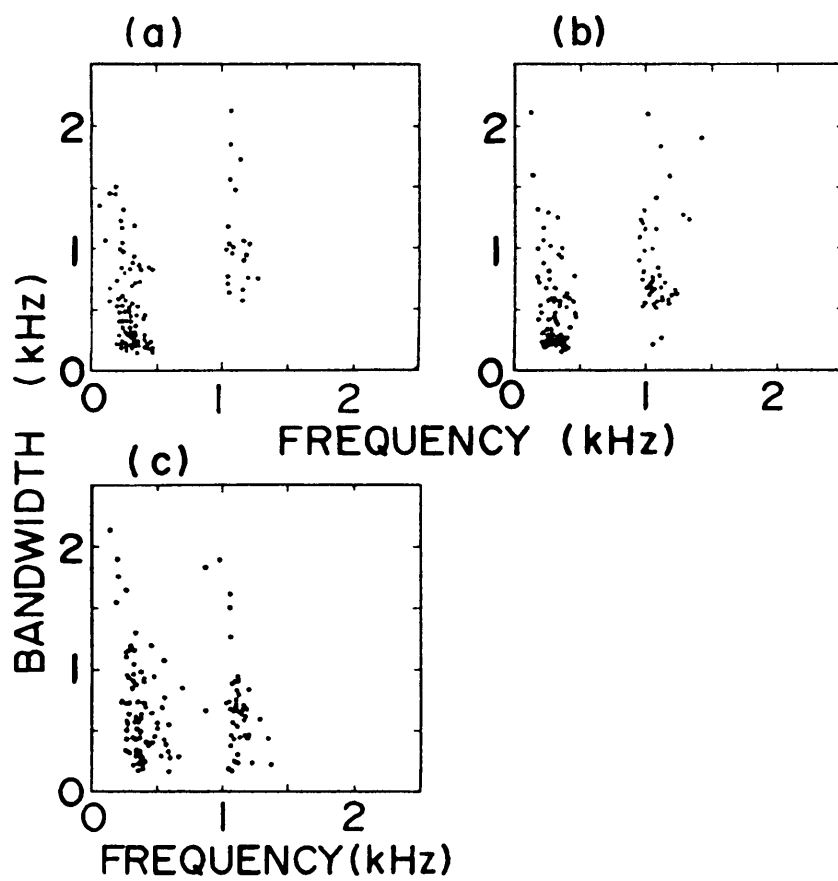


図 7.4: 段階 0、1、2 における各 2 秒間のドップラ信号の線形予測スペクトルの極周波数と帯域幅の分布 (a) 段階 0、(b) 段階 1、(c) 段階 2

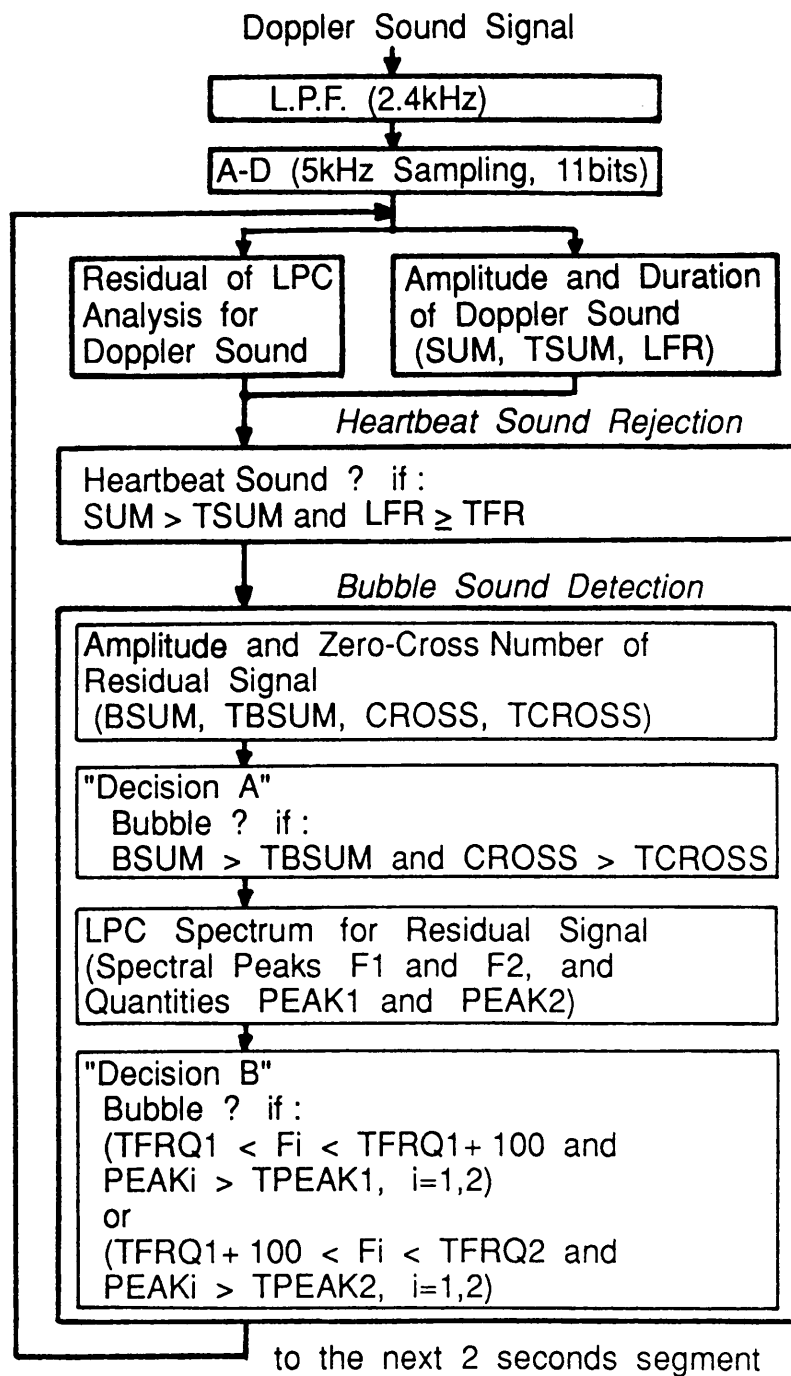


図 7.5: 気泡音検出手順の流れ図

下、「残差波形」という)には気泡音はほぼそのまま残り、心臓の鼓動による成分はかなり除去されることになる。線形予測の次数と線形予測を行うデータ長  $N$  は、鼓動による成分を軽減するためには大きい方が良く、一方処理の簡単さのためには小さい方が良いので、適当な値を実験的に見出す必要がある。

予備実験より、データ長  $N$  は鼓動音が少なくとも 2 回ないし 4 回生じる 10,000 とした。線形予測による正規化予測二乗誤差  $E$  (原波形の二乗和に対する残差波形の二乗和の比) は PARCOR 係数を  $k_i (i = 1, \dots, p)$  とすると以下の式で与えられる。

$$E = \prod_{i=1}^p (1 - k_i^2) \quad (7.1)$$

図 7.6 にデータ長 10,000 (2 秒) における正規化予測二乗誤差の分析次数による変化を示す。これを見ると分析次数を 4 より大きくしても正規化予測二乗誤差は減らず、分析次数は 3 ないし 4 でよいことがわかる。また、予備実験より鼓動の成分を軽減するが、気泡を残さなければならないので、多少予測誤差の多い 3 を分析次数とした。このようにすると鼓動のパワースペクトルは概して”への字形”をしているので残差波形には気泡音のある高域が多く残る。

図 7.7(a)、(b) はその一例で段階 1 の原波形 (a) と残差波形 (b) である。矢印 ↓ は熟練者が聴覚判断と波形ディスプレイの観察によって確認した気泡位置を示す。図 7.7 を見れば、(a) の原波形より (b) の残差波形のほうが鼓動音は軽減されているのがわかるが、気泡音の位置などを確定する事はできない。しかし、この残差波形を耳で聞くと気泡音があるか否かの判断はやりやすくなっている。

### 7.5.3 鼓動区間の除去

鼓動時の血流は乱流かそれに近い状態になっており、鼓動音のスペクトルは複雑に変化して気泡音と似たスペクトルになる場合もある。デモテープを試聴したところ、段階 3 以下では気泡音は鼓動区間にもあるが、そうでない区間と同じように存在し、また鼓動音と区別するのは難しいことがわかった。そこで前処理として鼓動区間を除去してしまうこととした。

鼓動区間を除去するには、鼓動区間は一般に振幅が大きく継続時間も気泡音より長いことを利用した。データ長は鼓動音が少なくとも 2 回は含まれる 2 秒 (サンプル数 10,000) とし、それを 100 個の小区間に分割した。各小区間のサンプルデータ 100 点の絶対値の和を  $SUM$  とする。 $SUM$  が閾値  $TSUM$  を越え、さらに、閾値を越える小区間の数  $LFR$  が  $TFR$  個以上連続する時それらの小区間を鼓動音と判定する。

小区間数の閾値  $TFR$  は 3 とした。その理由は低い段階での気泡音の持続時間はかなり瞬時的で、7.4 節で窓長 300 サンプルのフレーム 2 個連続と述べたように、単一気泡に対しては長くても 40ms (小区間数では 2 個) 程度だからである。

閾値  $TSUM$  は以下のようにして決めた。ドップラ信号の波形は、時間やセンサの位置のずれなどによって変化することは少ないが、平均強度はかなり変化する。したがって鼓

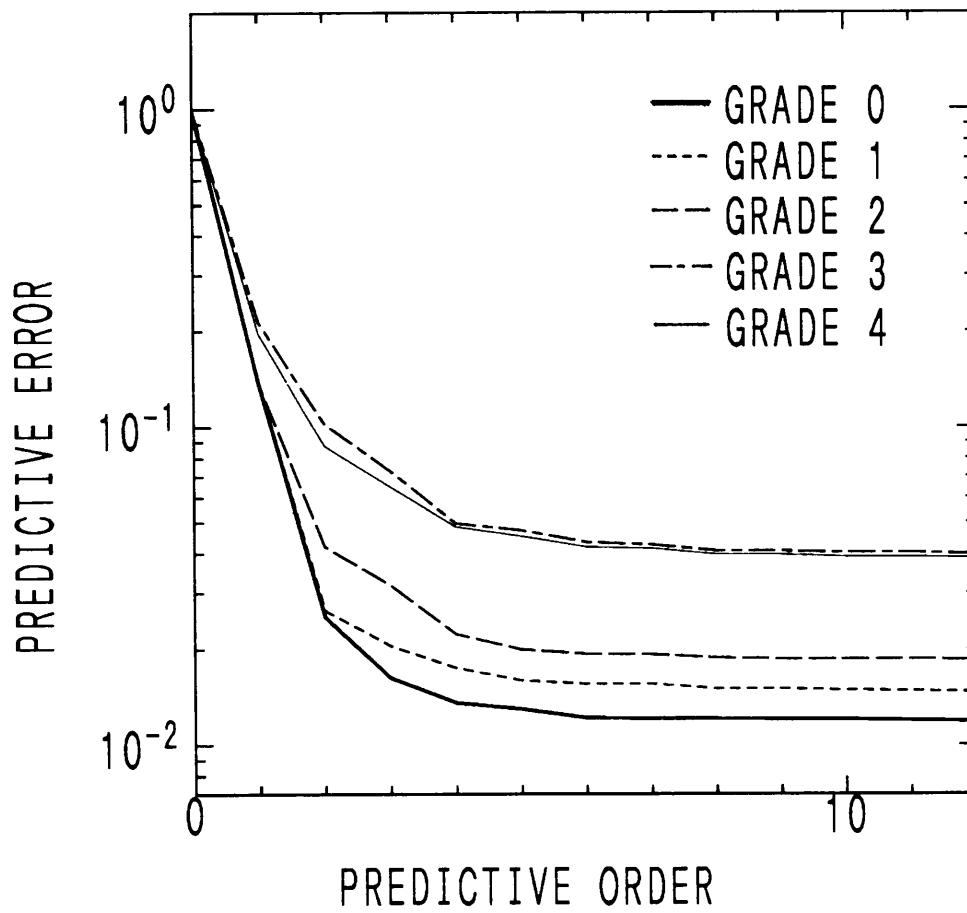


図 7.6: 2 秒の各段階のドップラ信号 (2 秒) に対する線形予測分析を行ったときの分析次数と正規化予測誤差の関係

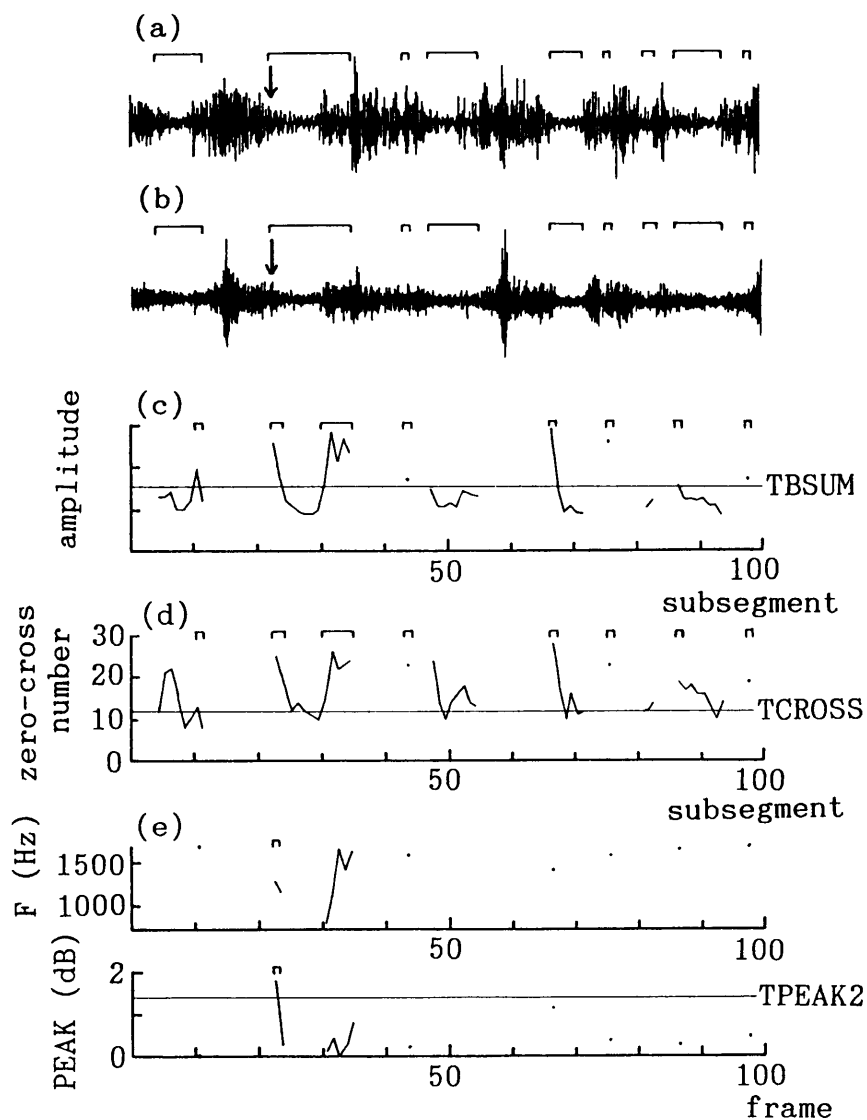


図 7.7: ドップラ信号に対する気泡検出過程の例

(a) と (b) は長さ 2 秒の原波形と次数 3 の線形予測分析の残差。矢印 ↓ は熟練者による気泡検出位置。(c) と (d) は鼓動区間を除去した残りの小区間に対する予測残差の振幅と予測残差のゼロ交差数の時間波形。(e) は判定 A で気泡候補とされた小区間における残差波形の 4 次の線形予測スペクトルのピーク周波数  $F$  と、その強度とピークより 100Hz 低い周波数における強度との差 PEAK。鍵括弧は気泡候補区間またはフレームを示す

動区間の判定の誤りを少なくするには平均強度の変化に応じて閾値  $TSUM$  を自動的に変化させる必要がある。そこで、閾値  $TSUM$  は  $SUM$  の平均値にある重み係数  $TKPER$  をかけた次式とした。

$$TSUM = \sum_{i=1}^{100} (SUM(i)/100) \times TKPER \quad (7.2)$$

重み係数  $TKPER$  の決定の仕方は 7.5.5 節で述べる。

#### 7.5.4 振幅、零交差数と線形予測による気泡音検出

7.5.2 節で得られた残差波形から 7.5.3 節で鼓動区間と判定されたものを除いた残りの区間の波形に対して、図 7.5 の判定 A、B の二段階の操作で気泡を検出する。判定 A では振幅と零交差数を見て気泡区間の予備選択をする。次にこの候補区間に対してさらに線形予測分析によるスペクトルの極を求め、判定 B で気泡区間であるか否かを判定する。

判定 A では、振幅のパラメータ  $BSUM$  と  $TBSUM$ 、零交差のパラメータ  $CROSS$  が用いられる。以下これらについて述べる。気泡音は 1 小区間（サンプル数 100）の半分ほどの長さのものが多いため、1 小区間内で開始点を 25 点ずつずらした長さ 50 点の微小区間を 3 つ取り（延べ長さ 100 点）、各 50 点のサンプルデータの絶対値の和を求め、3 つのうちの最大値をその小区間の振幅  $BSUM$  とする。 $TBSUM$  は振幅の閾値で、次式のように  $BSUM$  の平均値に重み  $TBPER$  をかけた値である。

$$TBSUM = \left( \sum_{i=1}^{INUM} BSUM(i)/INUM \right) \times TBPER \quad (7.3)$$

ここで、 $BSUM(i)$  は第  $i$  小区間の  $BSUM$  であり、 $INUM$  は鼓動区間検出アルゴリズムで鼓動区間と判定された小区間を除いた残りの波形の小区間数である。重み係数  $TBPER$  は 1.1 とした。

零交差数については、雑音による影響を避けるため、実際の零レベルより高い  $TZL$  に対する交差数を求める。

$$TZL = TBSUM/50 \quad (7.4)$$

$TZL$  に対する交差数（判定 A に関しては便宜上これを零交差数と記す）を 3 つの微小区間で求め、3 つの中の最大値を  $CROSS$  とし、零交差数の閾値を  $TCROSS$  とする。

判定 A では、 $BSUM$  と  $CROSS$  が次の条件を満たすとき気泡音候補小区間とする。

$$\text{判定 A: } BSUM > TBSUM \text{ かつ } CROSS > TCROSS \quad (7.5)$$

零交差数の閾値  $TCROSS$  は 12 としたがその理由はつぎの通りである。標本化周波数を 5kHz、零交差を求める微小区間のサンプル数を 50 とすると、 $F\text{Hz}$  の正弦波の零交差数は  $(2 \times 50 / 5000)F = 0.02F$  となる。気泡音の下限閾値周波数はダイバーによる個人差もあるが今回の実験では 600Hz としてよいので、 $F = 600$  とおき、 $TCROSS = 0.02 \times 600 = 12$  とした。

前処理と判定 A では気泡音は出来るだけ残さなければならないので、鼓動音を完全に除去できず、またその他のノイズも入っているので、判定 B で更にスペクトルに関する条件で気泡音を判定する。

判定 B においては、判定 A で気泡音の候補となった小区間に対して線形予測分析を行い、最終的に気泡音かどうかを判定する。線形予測には自己相関法を用い、予測次数は 4、1 フレームの窓長は候補小区間の前後に 100 サンプルずつ加えた 300 点、窓形はハミング窓である。分析で得られたスペクトル包絡から、極に対応する量としてピークピッキング法によるピーク周波数  $F_1$ 、 $F_2$  とその強度、および、帯域幅に関係する量として、ピーク周波数より 100Hz 低い周波数における強度とピーク強度との差  $PEAK_1$ 、 $PEAK_2$  を計算する。そして、次の式 (7.7) か式 (7.8) の条件を満足するフレームを気泡音と判定する。なおここでピークピッキング法や  $PEAK_i$  などを用いたのは、極の周波数と帯域幅を直接計算するより演算量を少なくするための便宜上である。

$$\begin{aligned} \text{判定 B: } TFRQ_1 < F_i < TFRQ_1 + 100 \quad \text{かつ} \\ PEAK_i > TPEAK_1, \quad i = 1, 2 \end{aligned} \quad (7.6)$$

または、

$$\begin{aligned} TFRQ_1 + 100 < F_i < TFRQ_2 \quad \text{かつ} \\ PEAK_i > TPEAK_2, \quad i = 1, 2 \end{aligned} \quad (7.7)$$

$TPEAK_1 = 2.4\text{dB}$ 、 $TPEAK_2 = 1.4\text{dB}$  である。周波数の閾値  $TFRQ_1$  と  $TFRQ_2$  については 7.5.5 節で述べる。

気泡音と判定されたフレームはその番号と中心周波数などの情報が表示される。この 2 秒のデータが終わったら次の 2 秒のデータの分析をする。この際、前の 2 秒のデータと 100 点分のデータを重複させる。これを与えられたデータについて次々に繰り返す。

### 7.5.5 重み係数と閾値の決定方法

ドップラ信号は人により多少異なるので、その人に適した重み係数と閾値 ( $TKPER$ 、 $TFRQ_1$ 、 $TFRQ_2$ 、 $TCROSS$ ) を設定しなければならない。 $TCROSS$  については 7.5.4 節で述べたので、それ以外の重み係数と閾値の決め方について次に述べる。

## 振幅の閾値に関する重み係数 $TKPER$

まず、気泡音の含まれていないデータ（デモテープでは段階 0、潜水模擬実験では潜水前のもの）の 2 秒間（小区間長を 100 点として、100 小区間分）について、 $TKPER$  に初期値として 0.8 をあたえ、式 7.2 により  $TSUM$  を計算する。次に、画面上に同じ区間の波形を表示し、 $TSUM$  を越える小区間をマークし、波形と照らし合わせる。 $TKPER$  を変え、波形の鼓動部とマークされた小区間番号が最もよく一致する  $TKPER$  を決定する。今回のデータでは  $TKPER$  は 0.65~0.75 であった。

## ピーク周波数の閾値 $TFRQ_1$ と $TFRQ_2$

2 秒間のデータを 1 フレームの窓長 60ms、フレーム周期 20ms の線形予測分析により、図 7.3、図 7.4 のようにスペクトル包絡とピーク周波数を求める。鼓動音以外の第 2 ピーク周波数を求め、視察によりそのほぼ中央を気泡の中心周波数とする。その値から、 $TFRQ_1 = (\text{気泡の中心周波数} - 400\text{Hz})$  とし、 $TFRQ_2 = (\text{気泡の中心周波数} + 800\text{Hz})$  とした。なお、デモテープと 4 名のダイバーの気泡の中心周波数は 1000Hz~1150Hz であった。

## 7.5.6 アルゴリズムの使用方法

まず、気泡のない段階 0 または潜水前の 2 秒間のデータにより各閾値を決定する。次に同じく気泡のないデータで 20~40 秒間アルゴリズムを実行し、閾値の適否を確認する。誤判定が多い場合は、式 7.2 の  $TKPER$ 、式 7.4 の  $TZL$  を再調整し、最終的に誤判定フレーム数が 1 分間に 4~5 フレーム以下になるようにする。閾値が決定されたら、次に、気泡があるか否かを判定したいデータについて分析を行う。段階区分により、1 分間のデータに対して検出される気泡音フレームの数が 10 個程度であれば、そのデータは段階 1 の状態でとられたデータと判定する。

## 7.6 結果と考察

### 7.6.1 デモテープによる気泡検出の実験

デモテープの場合は、段階 0 のデータにより各閾値を決定した後に気泡検出を行なった。

図 7.7 に 7.5 節で述べた判定の過程の具体例を示す。(a) と (b) はドップラ信号の原波形および 3 次の線形予測残差のサンプル列である。矢印は熟練者が聴覚と波形視察で求めた気泡位置を示す。図中の鍵括弧は原波形の振幅と持続時間により、3 小区間以上連続して閾値を越えるものを鼓動区間とみなして除いた残りの区間を示す。鍵括弧で示された小区間数は 48 である。(c) と (d) の折れ線はそれぞれ残差の振幅 ( $BSUM$ ) と零交差 ( $CROSS$ ) である。折れ線がところどころ途切れているのは鼓動音を除いたからである。判定 A では (c) と (d) の波形から閾値  $TBSUM$  と  $TCROSS$  を越える小区間を抽出し、気泡音の候補小区間を抽出する。鍵括弧はこの段階で残った小区間を示している。区間数は 13 であり (b) に比べかなり減っている。(e) は上で気泡音の候補とされた小区間に対する



4 次の線形予測スペクトラムのピーク周波数と、それより 100Hz 低い周波数における強度差 ( $PEAK_i$ 、この例では  $PEAK_2$ ) である。(e) の鍵括弧は閾値  $TPEAK_2$  を越えたフレームの位置を示す。このフレームは判定 B により気泡音と判定され、(a) と一致している。この例は鼓動 4 回で気泡が 1 箇所なので段階判定は 1 とされた。

表 7.1 にデモテープのデータに対する結果をまとめた。この表の熟練者とは筆者の一人であるが気泡検出にはかなり習熟している。段階判定結果を見ると、段階 3 までは正確に気泡判定がなされている。段階 4 のデータに対しては段階 3 と誤判定されているが、これは鼓動音を覆い隠すような大きな気泡音は鼓動音として除外されてしまうからである。また段階 3、4 で熟練者の気泡数が空白になっているのは、それらの段階では気泡は群生することが多く、いちいち数える事は出来なかったからである。しかし実際の気泡検出では、自覚症状がないかまたは少ない段階 1 や 2 の気泡音をうまく検出することが大切である。今後は段階 4 を 3 とみなすような誤判定を改良していく必要がある。

### 7.6.2 有人潜水模擬実験における気泡検出実験

表 7.2 に、1984 年 10 月に海洋科学技術センターで行われた有人潜水実験で録音したドップラ信号のデータについての判定結果を示す。潜水前、屈伸運動を 4 回行った直後のデータにより各閾値を決定し、深度 300m に 7 日間停留して各種の実験をした後、浮上する期間 (1 日につき約 28m 浮上) の深度 116m におけるデータを分析した。ドップラ信号の波形は潜水前と比べて深度 116m では多少複雑になっていたが、判定に影響はなかった。表 7.1 と同様本方法による気泡数は低い段階でやや多いという傾向があるが熟練者による聴覚判定と大差は無い。

### 7.6.3 気泡数の増大とドップラ信号のスペクトルピーク周波数の低下に関する考察

気泡音の周波数は気泡の数が少ない時は高く、多くなると低くなるという現象は古くから知られていたが、そのメカニズムについては解明されていなかった。図 7.8 は今回の実験で検出された気泡音の周波数のヒストグラムである。各段階とも気泡によるスペクトルピークは 1000~1400Hz の周波数になるものが多いが、段階 3、4 になると、600~800Hz の低い周波数を持つ気泡の割合が大きくなっていて、上述の現象が現れている。以下この現象のメカニズムについて考察する。

気泡の内圧は次式のように表面張力に比例し半径に反比例する [1]。

$$P_b = P_h + 2\gamma/a \quad (7.8)$$

$P_b$  : 気泡の内圧       $P_h$  : 外圧  
 $a$  : 気泡の半径       $\gamma$  : 表面張力

したがって気泡は半径が小さいほど内圧が大きく壊れやすい。また気泡にガスが供給され

表 7.1: 気泡検出用デモテープを判定した結果

段階	データ長 (秒)	気泡数		判定段階
		本方式	熟練者	
1	39.64	9	8	1
2	39.64	40	35	2
3	19.82	48	-	3
4	19.82	36	-	3

表 7.2: 有人潜水模擬実験の 116m におけるドップラ音を判定した結果

潜水士	データ長 (秒)	気泡数		判定段階
		本方式	熟練者	
E	19.82	0	0	0
F	19.82	2	0	0
G	49.54	7	7	1
H	19.82	9	7	1

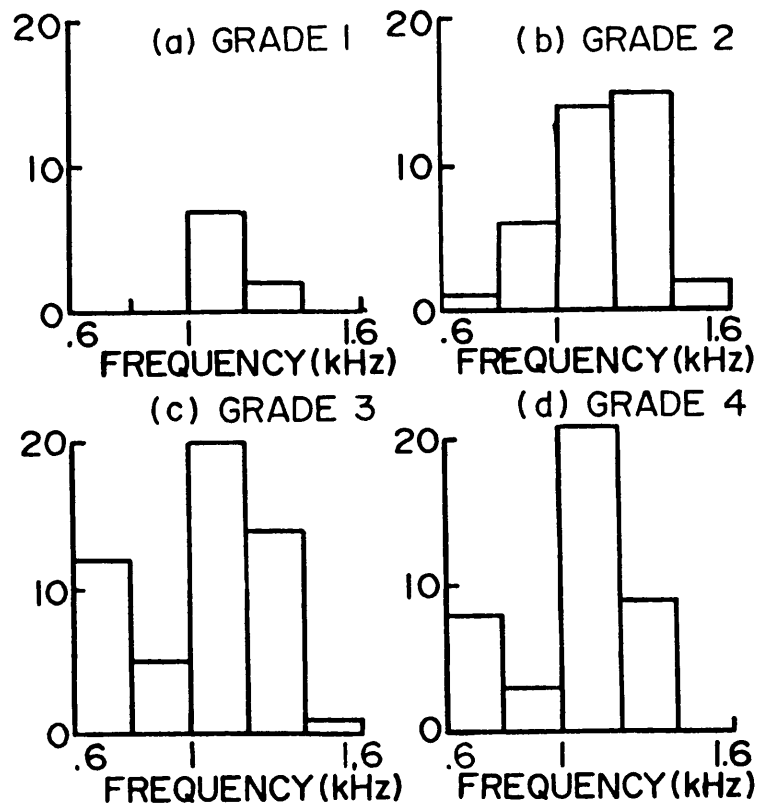


図 7.8: 気泡音のスペクトルのピーク周波数の頻度

ると半径は大きくなろうとするが内圧は減少するので萎縮する力も働く。その結果ある一定の大きさの気泡が存在する傾向がある。これらの気泡は循環系を流動するうちに集合して一つの塊になったり、大きく成長したりする [1]。

血液のように粘性のある液体が血管壁を流れる時は、中心部の流速は速く壁側部の流れは遅くなる。中心軸から離れた所を流れる気泡を考えると、血管内の流速方向に向かって気泡の左側と右側とで流速が違うので、図 7.9(a) のように血管の右側にある気泡は時計回りに、左側にある気泡は反時計回りに回転する力が働く。回転している気泡のまわりには図 7.9(b) の点線のような循環流が生じ、これが回転のない場合の流線と合わさって図 7.9(c) のような流線ができる。2つの速度が同じ方向に重なる側では流線は密になり、速度が大きくなってベルヌーイの定理によって圧力が減少する。逆に2つの速度が反対方向に重なる側では圧力が増加する。このため気泡は流れにも回転軸にも直角な中心軸に向かう方向に力を受ける。その結果気泡は血管の中心軸部に集まりやすく、そこでは流速が速いので観測されるドップラ信号の周波数は高くなる。多数の気泡が存在する場合は、血管の周辺にも存在するようになるので周波数の低いものが現われる。また、いくつもの気泡が融合して大きくなった気泡は扁平し、速度は気泡の各部の流速の平均的な値となるので遅くなる。気泡検出を行う比較的大い肺動脈での流れは乱流になっている [47] ので流速分布はもっと複雑になるが、その場合でも気泡が大きくあるいは多数になったとき平均流速が低下することは避けられない。このように、気泡の発生頻度が高くなると平均速度が低くなり、ドップラ信号のスペクトルのピーク周波数は下がることになる。

## 7.7 まとめ

気泡検出用デモテープを分析することにより気泡音の特徴を明らかにした。次に気泡発生の不規則性と線形予測分析における残差信号の性質を利用した新しい気泡音検出アルゴリズムを導き、気泡の少ない段階ではほぼ正確に気泡音を検出できることを示した。すなわち肺動脈付近からの超音波ドップラ信号を用い、2秒ずつのデータについて3次の線形予測分析法により求めた係数で逆フィルタリングして残差波形を求める。一方、心臓鼓動音の振幅と持続時間の特徴から鼓動区間を検出し、その区間を残差波形から除去する。次に、その残差信号の振幅と零交差数から気泡音の候補を求め、その候補について更に4次の線形予測分析を施して極周波数と帯域幅に関する性質から気泡音と判定する。この方法により気泡音のデモンストレーション用資料と潜水模擬実験の際の資料に対して気泡検出の専門家と同程度の判定ができた。また段階が増すと気泡音の周波数が低くなる現象についても考察を加えた。

本章で述べたアルゴリズムでは、判定式の重み係数と閾値を決定する所に目視による部分と再調整をする部分がある。実用上の便宜のためにはこれらを自動化する必要がある。これらに、については次章で述べる。

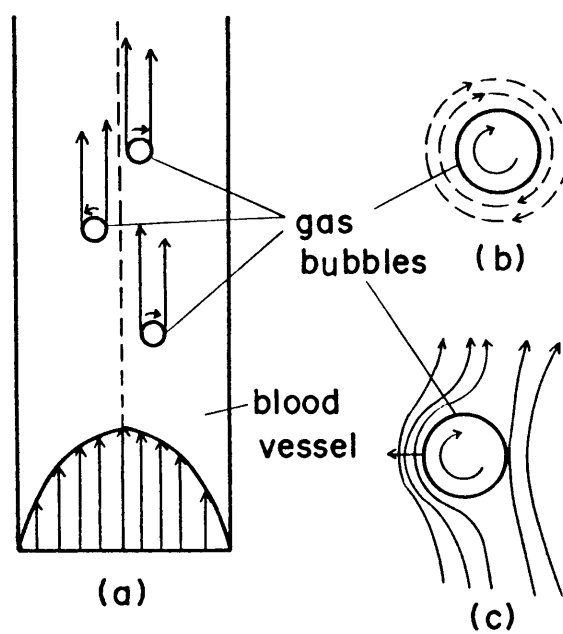


図 7.9: 血管内の気泡の流れと回転

## 第 8 章

# DSP による気泡検出とアルゴリズムの改良

### 8.1 はじめに

前章では計算機による信号処理によってこの気泡を検出するアルゴリズムを述べ、その有効性を確認した。ここでは、この気泡検出アルゴリズムを DSP (デジタル信号用プロセッサ) を用いて実行し、更にアルゴリズムの一部を改良して自動化し、ほぼ実時間で処理できるシステムを構成したので以下に述べる。また、実際の潜水士からのドップラ音のデータよりさらにアルゴリズムを改良したのでそれについて述べる。

### 8.2 DSP を用いた気泡検出システム

#### 8.2.1 DSP

DSP は、デジタル信号処理を効率よく高速に実行できるように、一般のマイクロプロセッサとは異なったアーキテクチャで作られた一種の CPU である。今回使用するものは、富士通の MB-8764 である。この DSP は、1983 年の発売で、マシンサイクルが最小 100ns の 16 ビット固定小数点である。この DSP には市販のハードウェアのエミュレートボードがあり、また、ヘリウム音声修復装置ですでに開発したプログラムも大きな手を加えないで利用できるという理由でこれを使用することにした。ただし、命令用メモリが 1024 ワード、外部 RAM が 1024 ワードまでなので、拡張用の D ラッチを用意することにより 32K ワード外部 RAM を外付して容量を拡張して使用した。

#### 8.2.2 システム構成

デジタル信号処理による気泡検出のアルゴリズムの開発については前章で述べた。そのアルゴリズムは、録音されたドップラ信号から気泡音をよく検出できるが、改良を要する点もあった。すなわち、そのアルゴリズムでは気泡信号の分析の中にまだ人間の目視

に頼った部分があり、これを自動化すること、および実時間での動作を可能とし、かつ潜水支援の現場に持ち運びができるように小型化すること、である。

図 8.1 に、今回新しく気泡検出のために構成したシステムのブロック図を示す。DSP のハードウェアエミュレータ・FDSP KIT-8764 (富士通) を中心に、ホストコンピュータとして、パーソナルコンピュータ FM 16 $\beta$  FDII (CPU: i80286、5 インチ FDD2 台内蔵) を使用した。また波形メモリ用に、外部 RAM の増設用メモリボード (32KW  $\times$  16 ビット SRAM) を製作して、KIT-8764 に接続した。その他に、周辺音響機器などがある。

### 8.3 気泡検出アルゴリズムと DSP での計算法

ドップラ信号分析の大まかな流れを図 8.2 に示す。取り込んだデータはフレーム単位で扱う。フレームに分けるのは、信号を短時間の特性の変化として捉えるためである。まず、3 次の線形予測を行い残差信号を得る。また、鼓動区間検出で鼓動音の大きいフレームを気泡区間候補から除く。次に気泡音検出 A で振幅と零交差数により、気泡の候補フレームを絞り込む。さらに、気泡音検出 B で気泡音の周波数に関連した分析を行い、最終的に気泡フレームを決定する。この方法の原理は、前アルゴリズムにならっている。以下 DSP システムに移植して完全自動化するための主な変更点を列記する。

- (i) データは外部 RAM の個数の関係で前アルゴリズムの 10000 点よりやや少ない 9600 点とし、100 点を 1 フレームとして全体で 96 フレームの構成とする。
- (ii) 前アルゴリズムでは鼓動区間検出の際に絶対値の和を用いたが、ここでは二乗和を用いる。
- (iii) 鼓動区間であるフレームを判定する際に、個人ごといき値を波形を見ながら決めていた。これを自動化するため、鼓動区間フレーム数が 50 フレーム以下になるようにいき値を決めるように改めた。それは、平均的な鼓動区間のフレームの割合が約 50%であったからである。
- (iv) 気泡音判定 B のうち極の位置とピークの大きさを求めるのに、前アルゴリズムでは FFT を用いてピークピッキング法で求めたが、ここでは高速にすることを目標とし、多くのデータを処理した結果からみて線形予測の次数は 4 次で十分であったことを考慮して 4 次の多項式を直接解くことに改めた。それに伴いピークの大きさは二次関数として近似し帯域幅に換算した。
- (v) 気泡音のある周波数範囲を気泡のない段階 0 から決める時にフレームごとの 4 次の線形予測分析の高い方の極周波数から目視で決めていたものを、9600 点の 4 次の線形予測分析の高い方の極周波数から決めることにした。

### 8.4 処理時間と結果

表 8.1 に演算処理数とプログラムステップ数などを示す。精度を保つために、DSP は 16 ビットの固定小数点演算ではあるがプログラムにより倍精度固定小数点 (32 ビット) や

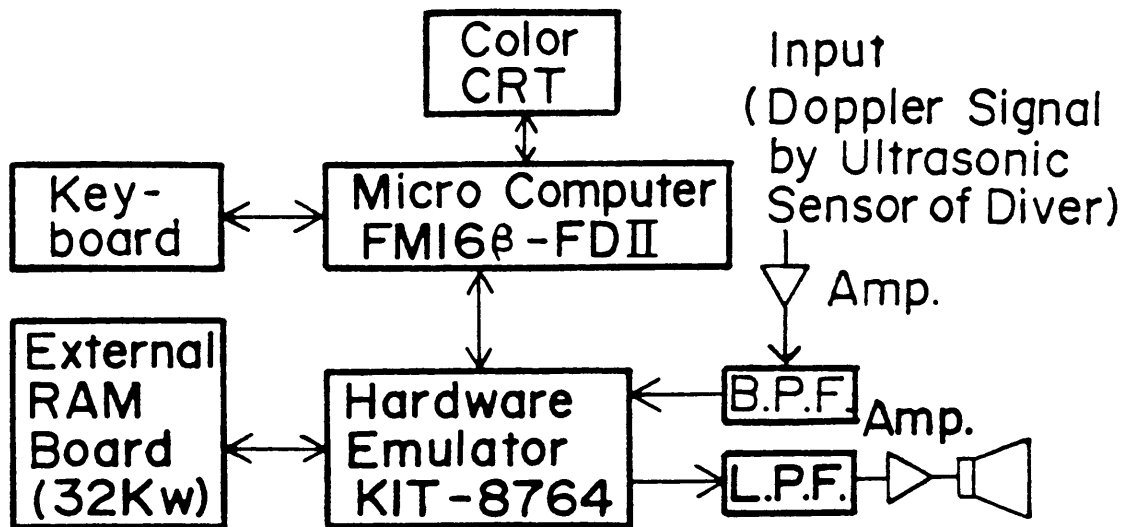


図 8.1: DSP を用いた自動気泡検出システムの概略

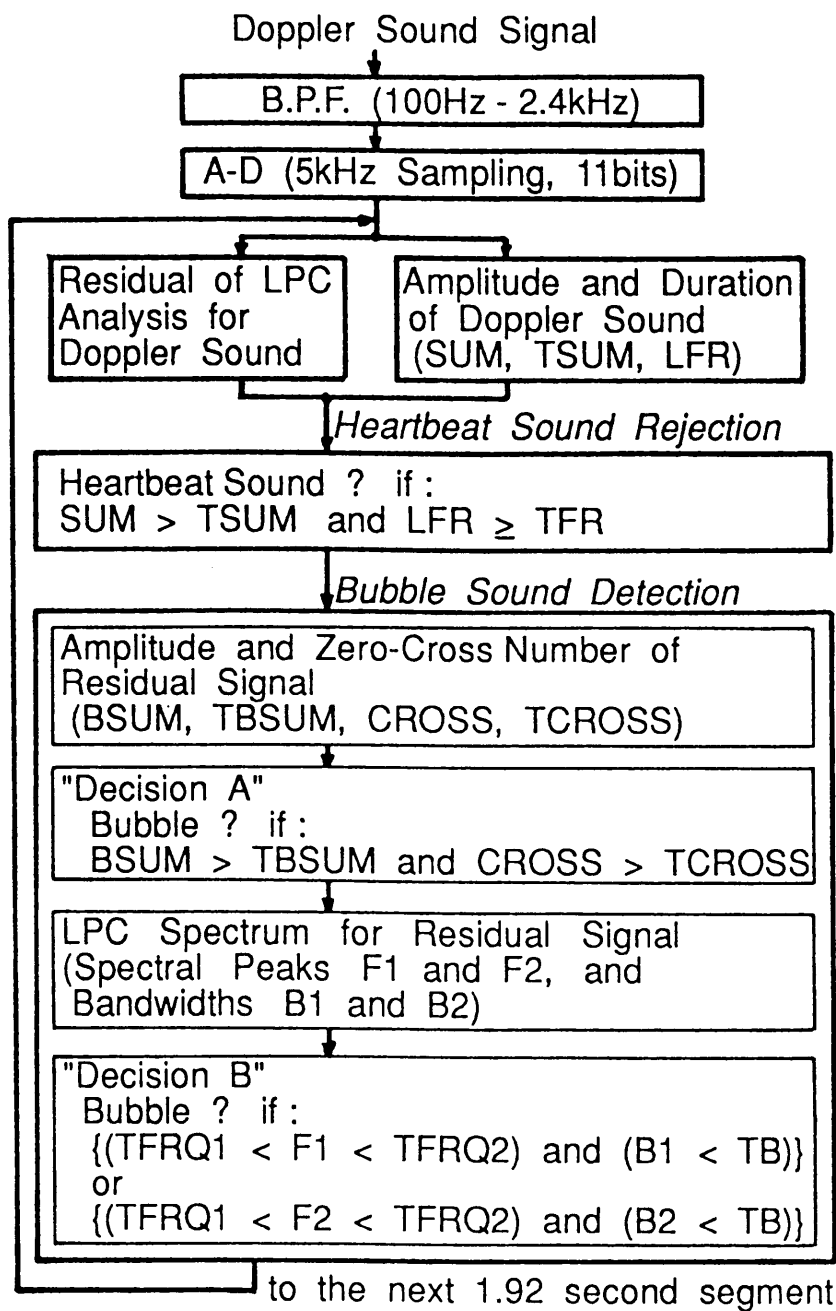


図 8.2: 気泡検出手順の流れ図



浮動小数点演算など行っている部分がある。倍精度固定小数点は 9600 点の自己相関を求めるときに、浮動小数点演算は気泡音判定 B で多項式の根を求めるときに使用した。この表は 1.92 秒 (96 フレーム) あたりの処理時間を示している。気泡音検出 A の零交差数と気泡音検出 B のフレーム数は平均的な値である。処理時間は 0.183 秒で、波形データの外部 RAM への転送、DSP プログラムの転送等を含めると約 1 秒である。気泡検出はドップラ音のデータ 1 分程度を用いれば判定できる。それに要する処理時間は 30 秒程度であるから、本システムは気泡検出用として十分な処理速度である。

このシステムにより気泡音のデモンストレーション用のテープを分析した結果を表 8.2 に示す。段階 1 の気泡検出頻度は、段階 2 から 4 に比較して著しく小さく、段階 1 を検出するという当初の目的を果たすことができた。ただし、段階 3 より段階 4 の方が気泡数が少なく区別できないのは前章の判定と同じである。

なお、このプログラムでは、鼓動音の判定、残差の振幅による判定、零交差数の判定、そして極の判定、の各過程における判定結果や、全データについて計算した残差波形が分析後にも残るようにしてある。これらのデータは必要に応じてシステムプログラムから呼び出せるような機能を持たせている。

## 8.5 アルゴリズムの改良

前アルゴリズムでは気泡音を鼓動区間を除いた部分で検出していた。しかし、鼓動区間とそれ以外とに同じように気泡音が存在するかは実際のドップラ音では確認されていなかった。ここでは、次節に示すように実際に漁業で潜水しているダイバーと 300 メートル有人潜水模擬実験のダイバーから得られたドップラ音のデータにより検討を行った。図 8.3(a) に気泡音を含むドップラ音を、(b) に 3 次の線形予測の残差波形を示す。気泡音の波形は紡錘形をしていて、振幅は他の部分より大きく持続時間は約 10ms であることがわかる。

図 8.4 に鼓動区間に気泡が存在するドップラ音の例を示す。図 8.4(a) は元の波形、(b) は (a) の一部を拡大した波形、(c) は 3 次の線形予測の残差波形を示す。この波形は図 8.3 と同じようであり、気泡音である。

前アルゴリズムでは鼓動区間を除く信号のみから気泡音を検出するように設計した。しかし、実際は図 8.4 に示したように鼓動区間に気泡音が存在する。したがって、鼓動音を含むすべての信号から気泡音を検出することを試みる。この目的のため図 8.5 の破線で示すように「気泡音判定」の判定 B で取り扱う区間に鼓動区間の分析を追加するようにアルゴリズムを変更する。

## 8.6 ドップラ音の資料

3 つの録音テープについて分析した。1 つは気泡検出の指針となるデモンストレーションテープ (前アルゴリズムと同じテープ, テープ A) である。2 つ目は 1991 年の 2 月に長崎県のたいらぎ貝採取ダイバー 3 人から録音されたテープ (テープ B) である。採録条件は潜水する前の安静時と屈伸運動中と、潜水終了の後の安静時と屈伸運動中の 4 つであ

表 8.1: 演算処理数と演算時間 (クロック周波数 8MHz)

処理過程	フレーム数	処理内容	積和の数	演算 ( $\mu$ s)	演算時間
鼓動区 間検出	96	二乗和	9600	I	6163
		その他	1500	I	2325
気泡音 検出 A	96	自己相関	38400	D	34125
		線形予測	20	I	25
		残差	15000	I	22663
	50	絶対値和	5000	I	4800
	30	零交差	4000	I	7600
		その他	200	I	338
気泡音 検出 B	20	窓掛け	15000	I	6850
		自己相関	30000	I	25413
		線形予測	600	I	488
		極の計算	3000	F	66575
		その他	200	F	6112
合計			122520		0.183 秒

I : 固定小数点演算、D : 倍長固定小数点演算、F : 浮動小数点演算

表 8.2: 気泡検出用デモテープを判定した結果

段階	データ長 (秒)	気泡フレーム数	判定段階
0	25.76	3	0
1	25.76	14	1
2	25.76	49	2
3	25.76	116	3
4	17.84	69	3

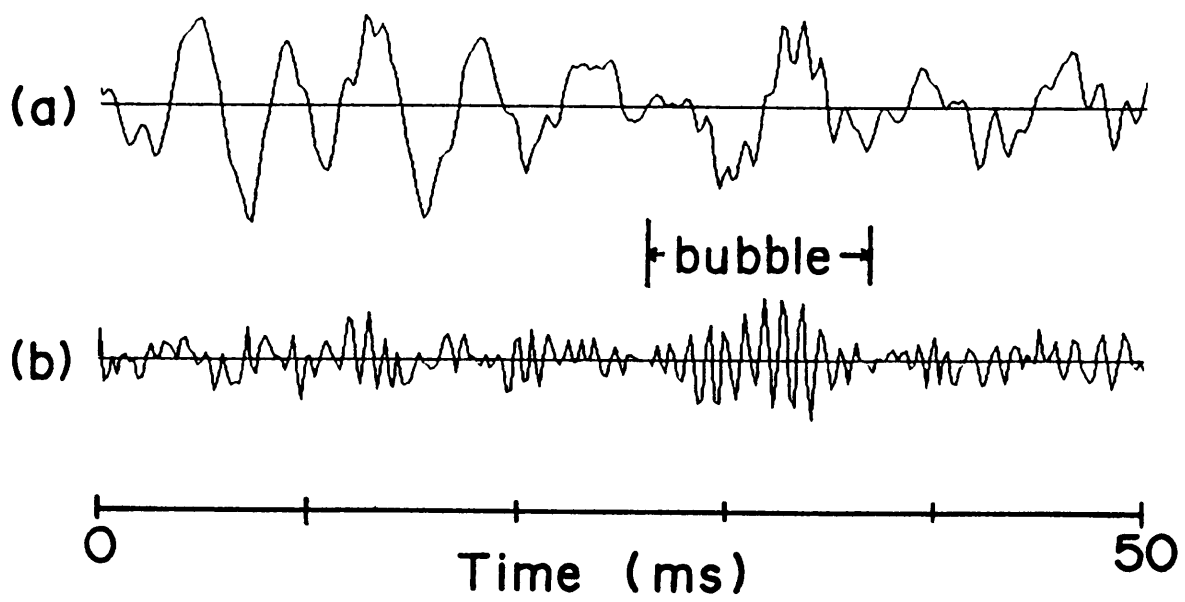


図 8.3: (a) 気泡音を含むドップラ音の波形と (b) その 3 次の線形予測の残差波形

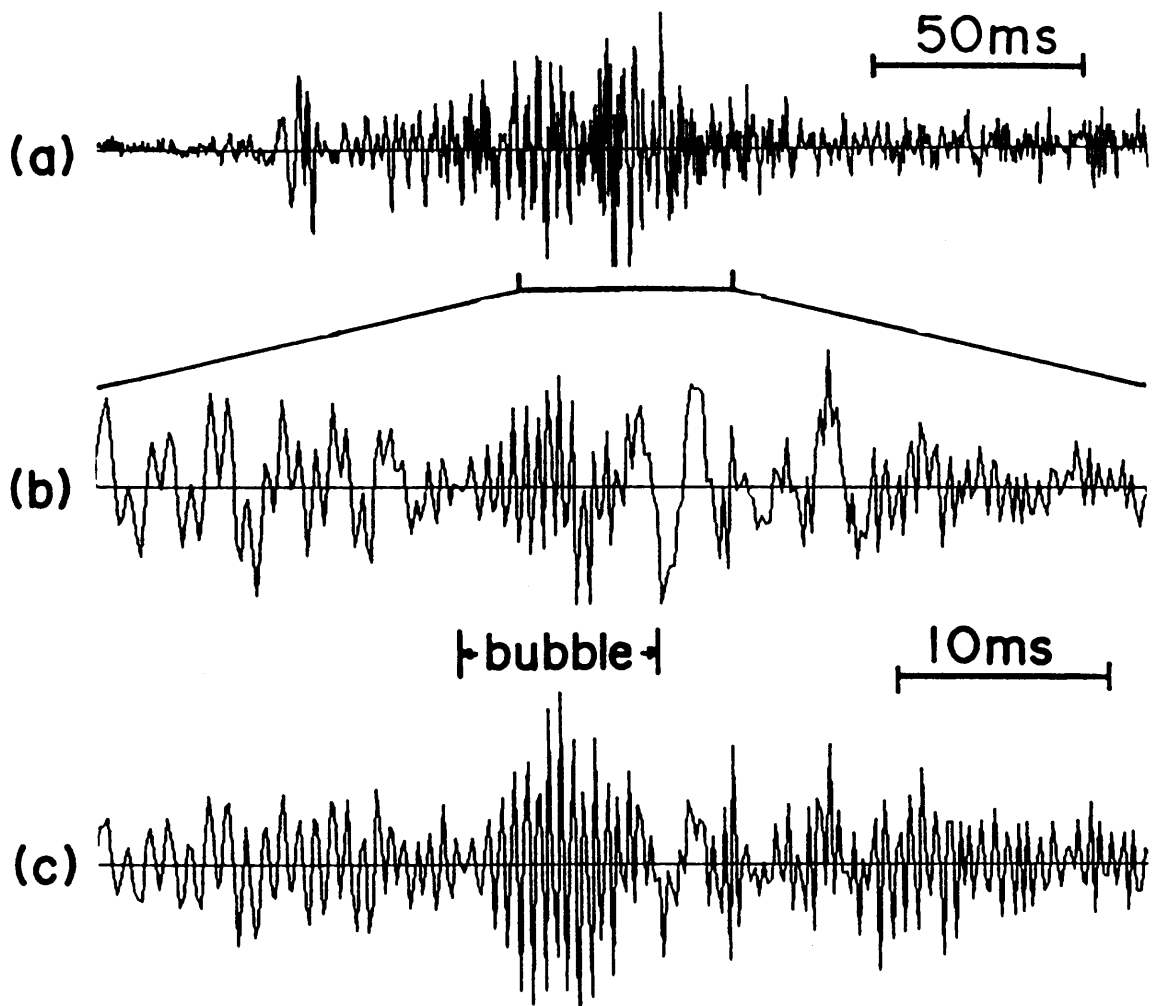


図 8.4: (a) 鼓動区間に気泡音を含むドップラ音の波形, (b)(a) の一部 50ms の拡大波形と (c)(b) の部分の 3 次の線形予測の残差波形

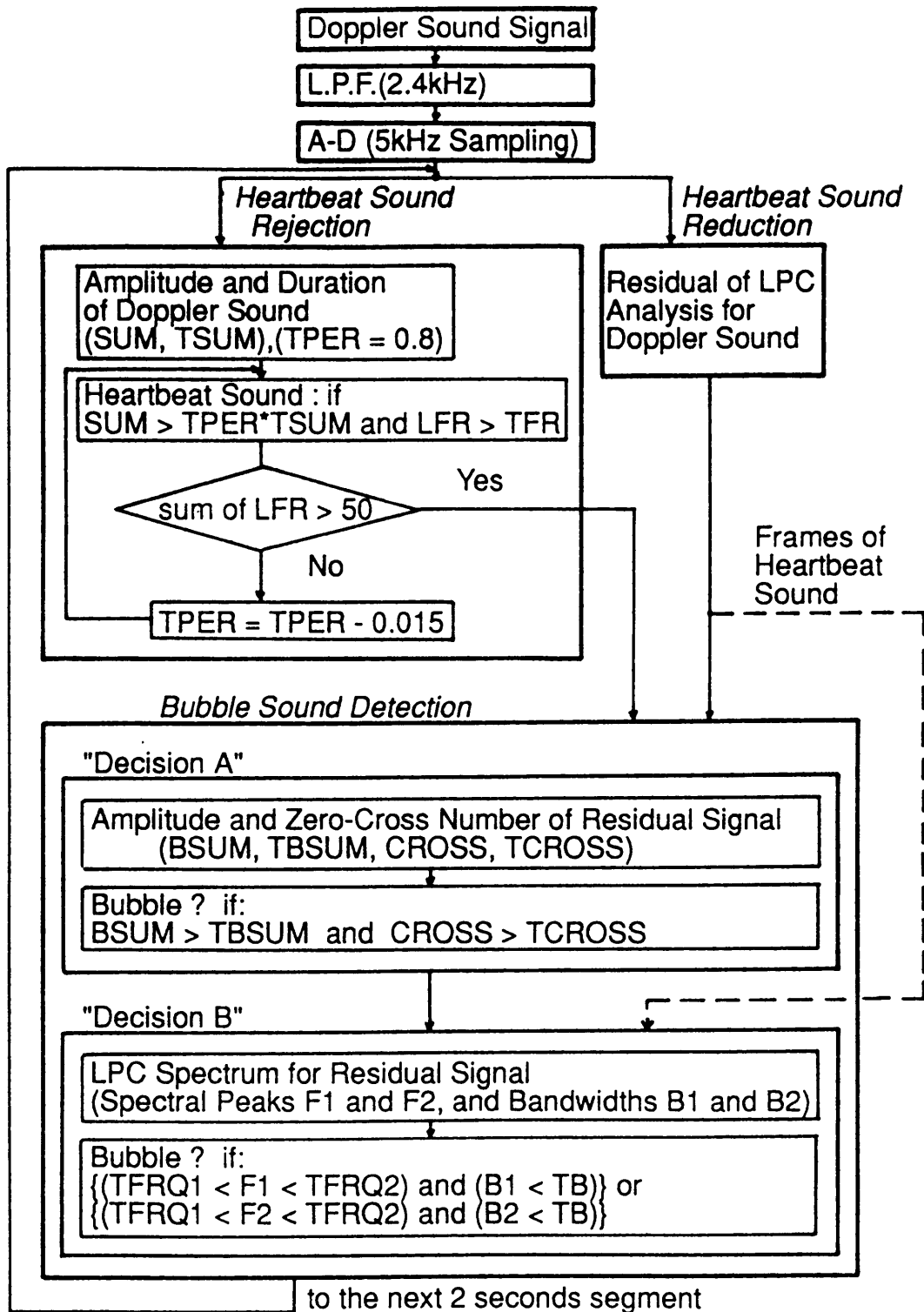


図 8.5: 改良アルゴリズムによる気泡検出手順の流れ図

る。したがって、テープ B では 12 例である。3 つ目は 1991 年 11 月海洋科学技術センター (JAMSTEC) で行われた 300 メートル有人潜水模擬実験の際に 4 人のダイバーから録音したテープ (テープ C) である。深度は 0、12、33、49、59 メートルで、それぞれ安静時と屈伸運動中に録音した 2 種類である。データは一人のダイバーの録音状態が悪かったので、40 例のうちの 28 例を使用した。

## 8.7 改良アルゴリズムによる結果

表 8.3 にテープ A における 1 鼓動あたりの気泡候補フレーム数とそのピーク周波数の分布を示す。(a) が鼓動区間を除いた場合 (前アルゴリズム) の 1 鼓動音あたりの気泡候補フレーム数であり、(b) が鼓動区間を含めた場合 (改良アルゴリズム) である。今回のアルゴリズムの方が気泡候補数が 2~3 倍多いことがわかる。しかし、両アルゴリズムとも段階 4 を除き段階が高くなれば気泡候補フレームが増えることには変わりはない。

表 8.4 にテープ B と C における 1 鼓動あたりの気泡候補フレーム数、そのピーク周波数の分布と判定結果を示す。1 つはテープ A と同じようであり、どちらのアルゴリズムで行っても判定される段階は変わらないもの (タイプ 1) と、テープ A と異なり、鼓動区間の気泡候補フレームが非常に多く判定される段階が高くなるもの (タイプ 2) の 2 つのタイプがあることがわかる。結果として、テープ B と C の 40 例のうち 19 例がタイプ 1 であり 21 例がタイプ 2 である。よって、段階の判定に鼓動区間を含めないと低い段階に判定する場合があります、改良アルゴリズムのように鼓動区間を含めて分析する必要がある。

また、表 8.3 を見ると、段階 3 と段階 4 とでは 1 鼓動あたりに含まれる個数では段階 3 の方が多いが、気泡音の周波数が 1500 から 2000 Hz にある個数は段階 4 の方が多い。よって、デモテープでは全体の個数ではなく、周波数の高い部分の個数により段階 3 と 4 の判定ができる。しかし、実際のデータは今のところ入手できていないので、今後、確かめる必要がある。

## 8.8 まとめ

前章で開発した気泡音検出アルゴリズムを自動化し、DSP を用いたほぼ実時間のシステムを開発した。すなわち、鼓動区間フレームを 50 フレーム以下になるように自動的にいき値を決め、鼓動区間フレームとする。一方、3 次の線形予測分析の残差を求める。この残差のうち鼓動区間フレーム以外について 2 段階で気泡音が存在するか否か判定する。まず、振幅と零交差数で予備選択した後、潜水前のデータから求めた気泡音の存在する周波数範囲と 4 次の線形予測分析の根から気泡音を検出するというアルゴリズムである。このシステムは、パソコンとハードウェアのエミュレータに外部 RAM32K ワードを追加した構成である。この装置により、ダイバーごとに潜水前に心臓大動脈付近に当てた超音波のドップラ変調信号を分析 (1.92 秒のデータを入力し約 1 秒で分析) しておけば、ドップラ変調信号を入力するだけで気泡判定が自動的にできる。処理時間は 1.92 秒のデータにつき約 1 秒であり、前章のアルゴリズムと同等の正確さで判定できることが確かめられた。

表 8.3: 気泡検出用デモテープの場合の 1 鼓動当たりの気泡候補数とその周波数分布

(a) 前アルゴリズム

段階	600-1000Hz	1000-1500Hz	1500-2000Hz	合計
0	0	0.09	0	0.09
1	0.09	0.29	0.04	0.41
2	0.14	0.92	0.31	1.36
3	0.26	2.86	0.20	3.31
4	0.15	2.15	0.35	2.65

(b) 改良アルゴリズム

段階	600-1000Hz	1000-1500Hz	1500-2000Hz	合計
0	0.25	0.19	0.13	0.56
1	0.24	1.20	0.04	1.48
2	0.53	2.75	0.69	3.94
3	0.95	11.11	0.60	12.66
4	0.77	8.00	1.15	9.92

表 8.4: テープ B と C の場合の 1 鼓動当たりの気泡候補数とその周波数分布と判定段階の例 (ここで、b:潜水前, a:潜水後, r:安静時, e:運動中, p:前アルゴリズム, i:改良アルゴリズム, 0m:深度 0m, 12m:深度 12m である)

(a) テープ B

潜水土	条件	アルゴリズム	600- 1000Hz	1000- 1500Hz	1500- 2000Hz	合計	判定 段階
1	b,r	p	0.06	0.03	0	0.09	0
		i	0.60	0.03	0.60	1.23	1
1	b,e	p	0	0.02	0	0.02	0
		i	0.34	0.14	1.00	1.48	1
1	a,r	p	0.36	0	0.10	0.46	1
		i	1.23	0.28	1.79	3.30	2
1	a,e	p	0	0.02	0.06	0.09	0
		i	1.91	0.09	0.96	2.95	2
2	b,e	p	0.03	0	0	0.03	0
		i	0.08	0	0.02	0.10	0
2	a,r	p	0.24	0.21	0.06	0.50	1
		i	0.47	0.44	0.20	1.11	1

(b) テープ C

潜水土	条件	アルゴリズム	600- 1000Hz	1000- 1500Hz	1500- 2000Hz	合計	判定 段階
3	0m,r	p	0	0.10	0	0.10	0
		i	0.73	0.10	0.07	0.90	0
3	0m,e	p	0	0.22	0.15	0.37	1
		i	1.02	0.67	0.85	2.54	2
3	12m,r	p	0.12	0.03	0	0.15	0
		i	0.55	0.18	0	0.73	0
3	12m,e	p	0.45	0.18	0.11	0.74	1
		i	2.13	0.66	0.82	3.63	2
4	0m,e	p	0.03	0	0	0.03	0
		i	0.38	0.03	0	0.41	0
4	12m,r	p	0.07	0	0	0.07	0
		i	0.63	0.03	0.03	0.70	0



さらに、気泡音の波形の特性を明らかにした。その観測結果より、気泡音はドップラ音の鼓動区間にも存在することを明らかにした。そこで前アルゴリズムに鼓動区間を含めた改良アルゴリズムで判定を試みた。その結果、デモテープでは判定結果は変わらなかったが、実際の潜水土からの収録の場合では、鼓動区間以外より鼓動区間内に多くの気泡が検出される場合があり、このときに前アルゴリズムより判定される段階が高くなることがわかった。結論として、気泡音の検出アルゴリズムは、3次の線形予測の残差を求め、振幅の大きい鼓動区間以外を振幅と零交差によって予備選択する。その後、鼓動区間と予備選択された区間を含めて残差波形をさらに4次の線形予測により気泡音がある区間であるかを決定する。このようにアルゴリズムは多少複雑になったが、よりよく検出されるようになった。また、段階3と段階4の区別については、気泡音の数ではなく気泡音の周波数の分布に差があることを示したが、実データが存在しないので、今後確かめたい。

今後は多くのデータでこのシステムの有効性を確かめ、より使い易いシステムに改良する予定である。

## 第 9 章

### 結論

本論文では環境圧潜水において潜水活動を支援する 2 つのシステムについて述べた。

それは、音声通信に必要なヘリウム音声の修復と潜水病予防のための潜水士の血液中に発生する気泡の自動検出についてである。

第一の課題としてヘリウム音声の修復について述べた。

まず、ヘリウム音声のデータ収集とそのときに使用するマイクロホンの特性について述べた。音声資料はヘリウム音声の特性や装置のよさを調べるため通常電話で使用される明瞭度音節表、単語リスト等を用いた。深度 300m という高圧ではマイクロホンの特性も変わることが予想されるので、標準マイクロホンであるコンデンサマイクロホンの他、エレクトレット型マイクロホン、ダイナミックマイクロホンを用いて特性を検討した。1/2 インチ・コンデンサマイクロホンでは感度が下がるものの周波数は 10kHz 以上まで平坦であり、ヘリウム音声の録音には支障がないことを示した。また、エレクトレット型マイクロホンも使用可能であることを示した。ダイナミックマイクロホンは 200Hz 付近にピークが、8kHz 付近にデップができ高圧では使用できないことがわかった。これらのマイクロホンでヘリウム音声を収録した。

次に、音声資料に基づきヘリウム音声の性質を分析した。母音だけでなく無声子音についてもほぼ Fant らの関係式が成立し、ホルマント周波数は音速に比例して上昇するとともに、声道壁が剛体とみなせなくなるために低いホルマント周波数で相対的により高く周波数が移動することが確かめられた。また、深度が増すと子音は一般的に母音に比べ音圧が低くなること、ピッチは高くなるが高々 30% であることを示した。

このようなヘリウム音声を修復することを目的として、残差のダウンサンプリング以外はデジタル回路である装置 (1 号機) の開発について述べた。ここで提案したアルゴリズムは深深度のヘリウム音声を修復できるようにホルマント周波数を非線形に変換できるものである。それは、LPC 分析で抽出した声道の極周波数を LPC の一種である LSP 係数を用いて非線形変換し、この LSP 係数と残差波形を用いて再合成するという方式である。この変換に伴うホルマント周波数の誤差を求めた。第 1 ホルマント周波数が低いときに誤差が大きいのが、 $\pm 25\text{Hz}$  以内におさめることができることを示した。次に、潜水実験の現場に持ち込んで実際のヘリウム音声を修復して正確に評価するために、実時間で動作可能なシステムの設計と製作についてその概要を述べた。演算は固定小数点演算とした。その演算語長と処理方式について述べた。全体を前処理、分析、変換、残差のダウンサン

プリングと合成の5つの処理ユニットに分け、パイプライン処理で実現した。入力された音声は修復されて出力されるまでに50msの遅れがある。ICは約800個で、幅51cm、奥行き40cm、高さ35cmの箱に収納されている。

この試作システムについて市販の修復装置(アナログ方式のピッチ同期型波形伸張方式)とともに評価した。修復に用いる変換方式と実時間動作のデジタル装置としての設計・製作上の問題点、およびヘリウム音声の物理的性質などについても検討した。その結果、ここで採用した修復法によって深度200mまでのヘリウム音声については市販の修復装置より明瞭度が高く、単語理解度は90%以上に修復された。深度300mのヘリウム音声では市販の修復装置より明瞭度が多少低かった。ただし、シミュレーションの修復音では市販の修復装置より明瞭度は高かったため、装置の設計・製作上の問題点がありその点について考察を加えた。非線形に変換した音声の明瞭度は、線形変換した音声の明瞭度より多少低かったが、音質の自然性は増した。このアルゴリズムにより深度300mまでの音声は了解度、声質とも実用上十分な程度に修復できることを示した。

次に、この試作機の結果を踏まえ、第2世代の高速の固定小数点タイプのDSPを用いてアルゴリズムの改良と種々の使用条件に耐えられる全デジタル回路のシステムを試作(2号機)した。その結果、LSP係数を変換するアルゴリズムを改良すれば、変換に伴う誤差は減少することを示した。しかし、了解度試験の結果では、線形変換に比べ非線形変換が明確によくなるとは言えなかった。ただし、突然の雑音による発振とか係数がうまく求まらなかった場合のエラー処理を強化し実用機として十分動作するシステムとした。このシステムは、試作1号機と基本構成は同じで、全体は5つの処理ユニットから構成されている。DSPを5個用い、ICの総数は約200個で、A3の大きさ1枚の基板に納められている。消費電力は約40W、試作1号機の約1/4の大きさである。メモリをROM化すれば約160個に削減できさらに小型化される。

今後の課題としては、有声子音における第1ホルマント周波数の低下の原因とそれに対応するアルゴリズムを開発し、より自然で音質のよいシステムを構築することである。また、騒音が大きい環境でもあるので、その対策も必要である。

二つの課題のうちのもう一つの課題である潜水病の予防および治療のために減圧時に潜水士の血液中に発生する気泡を検出し、自動診断することについて述べた。身体を循環した血液は必ず肺動脈を通るので、従来法では肺動脈に超音波を照射しドップラの原理による変調信号(ドップラ信号)を聴覚的に判断して気泡音を検出する。この信号には鼓動音等の気泡音以外の雑音が混入しているため専門家でも判定は容易ではない。本研究では、ドップラ法によって収録された信号より、まず気泡音の性質を明らかにした。次に、線形予測法により周期的である鼓動音を予測し、その残差信号を用いることで突発的に発生する気泡音をほぼ残したまま鼓動音を軽減した。さらに、この残差信号を用いて振幅と零交差数、周波数分析を用いて気泡音を検出するアルゴリズムを開発した。その結果、デモテープではSpencerの5つの段階区分のうち気泡のない段階0から段階3までの区分を正しく判定できた。気泡が少ないときの判定は専門家でも難しく、本装置の性能は潜水病の予防の観点からは十分であるといえる。また、海洋科学技術センターでの模擬実験における潜水士からのデータでは気泡検出の専門家と同一の段階区分で、正しく判定された。

次に、DSPを用いた準実時間の気泡検出装置を試作した。鼓動区間の設定等を自動化

するようにした。これらの設定を自動化しても上述の計算機のシミュレーションと同じ段階区分の判定ができた。処理時間はデータ 1.92 秒について約 1 秒であり、血液が全身をめぐり戻ってくる時間はおよそ 1 分間なので、この 1 分間のデータを用いて診断するとすると、データを取り終えた後、約 30 秒で診断が可能となった。

また、アルゴリズムを改良し、鼓動区間を含めて分析した方がよいことを実際のデータで示した。段階 3 と 4 の判定は気泡音の周波数 1500 から 2000Hz にある個数で判定すればよいことをデモテープのデータにより示した。

今後の課題は、さらに多数の人で判定を確かめることと、段階 3 と 4 を実際のデータで判定することである。

## 今後の展望

ヘリウム音声の修復に関しては、深度 300 メートルのヘリウム音声でほぼ了解できるまでに修復された。ただし、その音質の問題、話者自身の発声の問題、マイクロホンの問題等いくつか残された問題があるので、今後の展望とともに述べる。

開発したヘリウム音声修復装置はある程度の明瞭度が得られたが、その音質はまだヘリウム音声に近い。潜水士同士の会話や地上との通信ではこの修復装置に頼らざるを得ないので、できるだけ自然に聞こえることが利用度を高め、また通信による疲労度を減少させる。このためには、ホルマント周波数の非線形変換による降下が極めて重要である。非線形変換の度合いを強くできない原因については 6 章で述べた。今後、この非線形変換の度合いを強めるには、大量のデータによる分析と変換方式の検討が必要である。我々が構築し、データを登録しているヘリウム音声データベース (付録 D)[48],[49] は今後のヘリウム音声のよりよい修復方法を検討するのに重要であり有用であるので、できるだけ早く完成させたい。

また、ヘリウム空気中ではヘリウム音声データベースの結果から明らかになったように、語頭で有声音のバズバーがなくなり、中国語で見られる無声無気破裂音 [51] に近くなるものが増える。この音声は有声音の度合いが低く、場合によっては無声音に聞こえるものも現れる。この原因を実験と理論の両面から明らかにする必要がある。

今まで述べたヘリウム音声の修復では、発声は通常空気と同じ声道の形で行われていることを前提条件としている。しかし、話者自身もヘリウム空気中ではヘリウム音声が入耳からフィードバックされている。この声は話者にとっても理解できる音声ではなく、発声する時は無視するしかない。この結果、本来発声を制御している耳からのフィードバックがないのと同様なので、曖昧に発声される場合がある。これが、ヘリウム音声を修復してもよくなる原因の一つであるので、今後修復音を発声の時に耳にフィードバックさせて、明瞭度が改善されることを確かめたい。

通信に使用可能なマイクロホンの特性は、ヘリウム音声では 10kHz までほぼ平坦であることが要求される。しかし、現在までにこの特性を満たすものは、標準マイクロホンとしてのコンデンサマイクロホンとエレクトレット型コンデンサマイクロホンに限られている。コンデンサマイクロホンは高価で、しかも取り扱いに注意が必要であり、また湿度の管理も要求される。エレクトレット型コンデンサマイクロホンもコンデンサマイクロホン

ほどの取り扱いの注意は必要ではないが、電池が必要であり、水に濡らすことは厳禁である。潜水という環境では電池が不要で、水に濡れる程度では故障しないマイクロホンの開発が必要である。周波数特性が平坦でないマイクロホンで収録した音声ではヘリウム音声の修復はその実力を発揮できないので、この開発はヘリウム音声修復装置の普及の鍵になる。

潜水土同士や地上との通信では潜水土一人一人が修復装置を携帯し携帯電話のように交信できることが理想である。これは通信だけでなく潜水土同士が通じないという精神的なストレスも減少させるので装置のいっそうの小型化が急務である。

音声通信を行うにはこの他に解決しなければならない課題がある。修復音声の出力であるヘッドホンやスピーカの特性が変化する [50] ので高圧で使用できる物を見いだすか開発する必要がある。また、聴覚の特性が変化する [52] のでその特性も考慮しなければならない。更に高圧環境では騒音が大きい。これを減少させることもヘリウム音声の修復には重要な要素である。騒音源自身の騒音を減少させるとともに、吸音材の開発が必要である。一般に通常空気より吸音率が減少することを測定で確かめている [53]。今後、測定とともに高圧で吸音率の高い吸音材の開発について検討したい。

ヘリウム音声は、密度が高く、音速が速い気体中で発声されてはいるが、音声には変わりはない。従って、音声の生成モデルの検討を行うときに通常空気中で良く一致するだけでなく、ヘリウム空気中で一致すれば、そのモデルの信頼性は高いといえる。また、気体密度が高いので声道壁は剛体という仮定が成り立たないので、通常空気ではよくわからない現象が際だってくる可能性がある。この意味でもヘリウム音声のデータベースは必要と考えられるので早期に完成させたい。

ヘリウム音声の修復の技術は、音声の性質変換のひとつであり、男声を女声にする変換や男声を子供の声にする変換等への応用が考えられる。また、実時間の音声分析を用いた聴覚障害者への補助機器の応用等も考えられる。

潜水浮上時に血液中に発生する気泡検出では、減圧症 (潜水病) の予防が最も大切である。システム潜水と呼ばれる大規模な装置で潜水後の減圧を十分管理していても、数%の人には長年の潜水により骨壊死 (特に大腿骨や肘の関節付近の骨が腐り歩行等が困難になる) [3] という症状が発生する。圧搾空気を使用したヘルメット潜水では十分な管理がなされてはいないのでさらに発症率は高い。この原因は体内に発生する気泡による。これを防ぐには潜水浮上時に毎回モニターし、気泡の発生の場合には何らかの処置を施す必要がある。すべての潜水において気泡を判断できる専門家が常時いるわけではない。また、専門家にとっても段階 1 程度ではその判定も根拠が必要である。減圧時に自動的に気泡判定ができれば早期に処置でき、潜水の安全性が確保される。アルゴリズムとともに現場に持ち込めるようなさらに簡便な装置が今後要求される。

気泡検出では超音波が肺動脈にきちっと当たっていることが前提条件となる。現アルゴリズムではこの条件が満たされているドップラ音について判定が可能である。この条件が満たされていない場合には、気泡検出は行わず、データを棄却する必要がある。このアルゴリズムを開発することも今後の課題として残されている。

また、この気泡検出法では、体内にある静止した気泡の検出は不可能である。超音波画像を利用した方法が開発されてはいる [54],[55] が、気泡は小さいものは  $1\mu\text{m}$  以下 [56] であり今のところ検出できる装置はない。厳密な診断には、この静止した気泡の検出も必

要であり、今後の課題として残されている。

集積回路技術の進歩に伴い、高速で多機能のデジタル IC が開発された。本論文では、それらの IC を使用して安全に潜水するために開発した支援システムについて述べた。高速積和器やビットスライス・マイクロプロセッサ、および固定小数点演算タイプの高速 DSP(第 2 世代)の IC を用い、演算数、メモリ数、語長、構成、電源などに非常にこまかい検討をして実時間修復装置を作り上げた。ここで検討し、実際開発したハードウェアとソフトウェアの技術の集積は、極めて貴重なものであり、本研究によって確立された設計技術は、今後もっと大規模で複雑な処理装置を設計したり、今後開発される高機能の DSP の開発や改良の際に大きな貢献をするであろう。

# 謝辞

この研究を進めるにあたり、研究当初よりご指導いただいた鈴木久喜教授に深謝いたします。

論文審査にあたり熱心に御討論いただいた水品静夫教授、福田明教授、深林太計志教授、北澤茂良教授の各先生方に感謝いたします。

ヘリウム音声の収集や気泡検出のデータ収集に際しては、海洋科学技術センターの毛利元彦博士、説楽文朗氏、水嶋康夫氏をはじめ多くの方々の協力をいただきました。ここに、深謝します。また、気泡検出のデモテープを提供していただきこの研究のきっかけを与えてくださった元海洋技術センターの関邦博博士(現神奈川大学)に深謝します。

文部技官・石田敬三氏には日頃から多大な協力をいただきました。ここに深謝します。研究室の学生諸氏には、この研究を遂行するにあたり多大の協力をいただいた。ヘリウム音声の性質については加藤真一、清水啓司、山田安男、鈴木忠氏の、マイクロホンについては、吉田慎介、藤本敦、大橋正尚、内田義幸、小林守、河西康仁氏の、ヘリウム音声修復装置(1号機)では川本雅之、大久保克彦、高畠正人、斉藤洋氏の、ヘリウム音声修復装置(2号機)では笹間昭氏の、ヘリウム音声データベースでは千葉哲央、城内良直氏の、気泡検出では渡辺政博、木村慰作氏の協力をいただいた。ここに、感謝いたします。その他、研究室を卒業された多くの学生の協力を得たことを記して、感謝の意を表します。

# 付録 A

## ヘリウム音声収録用テキスト

海洋科学技術センター有人潜水模擬実験にて収録した音声データのテキストの例を以下に示す。



明瞭度試験用音節表の例（平等率 B-3）

	1	2	3	4	5
1	レ パ	ロ ビヤ	ビヤ キョ	オ ド	ミ リユ
2	キャ テ	ゴ ニヤ	ラ ギヤ	ル ア	ブ キユ
3	ピユ メ	リ シャ	ガ ト	ピョ マ	サ ニユ
4	フ ビユ	ヒ ヒョ	ゼ ジ	ス ミョ	セ ダ
5	エ ギユ	グ ミヤ	ゲ ヤ	ビ ビョ	チ ジョ
6	ニョ ズ	ク ホ	チャ ム	モ リヤ	ギ カ
7	ニ ギョ	ブ ボ	ペ チョ	ゾ ケ	イ ヒヤ
8	ジャ ウ	リョ ヘ	チュ コ	ツ ソ	ジュ バ
9	ミュ タ	ショ ハ	ザ ビ	デ ノ	シ ベ
10	ヌ ワ	ユ ネ	キ ポ	シュ ナ	ヒユ ヨ

2音節ワードリスト

	1	2	3	4	5
1	彼の カノ	雨 アメ	医者 イシャ	浮く ウク	演技 エンギ
2	追う オウ	帯 オビ	買う カウ	壁 カベ	絹 キヌ
3	濃い コイ	呼吸 コキュー	こんな コンナ	午後 ゴゴ	御覧 ゴラン
4	七 シチ	姿勢 シセー	為る スル	相談 ソーダン	ずっと ズット
5	全部 ゼンブ	縦 タテ	土 ツチ	当然 トーゼン	通る トール
6	共 トモ	だって ダッテ	同時 ドージ	何度 ナンド	煮る ニル
7	日本 ニホン	端 ハシ	はっと ハット	引く ヒク	普通 フツー
8	父母 フボ	不満 フマン	バス バス	前 マエ	丸 マル
9	無理 ムリ	有効 ユーコー	雪 ユキ	廊下 ローカ	脇 ワキ
10	主婦 シュフ	昭和 ショウワ	助手 ジョシュ	注射 チューシャ	病人 ビョーニン

3音節ワードリスト 3-A

	1	2	3	4	5
1	愛児 アイジ	頭 アタマ	嵐 アラシ	急ぐ イソグ	一体 イッタイ
2	浮かぶ ウカブ	上衣 ウワギ	男 オトコ	及ぶ オヨブ	大きい オーキイ
3	飾る カザル	構う カマウ	関する カンスル	刻む キザム	暮らす クラス
4	クリーム クリーム	決して ケッシテ	言葉 コトバ	高い タカイ	卵 タマゴ
5	尖る トガル	どちら ドチラ	大根 ダイコン	机 ツクエ	詰まる ツマル
6	絶対 ゼッタイ	ラジオ ラジオ	醒める サメル	さっぱり サッパリ	次第 シダイ
7	知れる シレル	親切 シンセツ	座る スワル	俳句 ハイク	話す ハナス
8	浸す ヒタス	深い フカイ	ほめる ホメル	不可能 フカノー	招く マネク
9	認め ミトメ	燃える モエル	全く マッタク	なさる ナサル	濡れる ヌレル
10	述べる ノベル	やがて ヤガテ	豊か ユタカ	漸やく ヨーヤク	状態 ジョータイ

101 音節表

ア	イ	ウ	エ	オ			
カ	キ	ク	ケ	コ	キャ	キュ	キョ
サ	シ	ス	セ	ソ	シャ	シュ	シヨ
タ	チ	ツ	テ	ト	チャ	チュ	チヨ
ナ	ニ	ヌ	ネ	ノ	ニャ	ニユ	ニヨ
ハ	ヒ	フ	ヘ	ホ	ヒャ	ヒユ	ヒヨ
マ	ミ	ム	メ	モ	ミャ	ミユ	ミヨ
ヤ		ユ		ヨ			
ラ	リ	ル	レ	ロ	リャ	リュ	リヨ
ワ							
ガ	ギ	グ	ゲ	ゴ	ギャ	ギユ	ギョ
ザ	ジ	ズ	ゼ	ゾ	ジャ	ジュ	ジヨ
ダ			デ	ド			
バ	ビ	ブ	ベ	ボ	ビャ	ビユ	ビヨ
パ	ピ	プ	ペ	ポ	ピャ	ピユ	ピヨ
ン							

まぎらわしい音節

1	蝸牛・粥・湯 ある・ギャル・やる 企業・器用・気宇 牛・鶴・湯 要請・行政・王政 火薬・可逆	5	客・百 今日・表 ヒュー・九
2	丸・成る うに・海 物・桃 目・根 無視・主	6	下記・破棄 聞く・引く 節・櫛 変・県 腰・星
3	尿・妙・用 ヤ・ニヤ・ミヤ ミユ・ニユ・ユ	7	瓶・ジン ブック・ズック 伝統・弁当・全糖 銅・某・象 残・段・番
4	終夜・昼夜 課長・歌唱 チャコ・車庫		

文章了解度試験用単文表の例（単文表1）

2648と書け

「宵待草」という詩の作者は竹久夢二ですか

日本の川の名を一つ記せ

明治座は映画館ですか

100メートルを10秒で走れますか

塩は甘いですか

ラジオ放送で音楽を放送しますか

春夏秋冬のうち何れが暑いでしょうか

「金色夜叉」の主要人物は誰でしょう

中川一郎は大蔵大臣ですか

労働者の祭を何と言いますか

6月は秋ですか

「三ツ矢」がサイダーなら「キリン」は何ですか

スキーは南国で出来ますか

和気清麻呂と源義経とはどちらが古いですか

偶数に奇数を掛けると偶数ですか

「二見が浦」は日本海にありますか

野球場名を一つあげよ

現代の日本の小説家を一人あげよ

伊東駅は東海道線にありますか

映画俳優を一人あげなさい

西洋音楽は昔から日本にありましたか

議事堂はスポーツ場ですか

西郷さんの銅像はどこにありますか

7月7日は何の日ですか

## 付録 B

# 音速と密度の求め方および声道の最低共振周波数の推定

### 音速と密度

音速と密度は以下の式で求めることができる。  
理想気体として算出する。深度を  $d$ メートルとすると、  
密度  $\rho_d$ は、

$$\rho_d = \frac{\sum_i x_i m_i}{V_0} \frac{K_0 P_d}{K_d P_0}$$

ただし、

$x_i$ : 気体元素  $i$  の分圧 (ATA)

$m_i$ : 気体分子  $i$  の分子量

$V_0$ : 1 気圧下、1mol の理想気体  
( $22.41 \times 10^3 \text{ cm}^3$ )

$K_d$ : 気体の絶対温度

$K_0$ : 273.16K (摂氏 0 度)

$P_d$ : 混合気体の圧力  $\{(d/10 + 1)\text{ATA}\}$

$P_0$ : 0m の大気圧 (1ATA)

音速  $c_d$ は、

$$c_d = \sqrt{\gamma_d P_d / \rho_d}$$

で算出される。ここで  $\gamma_d$ は、深度  $d$ メートルの気体の定圧比熱と定積比熱の比であり、次式で算出される。

$$\gamma_d = \frac{\sum_i x_i C_{pi}}{\sum_i x_i C_{vi}}$$

$C_{pi}$ : 気体元素  $i$  の定圧モル比熱

$C_{vi}$ : 気体元素  $i$  の定積モル比熱

ただし、計算においては、酸素、窒素、ヘリウムのみとする。酸素および窒素は 2 原子分子であるので、定圧モル比熱は  $(7/2)R$ 、定積モル比熱は  $(5/2)R$  であり、ヘリウムは 1 原子分子であるので、定圧モル比熱は  $(5/2)R$ 、定積モル比熱は  $(3/2)R$  である。ここで、 $R$  は気体定数である。また、分子量はそれぞれ 32、28、4 である。

これらの値から音速比、密度比が求められる。なお、高压気体環境では酸素分圧が 0.3 から 0.4ATA、窒素が 0.79ATA、それ以外がヘリウムとして算出している。なお、ATA は絶対気圧であり 1ATA は 1013 hPa である。深度が 10 メートル増すごとに 1ATA 環境圧力は増す。

## 声道の最低共振周波数

文献 [36] によれば、通常空気中での声道の最低共振周波数  $F_{wa}$  は以下のように推定できる。

$$F_{wa} = \frac{1}{2\pi} c \sqrt{\rho} \sqrt{\frac{Sl}{Vm_w}}$$

ここで、

$c$ : 空気の音速 (35,000 cm/s)

$\rho$ : 空気の密度 ( $1.18 \times 10^{-3}$  g/cm<sup>3</sup>)

$S$ : 声道壁の周囲長 (半径  $r$ cm の円筒とすれば  $2\pi r$ )

$l$ : 声道の長さ (17cm)

$V$ : 声道の体積 (半径  $r$ cm, 長さ  $l$  の円筒とすれば  $\pi r^2 l$ )

$m_w$ : 声道壁の単位面積当たりの重さ (1.17~1.71g/cm<sup>2</sup>)

$r$ : 声道の半径 (1.5cm)

である。

上記のように、声道を等しい断面の円筒とすれば  $F_{wa}$  は  $m_w = 1.17$  g/cm<sup>2</sup> のとき 204Hz、1.71 g/cm<sup>2</sup> のとき 169Hz になる。

## 付録 C

# ヘリウム音声修復装置のアルゴリズムとハードウェアの詳細

### 前処理部

分析する前に行う処理としては、低域ろ波、A-D変換、プリアンファシス、最大値検出、左シフト、窓掛け、パワーの計算がある。

プリアンファシスは以下の式で行う。

$$y(n) = s(n) - s(n-1) \quad (\text{C.1})$$

ここで、 $y(n)$ : 時刻  $n$  におけるプリアンファシスされたデータ

$s(n)$ : 時刻  $n$  におけるデータ

$s(n-1)$ : 時刻  $n-1$  におけるデータ

左シフトは最大値検出によって得られた値により、この最大値が最下位ビット (LSB) から 11 ビット目になるようにシフトする。この理由は分析部で説明する。

窓掛けは窓長 30ms について行う。

$$\begin{aligned} x(n) &= y_l(n)Hw(n) \\ Hw(n) &= 0.54 - 0.46 \cos \{2\pi(n-1)/(N-1)\} \end{aligned} \quad (\text{C.2})$$

ここで、 $y_l(n)$ :  $y(n)$  を左シフトしたデータ

$x(n)$ : 窓掛けされたデータ

$Hw(n)$ : ハミング窓

$N$ : 3 フレーム当たりのデータ数

$n = 1, 2, \dots, N$

パワー  $U_0$  の計算は以下の式で求める。

$$U_0 = \sum_{n=1}^N x(n)^2 \quad (\text{C.3})$$



## 前処理の回路とタイミング

### 最大値検出部

図 C.1に示すように入力データの絶対値（ただし、負の数は回路の簡単化のため、1の補数をとっている）をとり、プライオリティ・エンコーダで上位何ビット目に1があるかを調べ、比較器によりそれ以前の信号と比べて大きい場合にはDラッチに格納する。比較する区間は窓長ごとである。

### プリアンファシス、左シフト、窓掛け、パワー計算

この部分の回路の概略を図 C.2に示す。プリアンファシスはDラッチと全加算器からなり、引き算は前データを反転し、全加算器のLSB（最下位ビット）への桁上りを行って行う。左シフトはシフトレジスタにより、窓掛けとパワーの計算は積和器であるTRW社のTDC1010Jで行う。求めた $x(n)$ は分析部での $\varepsilon_0^f(n)$ 、 $\varepsilon_0^b(n)$ となり、パワー $U_0$ はDラッチに格納される。

$s(n)$ のメモリはA-D変換器からは一定間隔で入力され、プリアンファシス以降は左シフトの数により処理時間が異なるため、メモリの入出力が同時に発生しないように考慮した。

また、窓データのROMは語長16ビット、2048点なので、標本化周波数により下記に示す番地 $ad$ から窓データを得ることにした。

$$ad = [2048(n-1)/(N-1)] \quad n = 1, 2, \dots, N \quad (C.4)$$

[ ]はガウスの記号、 $N$ は窓長あたりのデータ数 $N = 30f$ 、 $f$ : 標本化周波数(kHz)

この窓データによる誤差は、シミュレーションの結果、ホルマント周波数、帯域幅ともに2Hz以内である。

前処理の演算時間は、1データあたり、プリアンファシスに600ns、データシフトに最大1 $\mu$ s、窓掛けに1 $\mu$ s、 $U_0$ の計算に600nsかかる。よって標本化周波数30kHzのとき最大約2.9msとなり1フレームが10msなので仕様は満たされる。

## 分析部のハードウェア

前処理を含めた概略のタイミング図は図 C.3である。3フレーム分のデータのが得られた後に、次のフレームでプリアンファシス、左シフト、窓掛け、パワー計算を行い、さらにその次のフレームでPARCOR分析を行う。

修復音声の周波数帯域を5kHzとすると、最大標本化周波数は30kHz、分析次数は12となる。式4.6~4.10の演算回数は標本化周波数を30kHzとすると、式4.6は除算12回、式4.7は積和が $900 \times 12$ 回（窓長を30msとすると $N=900$ ）、式4.8は乗算が24回（ただし、 $U_0$ は前処理で計算済）、式4.9と式4.10は積和が $900 \times 12$ 回である。よって、最大積和32,400回、乗算24回、除算12回である。式4.6と式4.8（以下、「 $U&k_i$ 部」という）は語長が32ビットであり、演算回数もそれほど多くないのでマイクロプロセッサZ-80を

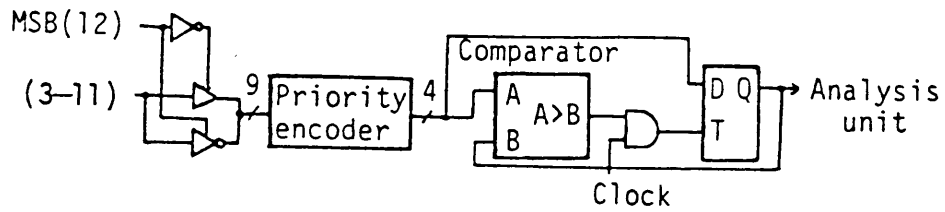


図 C.1: 最大値検出回路の概略図

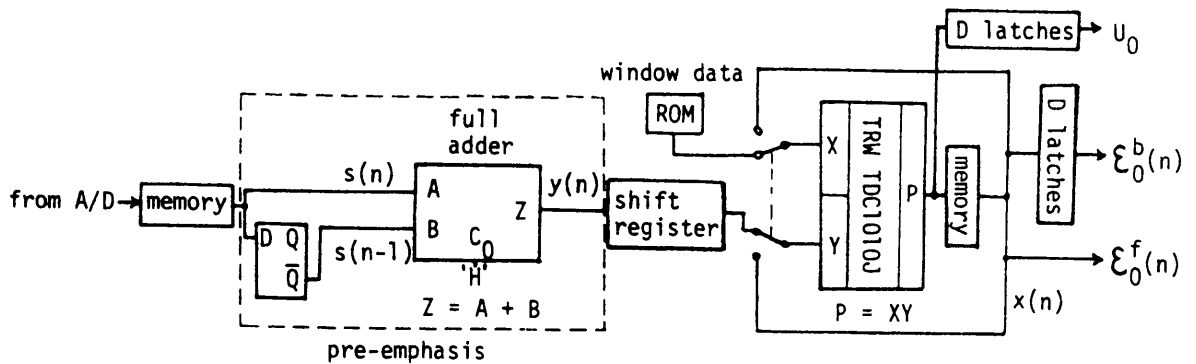


図 C.2: 前処理部（最大値検出部を除く）の概略図

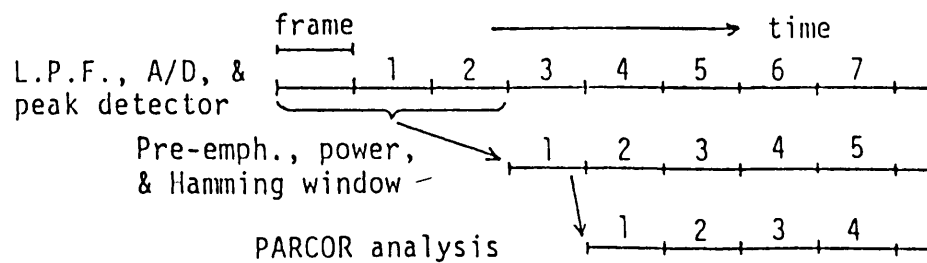


図 C.3: 前処理部と分析部のタイミング図

用いる。残る式 4.7、式 4.9、式 4.10（以下、「 $W&\epsilon$ 部」という）で最大の積和が 32,000 回となる。1 フレームを 10ms とすると 1 回あたりの積和に許される演算時間は 308ns となり、メモリへのアクセス速度を考慮すれば、この時点で入手できる LSI の高速積和器 TDC1010J が 1 個だけでは無理なので、その積和器を 2 個用いて並列処理することとした。

## $W_i&\epsilon_i$ 部

この部分は積和器を 2 個用いたので、並列処理に向く方法が必要である。以下この方法について述べる。

$\epsilon^f$ と $\epsilon^b$ のメモリは二つずつある。まず、前処理部により最大値検出、左シフトによって正規化され、さらに窓掛けされたデータが $\epsilon_{odd}^f(n)$ と $\epsilon_{odd}^b(n)$ に入る。ただし、 $\epsilon_{odd}^f(n)$ と $\epsilon_{odd}^b(n)$ のデータは定義通り 1 標本化周期分ずらせて格納する。このずれを利用してアドレスを 1 つおきに与えることにより  $W_1$  が計算できる。その概略を図 C.4 に示す。 $W_1$  の計算が終わると後述する  $U&k$  部で  $k_1$  と  $U_2$  を計算する。この  $k_1$  を用いて、 $\epsilon_1^f(n)$ 、 $\epsilon_1^b(n)$  を計算する。図 C.5 にその概略を示す。次数が奇数のとき  $\epsilon_{odd}^f$  と  $\epsilon_{odd}^b$  のデータにより  $\epsilon^f$  と  $\epsilon^b$  を計算し、 $\epsilon_{even}^f$ 、 $\epsilon_{even}^b$  のメモリに書き込むとともに、1 アドレスおくらせて同じデータを  $\epsilon_{odd}^f$ 、 $\epsilon_{odd}^b$  に書き込む。これは次の  $W$  を計算するのに用いられる。次数が偶数のときは、前文の odd と even を入れ換えた関係となる。この値により  $W_2$  を計算するのがこの概略を図 C.6 に示す。以下同様にして行うが、 $\epsilon_{12}$  のときのみ、 $\epsilon_{odd}^f$  と  $\epsilon_{odd}^b$  には前処理より次の正規化されたデータが入力される。

演算時間は  $W$  が 1 データあたり 300ns、 $\epsilon$  が 650ns である。

## $U_i&k_i$ 部

この処理にはマイクロプロセッサ Z-80 と乗除算器 CDP1855C を用いた。この部分は  $U_0$  を前処理より、 $W_i$  ( $i = 1, 2, \dots, 12$ ) を  $W&\epsilon$  部より与えられたとき、 $k_i$  と  $U_i$  ( $i = 1, 2, \dots, 12$ ) を計算する。この部分のバスの概略を図 C.7 に示す。計算のフローチャートを図 C.8 に示す。乗除算器 CDP1855C は 1 個あたり 8bit であり、 $U_0$  の値はすべての入力 が 12bit になったときでも最大 33bit なので、これを 4 つ直列につないで 32bit にして使用した。取扱いできるのは正の値のみである。乗算は  $X(32\text{bit})$ 、 $Y(32\text{bit})$ 、 $Z(32\text{bit})$  を入力すると  $XZ + Y$  の計算結果が  $Y$  (上位 32bit)、 $Z$  (下位 32bit) に出力され、除算の場合は被除数を  $Y$  (上位 32bit)、 $Z$  (下位 32bit) に、除数を  $X$  に入力すると、商は  $Z$ 、余りは  $Y$  に出力される。演算時間は 1bit あたり  $2\mu\text{s}$  であり、32bit では  $64\mu\text{s}$  かかる。このため乗算、除算ともに計算が始まると Z-80 はこの時間をカウントするループ (図 C.8 「TIME SUB」) にはいる。また、 $k_i$  の絶対値は 1 以下なので、これを満たさない場合は強制的に  $k_i > 1$  のときは 7FFFH、 $k_i < -1$  のときは 8000H にしている。

演算時間は  $W$  の符号によって異なるが、 $k$  を求めるために最大  $129.5\mu\text{s}$ 、 $U$  を求めるために  $258\mu\text{s}$  である。

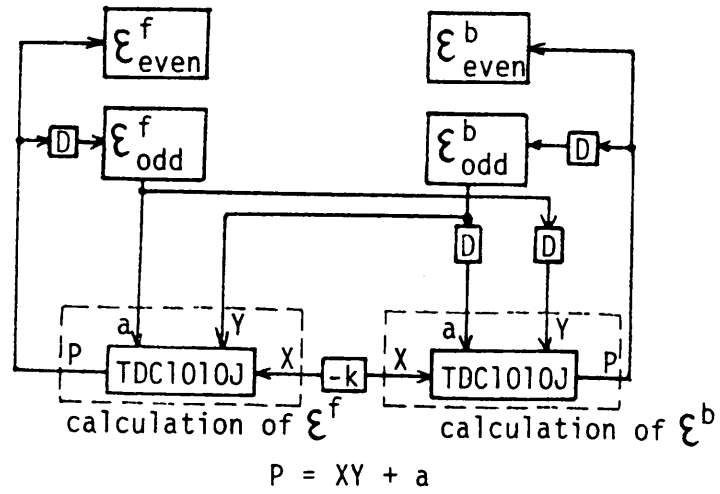


図 C.4:  $W_1$  の計算

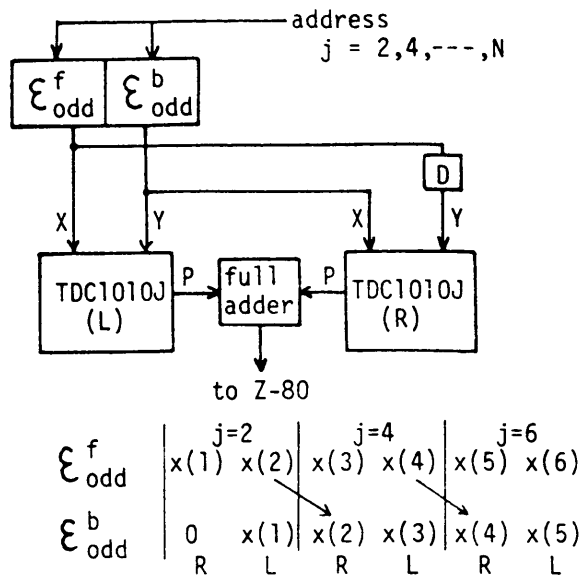


図 C.5:  $\epsilon$  の計算 (分析次数  $i$  は奇数)

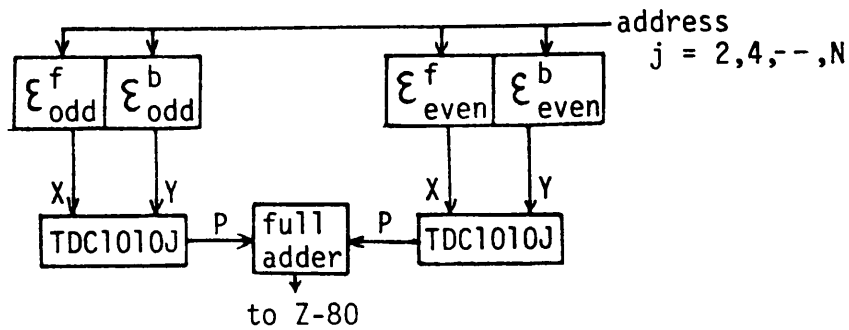


図 C.6:  $W_j$  の計算 (ただし分析次数  $i \geq 2$ )

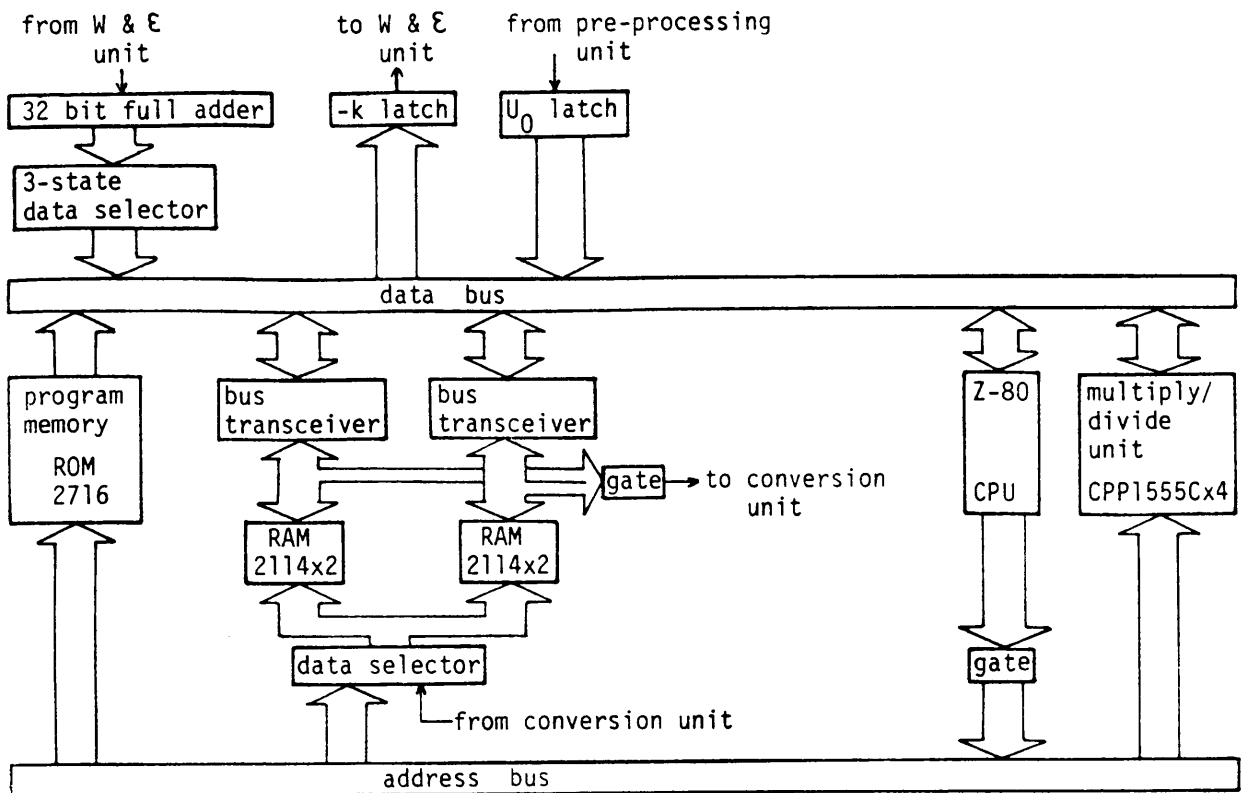


図 C.7: Z-80 周辺の概略図

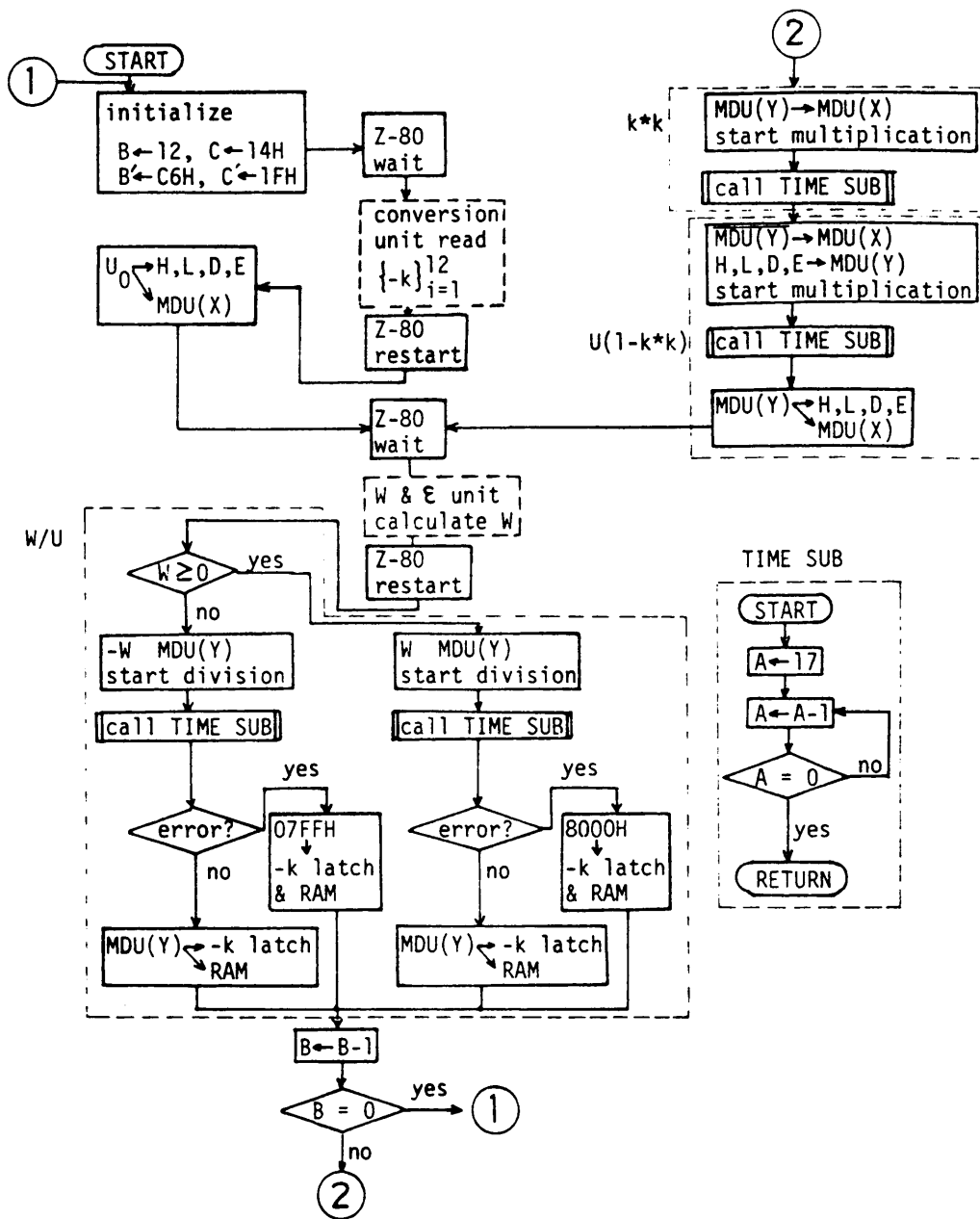


図 C.8: Z-80 プロのグラムフローチャート

## LSP 係数変換部の設計方針

設計にあたっては以下のことを考慮した。

- (a) できる限り、ハードウェアの構成を簡素化すること。
- (b) 演算回路におけるデータ語長は 24bit。
- (c) 24bit × 24bit の乗算が約 10 $\mu$ sec で演算可能であること。

(a) については、ハードウェアの構成が簡単なほど製作しやすく、信頼性の高くなるからである。(b) については上述した。(c) については以下の理由による。Newton-Raphson 法による各次数での収束回数を 5 回とすると、積和演算 350 回、除算演算 54 回となり、そのほかに PARCOR 係数から  $\alpha$  係数を求める部分での積和演算 66 回、LSP 係数の変換の表引きもある。よって、この演算を 10ms (1 フレーム) 以内で処理するために、乗算の演算は 10 $\mu$ sec を目標とした。

上記の条件を具体化するには、

- (ア) すべてを TTL-IC で構成する。
- (イ) 既存の 8bit や 16bit 系のマイクロプロセッサを使用する。
- (ウ) 既存のビットスライス・マイクロプロセッサを使用する。

の三つの方法が考えられる。(ア) については演算速度に問題はないが、構成の複雑さ、煩雑さによって信頼性に欠ける恐れがあり、(イ) についてはマイクロプロセッサの語長が 24bit より短く、見かけ上 24bit にしなければならないので演算速度に問題がある。よって、(ウ) で語長を 24bit とし、乗算はハードウェアを多少付加することで満足させることにした。

### ビットスライス・マイクロプロセッサの選択

今回は AMD 社 Am2900 ファミリを選んだ。その理由を以下に記す。

- (A) 演算部は 4bit 単位であり、ユーザはこれを  $n$  個直列接続して  $4n$  bit の語長のプロセッサを構成できる。
- (B) マイクロプログラム制御方式であり、プログラムがユーザに解放されている。
- (C) 4bit ビットスライスの標準品種で、一般的であり、マニュアルも完備している。
- (D) サイクルタイムが 100nsec と比較的高速である。
- (E) このファミリは TTL であり、周辺回路には TTL がそのまま使用できる。
- (F) 種々の機能を持つ LSI がファミリとして供給されている。

(A)、(B) の理由により、ユーザがシステムを目的に合ったように最適化を計ることができる。しかし、構成の自由度が大きいのでシステムの細部までユーザが設計できる。今回はハードウェアの比重を極力減らし、ほとんどの制御をマイクロプログラムに任せることでシステムを簡素化し、信頼性を高めることにした。以下に今回使用した Am2900 ファミリの三種類とその個数を示す。

Am2901B	4bit バイポーラマイクロプロセッサ	6 個
Am2904	ステータスとシフトコントロールユニット	1 個
Am2910	マイクロプログラムコントローラ	1 個

## LSP 変換部のハードウェアの構成

図 4.6 にこの変換部のブロック図を示す。変換部は分析部より送られてくる 10ms ごとのスタートパルスによって処理を開始する。

この部分は大きく分けて制御部と演算部に分かれている。制御部は中心にマイクロプログラムコントローラ Am2910 があり、マイクロプログラムの実行番地を制御している。マイクロプログラムの記憶にはアクセスタイム 85ns の C-MOS RAM HM6148P-6 を用いている。RAM を用いたのは、マイクロプログラムの開発が当研究室のミニコン M-70B で支援でき、またマイクロプログラムの内容をバッテリバックアップできるからである。また、このメモリの出力にはパイプラインレジスタと称するポジティブエッジトリガの D ラッチを置くことにより、フェッチサイクルと実行サイクルが 1 クロックででき、変換部全体が高速になる。

演算部は Am2901B を 6 個直列に接続して語長を 24bit としたものを中心に、演算を高速化するルックアヘッドキャリジエネレータ 74S182、およびステータスの制御とデータのシフトを行うステータスとシフトコントロールユニット Am2904 が周辺にある。

内部バスには Am2901B の Y 出力、バッファメモリ、テーブルメモリ 1 (逆余弦の二乗)、テーブルメモリ 2 (余弦の平方根)、パイプラインレジスタ定数フィールド、式 4.15 の  $C_1$ 、 $C_2$  が接続され、バッファメモリを除く出力ゲートは、マイクロ命令の 3bit で切り換えられる。また、このバスはメモリアドレスラッチの出力とともに外部に引き出され、分析部と合成部に接続される。マルチプレクサは乗算の場合に使用される。メモリアドレスラッチは構成を簡素化するために 1 組だけ設けた。このラッチは内部バスの下位 11bit を入力にしている。この 1 組のラッチで 5 つのアドレス (バッファメモリ、テーブルメモリ 1、テーブルメモリ 2、分析部の PARCOR 係数メモリ、合成部の LSP 係数メモリのアドレス) を指定しているので、必要に応じてこのラッチにセットしなければならない。

内部バスには 24bit のデータ表示用のラッチが接続してある。このラッチの出力には LED が接続してあり、常時、このラッチの内容を表示している。この部分は、実際に動作させる場合には不用であるが、動作試験の段階で、データを読みとり、デバッグを行うためのものである。



次に各部のタイミングを示す。変換部は後述するようにマイクロ命令を水平方式としているので、単相のクロックに同期してすべて処理を行い複雑なタイミングはない。1クロックサイクルでの最大時間は演算だけでなくその結果による条件分岐を伴う場合である。この場合のタイミング図を図 C.9に示す。これより、1クロックは 316.5ns 以上必要なことがわかる。そこで、余裕をみて、3MHz のクロック（1クロック約 333ns）とした。その結果、24bit × 24bit の乗算（1演算あたり 33クロック）は 11 $\mu$ sec となり、ほぼ目標通りの値である。

なお、パワオン時には自動的に初期化する回路を設け、パイプラインレジスタのネクストアドレスフィールドの内容を 0 としているので、最初のスタートパルスが入った時点で、変換部はマイクロプログラムの 0 番地からスタートする。

## マイクロ命令

図 C.10にマイクロ命令の構成を示す。一語は 54bit で水平方式を基本にして構成され、一語は 19 フィールドに分かれている。水平方式を採用したのは、高速性を重視したからである。これにより一語のビット数は多くなったが、プログラムの自由度が増し、垂直方式よりも高速処理が可能となっている。

## マイクロプログラムの概要

変換部のマイクロプログラムはサブルーチンを含めて 14 個のブロックで構成されている。全体を機能別にブロック化したのは、各ステージ、各サブルーチンごとにデバックできるようにし、マイクロプログラムの開発のしやすさを狙ったためである。表 C.1に各ステージ、各サブルーチンの主な概要を示す。Newton-Raphson 法で解を求める部分の演算が処理時間の大部分を占めていることがわかる。

このプログラムを動作させたところ約 9ms で処理が終わることを確認した。

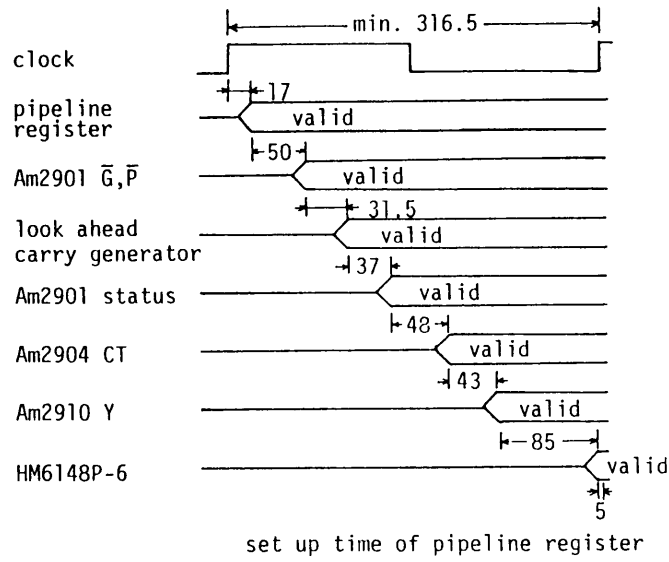


図 C.9: LSP 変換部のタイミング図

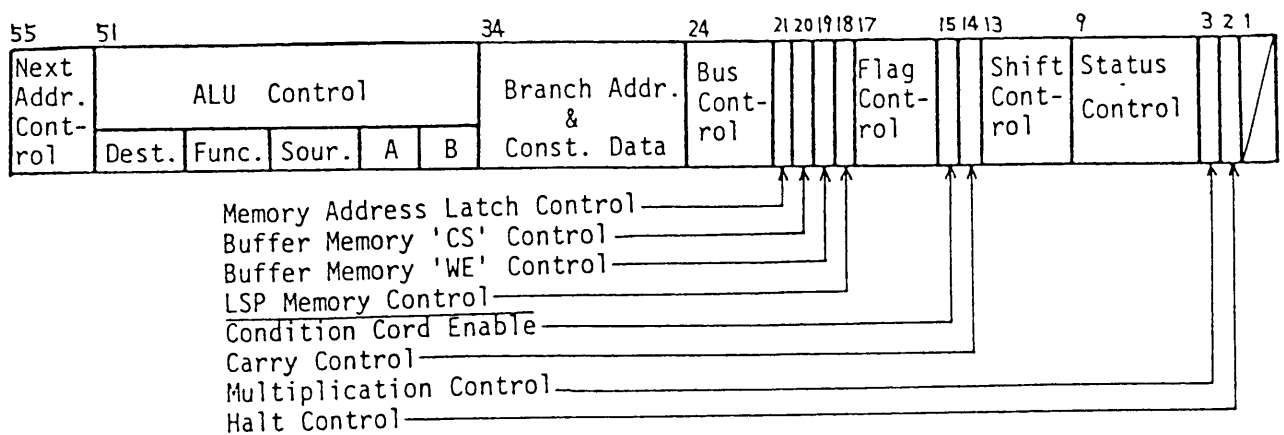


図 C.10: LSP 変換部のマイクロ命令

表 C.1: LSP 変換部のマイクロプログラムの内容

ルーチン名	番地	プログラム サイズ	プログラムステップ数	処理
<i>INIT</i>	0	16	$48 + SET$	初期化
<i>STAGE1</i>	70	9	53	PARCOR 係数を分析部から読み込む
<i>STAGE2</i>	80	33	3093	PARCOR 係数から $\alpha$ 係数を求める
<i>STAGE3</i>	B0	22	87	$a \sim f$ の係数を求める
<i>STAGE4</i>	D0	18	$48 + 2COCAL$ $+12NEWTON$	LSP 係数の計算
<i>STAGE5</i>	F0	32	776	LSP 係数の変換
<i>STAGE6</i>	110	26	1436	変換した係数を合成部へ出力する
<i>COCAL</i>	130	59	215	6 次の多項式の係数を求める
<i>ESCAPE</i>	170	14	113	エラー処理
<i>DIV2</i>	300	22	103	割算 (NEWTON で使用)
<i>NEWTON</i>	370	58	$n \neq 1$ $79n - 52 + m(76n + 82)$ $n = 1$ $21 + 158m$	6 次の多項式の根の計算 ( $n$ : 次数, $m$ : 収束回数)
<i>SET</i>	3B0	29	85	定数をセットする
<i>DIV</i>	3D0	21	103	割算 (NEWTON 以外で使用)
<i>MPY</i>	3F0	6	33	掛算

## 付録 D

# 日本語ヘリウム音声データベース

### はじめに

音声研究を進める上でデータベースが重要であることが認識され、日本でも種々のデータベースが開発されている [57]-[61]。また、パーソナルコンピュータの高性能化に伴い、データベースをパーソナルコンピュータで取り扱うことも容易になってきた。

6章までにヘリウム音声という特殊な音声の修復法を検討し、実時間で動作する修復装置の開発とその評価について述べた。この研究のため、海洋科学技術センター (JAMSTEC) で行われた有人潜水模擬実験において 31ATA (深度 300m 相当) までの日本語の音声データを 1980 年より収録してきた。これまでに集められたデータは第 2 章で述べたもので録音テープで約 60 巻になる。特性のよいマイクロホンで収録された日本語のヘリウム音声は、世界的にみてもこれらのデータが唯一である。また、有人潜水は莫大な費用がかかり、しかも 300 メートルの潜水では、300 メートルの滞在はわずか一週間でも全日程は一月もかかるので、ヘリウム音声を収録するのは簡単ではない。よって、このデータを整理して保存し、汎用性をもたせるとともに統計的な解析ができるようにすることは非常に価値のあることである。

そこで、この目的のため 1990 年よりデータベースの構築を計画し、パーソナルコンピュータで簡便に使用できるシステムを作成した。音声資料については第 2 章に述べたものである。ここでは、本データベースの設計理念とそのハードウェアとソフトウェアの概要を述べる。

### 日本語ヘリウム音声データベースの仕様

日本語の音声データベースはこれまでいくつか作成されている [57],[61] ので、これらのデータベースを参考にヘリウム音声に必要な機能等を追加した。たとえば、ヘリウム音声は了解度が低いので、ヘリウム音声をそのまま聞いて理解するのはむずかしく、発声者自身も耳からのフィードバックはヘリウム音声となっていて困難なので発声が正しく行われたかは習慣にたよるしかない。そこで、原波形、第 6 章で述べたヘリウム音声修復装置による処理波形、または原波形の低速再生によってヘリウム音声をわかりやすく聞くこと

ができるように設計し、正しく発声されているかの確認とともに、まちがって発声された場合は聞こえる音素に変更し、誰にでも利用できるように音素ラベル情報を付加することにした。典型的なデータベース [57] を参考に、話者名、発声テキスト、採録年月日等の採録条件を定めるインデックス情報を選んだ。これに、ヘリウム音声特有のインデックス情報として、潜水（模擬）深度の情報を付加した。インデックス情報は単語または文章ごとにインデックスファイルとして登録する。音素のラベリング時のディスプレイ表示については文献 [62] を参考にした。

計算機としては、一般に普及している NEC98 シリーズと互換性のあるパーソナルコンピュータ EPSON PC-386GS を使用した。その理由は、経済的であり、データベースを作り、使用するのに十分な処理速度および周辺機器があるからである。

## システムのハードウェア

このシステムは、図 D.1 に示すようにエプソン PC-386GS(数値演算コプロセッサ 80387 搭載、マウス付き)、光磁気ディスク MO-7616 (アイシーエム、600MB/両面)、DAT デッキ (光または同軸デジタル入出力付き)、DAT インターフェイスボード IS-3690 (岩通アイセル)、磁気ディスク (SCSI タイプ) からなる。

磁気ディスクに登録するには音声データ量が多いので、数枚の光磁気ディスクカートリッジに登録することにした。また、インデックス情報はいつでも使用できるように磁気ディスクにインデックスファイルとして登録することにした。

## システムのソフトウェア

このソフトウェアは以下の 8 つからなる。

- 波形切り出しプログラム HSCUT.EXE
- ソナグラムデータ生成プログラム HSFFTDMS.EXE
- 音素ラベリングプログラム HSLABELS.EXE
- 検索性データ作成プログラム MERGE.EXE
- 音声データ検索プログラム HSDB.EXE
- 音声波形観測、分析プログラム PLOTM98.EXE
- バックアップ作業プログラム HSWORKS.EXE
- 音素継続時間分布作成プログラム PHONEME.EXE

最初の 4 つはデータベース構築のためのプログラムであり、5 つ目は必要なデータを検索するプログラムである。6 つ目は波形の観測、FFT や LPC 分析法による処理とディスプレイ画面表示および音声の試聴ができる。7 つ目はデータを二重化し保存するのを効率よく行うプログラムである。

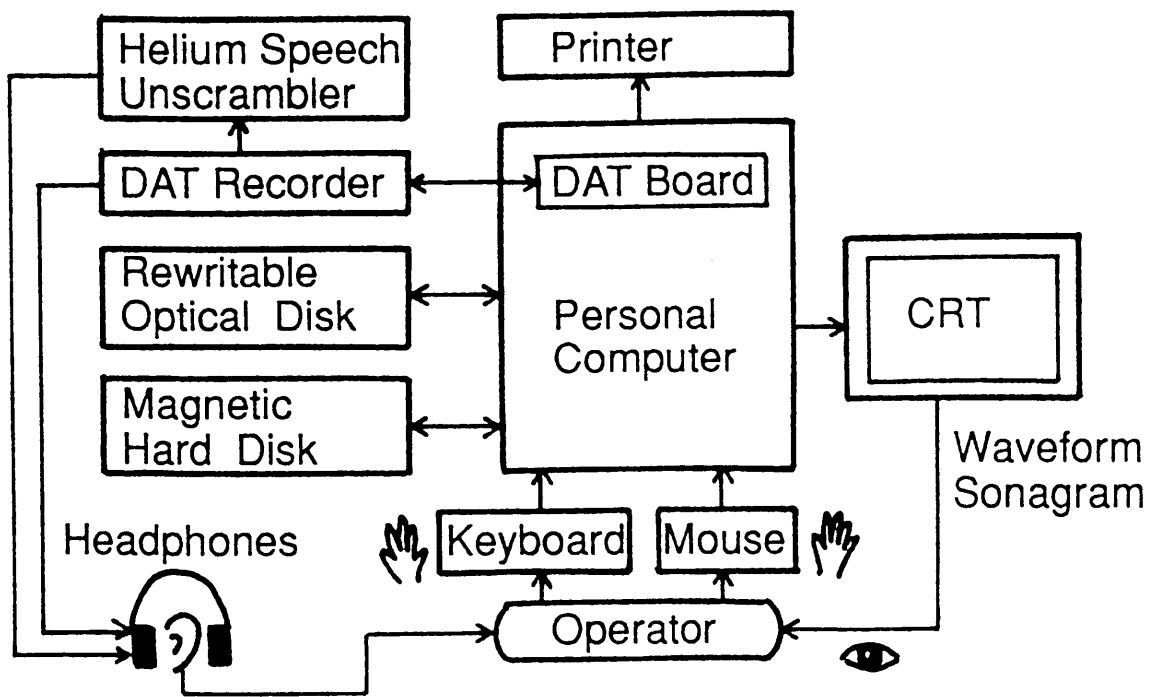


図 D.1: ヘリウム音声データベース構築システム

データベースへの登録作業はできるだけ効率的に行われなければならないので、キーボードによるデータ入力が必要最低限とし、マウスによるメニュー選択でほとんどの操作を行えるようにした。

## 音声データ、インデックスファイルと検索時に使用するファイルの構造

### 音声データファイル

音声データは、おのおの単語または文章ごとに一つのファイルとして格納される。そのファイルは、1サンプルあたり16bit、インテル8086形式で登録されている。標本化周波数は48kHzである。標本化周波数48kHzは通常の空気中の音声には高過ぎるが、ヘリウム音声では、ヘリウムの音速が速いために周波数帯域が広がるので必要な周波数である。したがって、1秒あたりの音声に必要な記憶領域は96kBである。

音声データファイル名は各々排他的になるように、採録年月日とマイクロホン番号とでできるディレクトリの下に、採録年月日、潜水深度、テキストの番号、単語の番号、マイクの番号、発声者の番号によりできる名前に拡張子の最後に“d”がついたものが登録される。

### インデックスファイルの内容

インデックスファイルはどんな条件でどの話者がどの音声データを発声したかを特定する全ての情報を与えるファイルである。表D.1にインデックスファイルの内容を示す。各々のインデックスファイルの内容は話者名、採録年月日、深度、マイクロホン番号、単語または文章の内容、各音素の音声ファイル中の開始終了位置等である。インデックスファイルは各々の音声データごとにある。

このインデックスファイル名も音声データファイル名と同様に、各音声データファイルに固有の名前でなければならない。そこでこのファイル名は、採録年月日、潜水深度、テキスト番号、単語番号、マイク番号、発声者番号からファイル名を作り、拡張子の最後は“Z”とした。

### データ検索ファイル

インデックスファイルはデータごとに作られるので、多数のデータが格納されているデータベースから特定のデータを検索するには時間がかかる。そこで、インデックスファイルから検索に必要な情報のみを抜きだしたものを「検索.dat」という名前の1つのファイルに作る。この「検索.dat」はMERGE.EXEにより作られる。

表D.2はおのおの単語または文章ごとの「検索.dat」の内容を示す。この内容は、採録年月日、深度、マイクロホン番号、話者名、テキスト内容等で1単語または文章あたり500バイトである。

表 D.1: インデックスラベルの構造 (番地はバイト表示)

番地	データの型	データの内容
0	unsigned short	発声者の番号
2	unsigned short	テキスト中の単語の番号
4	unsigned short	テキストの番号
6	unsigned short	実験年
8	unsigned short	実験月
10	unsigned short	実験日
12	unsigned short	潜水深度 [m]
14	unsigned short	使用したマイクの番号
16	unsigned short[5]	予約
26	unsigned short	コメント無 = 0, 有 = 1
28	char[124]	コメントの内容
152	unsigned long	データの長さ
156	unsigned short	男 = 0
158	unsigned short	標本化周波数 (48000Hz)
160	char[22]	テキスト
182	char[22]	発声者の名前
204	char[22]	発声者の名前の読み
226	char[100]	単語や文の内容
326	char[100]	単語や文の読み
426	unsigned short[3]	予約
432	char[80]	音声ファイル名
512	char[256]	ローマ字による音素表現
768	long[512]	音素の先頭からの位置



表 D.2: 検索用データ “検索.dat” の 1 音声データ当たりの情報 (番地はバイト表示)

番地	データの型	データの内容
0	unsigned short	年度 (西暦)
2	unsigned short	月
4	unsigned short	日
6	unsigned short	潜水深度 (m)
8	unsigned short	マイク番号
10	char[10]	発声者の名前の読み (カナ)
20	char[10]	テキスト区分
30	char[50]	単語 (文) の読み
80	char[256]	単語 (文) のローマ字表現
336	char[80]	音声ファイル名
416	char[80]	インデックスファイル名
496	long	コメントの有無

## データの収集と蓄積の手順

### 録音テープからのデータの格納 (HSCUT.EXE)

データの収集と蓄積の手順を図 D.1のブロック図を利用して説明する。

最初に、HSCUT.EXEにより、DATデッキのテープから21秒の波形データをパソコンの主記憶に読み込む。このデータに対してCRT上で波形を見ながらマウスのクリックにより単語または文章の最初と最後にマークを付ける。また、その位置が正しいかどうかを確かめるために、録音された速さでの再生や1/2の低速再生やヘリウム音声修復装置により、マークした位置間の音を聞くことができる。こうした確認の後、各音声ファイルは光磁気ディスクに登録される。これらのファイルの音素ラベル以外のインデックス情報はおのおの「一時インデックスファイル」として磁気ディスクに登録される。

### 音素ラベリング (HSLABELS.EXE)

データに音素記号をラベル付けするため、作業者は図 D.2に示すような波形、パワ、隣接するフレームのスペクトル変化率、およびサウンド・スペクトログラムを視察し音声を試聴しながら操作を行う。このプログラムがHSLABELS.EXEである。

HSLABELS.EXEには、さらに音素ラベルを正確に付けるための波形の時間に対する拡大、この拡大時における再生（録音された速さ、1/2および1/3の低速再生）がある。また、サウンド・スペクトログラム（ソナグラム）や表示された図のコピーなどいくつかの機能がある。

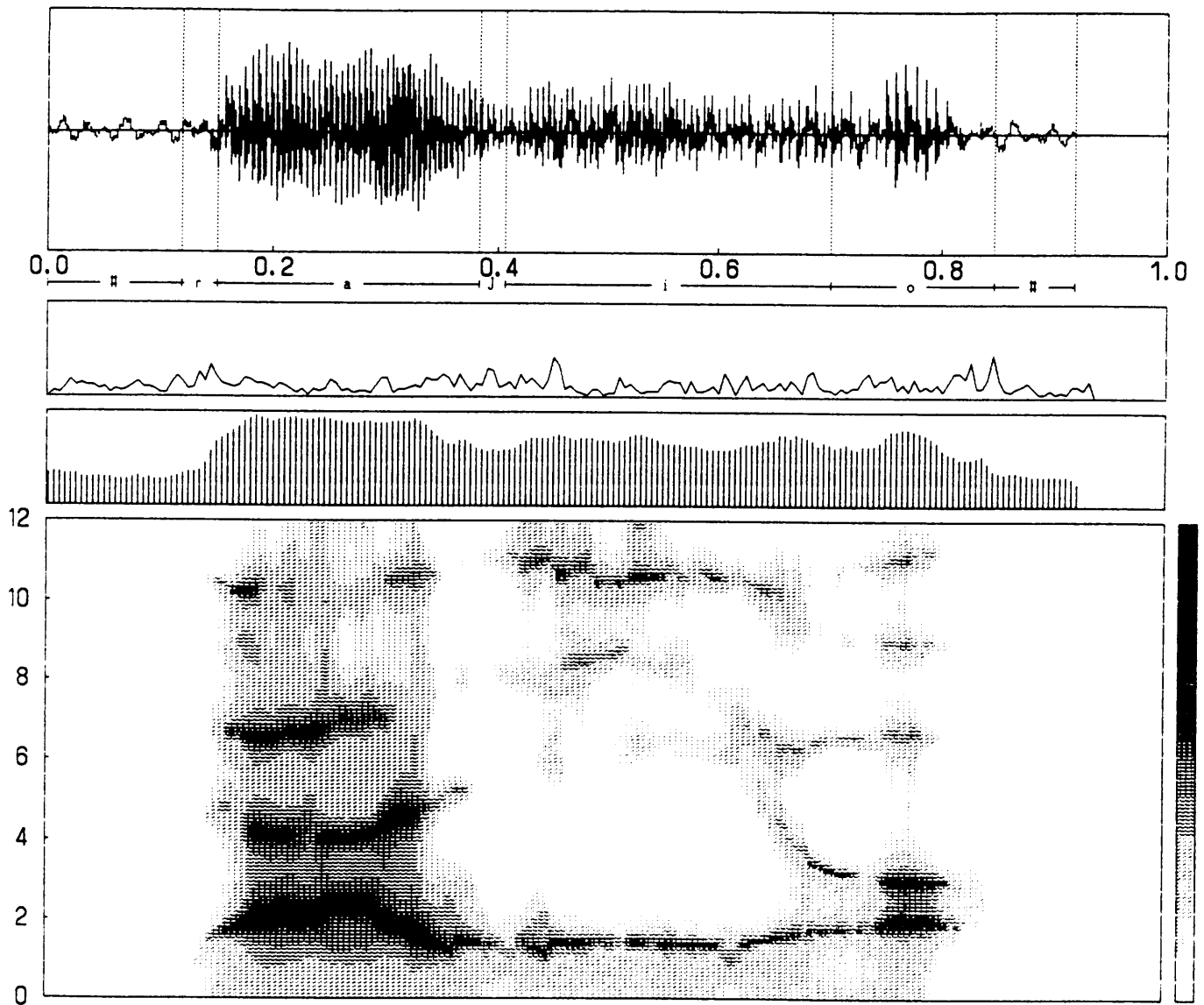
インデックス情報としてはHSCUT.EXE実行時に書き込んだ一時インデックスファイルを参考にさらに「シ無声化」、「バズバーなし」等のコメントの追加、読み通りには聞こえないときの音素名の変更を行う。

また、HSLABELS.EXEを実行する時は音声波形データのスペクトログラムを使用するので、それをあらかじめ求めておくと都合がよい。このスペクトルはHSFFTDMS.EXEにより線形予測（LPC）スペクトルとして求めることができる。しかし本データベース用のパソコンでは1秒間のデータのLPCスペクトルを計算するのに2分かかるので、ネットワーク上の別のワークステーション（三菱ME-R7100）にデータを転送して計算できるようにしてある。これにより1秒のLPCスペクトルの計算時間は約10秒となる。

これらの機能を使用して正しい音素名を付けた後、インデックスファイルに書き込む。

## データベースの使用

データベースから必要な音声データを検索するときの検索条件には、年度、月、日、潜水深度、マイク番号、発声者名、テキストの種類、音素または音素列がある。これは全部指定してもよいが全部は指定しなくてもよい。また音素の指定には論理和がとれる。たとえば、音素列/aiu/とも指定できるし、/a/または/i/または/u/のいずれかの音素を含むものとも指定できる。日本語の仮名表現での検索も可能である。



1982/11/28 他谷 300m Mic-3 3音節 - ラジオ

図 D.2: ヘリウム音声データベースの音素ラベル付け用の表示例

与えられた条件により、データ検索プログラム HSDB.EXE は「検索.dat」を検索する。検索し該当するデータ数を CRT 上に表示する。このデータを観測したり分析したりするには、システムの HSLABELS.EXE か PLOTM98.EXE を起動する。ただし、この段階で使用される際は HSLABELS.EXE の音素ラベルをつける機能は外されている。PLOTM98.EXE では FFT や LPC によるスペクトル分析や音声の再生ができる。また、必要ならば該当する音声データを任意のディスクにコピーすること、それらの音声データに関するインデックス情報を、テキスト形式で出力させることができる。

データベースから任意のデータの検索にかかる時間はおよそ 10 秒である。

## まとめ

ここでは日本語ヘリウム音声データベースについて述べた。現在、登録され、音素ラベル付けが終了しているデータは 1982 年の各 3 名 6 深度、1991 年の各 3 名 7 深度の明瞭度試験用音節表、ワードリスト（1982 年 3 音節、1991 年 2 音節）を含む約 4,300 個である。今後さらに多くのデータを登録後、CD-ROM 化し配布できるようにする計画である。

また、このデータベースにより 6 名の話者について調べたところヘリウム空気では有声子音のバズバーがなくなり、破裂のみになりやすいことがわかった。このように、従来よく知られていなかったヘリウム音声の音響的性質が定量的に示された。今後、音圧レベル、ピッチ、スペクトル等の分析を行う予定である。

## 参考文献

- [1] Y. Berry, P. Gavarry, J. P. Hubert, J. Le Chuiton, and J. Parc 著, 関 邦博 訳, 「潜水学」 マリン企画 (昭和 57 年) .
- [2] 「潜水作業技術の研究開発 昭和 56 年度 成果報告書」海洋科学技術センター (昭和 58 年) .
- [3] たとえば、川 島整形外科病院五周年記念論文集「高圧医学・潜水医学論文集」 川 島整形外科病院 (昭和 61 年) .
- [4] J.D.Markel, A.H.Gray, Jr. 著, 鈴木久喜 訳 「音声の線形予測」 コロナ社 (昭和 55 年).
- [5] G. Fant and J. Lindqvist, "Pressure and Gas Mixture Effects on Diver's Speech", STL-QPSR (1968).
- [6] K. Holywell and G. Harvey, "Helium Speech", J. Acoust. Soc. Amer., Vol.36, pp.210–211 (1964).
- [7] 鈴木誠史, 中津井護, 高杉敏男, 田中良二, "ヘリウム音声の研究 —高圧環境で発声された音声の性質とその了解性の改善—", 電波研究所季報 vol.24, No.129, (1978).
- [8] W. R. Stover, "Technique for Correcting Helium Speech Distortion", J. Acoust. Soc. Amer. Vol.41, pp.70–74 (1967).
- [9] 鈴木久喜, 大山玄, "線形予測とインパルス応答変換によるデジタルフィルタを用いたヘリウム音声修復系", 信学論 (A), 58–A, 6, pp.377–384 (1975).
- [10] R. M. Golden, "Improving Naturalness and Intelligibility of Helium-Oxygen Speech, Using Vocoder Techniques", J. Acoust. Soc. Amer, Vol.40, pp.621–624 (1966).
- [11] D. A. A. Roworth, "Helium-Underwater-Speech Translating Equipment", Electrical Communication, Vol.45, No.1, pp.18–23 (1970).
- [12] M. A. Richards, "Helium Speech Enhancement Using the Short-Time Fourier Transform", IEEE Trans. Acoust., Speech & Signal Process., ASSP–30, No.6, pp.841–853 (1982).

- [13] H. Suzuki, G.Ooyama and S. Wakayama, "Conversion of Vocal Tract Shape for Spectral Warping by PARCOR analysis-synthesis system", Acoust. Soc. Am. Acoust. Soc. Japan Joint Meeting (1978).
- [14] M. Copel, "Helium Voice Unscrambling", IEEE Trans., AU-14, pp.122-126 (1966).
- [15] R. A. Flower, "Final Technical Report on Helium Speech Investigations", Singer-General Precision, Inc. (Dec. 1969).
- [16] J. L. Stewart, "Helium Speech Processing", Proc. of 21st Aerospace and Electronics Conference, pp.271-274 (1969).
- [17] R. F. Quick, Jr., "Helium Speech Translation Using Homomorphic Techniques", Physical Scis. Res. Papers, No. 425, United States Air Force (1970).
- [18] S. Helmut, "A Transceiver for Helium Atmosphere", Technische Mitteilungen AEG-Telefunken, 61, pp.378-381 (1971).
- [19] M.F.Gillis, "In vivo detection of circulating gas emboli associated with decompression sickness using the Doppler flowmeter", Nature, 217, pp.965-967 (1968).
- [20] M.P.Spencer, S.D.Campbell, and J.L.Lindbergh, "Experiments on decompression bubbles in the circulation using ultrasonic and electromagnetic flowmeters", J.Occupational Med., 11, pp.238-244 (1969).
- [21] M.P.Spencer, "Decompression limits for compressed air determined by ultrasonically detected blood bubbles", J.Appl.Phys., 40, pp.229-235 (1976).
- [22] E.O.Belcher, "Quantification of bubbles formed in animals and man during decompression", IEEE Trans., Biomed.Eng., BME-27, 6, pp.330-338 (1980).
- [23] 電子通信学会編, 「聴覚と音声」 p.342, pp.344-346 電子通信学会 (昭和 48 年).
- [24] 三浦甫, 松井英一, "標準マイクロホンの音圧校正におけるカプラ内波動の解析", 日本音響学会誌, 30, pp.639-646 (1974).
- [25] P.V.Brüel, "The accuracy of condenser microphone calibration methods, Part 2", Brüel & Kjaer Tech. Rev. No. 1, pp.11-17 (1965).
- [26] 鈴木久喜, 中井孝芳, 吉田慎介, 藤本敦, 高橋多助, "高圧気体環境におけるコンデンサマイクロホンの特性", 電子通信学会電気音響研究会, EA83-19, pp.49-56(1983).
- [27] 鈴木久喜, 藤本敦, 高橋多助, "高圧環境における音響機器の動作の研究 —コンデンサマイクロホン, スピーカ, 騒音計について—", 電子通信学会電気音響研究会, EA84-8, pp.7-14(1984).

- [28] 鈴木久喜, 中井孝芳, 吉田慎介, 藤本敦, 高橋多助, "潜水時の高圧気体環境におけるコンデンサマイクロホンの特性", 日本音響学会誌, Vol.44, 3, pp.201-209 (1988).
- [29] H. Suzuki, M. Ohashi, T. Nakai, and M. Kobayashi, "Measurement of Diaphragm Movement of Condenser Microphone in High Pressure Ambient Gases by Means of Laser Heterodyne Interferometer", 13th International Congress on Acoustics, Vol.4, pp.11-14, (1989).
- [30] 鈴木久喜, 内田義幸, 大橋正尚, 中井孝芳, "高圧気体環境におけるコンデンサマイクロホンの特性の数値シミュレーション", 信学論 A, J72-A, 5, pp.743-751 (1989).
- [31] 鈴木久喜, 中井孝芳, 清野堅司, "潜水における高圧気体中のコンデンサマイクロホンの膜振動の数値シミュレーション", 電子情報通信学会応用音響研究会, EA89-31, pp.95-102 (1989).
- [32] 河西康仁, 中井孝芳, 鈴木久喜, "有人潜水実験用潜水シミュレータ内の騒音の測定", 電子情報通信学会応用音響研究会, EA91-12, pp.9-15 (1991).
- [33] J.D.Markel and A.H.Gray 著, 鈴木久喜訳, 「音声の線形予測」, p.247 (1980).
- [34] H. Suzuku, T. Nakai, and K. Shimizu, "Measurement and Analysis of Speech Sound Radiated from Vocal tract Wall", Proc. of ICASSP 86, pp.1625-1628 (1986).
- [35] 党建武, 中井孝芳, 鈴木久喜, "声道内音圧及び頬の振動加速度による頬インピーダンスの測定", 日本音響学会誌, 48, 9, pp.621-628 (1992).
- [36] G. Fant, "Vocal Tract Wall Effects, Losses, and Resonance Bandwidths", STL-QPSR, 2-3, pp.28-52 (1972).
- [37] 中井孝芳, 鈴木久喜, "最大可能標本化周波数 30kHz の実時間 PARCOR 音声分析装置", 静岡大学工学部研究報告, pp.17-30 (1983).
- [38] 管村, 板倉, "線形予測係数の線スペクトル表現とその統計的性質", 信学論 (A), J64-A, 8, pp.323-330 (1981).
- [39] 斉藤修三, 中田和男著, 「音声情報処理の基礎」, pp. 109-110, オーム社,(1981).
- [40] J. Le Roux and C. Gueguen, "A Fixed Computation of Partial Correlation Coefficients", IEEE Trans., ASSP-25, pp.257-258 (1977).
- [41] 小林, 板倉, "変形ラティス形 PARCOR 分析法の簡単化", 音響学会講論集, pp.347-348 (1978-10).
- [42] 北脇, 板倉, 斉藤, "PARCOR 形音声分析合成系における最適符号構成", 信学誌 (A), J61-A, pp.119-126 (1978).

- [43] 東倉, 板倉, "PARCOR 帯域圧縮方式における音声品質向上", 信学誌 (A), J61-A, pp.254-261 (1978).
- [44] J. H. McClellan, T. W. Parks, and L. R. Rabiner, "A Computer Program for Designing Optimum FIR Linear Phase Digital Filters", IEEE Trans., AU-21, pp.506-526 (1973).
- [45] 鈴木久喜, 党建武, 中井孝芳, "音声発声時の口唇, 鼻孔, 咽頭部の音と振動の測定および非鼻音化音声における鼻腔口腔間の音響漏えいのシミュレーション", 信学論 (A), Vol.J74-A, 12, pp.1705-1714 (1991).
- [46] Chengxiang Lu, Takayoshi Nakai, Hisayoshi Suzuki, "Finite Element Simulation of Sound Transmission in Vocal Tract", J. Acoust. Soc. Jpn., (E)14, 2, pp.63-72 (1993).
- [47] 高田明和, 本田西男, 森田之大編, 「臨床生理学」, p.395, 医歯薬出版 (1983).
- [48] 千葉哲央, 城内良直, 中井孝芳, 鈴木久喜, "ヘリウム音声データベースの構築", 電子情報通信学会応用音響研究会, EA92-22, pp.39-46 (1992).
- [49] 城内良直, 中井孝芳, 鈴木久喜, "ヘリウム音声データベースの応用 —音素の継続時間とバズバーについて—", 電子情報通信学会応用音響研究会, EA93-6, pp.1-8 (1993).
- [50] 鈴木久喜, 中井孝芳, 薩川芳之, "高圧環境におけるヘッドホンの特性", 電子情報通信学会応用音響研究会, EA91-13, pp.17-24 (1991).
- [51] 易傑, 鈴木久喜, "中国語の無声無気破裂音と零声母の認識", 音響学会論文誌, 44, 5, pp.361-368 (1988).
- [52] 河西康仁, 中井孝芳, 鈴木久喜, "高圧ヘリウム空気中における最小可聴値測定", 音響学会講論集, pp.445-446 (1992.10).
- [53] 倉本克治, 河西康仁, 中井孝芳, 鈴木久喜, "高圧気体環境における吸音材の吸音率について", 電子情報通信学会応用音響研究会, EA93-7 (1993).
- [54] S. Daniels, J.M. Davies, W.D.M. Paton, and E.B. Smith, "The detection of gas bubbles in guinea-pigs after decompression from air saturation dives using ultrasonic imaging", J. Physiological Society, 308, pp/369-383 (1980).
- [55] T. Ikeda, S Suzuki, K Shimizu, Y Okamoto, and M.E. Llewellyn, "M-mode ultrasonic detection of microbubbles following saturation diving: a case report and proposal for a new grading system", Aviation, Space, and Environmental Medicine, pp.166-169 (1989).
- [56] R.S. Mackay and G.Rubissow, Jr., "Decompression studies using ultrasonic imaging of bubbles", IEEE trans. Biomed. Eng., BME-25, 6, pp.537-544 (1978).



- [57] 匂坂芳典, 浦谷則好, "A T R 音声・言語データベース", 日本音響学会誌, Vol48, 12, pp.878-882 (1992).
- [58] 田中和世, 速水悟, "電総研の研究用音声データベース", 日本音響学会誌, Vol48, 12, pp.883-887 (1992).
- [59] 小林哲則, 板橋秀一, 速水悟, 竹沢寿幸, "日本音響学会研究用連続音声データベース", 日本音響学会誌, Vol48, 12, pp.888-893 (1992).
- [60] 板橋秀一, "文部省「重点領域研究」による音声データベース", 日本音響学会誌, Vol48, 12, pp.894-898(1992).
- [61] 牧野正三, 二矢田勝行, 真船裕雄, 城戸健一, "東北大-松下单語音声データベース", 日本音響学会誌, Vol48, 12, pp.899-905 (1992).
- [62] 武田一哉, 匂坂芳典, 片桐滋, 桑原尚夫, "音韻ラベルを持つ日本語音声データベースの構築", 電子情報通信学会音声研究会, SP87-19 (1987).

# 発表論文

## 論文

鈴木久喜, 中井孝芳, 吉田慎介, 藤本敦, 高橋多助, ”潜水時の高圧気体環境におけるコンデンサマイクロホンの特性” 日本音響学会誌, Vol.44, 3, pp.201-209 (1988) 2章

中井孝芳, 鈴木久喜, ”LSP パラメータに基づく実時間ヘリウム音声修復装置の設計・製作”, 信学論 D, J71-D, 9, pp.1660-1668 (1988) 4章

鈴木久喜, 中井孝芳, ”LSP パラメータに基づく実時間ヘリウム音声修復装置の評価”, 信学論 D, J72-D-II, 4, pp.467-475 (1989) 5章

鈴木久喜, 内田義幸, 大橋正尚, 中井孝芳, ”高圧気体環境におけるコンデンサマイクロホンの特性の数値シミュレーション”, 信学論 A, J72-A, 5, pp.743-751 (1989) 2章

鈴木久喜, 党建武, 中井孝芳, ”音声発声時の口唇, 鼻孔, 咽頭部の音と振動の測定および非鼻音化音声における鼻腔口腔間の音響漏えいのシミュレーション”, 信学論 (A), Vol.J74-A, 12, pp.1705-1714 (1991) 6章

中井孝芳, 木村慰作, 鈴木久喜, ”DSP を用いた線形予測分析法による血液中の自動気泡検知 (ABD) システム”, 信学論 D-II, J75-D-II, 2, pp.414-416 (1992) 8章

中井孝芳, 渡辺政博, 鈴木久喜, ”線形予測分析法を用いた潜水浮上時における血液中の気泡検出”, 信学論 D-II, J75-D-II, 4, pp.816-824 (1992) 7章

党建武, 中井孝芳, 鈴木久喜, ”声道内音圧及び頬の振動加速度による頬インピーダンスの測定”, 日本音響学会誌, Vol.48, 9, pp.621-628 (1992) 3章

## 国際学会

H. Suzuki and T. Nakai, "Analysis–Conversion–Synthesis System Based on LSP Parameter for Helium Speech Unscrambling", Proc. of 11th International Congress on Acoustics, Vol.4, pp.133–136 (1983) 4章

H.Suzuki, T.Nakai, M.Watanabe and F.Sidara, "Digital Signal Processing for Detecting the Gas Bubbles Generated in Blood Streams during a Decompression Procedure from Submergence", Proc. of China-Japan Joint Symposium on Acoustics, pp.118–121 (1985). 7章

H. Suzuki, T. Nakai, and K. Shimizu, "Measurement and Analysis of Speech Sound Radiated from Vocal Tract Wall", Proc. of ICASSP 86, pp.1625–1628 (1986). 3章

H. Suzuki and T. Nakai, "A New Helium Speech Unscrambler Using DSP", Proceedings of the 10th Meeting of the United States-Japan Cooperative Program in Natural Resources(UJNR) Panel on Diving Physiology and Technology, pp.261-275, (1989). 6章

H. Suzuki, M. Ohashi, T. Nakai, and M. Kobayashi, "Measurement of Diaphragm Movement of Condenser Microphone in High Pressure Ambient Gases by Means of Laser Heterodyne Interferometer", 13th International Congress on Acoustics, Vol.4, pp.11-14, (1989). 2章

H. Suzuki and T. Nakai, "Automatic Bubble Detection System of Gas Bubbles Generated in Blood Streams by Linear Predictive Analysis Method", Proceedings of the 11th Meeting of the United States-Japan Cooperative Program in Natural Resources(UJNR) Panel on Diving Physiology and Technology, pp.19-35, (1991). 8章

T. Nakai and H. Suzuki, "Detection of Gas Bubbles Generated in Blood Streams in Decompression Process", Proc. of China–Japan Joint Meeting of Underwater Medicine and Technology, pp.28–39 (1992). 8章

H. Suzuki and T. Nakai, "A Database of Japanese Helium Speech", Proc. of China–Japan Joint Meeting of Underwater Medicine and Technology, pp.115–122 (1992). 付録 D

## 口頭発表

加藤真一, 佐藤雅裕, 中井孝芳, 鈴木久喜, "300m 潜水模擬実験における音声の音響分析", 音響学会講論集, pp.481–482 (1981) 3章

中井孝芳, 鈴木久喜, ”深々度潜水時の音声のデジタル処理による修復”, 電気関係学会東海支部連合大会論文集, S74-S77 (1985) 4章, 6章

中井孝芳, 鈴木久喜, ”DSP を用いた実時間ヘリウム音声修復装置”, 電子情報通信学会応用音響研究会, EA90-63, pp.9-16 (1990) 6章

千葉哲央, 城内良直, 中井孝芳, 鈴木久喜, ”ヘリウム音声データベースの構築”, 電子情報通信学会応用音響研究会, EA92-22, pp.39-46 (1992) 付録 D

城内良直, 中井孝芳, 鈴木久喜, ”ヘリウム音声データベースの応用 —音素の継続時間とバズバーについて—”, 電子情報通信学会応用音響研究会, EA93-6, pp.1-8 (1993) 6章

## その他

中井孝芳, 鈴木久喜, ”最大可能標本化周波数 30kHz の実時間 PARCOR 音声分析装置”, 静岡大学工学部研究報告, pp.17-30 (1983) 4章