

修士論文

音響的臨場感の  
記録・再生に関する研究

平成 6 年 2 月 10 日 提出

指導教官 廣瀬 通孝 助教授

東京大学大学院工学系研究科  
情報工学専攻修士課程

学生証番号 26566 里見 充則

# 目 次

1 序論	1
1.1 本研究の目的	2
1.2 内容概説	3
2 臨場感に関する考察	5
2.1 はじめに	6
2.2 視覚的臨場感	6
2.2.1 立体感	6
2.2.2 広視野効果	8
2.2.3 広視野効果の定量化	9
2.3 聴覚的臨場感	9
2.3.1 音像定位の仕組み	11
2.3.2 脳内機構からの見た音像定位	15
2.3.3 音像の記録・再生方法	17
2.4 臨場感の伝送と記録・再生	19
2.4.1 従来のメディアとの関係	19
2.4.2 臨場感の伝送	21
2.4.3 臨場感の記録・再生	22
3 臨場感と人工現実感	26
3.1 はじめに	27
3.2 人工現実感におけるインターフェース	29
3.3 立体音響提示システム	30
3.3.1 合成音場方式	31

3.3.2 頭部音響伝達関数方式 . . . . .	32
3.4 立体音響ディスプレイ . . . . .	33
3.4.1 RSS システム . . . . .	34
3.4.2 Convoltotron . . . . .	36
3.5 当研究室の人工現実感システム . . . . .	41
3.5.1 人工現実感システムのハードウェア . . . . .	41
3.5.2 人工現実感システムのソフトウェア . . . . .	49
<b>4 音響的臨場感の記録・再生</b>	<b>56</b>
4.1 はじめに . . . . .	57
4.2 Convoltotron . . . . .	57
4.2.1 頭部音響伝達関数 (HRTF) . . . . .	57
4.2.2 Convoltotron の定位精度 . . . . .	58
4.3 音響的臨場感の記録 . . . . .	66
4.3.1 2点法による音源位置推定 . . . . .	68
4.3.2 3点法による音源位置推定 . . . . .	73
4.4 Virtual Dome . . . . .	80
4.4.1 Virtual Dome の構成 . . . . .	80
4.4.2 カメラヘッドシステムおよび通信回線 . . . . .	81
4.4.3 画像提示システム . . . . .	82
4.5 視覚的臨場感と音響的臨場感の融合実験 . . . . .	83
4.6 本章のまとめ . . . . .	86
<b>5 VSS による音響的臨場感の記録・再生</b>	<b>89</b>
5.1 はじめに . . . . .	90
5.2 原理および構成方法 . . . . .	90
5.2.1 2つのマイクロフォンによる方法 . . . . .	90
5.2.2 4つのマイクロフォンによる方法 . . . . .	91
5.2.3 VSS のシミュレーションによる性能 . . . . .	95
5.3 点音源の音像定位 . . . . .	98
5.4 VSS による状況認識実験 . . . . .	103

5.5 本章のまとめ . . . . .	103
<b>6 今後の展望</b>	<b>106</b>
6.1 はじめに . . . . .	107
6.2 空間情報を用いた音響空間の再構成 . . . . .	107
6.3 CAVE System . . . . .	109
6.4 KK System . . . . .	110
<b>7 結論</b>	<b>113</b>
<b>謝辞</b>	<b>116</b>
<b>参考文献</b>	<b>117</b>

# 第 1 章

## 序論

## 1.1 本研究の目的

人類は文字というものを発明し、それを用いて記録することによってはじめて、人間の記憶を介在することなしに、知識を保存することが可能になった。アッシリアの首都ニネヴアの紀元前8世紀頃の図書館の跡から発見された、アッシリア語を楔型文字で記録した「ギルガメッシュ叙事詩」は、はるか古代の文化的知識であるこの物語を現代の我々に伝え、この事実の重要性を如実に物語っている[22]。

今のあるままを記録したいという人類の欲求はこの文字の発明だけに留まらず、蓄音器、写真や映画のようなさまざまなメディアを生み続けてきた。今世紀に入ってようやく映像と音響信号の双方を含み、個人ベースで手軽に利用可能なメディアとしてVTR(Video Tape Recorder)の技術が開発され普及してきたが、情報を再生した時に、第一人称的な立場から情報を体験できているとは言い難い。

この次のステップとしては、第一人称的つまり、より高い臨場感<sup>1</sup>のある記録メディアが望まれている。このようなメディアができれば、記録された臨場感を他人が体験することによって複数の人間での体験の共有ができるであろう。

このような高い臨場感を生成する技術として、人工現実感(VR: Virtual Reality)の技術をあげることができる。この人工現実感の技術を用いると、複数の感覚情報を仮想空間を通して知覚することによって感覚の整合性を容易に保つ事ができるようになる。

臨場感にとって最も重要な感覚情報が視覚情報であることに疑問の余地はないが、音響情報を欠いた映画がどれほど迫力のないものになるかを見ればわかるように、聴覚情報も臨場感には重要である。

視覚情報と聴覚情報に関してその指向性の違いを比較すると、次のようになる。

- 視覚情報は、頭部の前方に強い指向性を持っている。つまり前の情報しか見えない。

---

<sup>1</sup>ここで述べている臨場感とは、自分がその場に存在するような知覚を生じさせるという意味で用いている。

- 聴覚情報は、基本的に全方向性の情報である。

また、時間軸方向に関しての差分的な情報量に関して比較すると、次のようなになる。

- 視覚情報は、時間軸についての差分情報量が比較的少ないので、情報の更新速度が遅くても臨場感は損なわれにくい。
- 聴覚情報は、時間軸についての差分情報量が大きく、臨場感を保つためには常に情報を更新していかなくてはならない。

つまり、画像的には静止画像(時間的な差分情報は零)でもある程度の臨場感を持ち得るのに対して、時間的に変化のない音響信号では、あまり臨場感を望む事はできない。

上述のような特徴から、聴覚的な臨場感を記録・再生するには、視覚的な臨場感を記録・再生するのとは違った面での問題が存在することが分かる。

本研究は、この問題点を克服し音響的な臨場感を記録・再生することに関する研究である。

## 1.2 内容概説

本論文は、「音響的臨場感の記録・再生に関する研究」と題し、次の7章から構成される。

第1章は、序章であり、この研究の目的をまとめてある。

第2章では、視覚的 / 聴覚的臨場感についてまとめ、臨場感の伝送と記録・再生に関する問題点に関して考察する。

第3章では、今回の臨場感記録・再生システムのベースとなる人工現実感システムと、聴覚ディスプレイの役割を果たす立体音響装置に関して述べ、実際の開発環境などに関しても触れることがある。

第4章では、実際の音響的臨場感の記録・再生システムに関して述べる。

第5章では、音源の位置情報を推定することなしに、音響的臨場感を記録・再生することのできるシステムとしてVirtual Speaker System(VSS)を提唱する。

第6章では、本論文での研究結果を踏まえて今後の展望に関して述べる。

第7章では、本研究の結論についてまとめる。

## 第 2 章

### 臨場感に関する考察

## 2.1 はじめに

この章では、臨場感という概念に関しての理解を明確にすることを目的としている。人間の理解する臨場感を、視覚的な臨場感と聴覚的(音響的)な臨場感に分けて考え、それぞれの臨場感を構成する要素に関してまとめる。また、臨場感の伝送と記録・再生の社会的な必要性を述べ、視覚的な臨場感と聴覚的な臨場感を比べた場合にどのような技術的な困難が存在するかを認識する。

## 2.2 視覚的臨場感

### 2.2.1 立体感

臨場感を国語辞典で索くと、”現実に何かが行なわれている場面に身を置くような感じを与えられる場合を示す”とある。臨場感という言葉をさらに厳密なかたちで定義しようとするのは非常に難しいことであり、満足のいく定義をみつけることはできない。しかし、どのような要因により臨場感が生起されるかという事柄については比較的多くの研究結果を見い出すことができる。

樋渡は画像の物理的な情報と、それによって引き起こされる視覚心理的な反応との対応関係をまとめている。表2.1は畠田らがこれをさらに簡単に整理したものである[12]。この表の中で、まず、(B)の(c)立体感という要因に注目してみたい。単純に考えれば、もし、我々の周囲の全波面を合成することができれば、そこには同時に臨場感も生成されているはずである。そのように、「自然により近い空間を体験する」という観点から考えてみると、我々の見る画像に奥行き方向の空間情報がどの程度再現されているかということを視覚的臨場感生起の第一の基準としてよいであろう。

そのような、立体感を知覚する要因は以下のようにまとめることができる。

- (1) 運動視差：観察者の動きと表示画面との相互作用を考慮し、観察位置(視点)の変化に応じた画像を生成して表示することができれば、利用

表 2.1: 画像情報と視覚心理特性 (文献 [12] から引用)

	物理情報	感覚・知覚レベル	認識・情緒レベル
(A) 表示面	$\left\{ \begin{array}{l} 1) \text{画面サイズ} \\ 2) \text{表示面周囲条件} \\ 3) \text{表示像 (実像・虚像)} \end{array} \right\}$	$\left\{ \begin{array}{l} i) \text{視野} \\ ii) \text{調節} \\ iii) \text{観察状態} \end{array} \right\}$	$\left\{ \begin{array}{l} i) \text{臨場感} \\ ii) \text{疲労} \\ iii) \text{見やすさ} \end{array} \right\}$
(B) 空間情報	$\left\{ \begin{array}{l} a) \text{点・線} \left\{ \begin{array}{l} 1) \text{幾何ひずみ} \\ 2) \text{OTF高域特性} \end{array} \right\} \\ b) \text{面} \left\{ \begin{array}{l} 1) \text{階調} \\ 2) \text{色調} \\ 3) \text{形状} \end{array} \right\} \text{OTF低特性} \\ c) \text{立体} \left\{ \begin{array}{l} 1) 2\text{眼像} \\ 2) \text{多眼像} \\ 3) \text{空間像} \end{array} \right\} \end{array} \right\}$	$\left\{ \begin{array}{l} i) \text{錯視} \\ ii) \text{MTF高域特性} \\ \quad (\text{視力・エッジ現象}) \\ iii) \text{鮮鋭度} \\ \quad (\text{MTF低域特性}) \\ \quad (\text{明暗・色弁別}) \\ ii) \text{対比現象} \\ iii) \text{図形知覚} \end{array} \right\}$	$\left\{ \begin{array}{l} i) \text{識別性} \\ ii) \text{視認性} \\ iii) \text{誘目性} \\ iv) \text{可読性} \end{array} \right\}$
(C) 時間情報	$\left\{ \begin{array}{l} 1) \text{時間変化} \\ 2) \text{点滅} \end{array} \right\}$	$\left\{ \begin{array}{l} i) \text{感覚時間} \\ ii) \text{時間周波数特性} \\ \quad (\text{フリッカー}) \\ iii) \text{順応・経時対比} \end{array} \right\}$	$\left\{ \begin{array}{l} i) \text{不快感} \\ ii) \text{疲労} \end{array} \right\}$
(D) 時空間情報	$\left\{ \begin{array}{l} 1) \text{運動} \\ 2) \text{形の変化} \end{array} \right\}$	$\left\{ \begin{array}{l} i) \text{運動視} \\ ii) \text{時空間周波数特性} \\ iii) \text{図形残効} \end{array} \right\}$	$\left\{ \begin{array}{l} i) \text{追力感} \\ ii) \text{運動感} \end{array} \right\}$

者に積極的な観察姿勢をとることを可能にし、空間知覚に効果的である。

- (2) 両眼視差：空間を見たとき左右の眼の網膜上に写る像のずれ(視差)は、奥行き知覚の中心的な情報となっている。したがって、左右の眼に正確な視差画像を提示できれば、空間知覚に効果的である。
- (3) 輻輳、調節：左右両眼が物体をその視線上に捉えるように反対方向に動く運動を輻輳・解放運動と呼んでいる。調節は、眼のレンズである水晶体の厚さを毛様筋の収縮により変え、水晶体を含む眼球全体の焦点距離を調節する。
- (4) 画面精細度：眼の視力には分布特性があり、注視しているところは高いが、その前後ではぼけが生ずるため低い。人は、この網膜上のぼけを奥行き知覚の手がかりのひとつにしている。したがって、3次元世界モデルの中で観察者が注視している部分については、充分精細に

画面に投影し、その周辺の精細度については、実空間を見たときに網膜上に投影される画像の精細度分布を考慮して投影することが望ましい。

- (5) 広視野：視野が狭いと能動的な情報探索ができず、表示空間の奥行き知覚効果も弱くなる。したがって、視野の広い表示系が望ましい。

(1)、(2)、(3)の実現のためには、3次元情報を再現するような手段が必要であり、2眼、多眼式立体画像などの両眼視差などの奥行き弁別機能を満足させる方式やホログラフィーなどの物理的空間像を再現する方式がある。

一方、(4)、(5)は、大画面、虚像表示などの心理的効果を利用した方

式により実現することが可能である。

### 2.2.2 広視野効果

立体感がない場合でも臨場的体験を味わうことはできる。例えば、劇場で映画を見た場合の迫力感などがそれに相当する。そのような意味では、(A)の表示面の状態が臨場感合成の第二の基準としてあげられるであろう。

表示面の状態によって観察者が受ける効果としては、1. 広視野効果、2. 観察の許容条件、3. 表示面周囲の影響、4. 表示面の効果(実像表示と虚像表示の相違)などがある。このうちで臨場感に与える影響が最も大きいのは、広視野効果である。

広い視野に画像を提示した際に感じる広視野効果には、次のような効果が含まれている。

- (1) 表示面効果：画枠が広くなったり、表示面の素材が目立たなくなると、画像表示の位置情報が弱くなり、平面画像が奥行きや広がりのある画像に感じる。
- (2) 画像空間との融合効果：表示画像内の物体の傾きや動きによって、観察者自身の姿勢が傾いたり、乗物に乗っているように感じる場合がある。これは、表示された画像空間と観察者のいる空間とが融合した状態になり、観察者の主観的座標系が画像成分によって誘導されるために生ずる。

- (3) 表示情報と受容能力の関係：各感覚系からの情報受容量を比較すると、視覚情報が最も多い。ところが、狭い画角の画像表示では、視覚の受容能力を充分利用していない。画角を広げていくことにより、視覚情報が増加し、自然な情報受容状態に近付いていく。
- (4) 観察時の制約緩和：表示面が広くなると、観察位置や姿勢などへの制約が少なくなり、観察者の能動的動作(注視点移動など)が自由に行なえる。

以上のような効果が一緒になって、いわゆる大画面映画などを見たときに感じる心理効果－臨場感、奥行き感、広がり感、融合感、迫力感、自然感、開放感など－を生み出す一因となっているわけである。

### 2.2.3 広視野効果の定量化

畠田らは広視野効果のうち、一般に使われる臨場感に近い(2)の融合効果を中心に調べている。すなわち、臨場感という概念を、『表示された画像空間と観察者の居る空間とが、同一空間のように感じ、観察者の主観座標系が画像情報によって影響され、傾いたり、移動しているように感じる状態』と考えて、その定量化を行っている[13, 11]。

これは、広視野効果を定量化する目的で方向感覚誘導効果の実験を行い、これと表示画角の変化による主観評価実験とを比較したものであり、その結果、人間の視野の特性について図2.1を得ている。

図中Aは弁別視野と呼ばれ、視力などの視機能の優れている範囲である。Bは有効視野と呼ばれ、眼球運動だけで注視し、瞬時に情報受容可能な範囲である。Cは誘導視野と呼ばれ、呈示情報の存在がわかる程度の識別能力しかないが、空間座標系の判定には影響を与える範囲である。Dは補助視野と呼ばれ、暗黒内での光感覚がみられる範囲である。

## 2.3 聴覚的臨場感

人々が臨場感と言った場合には、視覚的臨場感に加えて暗黙のうちに聴覚的臨場感も含まれている。では、聴覚的臨場感とは何を指すのであろう

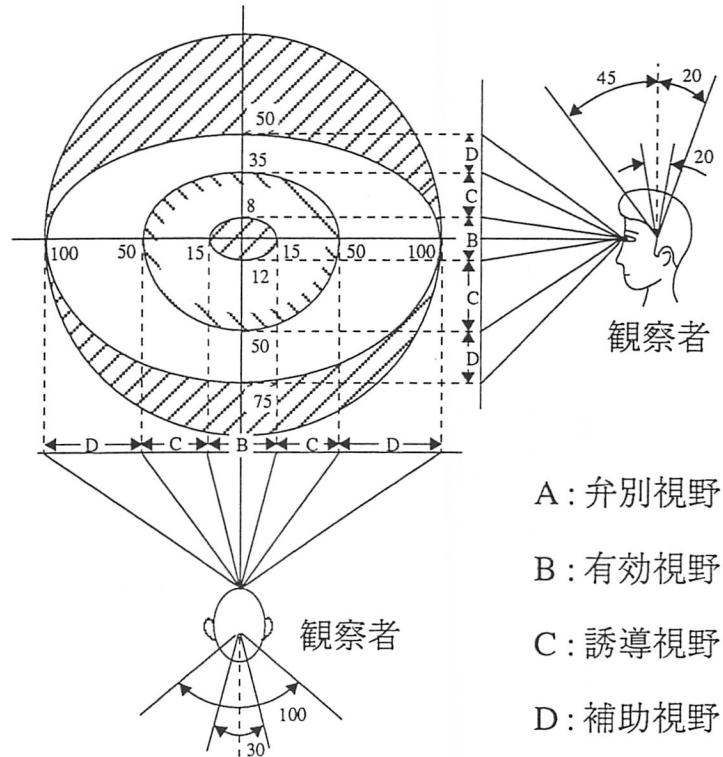


図 2.1: 視野内の情報受容特性 (文献 [11] から引用)

か。ここでは聴覚的な臨場感として、「第一人称的な音響体験ができる感覚」と定義することにする。

ヘッドフォンで普通の音楽や人の声などの音を聞いた場合に、聴覚的に高い臨場感を感じることは稀である。この理由として次の事柄が考えられる。

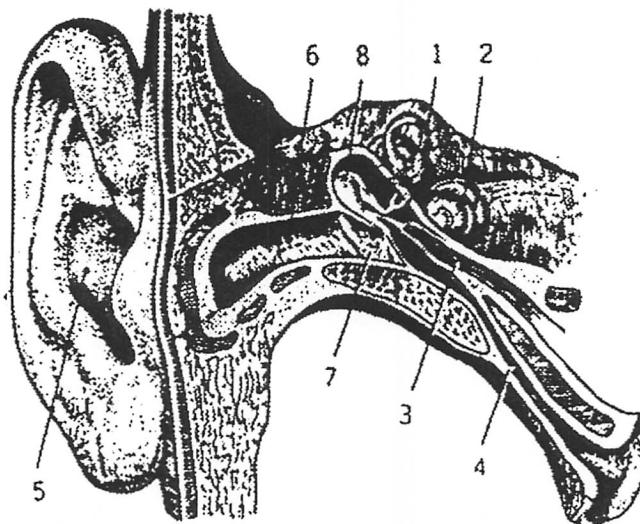
1. ステレオの音を聞いている場合でも、頭部の中か耳のすぐ側で鳴っているような知覚をする場合が多い<sup>1</sup>。
2. 頭部の運動に対して音の鳴っている方向が運動して動かないでの、実際の音との違和感を感じる。

そこで次に、1の現象の原因について考察してみることにする。

<sup>1</sup>頭内定位という。

### 2.3.1 音像定位の仕組み

多くの場合、人間は純粹に聴覚情報のみからその音源の位置を知る事ができる。この聴覚情報から心理空間に再構成された音源を音像という。そして、音像が認識されることを音像の定位 (localization) という。



1: 三半器官, 2: 蝸牛, 3: 耳内筋, 4: 耳管,  
5: 耳甲介腔, 6: 外耳道, 7: 鼓膜, 8: つち骨

図 2.2: 耳の構造 (文献 [3] から引用)

聴覚情報から音像の認識が行なわれるプロセスにおいては、人間の頭部と耳の構造による効果が重要な役割を果たしている。人間の耳は図 2.2 のような構造になっている。耳の構造を大きく分けると、外耳 (external ear)、中耳 (inner ear)、内耳 (inner ear) の 3 つの部分に分けることができる。音像の定位に重要な外耳は耳介 (pinna) と外耳道 (external ear canal) から構成されている。

耳介は頭部の側方に位置し顎の関節と側頭骨の間にあり、外耳道入口を包み込んでいる。耳介は堅い皮膚で覆われている軟骨格からなり、特徴的なレリーフ状の形をしており、はっきりと個人差が見られる。初期の研究

では、耳介の役割は集音装置としての役割だとされていた。しかし、現在では音響信号の方向と距離によって定まる線形フィルタの役割をはたしており、空間音響にとって不可欠なものである。

外耳道は、堅い皮膚を内壁とするやや曲がった管でその長さは平均 25 [mm] 程度である。また、外耳道の終端は鼓膜で、強い音圧の場合、反射的に筋肉の収縮が起こり、鼓膜の張力が増加して変位し難くなる。その結果、耳の感度が低下（この現象を音響弛緩という）してリミッターの役割を果たす。

聴覚情報から音像を認識する時に、どのような手がかり（キュー：Cue）を人間が利用しているかということについては、いくつかのモデルが提唱されている。

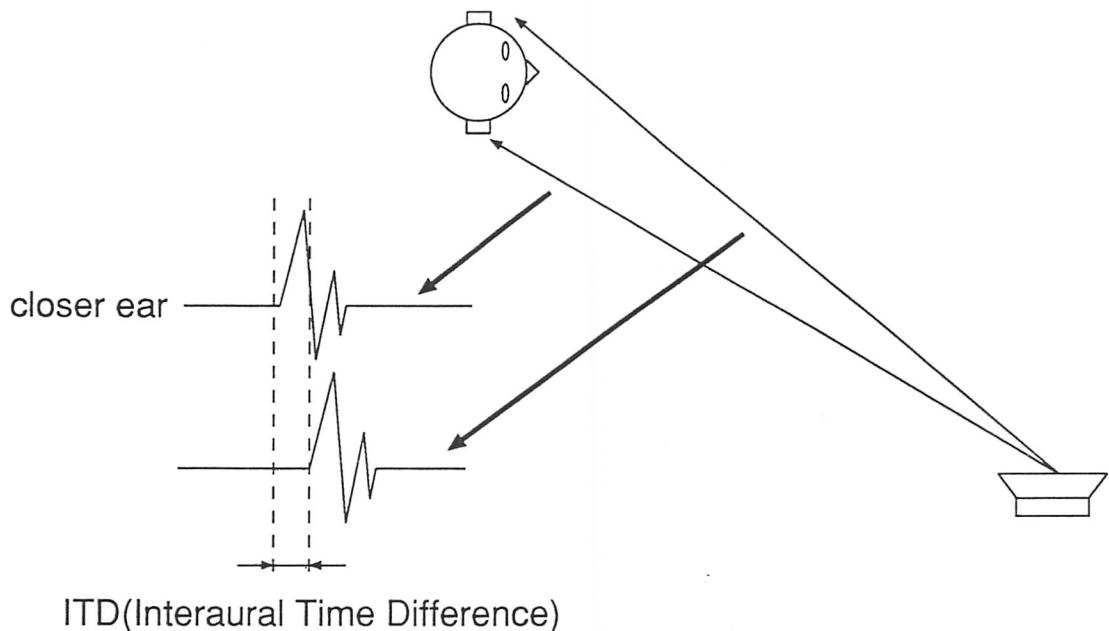


図 2.3: ITD の定義

一番単純なモデルは、Lord Rayleigh によって提唱された “Duplex Theory” である [27]。このモデルでは、左右の耳に到達する音響信号の時間差 (Interaural Time Differences:ITD, 図 2.3) と強度差 (Interaural Intensity Differences:IID, 図 2.4) という 2 つのキューを考えている。このモデルの問題

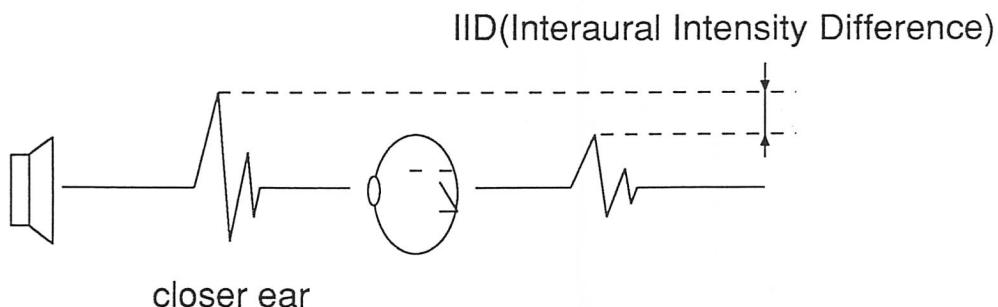


図 2.4: IID の定義

点は、左右の耳に到達する音響信号は相似と仮定しているために、あまり良い定位精度が得られない。この事実は、IID と ITD のキュー以外に波形の変化が音像定位において重要な役割を果たしていることを示唆している。

その後の研究結果から、現在では音像定位に重要なキューとして次のようなものがあることが分かっている [3]。

1. 両耳間時間差：これは Duplex Theory の ITD に相当する。厳密には、搬送波 (carrier) の位相差と包絡線の時間差の 2 つに分けて考える必要があり、前者はほぼ 1.6[kHz] 以下の周波数に対してのみ影響を与える、後者はだいたい 100[Hz] 以上においてキューとして作用する。
2. 両耳間レベル差：これは Duplex Theory の IID に相当する。予想されるように、可聴域全域でキューとなる。
3. 音響信号の波形変化：人間の頭部や耳介の影響により、音源の位置に依存する形で周波数的なフィルター効果を受ける。これは、可聴域 (Audio Frequency) での波長が耳介のスケールと同じオーダーなので、波の散乱・回折効果が重要になることに起因している。

これらの人間が認識しているキューの影響をまとめて評価するのには、伝達関数を用いるのが適切である。立体音響でよく用いられる伝達関数には、以下のようなものがある。

1. 自由空間伝達関数：音源と耳道内の1点(主として鼓膜上)との間の伝達関数。
2. 単耳伝達関数：ある方向と距離の音源に対する自由空間伝達関数を基準とした、自由空間伝達関数(原則として、正面より平面波が入射する場合を基準とする。)
3. 両耳間伝達関数：両耳間の測定可能な点における両耳音圧の関係。通常は音源側の耳における音圧を基準音圧とする。

自由空間伝達関数は、インパルス信号に対する鼓膜付近の音圧をプローブによって測定すれば計測可能である。この自由空間伝達関数をもちいれば(特定の個人に対しては)完全な音響信号の影響をシミュレートできる。

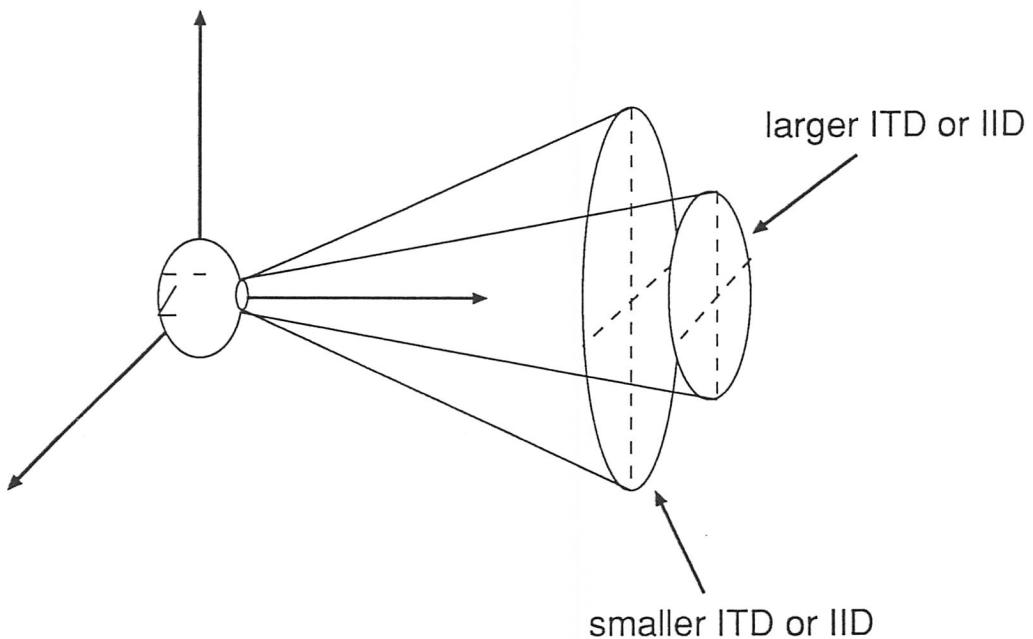


図 2.5: Cones of Confusion

次に、音像定位に関する問題点であるが、1つめとして、音像の定位位置が実際の音源と前後反対になる現象が良く見られる。これは、Duplex Theory を用いて考えれば図 2.5のように、ITD と IID が同じ音源の位置が Cone

状に存在することから容易に理解できる。前後の混乱に比べて上下の混乱が起こり難いのは、耳介の構造が上下非対称なためである。

2つめとして、音像の定位し易い音とし難い音が存在することである。物理的な面からは、周波数幅の極端に狭い音（純音など）は定位し難いことが分かっている。また、心理的な面では普段聞きなれない音の方が聞きなれた音よりも定位し難いことや、点音源であることが経験的に分かっている音（ノックの音や靴音）などは定位し易い事が分かっている。

3つめは、頭部音響伝達関数（Head-Related Transfer Function : HRTF）の個人による差が大きい点である。根本的にこの問題を克服するには、個人個人の HRTF をデータベース化しておくしかないが、多くの人が使ってみて良く定位するかなり一般的な HRTF が存在することが分かっている [3]。

4つめは、定位した音像には一般的に距離感が伴うが、色々な実験から距離感に関してはその人の存在する環境の残響音の影響が重要であることが分かっている。つまり、音像の定位位置（距離感）を制御するのには、その環境の残響効果を何らかの方法で考慮してやる必要がある。

### 2.3.2 脳内機構からの見た音像定位

動物の脳の内部で、両耳から伝わってくる音響信号から音像の位置を推定しているのであろうか。最近の研究によって、夜行性の猛禽メンフクロウ（図 2.6）における音響位置推定の脳内機構の仕組みが理解されるようになってきた [10]。

メンフクロウは、あらゆる動物の中で、最も正確に音源の位置を知ることができる。定位の誤差は、水平・垂直方向とも 1-2 度で、暗闇でも音だけを頼りに獲物をとらえることができる。メンフクロウは、2.3.1で述べた音像定位のための手がかりのうち、ITD を用いて音の水平位置を決め、垂直位置は IID を使って決めている。この鳥が音の垂直位置を知るのに IID を使える理由は、左右の耳朶の位置が上下方向にずれているからである。その結果、彼らが音源定位に使っている周波数領域（3-9[kHz]）では、左右の耳が上下方向によって異なる感度を持ち、IID が音源の垂直位置によって変化し、手がかりとなるのである。

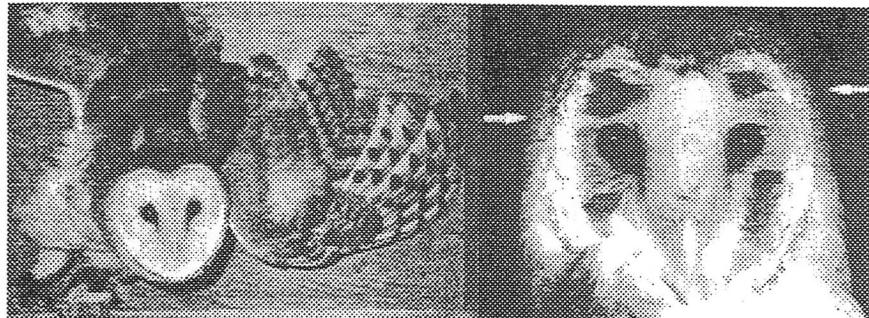


図 2.6: メンクロウ(英語名 barn owl, 学名 *Tyto alba*)(文献 [10] から引用)

内耳の音受容細胞は、周波数に選択性を持つが、音の位置はコードしていない。しかし、メンクロウの中脳の下丘外側核の細胞は空間内の特定の領域から発した音にのみ反応する。しかも、これらの細胞は、そのコードしている音の位置に従って、下丘外側核の中で整然と並び、聴覚空間の地図を形成している。受容細胞は音の位置を“知らない”のだから、脳が音の位置を算出し(聴覚受容野を持ち)、音の位置を表示した(聴覚地図を作った)のである。

メンクロウの脳内では、左右の耳から伝わってくる情報から、両時間の位相差を検出している。その仕組みを図示したのが、図 2.7 左である。

左耳と右耳からの位相情報を時間的なディレイを掛けて図のように A から E までの細胞に入力したとする。それぞれの細胞は、両耳からの信号が同時に到達した時に最大の活動電位を示すとすると、A から E までのそれぞれの細胞は、位相選択性を持つようになり、しかも位相差の大きさに従って配列するようになる。この回路は、1948 年に仮説として提唱されていたが、層状核でこの仮説の通りのことが起きていることが、近年、証明されている [10]。

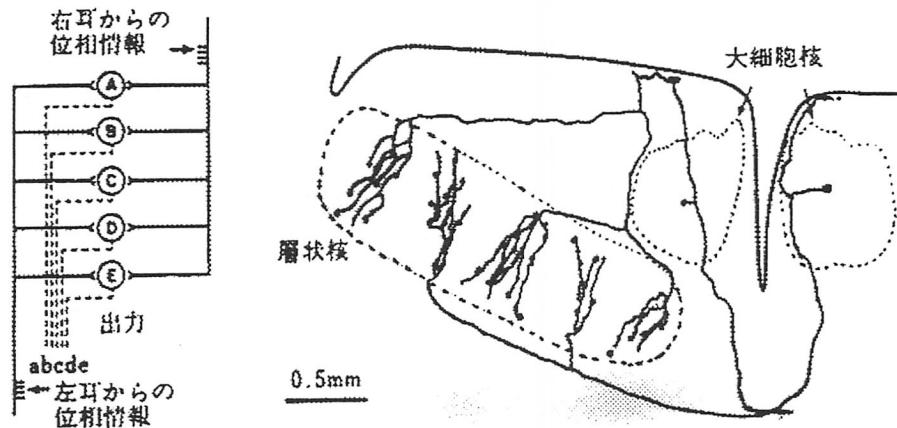


図 2.7: 両耳間位相差検出のための神経回路、モデル(左)とメンフクロウの層状核の構造(右)(文献[10]から引用)

### 2.3.3 音像の記録・再生方法

従来、音響的臨場感を記録・再生するのに大きく分けて次の二つのアプローチが考えられてきた。原音場を忠実に記録・再生する方法と、受聴者の両耳音圧を記録・再生する方法である。

次に、受聴者の両耳音圧を記録・再生する方式であるバイノーラル記録・再生方式について述べる。

バイノーラル方式とは、精巧に人間の頭部を真似た人形の耳の部分に小型マイクを組み込んだダミーヘッド(図2.8はダミーヘッドの一例)で、原音場を忠実に録音、そして(基本的に)ヘッドフォンによって音場を再生するシステムである。この方法の原型的なものは、すでにパリ国際博覧会(1881)でバイノーラル電話として登場し会場の話題をさらったという記録が残っている[30]。基本的には1対(2本)の電話を使って、オペラ座の音を博覧会会場まで伝送したという簡単な原理(バイノーラルというよりはむしろ、現在のステレオに近い)であるが<sup>2</sup>、あたかも自分がオペラ座にいるような錯覚を起こす程の臨場感を伝えたと記録には残っている。

<sup>2</sup>ダミーヘッドは用いていない。

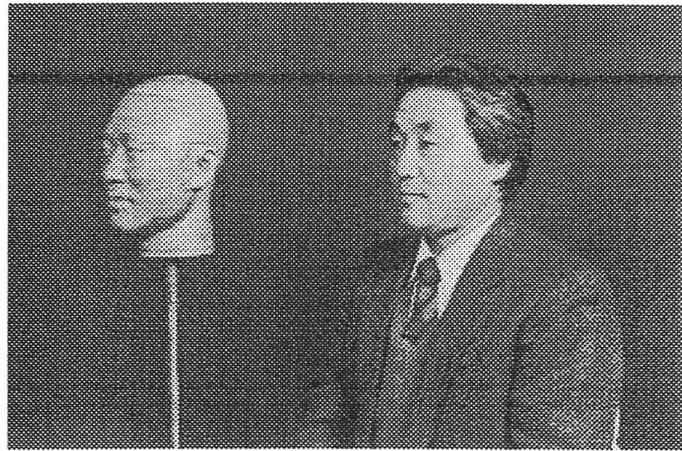


図 2.8: ダミーヘッドの例 (文献 [30] から引用)

バイノーラル録音した音響信号を再生するには、録音した2チャンネルの信号をそれぞれ左右の耳にヘッドフォンで再生するだけで良い。しかし、マイクロフォンをダミーヘッドの鼓膜位置に置いた場合には、両耳音響信号の収音を行なう時、音源からの音響信号はダミーヘッドの外耳道を一度通過するため、ヘッドフォンで再生すると管状の外耳道の持つ共振特性が2度かかった音響信号を受聴者は聞くことになる。従って、ヘッドフォン受聴での外耳道の共振の影響を避けるためには、マイクロフォンユニットを外耳道入口に配置したダミーヘッドを用意することが多い。

もう少し考察をすすめてみると、音響電気信号入力からヘッドフォンをかけた時の受聴者の鼓膜上の音圧までの伝達関数は、ヘッドフォンの種類によって異なってくる。ヘッドフォンには、外耳道に挿入するタイプ(インサートタイプ)、外耳を覆うタイプ(クローズタイプ)と、外耳の上に乗るタイプ(オープンタイプ)があるが、いずれも自然な受聴に比べて外耳部の働きが著しくゆがめられる。そこで、ヘッドフォンをかけた時、人間の頭部音響伝達関数に近い特性が自然に出るようなヘッドフォン(自由音響特性ヘッドフォン)が、一つのあり方として最近の主流になってきている。

しかし、この方法でも、収音時に音源方向で決まった頭部音響伝達関数で重みづけられた音響信号を再び頭部音響伝達関数に似た特性を持ったヘッ

ヘッドフォンで重みづけられる問題が存在する。この伝達関数が2度かかる問題を解決するためには、

1. ダミーヘッド収音信号を基準化する(そして、自由音場特性ヘッドフォンで受聴する)。
2. ヘッドフォンの伝達関数を平坦にする(そして、ダミーヘッド収音信号をそのまま受聴する)。

などの方法が取られる。

特に後者の方法については、従来、ヘッドフォンとは無色透明に音響信号を再生するものと考えられていたが、ヘッドフォンで受聴すると音像が頭内に定位することが問題となり、それに関する研究の結果、ヘッドフォンの伝達特性が問題であることが判明した。特にクローズタイプのヘッドフォンでの、ヘッドフォンの振動板の内側および耳甲介腔内にできる空の容積(キャビティ)による伝達特性の高域低下などが大きくクローズアップされてきた。現在でも、高域まで周波数特性が平坦なヘッドフォンは実現していない。しかし、掛けた状態での伝達特性を平坦にするためには、幸い電気回路によって補正することが可能である。

## 2.4 臨場感の伝送と記録・再生

ここでは、現在の利用されている各種のメディアと臨場感伝送 / 記録・再生システムの関係、それから臨場感の伝送と記録・再生における技術的な問題点の違いについて議論することにする。

### 2.4.1 従来のメディアとの関係

臨場感記録・再生システムを現在存在する様々なメディアと比較検討してみることにする。

#### シンボルベースのメディア

図2.9は、記号ベースのメディアについてまとめた図である。

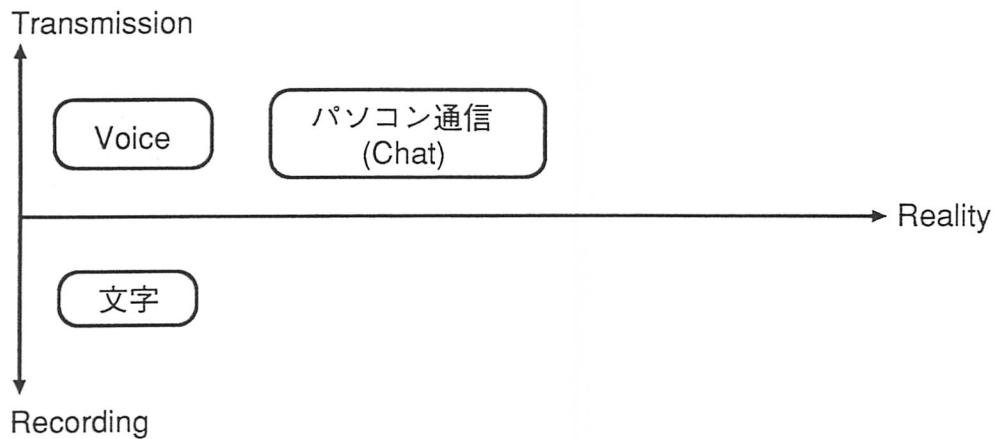


図 2.9: シンボルベースのメディア

シンボルベースのメディアとして一番古いのは間違いなく、音声的言語である。時代的にその次のメディアとして、記号的言語（文字）を上げることができる。

歴史上初めての文字が生まれたのはメソポタミアであるといわれている[22]。BC 第 6 千年紀から第 1 千年紀にかけて、この地域には南部にシュメール人、北部にはアッカド人がそれぞれの国を作っていた。初期の文字は、シュメール人が残したウルク市街の大神殿跡から発見された「ウルク書板」と呼ばれている BC 第 4 千年紀の粘土板からうかがい知ることができる。この粘土版には、穀物や家畜の数などが記録されており、神殿で使用されていた一種の会計帳簿だといわれている。このように人類が初めて記号を使ったのは、農業に関しての記録だったと推測されている。

最近では、パソコン通信の普及により“Chat”と呼ばれる、文字を用いた対話的なコミュニケーションも行なわれるようになってきた。

#### 音声ベースのメディア

次に、図 2.10は音声（音響情報一般を含む）を用いるメディアについてまとめた図である。図の横軸には臨場感をとり、縦軸には情報の伝送か記録かによって区分してある。

音声情報の伝送では、ラジオ（無線）や電話（有線）のようなメディアが存在し、音情報の記録では、一番古い蓄音器などを用いたモノラル録音、2チャンネルを同時に記録するようにしたステレオ録音、またより臨場感を損なわない記録方法として2.3.3で述べたバイノーラル録音方式がある。

この図中で、音響的臨場感伝送／記録・再生システムは一番右端に位置し、従来のメディアでは不可能だったより第一人称的な臨場感のあるシステムに位置する。

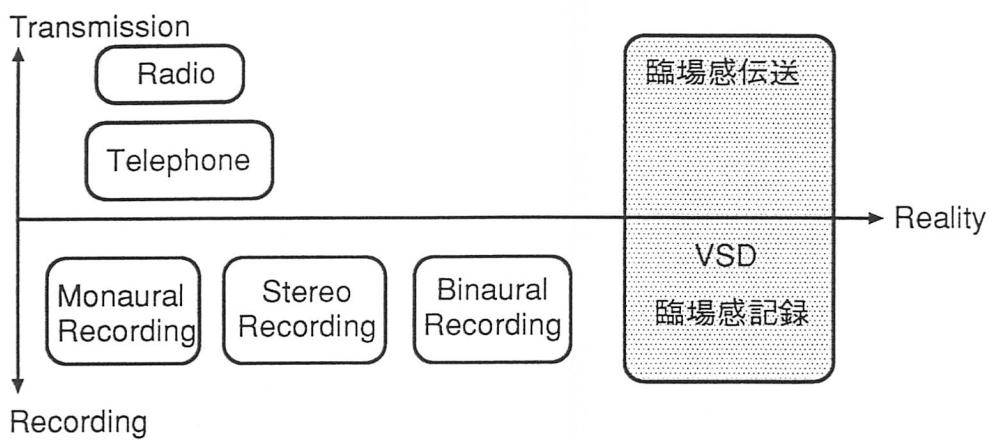


図2.10: 音声ベースのメディア

#### 画像ベースのメディア

図2.11は、画像ベースのメディアをまとめてみたものである。

一番左端には、静止画の写真が位置している。その次には、動画像のメディアであるテレビジョン、映画やビデオが位置している。

視覚的臨場感を伝送／記録・再生するシステムは、音響的臨場感での場合と同様に一番右に位置している。

#### 2.4.2 臨場感の伝送

視覚的臨場感の伝送する方法として、従来開発が行なわれてきたシステムの概念的構成図を示したのが、図2.13である。基本的な構造は、人間

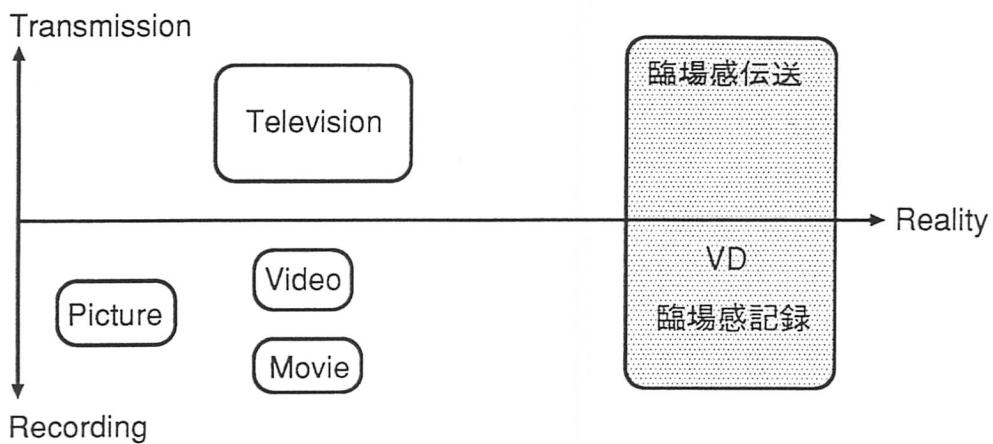


図 2.11: 画像ベースのメディア

の頭の動きをセンサで検出し、その動きと連動するように遠隔地のカメラを動作させる。そして、カメラからの画像を頭部搭載型ディスプレイ (Head Mounted Display:HMD, 図 2.12) で装着者に提示してやれば、非常に広い視野を提示することができる。

ただし、この場合に通信系に時間的な遅れが存在し、頭部の運動と画像の更新の間に大きな時間的な遅れができるようになると、見回し動作ができる場合でも臨場感が損なわれ、さらに時間遅れがある臨界点を越えてしまった場合には、利用者が酔いや眩暈等の症状を見せることも報告されている。

しかし、ここでの議論は通信系の時間的な遅れが無視できるものとして進める。

音響的な臨場感の伝送に関しても視覚的な臨場感の場合のアナロジーを適用すれば図 2.14 のようなシステムを考えることができる。

#### 2.4.3 臨場感の記録・再生

まず、視覚的臨場感の記録・再生について考えてみる。

2.4.2で考えた視覚的臨場感伝送システムを臨場感の記録・再生システムに応用することを考えると、画像の時間差分的な情報量が十分少なければ、



図 2.12: HMD(Head Mounted Display)

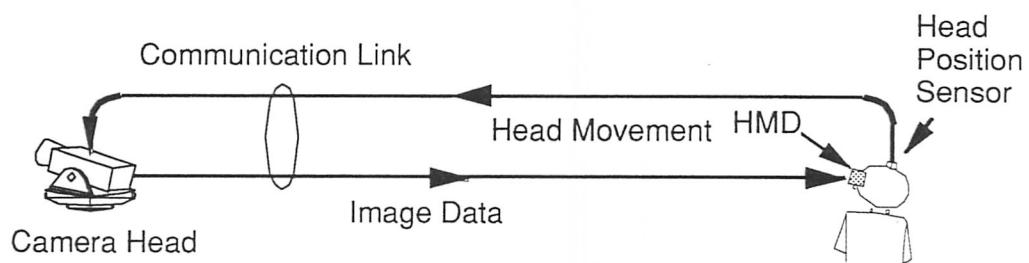


図 2.13: 視覚的臨場感伝送の概念図

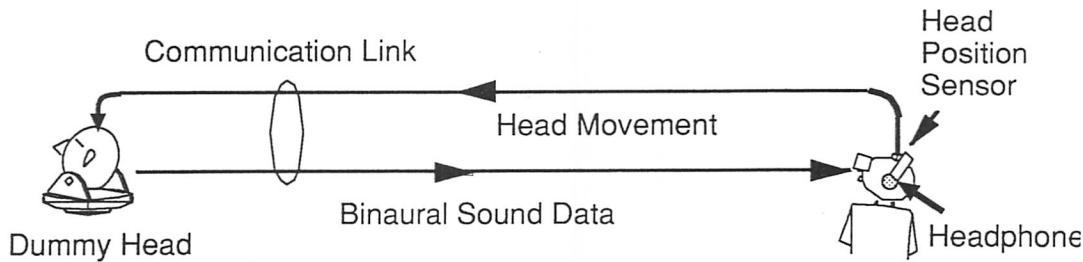


図 2.14: 聴覚的臨場感伝送の概念図

つまりあまり画像の変化がなければ、図 2.15 のように、画像を高速に検索して提示できるメディア (LD など) に全天周の画像を一度蓄積しておいて、再生する時には、頭部の動きと矛盾しない画像を提示してやることによってある程度の視覚的臨場感を再生することができる。

これに対して、聴覚的な臨場感を再生することを考えてみる。(ここでは、大がかりな装置を必要とする音場再生方式については除外して議論をすすめる。)

先ほど述べた視覚的な臨場感を記録・再生するシステムとのアナロジーで考えると、ダミーヘッドで録音した音をヘッドフォンで再生する 2.3.3 で触れたバイノーラル方式のようなシステムを考えてしまうが、この方式には重大な欠点が存在する。

それは、頭部の動きと音像の動きが一致しないということである。この欠点を解消するには、次の 3 つの方法が考えられる。

1. 全天周の静止画像を取るように、ダミーヘッドをすべての方向に向けて、全天周の静止音像を記録する。
2. 様々な方向を向けたダミーヘッドを多数配置して、様々な方向を向いた時の音をすべて記録する。
3. 一つのダミーヘッドまたは、有限個のマイクロфонのデータから臨場感を再生する時に任意の方向から聞いた時の音響信号を再構成する。

1は、定常的な音響的な臨場感しか記録・再生できないから実用的ではない。また2は、全天周を覆うには多数のダミーヘッドを(できるだけ同じ空間的位置に)配置しなければならず、また記録も非常に多数のチャネルを記録しなければならないので、非現実的である。そこで、3の方法であるが、もし有限個のマイクロフォンの音響データを、音源の位置(方向)と音源から放出されている音響信号自身の情報に分離して、何らかの方法によって頭部の方向を補正した音響信号を生成できれば、音響的な臨場感を記録・再生できると考えられる。

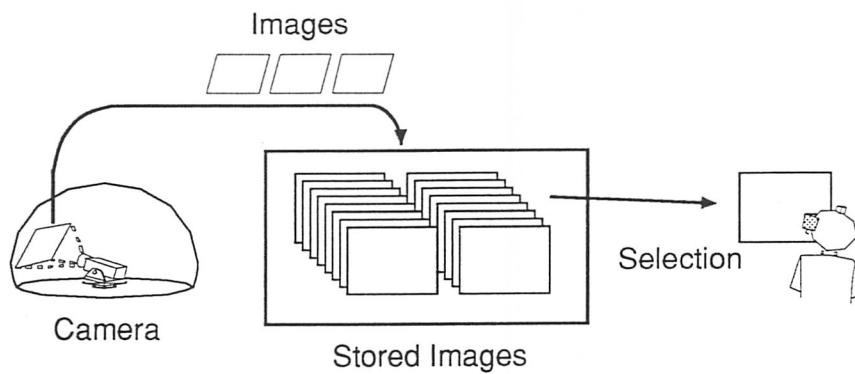


図 2.15: 画像蓄積式臨場感再生システム

# 第 3 章

## 臨場感と人工現実感

### 3.1 はじめに

人工現実感 (Virtual Reality) というキーワードが注目されるようになって久しいが、この間、人工現実感そのものの概念が大きく変化してきていると感じられる。すなわち、当初人工現実感システムと言えば体験者は VPL 社製の 頭部搭載型ディスプレイ (HMD) EyePhone を装着し、そこに立体的に提示される仮想の空間を体験し、また、場合によっては Headphone まで頭部に装着し立体音場まで体験することができ、更に人間の側からの入力装置として、これまた VPL 社の手形状計測装置 DataGlove を装着し、ときには DataSuits なる身体形状計測装置を着込み体験者がどのような格好をしているかを入力できるようにしてあり、そして頭部及び手の位置及び姿勢の計測装置として Polhemus 社の Polhemus Sensor をそれぞれ手おおよび頭につけている姿を思い浮かべる場合が多くかった。つまり、

- HMD による立体映像
- 3 次元空間センサによる頭部及び手の位置の検出
- 特殊手袋による手の形状の獲得

が揃ってはじめて人工現実感であるという見方すらされていたように思われる。

しかしながら、最近では人工現実感の概念がもっと拡大されて解釈されるようになってきた。すなわち、関連している技術としての、テレオペレーション・テレイゲジスタンス、3次元コンピュータグラフィックス、CAD/CIM なども人工現実感の一つとして捉えられてきている。

もちろん、これらには3次元表示、インタラクション 等の共通した性質がある訳で、先に挙げた例はむしろ、人工現実感の一側面でありあくまで狭義の人工現実感として捉えられるべきであろう。

また、今一つとして、先の条件を必ずしも全て満たさなくても構わないというような捉え方がされてきている。すなわち、例えば先に挙げた3つの条件が全て満たされない場合でも人工現実感と呼ばれているものもある。もちろん、ここにおいては、アミューズメントの分野による一般大衆に対

して Virtual Reality という言葉の新しいインパクトのあるキーワードを宣伝として(当時としては)拡大解釈しての使用し、そのマスメディアの流行化という背景の存在は否定できないと考えられる。しかし、その結果として Virtual Reality、人工現実感という言葉が逆に広く市民権を得たとも言えるであろう。

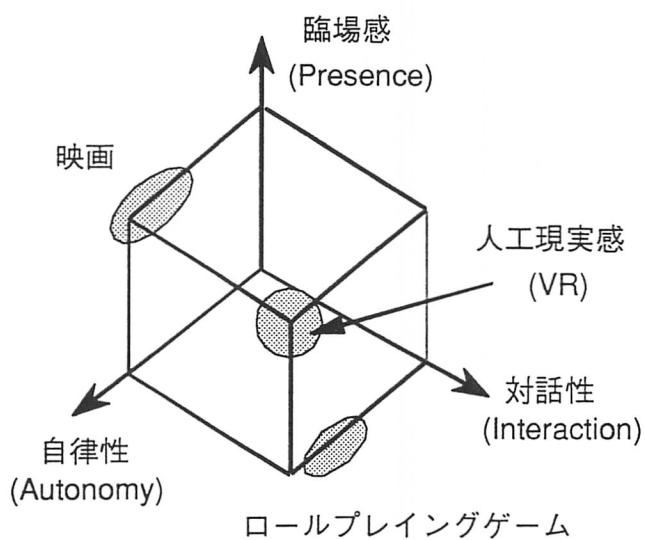


図 3.1: 人工現実感の三要素 (ゼルツァ提唱)

現在ではむしろ、MIT のゼルツァの提唱する図 3.1 の様な図を用いて人工現実感の説明される場合が多い。ここでは、3 要素として、

- 臨場性 (Presence)
- 対話性 (Interaction)
- 自律性 (Autonomy)

が挙げられている。先の 3 要素と比較してみると、HMD による立体映像が臨場性に、3 次元空間センサ及び特殊手袋の部分が対話性に、そして、対話による結果の世界への反映の部分が自律性と見ることができるであろう。

このように見てみるとことによっても、最初に挙げた人工現実感のイメージが、確かに典型例ではあるが、一例に過ぎなくなっていることも理解できるであろう。

### 3.2 人工現実感におけるインターフェース

次に、人工現実感システムにおける重要な要素である、人間とのインターフェース部分について詳しく見てみることにする。

人工現実感のシステムにおいて、人間とコンピュータシステムの間のインターフェースを受け持つサブシステムとして、コンピュータの内部状態(仮想空間の状態)を人間に理解できるような形で提示するディスプレイサブシステム(ここでディスプレイとは、視覚の情報を提示する普通の意味でのディスプレイも含む広い意味での提示装置のこと)と、人間の状態を推定するための計測サブシステム(人間の外部状態だけではなく脳波や心電のような内部状態を計測する装置も含む)がある。

人間に情報を提示するためには、当然人間の五感(視覚、聴覚、触覚、味覚、嗅覚)を利用するので、それぞれの感覚に応じて情報提示装置(Display)が考えられる。それぞれの感覚ディスプレイの現状をまとめると次のようになる。

- 視覚ディスプレイ：視覚ディスプレイとして有名なものにHMD(Head Mounted Display)がある。固定型の視覚ディスプレイでは困難な全方向への情報提示が可能なHMDであるが、一般の人が期待するよりもはるかに解像度が低い。高い臨場感を得るためにもさらに解像度の高いHMDが望まれる。
- 聴覚ディスプレイ：普通のステレオ方式の音響提示システムでは、左右の方向を提示するのは可能であるが、前後と上下の方向を提示するのは困難である。そこで、立体音響を利用した聴覚ディスプレイが注目を浴びている。
- 觸覚ディスプレイ：仮想空間内でさまざまな作業を行なうにあたって

触覚という感覚は想像以上に重要である。しかし、現在の段階として人間の体全体の触覚を実現するのは非常に困難なので、主に手に対する触覚ディスプレイが主な研究対象となっている。

- 嗅覚・味覚ディスプレイ：嗅覚と味覚への情報提示(刺激)には、何らかの化学的な作用が必要不可欠であり、視覚における三原色のようなものも確定していないので、今後の生理学的な研究が進歩する必要がある。

次に、人間の状態を推定するための計測サブシステムとしては大きく分類して人間の内部状態と外部状態を計測するシステムがある。

外部状態を計測するシステムとしては、位置センサーの Polhemus Sensor や DataGlove などが有名である。その他に、少しユニークなシステムとして、DATANOSE を上げることができる [15]。

また、人間の内部状態を計測するシステムとしては、指先の血流量を測定して緊張度を測るシステムや、心電から RRV を計測して集中度を測定するシステムなどがある。

### 3.3 立体音響提示システム

ここでは、音響的臨場感を伝送 / 記録・再生するために必要不可欠である受聴者に対して立体音響 (Spatial Sound) を提示できる立体音響システムについて述べる。

立体音響システムには、そのアプローチによって次の2つに区分することができる。

1. 受聴者の周囲の音場をそのまま再構成してしまう合成音場方式。
2. 受聴者の鼓膜に到達する音場をシミュレートして立体音響を提示する頭部音響伝達関数 (HRTF) 方式。

次に、個々の方式についてもう少し詳細に見てみることにする。

### 3.3.1 合成音場方式

音場をそのままある空間に再生するためには、原理的には直接音およびすべての個々の反射音、そして残響音をシミュレートしなければならないが、実際には不可能である。しかし、室内の収音する1つの場所での音場を十分な精度でシミュレートすることは可能である。例えば、コンサートホールのある場所の音場のシミュレートは可能である。この目的を達成するには、多くの再生スピーカを再生室(この目的では、無響室が最適)に配置する。これらのスピーカは、受聴者に対して最も重要な反射音の入射方向に配置する。そのほかに多くのスピーカを拡散、残響音場用として配置する。スピーカは、個々にまたは群として、フィルタ、レベルコントロール、遅延ユニット、残響装置の回路によって駆動される。この装置の一例を図3.2に示す。

この方法では当然、伝達関数の個人差は問題にならないが、再生装置が非常に大がかりなものとなってしまう事が欠点である。

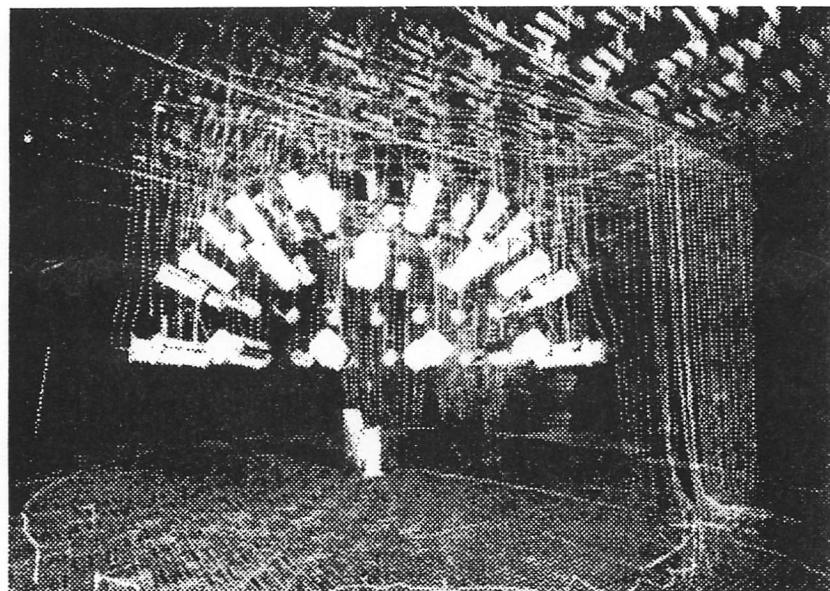


図3.2: ゲッティンゲン大学第3物理学研究所の合成音場装置; 被験者の周囲に65個のスピーカーが配置されている(文献[3]から引用)

### 3.3.2 頭部音響伝達関数方式

人間の頭部音響伝達関数が既知であれば、計算機などによって、鼓膜での音圧をシミュレートすることが可能である。

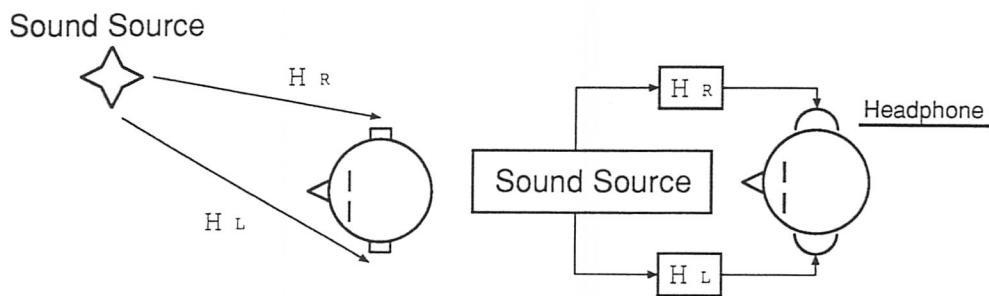


図 3.3: 頭部音響伝達関数方式による立体音響システムの原理

図 3.3 の左図のように音源で発生した音が、人間の鼓膜に達するまでの影響を考えてみる。無響室であると仮定すると、音源で発生した音は人間の頭部や耳の影響を受け波形が変化する。ここでこの波形の変化は一種のフィルターがかかったものと考えることができる。この波形の変化をモデル化するのに頭部音響伝達関数 HRTF(Head Related Transfer Function) を導入する。この頭部音響伝達関数は、被験者の耳の中に極めて小さいプローブマイクロフォンを挿入してインパルス応答を計測することによって測定することができる。

次に、立体音響を生成する方法であるが、簡単にいえば左右の耳に到達すべき音響信号を頭部音響伝達関数から合成すれば良い。これを数学的に表現すれば、音源の音とインパルス応答(頭部音響伝達関数の時間表現)の畳み込み積分(Convolution)を取ることに相当する。式で表現すると、

$$O(t) = \int_0^t I(\tau)H(t - \tau)d\tau \quad (3.1)$$

となる。ここで  $I(t)$  は音源の信号、 $H(t)$  はインパルス応答、 $O(t)$  は出力の音響信号である。図 3.3 の右図のようにヘッドフォンを用いて立体音響を生成する。

この方法を用いれば、原理的には(無響室での)音場を忠実に再現することが可能である。

但し、実際の立体音響システムでは頭部音響伝達関数を極座標でグリッド状の各方向で測定しておき、立体音響を生成する時には適当な方向の頭部音響伝達関数を既知の頭部音響伝達関数から推定(補間)し立体音場を合成する。

### 3.4 立体音響ディスプレイ

3.3では、受聴者に対して立体音響を提示するためのシステムについて概観したが、このなかで合成音場方式の立体音響システムを人工現実感のシステムとして組み込むには次のような利点と欠点が存在する。

利点としては、

1. 頭部の形状の個人差の問題、つまり頭部音響伝達関数の個人差が問題とならない。
2. 音場自身を再現しているので、頭部の運動をトラッキングする必要がなく、HRTF方式で問題となり易い時間遅れの問題が存在しない。

などが上げられる。それに対して欠点は、

1. 装置自身が大がかりなものになってしまい、ポータブルな人工現実感システムには組み込むことが困難である。
2. 音場を再生する空間として(理想的には)無響室が必要で、また再現しようとする音響場の時間的周波数が上がるにつれ、非常に多くのスピーカを必要とするようになる。

である。

これらの利点と欠点を総合して考慮すると、合成音場方式よりはHRTF方式を採用した方が良いと考えることができる。そこで、本研究ではHRTF方式を中心として議論をすすめることにする。

ここで、立体音響提示システムと立体音響ディスプレイシステムの違いについて明確にしておこう。人工現実感システムにおける音響ディスプレイとして要求される機能として、音像を提示することの他に提示するべき音響信号を合成する事も広義には含まれる。

次に紹介するのは、立体音響ディスプレイシステムの核となる HRTF 方式の立体音響提示システムである。これにサンプラー やシンセサイザーのような音源を組み合わせることによって、(広義の) 立体音響ディスプレイとしての機能を満たすことができる。

### 3.4.1 RSS システム

RSS システムは、ローランドが 1990 年に発売した 2 チャンネルスピーカによる立体音場録音装置である。このシステムは元来立体音響をともなった 2 チャンネルスピーカ再生用の音響信号を構成するためのシステムである。

RSS システムの全体構成を示したのが、図 3.4 である。全体の信号の流れは、A/D/A CONVERTER でサンプリングした信号を PROCESSOR (RSS-8048) で信号処理しその結果を再び A/D/A CONVERTER に戻しアナログ信号化してスピーカで再生したり、普通のステレオのレコーダで録音する。PROCESSOR は、方向に関する頭部音響伝達関数と入力信号の畳み込み積分 (Convolution) を行なうバイノーラルプロセッサーと、スピーカによるクロストーク信号の影響を排除するような信号を付加する (原理については後述) トランゾーラルプロセッサーから構成される。図中の一番左の CONTROLLER (SSC-8004) は MIDI で外部の装置と接続され RSS システム全体の制御を行なう。

RSS システムの特徴の一つに、2 チャンネルのスピーカによる立体音響の再生を挙げることができる。ステレオスピーカによる再生を行なう場合には、ヘッドフォンを使う場合と違い左耳に到達する音響信号は左のスピーカから出力された信号だけではなく、右のスピーカから出力された音響信号も (当然) 到達することになる (クロストークという)。このため、バイノーラル録音した音響信号や頭部音響伝達関数を畳み込み積分した信号はその

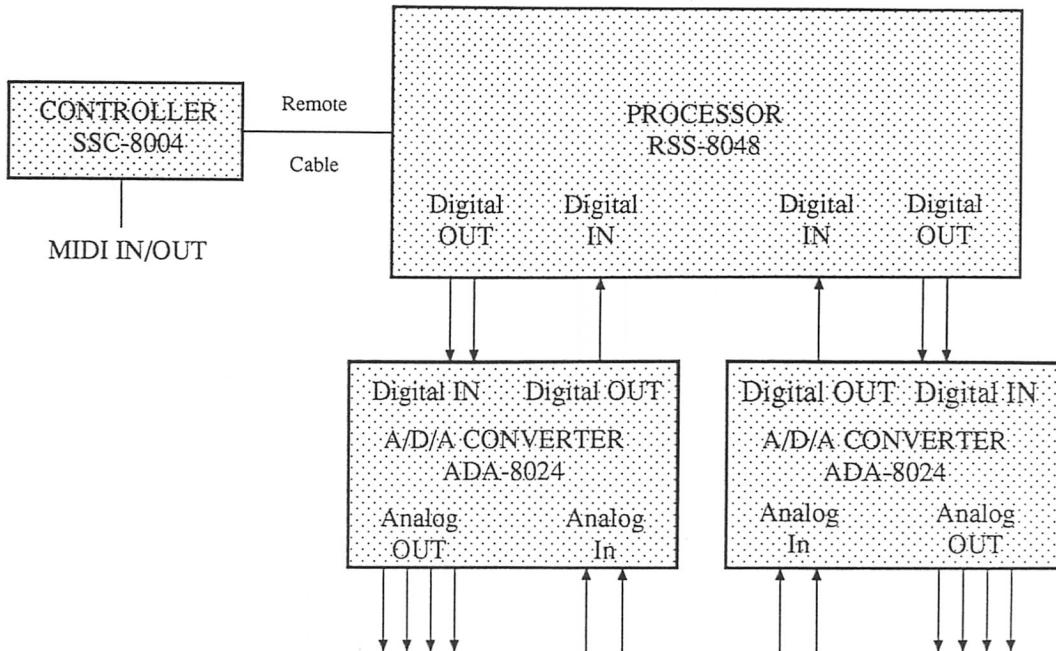


図 3.4: RSS システム

ままでは、スピーカシステムでは再生できることになる。

この問題点を解決するために、左右のスピーカから再生する音響情報にあらかじめクロストークを排除するような信号を補償しておくような方法を RSS システムでは採用している。

このクロストークを排除するための方法を示したのが図 3.5である。ここでは、議論を簡略化するために受聴者は 2 つのスピーカから等位置に存在し、スピーカから耳までの伝達関数には、左右による差はないものと仮定する。

図の一番左から、ダミーヘッドで収音した信号や伝達関数の影響を考慮してバイノーラル化した音響信号が伝わってくるものとする。このバイノーラル音響信号に、この図のように補償回路を入れたシステムで処理しステレオスピーカで再生してやればクロストークを(理論的には)完全に排除することができる。

次に、この補償システムでクロストーク信号を除去できることを数式を

使って確かめてみることにする。左側のバイノーラル音響信号を  $L$ 、右の音響信号を  $R$  とすれば、左のスピーカから放射される信号  $L_S$  と右のスピーカから放射される信号  $R_S$  は、

$$L_S = \frac{1}{1 - \left(\frac{H_{12}}{H_{11}}\right)^2} \frac{1}{H_{11}} \left( L + \frac{-H_{12}}{H_{11}} R \right) \quad (3.2)$$

$$R_S = \frac{1}{1 - \left(\frac{H_{12}}{H_{11}}\right)^2} \frac{1}{H_{11}} \left( R + \frac{-H_{12}}{H_{11}} L \right) \quad (3.3)$$

と記述される。この2つの信号から実際に左耳の鼓膜に到達する信号  $L_R$  は次のように計算される。

$$\begin{aligned} L_R &= H_{11}L_S + H_{12}R_S \\ &= \frac{1}{1 - \left(\frac{H_{12}}{H_{11}}\right)^2} \frac{1}{H_{11}} \left\{ \left( H_{11} - \frac{H_{12}^2}{H_{11}} \right) L + \left( H_{12} - H_{11} \frac{H_{12}}{H_{11}} \right) R \right\} \\ &= L \end{aligned} \quad (3.4)$$

同様に、 $R_R = R$  となって、受聴者の鼓膜に伝達したい音響信号 ( $R$ ,  $L$ ) がスピーカシステムを用いても伝達可能なことが分かる。

ただし、この方法では受聴者は頭部を軽く固定し、動かしてはならない。回路定数を設定するために行なったスピーカと受聴者の距離関係は厳しく守られる必要がある。受聴者が頭部を動かして、設定した値が変わると、信号の到達時間に左右の耳で違いが生じ、もはやクロストークは完全にキャンセルされなくなる。

前方のみを注視して、比較的頭部の運動が少ないテレビジョンやラジオ、ステレオの音楽のような分野ならまだしも、人工現実感のような臨場感を重視する分野で使用するには、この制約は強過ぎるようと思われる。

### 3.4.2 Convoltotron

Convoltotron は Crystal River Engineering 社製のリアルタイム立体音響提示システムである。Convoltotron を組み込んだシステムとして、Audio-

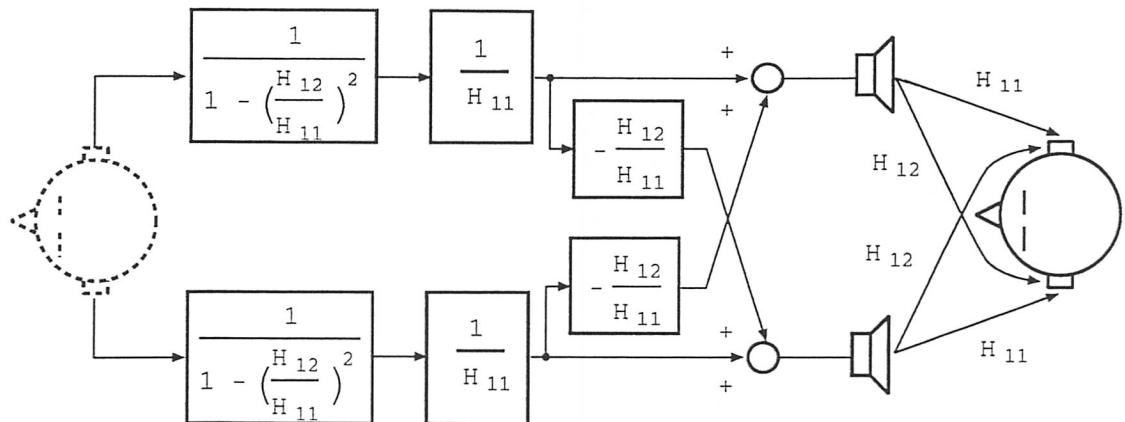


図 3.5: 2 チャンネルスピーカ再生時の電気的補正方法

Sphere がある。この AudioSphere とは SIGGRAPH '90 で発表された VPL 社製の立体立体音響ディスプレイシステムである<sup>1</sup>。AudioSphere は VPL 社の人工現実感システム (RB2) に組み込めるように設計されたシステムである。

AudioSphere の簡単な原理を示したのが図 3.6 である。インパルス信号を音源とする時の、耳の外耳道入口の信号 (FIR:Finite Impulse Response) を測定しておき、それと定位させたい音の Convolution をリアルタイムで計算し、ヘッドフォンで受聴者に提示する仕組みである。

AudioSphere の心臓部を占めているのが Convolvotron という Convolution engine である。Convolvotron を大きく分けると DSP(TMS320/C25) と並列乗算器 (128 個) から構成されている (図 3.7)。Convolvotron は、4 つの音源 (Sampling Rate : 50[kHz]) の音像を同時リアルタイムで制御することができる性能を持っている。

正確にいうと、Convolvotron とはプラットフォームとして IBM-PC/AT 互換機を用いる 2 枚の ISA ボードとそれを制御するソフトウェア群から構成されている (表 3.1, 表 3.2)。Convolvotron の外観を図 3.8 に示す。

古いバージョンのソフトウェアライブラリでは、プリミティブな関数群

<sup>1</sup>VPL 社はすでに解散している。

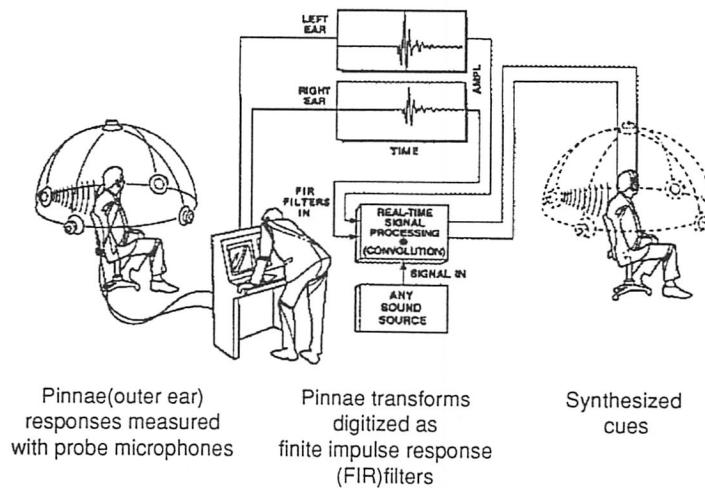


図 3.6: AudioSphere の原理図 (文献 [27] から引用)

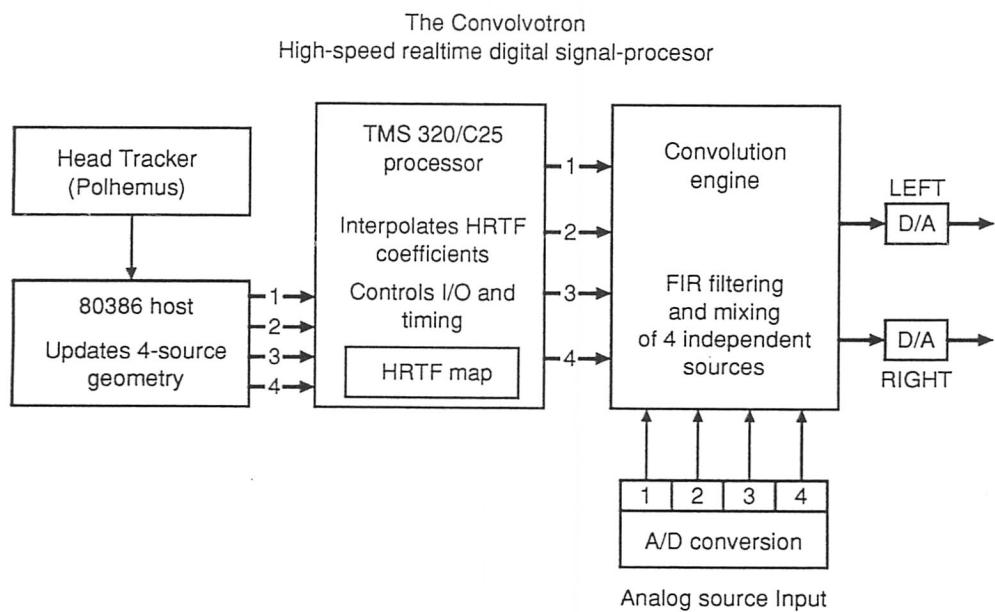


図 3.7: Convolvotron の構成図

しかサポートされていなかったが、新しいバージョンのソフトウェアライブラリでは、三次元空間内での受聴者との音源の位置関係から相対的な音源方向をライブラリ側で計算するよう変更されている。

また、Convoltotron を他の計算機と接続する時に Convoltotron をサーバ化するプログラムも添付されるようになった<sup>2</sup>。

Convoltotron の他に Crystal River Engineering 社では、DSP にモトローラ社の 56000 を用いた Beachtron と、8台の Convoltotron と 7台の Beachtron を制御するためのシステムとして Acoustetron を開発している。

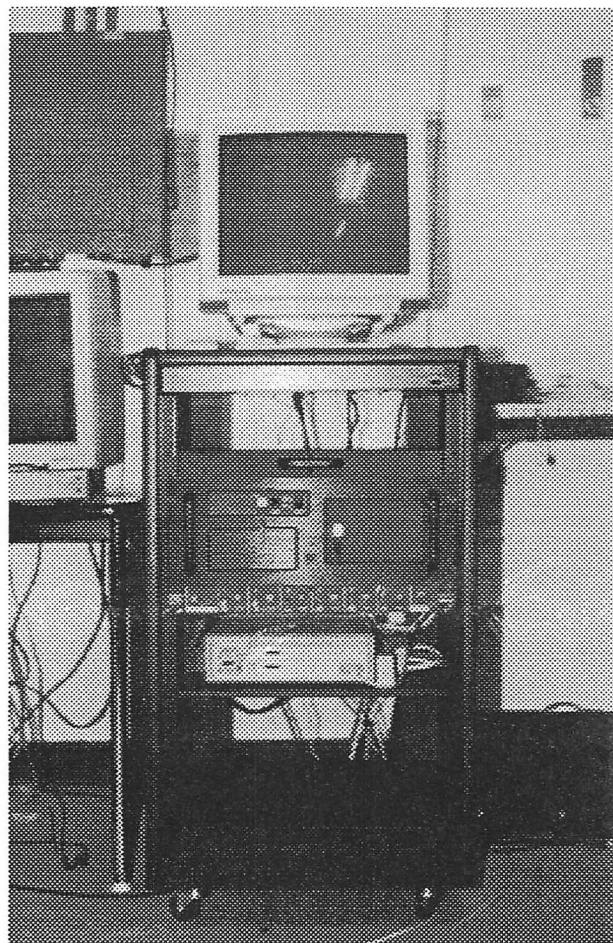


図 3.8: Convoltotron の外観

---

<sup>2</sup>但し、その分全体的にブラックボックス化されている。

表 3.1: Specification of Convolutron

Convolutron	
Platform	ISA 2-card set for IBM PC-AT compatible machine.
DSP	TMS320/C25
Inputs	4 inputs – synchronized 16-bit A/D converters running at up to 50[kHz].
Outputs	2 outputs – separate 16-bit D/A converters synchronized to the A/Ds.
Speed	Peak convolution speed of 320 million [taps/sec] (multiply, accum & shift)

表 3.2: Specification of DSP(TMS320/C25)

DSP(TMS320/C25)	
Instruction Cycle	100[ns] (10 MIPS)
Program Memory Area	64[Kword]
Data Memory Area	64[Kword]
Instruction Set	Compatible to TMS32020 DSP(Object code level)
Clock	40[MHz]

### 3.5 当研究室の人工現実感システム

ここでは、当研究室で研究用に使用している人工現実感システムをハードウェアから見た面とソフトウェアから見た面に分けて見てみる。

#### 3.5.1 人工現実感システムのハードウェア

当研究室の人工現実感システムの概要を図示したのが図3.9である。次に、システムを構成する個々の要素に関して次に述べることにする。

##### 頭部搭載型ディスプレイ

人間に対して画像を提示するためのシステムとして、頭部搭載型ディスプレイ (Head Mounted Display : HMD) が開発されている。

HMD の原理は空間位置センサからの情報に基づいて頭部の方向を感じし、それに応じた映像を小型ディスプレイに常に供給してやることである。そうすることによって、HMD を装着している人間は、広大な映像空間を見回しているのと同じような感じを得ることができる。

当研究室で利用できる HMD には次の2種類がある。1つ目は VPL 社が開発した EyePhone(図3.10, 表3.3) である。2つ目として、Visortron(図3.11) がある。

両者を比較すると、EyePhone は人工現実感用に開発されたデバイスであり広視野を確保しているが、使用している液晶ディスプレイの解像度が悪いのが難点で、装置自体も重く、装着感が良いとは言い難い。

それに対して、Visortron は人工現実感システム専用に開発されたデバイスではないので、視野は EyePhone ほど広くはないが、簡単に着脱可能で長時間装着していても EyePhone 程疲れない。

##### 空間位置センサ

人工現実感においては、人間の存在する実空間と仮想空間との間にインタラクションを持たせるために人間の動作を測定する必要性がある。

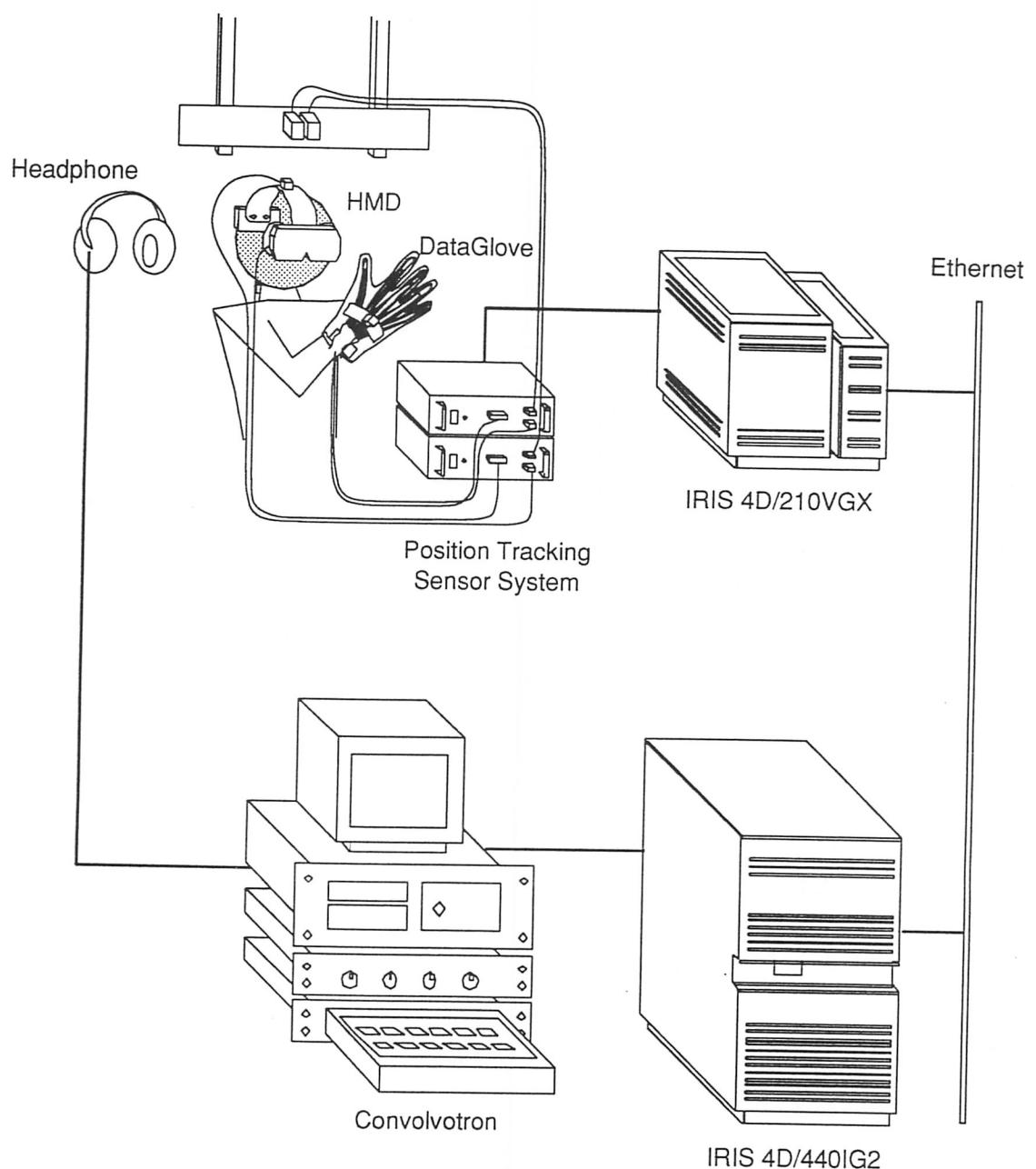


図 3.9: VR System



図 3.10: EyePhone の装着図

表 3.3: EyePhone の諸元

名称	数値	説明
画面間隔 (SCREENSEP)	0.875"	画面隙間間隔
画面幅 (SCREENWIDTH)	2.133"	片画面幅
画面高さ (SCREENUP)	0.860"	光学中心から画面上部までの距離
画面低さ (SCREENDN)	0.730"	光学中心から画面下部までの距離
画面距離 (SCREENDIST)	1.383"	眼球から画面までの距離
視野角 (垂直方向)	83 度 (片眼)	
視野角 (水平方向)	58 度 (片眼)	
解像度	360×240	



図 3.11: Visortron の装着図

この目的のセンサとして様々なタイプの空間位置センサが研究開発されているが、多自由度リンクのように身体を拘束することなく、非接触で空間内の位置を測定するために使用されるのが、磁気を使用した空間位置センサである。

空間にコイルを置いてそのコイルを電流  $I$  で励振すると、コイルの周囲の空間には次のような磁界が形成される。

$$\left. \begin{aligned} B_r &= a \cdot \frac{I \cos \theta}{r^3} \\ B_\theta &= a \cdot \frac{I \sin \theta}{2r^3} \end{aligned} \right\} \quad (3.5)$$

ここで、 $r$  はコイルからの距離、 $\theta$  はコイルの中心軸となる角であり、 $a$  は空間の透磁率などで決まる定数である。

この磁界中に別のコイルを置き、そのコイルに励磁電流を測定することによって空間中に置ける位置の情報を得ることができる。実際には、互いに直交した 3 つのコイルで磁界を時分割で作り、それを直交した 3 つのコ

イルで位置を感じて位置測定を行なうようになっている。

当研究室では、この磁気空間位置センサとして Polhemus 社製の Isotrak 空間位置センサ(図 3.12、表 3.4)、と Fastrak 空間位置センサ(図 3.13、表 3.5)が利用可能である。

Fastrak 空間位置センサは Isotrak 空間位置センサを改良したセンサであるので、サンプリング周波数や角度・位置精度ともに改善されている。

### Data Glove

前節で述べた、空間位置センサのように一点測定型のセンサで人間の形のような多点のデータを入力するのは効率的であるとはいえない。

そこで、手の動作のような複雑なデータを入力するために開発されたのが Data Glove である(図 3.14)。

Data Glove の原理は、薄手の手袋の指の部分に光ファイバを付けておいて、指が曲がった時の光ファイバの伝達ロスを測定することによって、逆に指の曲がり具合を測定する、という原理である。

ただ、光ファイバの特性が時間的に変化するので、時々キャリブレーションを行なう必要がある。

### 立体音響ディスプレイ

当研究室で利用可能な立体音響ディスプレイとしては、Convolvotron がある。Convolvotron に関しては、3.4.2で触れたのでそちらを参照のこと。

### グラフィックス・ワークステーション

当研究室で使用している、グラフィックス・ワークステーションは Silicon Graphics 社製のワークステーションである(表 3.6)。

ここで、表 3.6中の項目について簡単に述べておくことにする。CPU の性能を表す VAX MIPS 値は、Dhrystone Benchmark Test(Ver1.1+) を用いて実測した。FPU の性能を表す MFLOPS 値は、FLOPS Benchmark

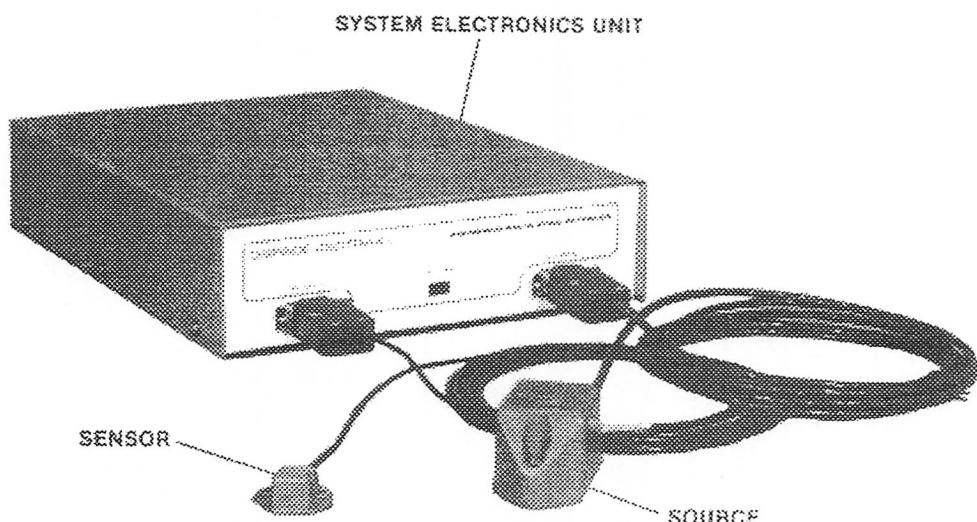


図 3.12: Isotrak 磁気空間位置センサ

表 3.4: Isotrak 磁気空間位置センサの諸元

	コントロールボックス	磁気ソース部	センサ部
長さ	12.25[inch]	2.4[inch]	0.9[inch]
幅	9.12[inch]	1.4[inch]	1.1[inch]
高さ	2.88[inch]	1.4[inch]	0.6[inch]
重量	96.0[oz.]	3.5[oz.]	0.8[oz.]
静的精度 (位置)		0.13[inch] RMS(ソースから 4~15[inch] の範囲) 0.25[inch] RMS(ソースから 30[inch] の場所)	
静的精度 (角度)		0.85[deg.] RMS(ソースから 30[inch]までの範囲)	
分解能 (角度)		0.35[deg.] RMS(ソースから 30[inch]までの範囲)	
データ出力		最高 60[Hz]	

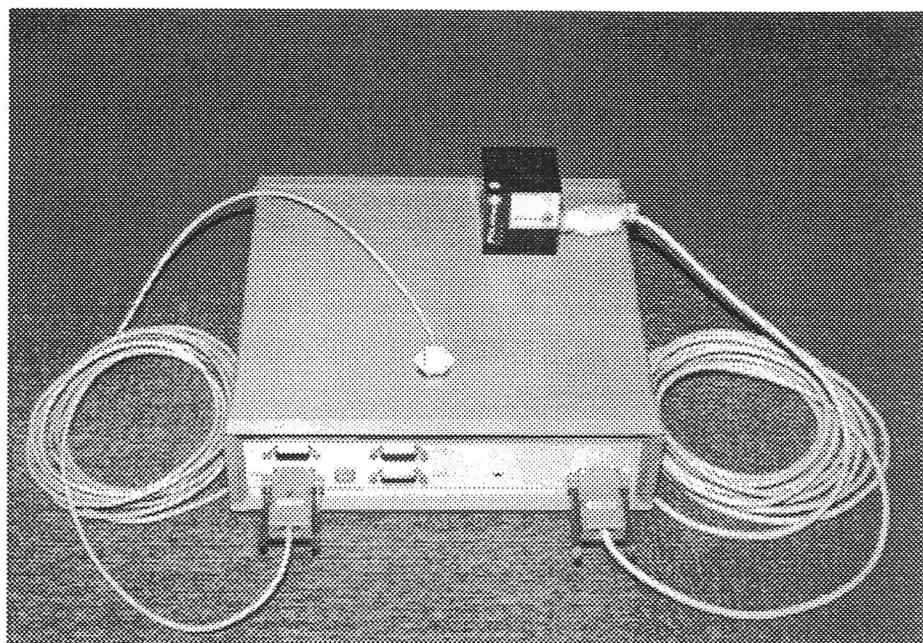


図 3.13: Fastrak 磁気空間位置センサ

表 3.5: Fastrak 磁気空間位置センサの諸元

	コントロールボックス	磁気ソース部	センサ部
長さ	11.38[inch]	2.1[inch]	1.1[inch]
幅	11.06[inch]	2.1[inch]	0.9[inch]
高さ	3.63[inch]	2.3[inch]	0.6[inch]
重量	5.0[lb.]	0.6[lb.]	0.6[oz.]
静的精度 (位置)	0.03[inch] RMS(ソースから 30[inch] の範囲)		
静的精度 (角度)	0.15[deg.] RMS(ソースから 30[inch] の範囲)		
分解能 (角度)	0.025[deg.] RMS(ソースから 30[inch] の範囲)		
データ出力	120[Hz](1 Receiver), 60[Hz](2 Receivers)		

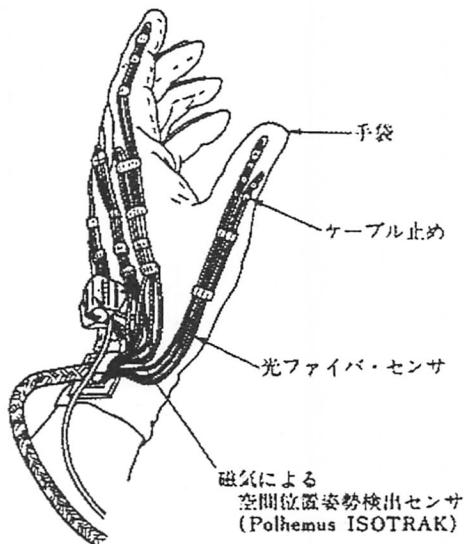


図 3.14: Data Glove(文献 [17] より引用)

TEST(Ver2.0)を用いて実測した。TexPix 値<sup>3</sup>と TexAA 値<sup>4</sup>は主にテクスチャマッピングの性能を表す。また、Tmech 値<sup>5</sup>は主にポリゴンの処理性能を表す[33]。

近年、ワークステーションの性能は RISC(Reduced Instruction Set Computer) 技術を採用した CPU(Central Processing Unit) の登場により、格段に向上してきた。

CPU の性能が向上するのと同時に、グラフィックスの性能も信じ難いほどに性能が向上した。特に、グラフィックス・ワークステーションの開発でトップの技術力を保持している米 Silicon Graphics 社製のワークステーションにおいて、ハードウェアのレベルで大容量のテクスチャマッピングをリアルタイムで可能な計算機 (IRIS Crimson/Reality Engine) が開発されてきてからは驚くほどのリアリティのある画像をリアルタイムで生成可能になった。

<sup>3</sup>RM ボード最大構成時のテクスチャマッピングを施したピクセルの描画速度

<sup>4</sup>50 ピクセル、三角形メッシュ、アンチエイリス処理後、ミップマップ、トライリニア補間、テクスチャマッピングを施した時の性能

<sup>5</sup>50 ピクセル、三角形メッシュ、24 ビット・フルカラー、光源処理なし、フラットシェーディング、Z バッファ処理、方向性なしの場合の性能

表 3.6: Graphics Workstation

名称	IRIS Crimson	IRIS 4D/440IG2	IRIS 4D/210VGX
CPU(個数)	MIPS R4000(1)	MIPS R3000(4)	MIPS R3000(1)
FPU(個数)	MIPS R4010(1)	MIPS R3010(4)	MIPS R3010(1)
Clock[MHz]	100(外部 50)	40	25
Memory[MB]	128	32	8
1 次キャッシュ	16[KB]	128[KB]	128[KB]
2 次キャッシュ	1[MB]	1[MB]	—
VAX MIPS	80.4	38.6(/CPU)	14.5
MFLOPS	15.5	9.9	4.2
Graphics System	Reality Engine	VGXT	VGX
TexPix	320M	50M	—
Tmesh	1.1M	1.1M	—
TexAA	600K	35K	—

### 3.5.2 人工現実感システムのソフトウェア

従来、当研究室では人工現実感用のアプリケーションを記述するために Vis-Age<sup>6</sup>という言語システムが開発されてきた。

この Vis-Age は以下のような基本的機能を提供する。

- オブジェクトの座標系
- オブジェクトの法則計算の管理
- カメラモデルの管理
- ハンドモデルの管理
- 行列計算等のユーティリティ

<sup>6</sup>開発者は当研究室博士課程 3 年の木島竜吾氏

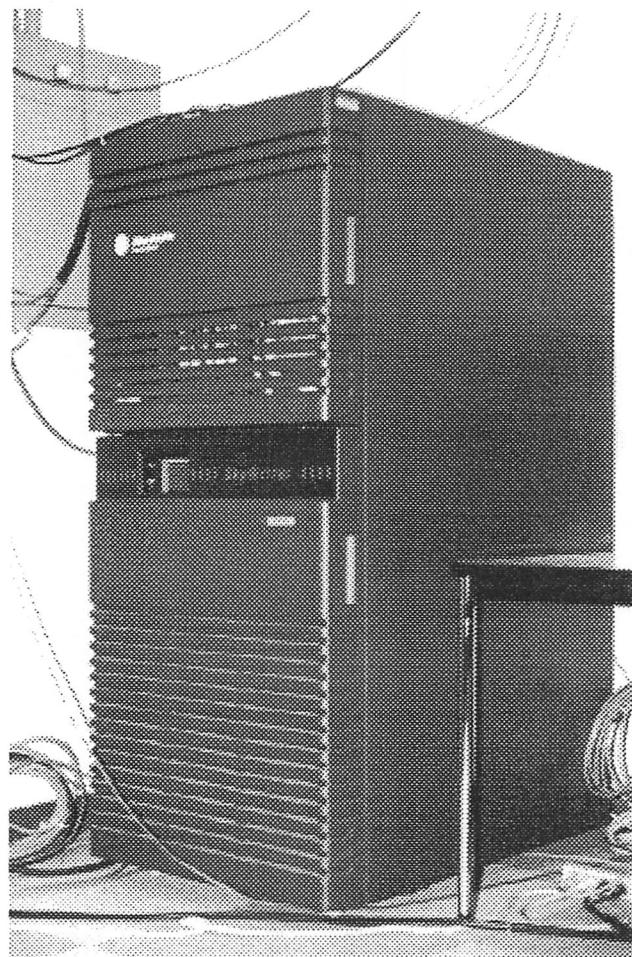


図 3.15: IRIS 4D/440IG2

従来のソフトウェア環境を階層 (Layer) 別に図示したのが図 3.16である。Vis-Age がかなり高度な機能まで提供するので、ユーザのアプリケーションは高次のプログラミングを行なえばいいという点で、画期的なシステムであるといえる。

しかし、現在当研究室の研究を行なっている人の Vis-Age の使い方を観察してみると、従来のシステムについての問題点が見えてくる。

1. Vis-Age の機能の中で、主に使用されているのは、空間位置センサや Data Glove に対するインターフェース部分であり、画面に対する描画

は自分で記述する場合が多い。

2. 外部のデバイスに対して直接駆動しているので、空間位置センサのようなデバイスを複数のアプリケーションで共有することができない。
3. 各種のデバイスがそれに接続されている計算機に拘束されてしまい、その資源を他の計算機から容易に利用することができない。
4. Vis-Age 自身が C 言語で記述されているので、X-Window System のような大規模なライブラリを構築する時の問題点、つまりライブラリの見通しが悪くなり、メンテナンス性が悪くなりがちである。

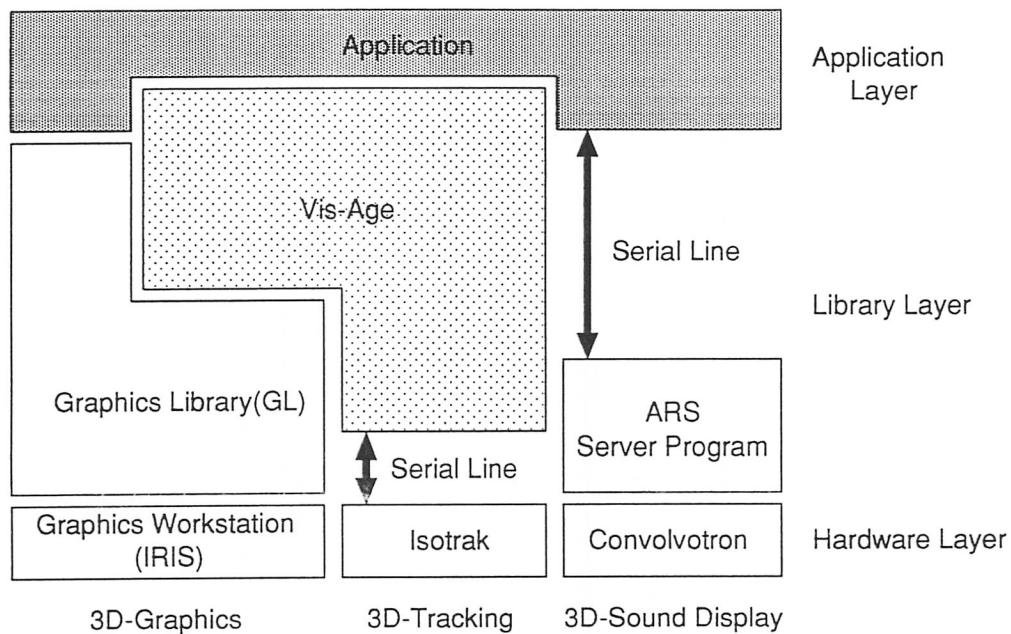


図 3.16: 従来のソフトウェア環境

このような問題点に対する、これらの問題を解決するようなシステムとして図 3.17のようなシステムを開発した。

今回、どのようなシステムを開発して前述のような問題を解決したかを次に詳しく述べることにする。

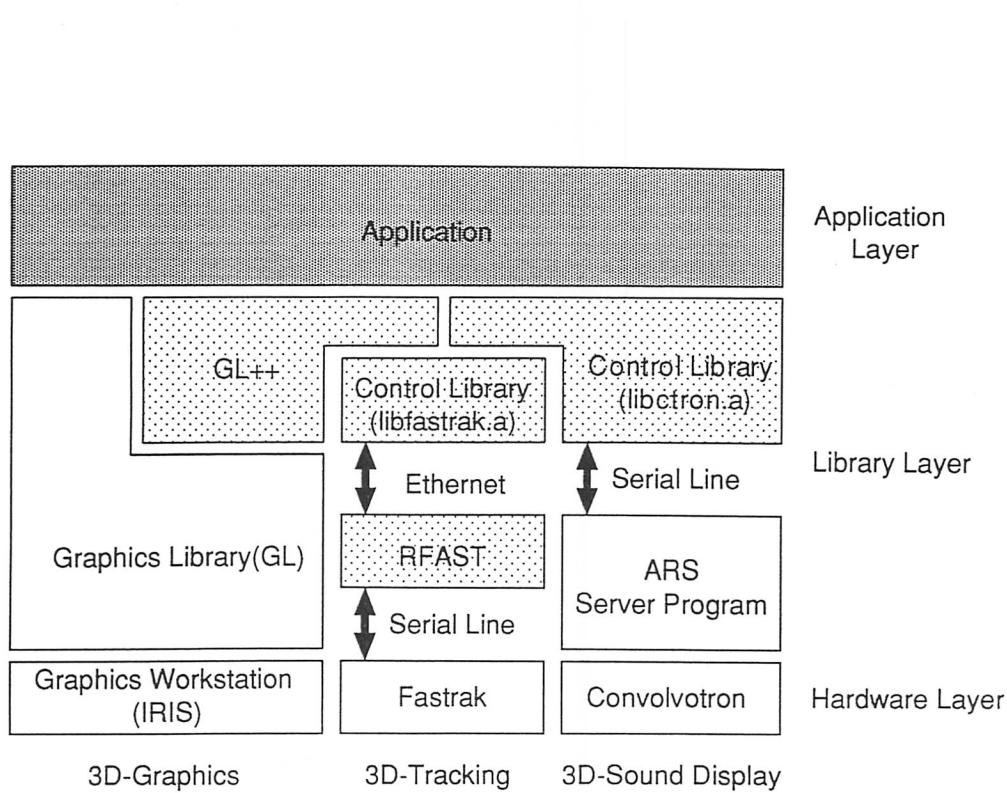


図 3.17: 新しく開発したソフトウェア環境

### GL++

GL++ は、IRIS の Graphics Library である GL をオブジェクト指向言語である C++ に対応するようにカプセル化したライブラリである。処理系としては IRIS 純正の C++ 処理系ではなく、GNU<sup>7</sup>の C++ の処理系<sup>8</sup>を用いている。

このクラスライブラリは大きく分けて、GL のプリミティブな機能をサポートするための基本クラス群と様々な立体オブジェクト（球体、直方体、平面など）を階層化して記述するための物体クラス群から構成されている。

基本クラス群は、次のクラスから構成される。

- 色に関するクラス (**GLColor**)
- 行列に関するクラス (**GLMatrix**)

<sup>7</sup>Gnu is Not Unix

<sup>8</sup>現在、GNU-C Ver2.5.7 & Libg++ Ver 2.5.3

- ベクトルに関するクラス (GLVector)
- 物体の性質に関するクラス (GLMaterial)
- 光源に関するクラス (GLLight)
- 光源モデルに関するクラス (GLLmodel)
- ウィンドウに関するクラス (GLWindow)

物体クラス群は、次のクラスから構成されている。

- すべての物体クラス群の仮想基底クラス (GLObject)
- 実体を持たないクラス (GL\_VOID)
- 平面クラス (GL\_Plane)
- 直方体クラス (GL\_Box)
- 球体クラス (GL\_Ball)

物体のインスタンスはすべて、演算子 (<<) を用いることによって階層的につなげることができる。

### RSLIB(Ver2.0)

このライブラリは、シリアル回線 (RS-232C) を TCP/IP のソケットを利用したライブラリ (libsocket.a) を利用することで、任意のマシンの RS-232C を利用するが出来るようにしたライブラリである。また、RS-232C ポートの排他制御を行なっているので、このライブラリを利用しているアプリケーションが重複してポートをオープンすることを防ぐ事ができる。

### RFAST & Interface Library

従来のシステムでは、Vis-Age が直接、Isotrak 空間位置センサとのデータの送受信を行なっていたので、他の計算機から利用することもできないし、複数のアプリケーションで共有することもできなかった。

それに対して、新しいシステムでは Fastrak 空間位置センサをサーバ (RFAST) を一段置いて接続することによって、ネットワークにつながった任意の計算機から複数のアプリケーションがデータを共有する事が出来るようになった。

図 3.18は、空間位置センサの位置をリアルタイムで監視できるように開発した rfmonitor の画面である。このように、実験を行なっている最中に、正常に動作しているかどうかを確認できるようになっている。

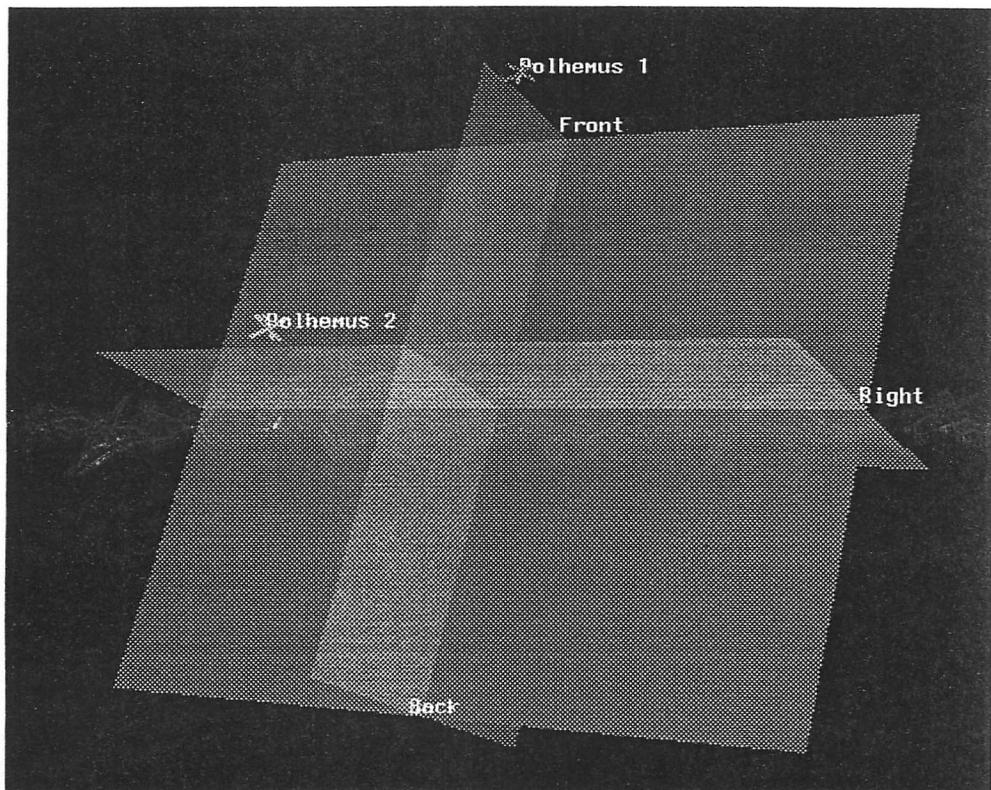


図 3.18: rfmonitor の画面

### Convolvotron Control Library

このライブラリは、Convolvotron のサーバ(ARS Server)をワークステーション上のアプリケーションが制御するためのライブラリであり、前述の RSLIB を使用することによってネットワークに接続されているどの計算機からでも利用できるようになっている。

## 第 4 章

音響的臨場感の記録・再生

## 4.1 はじめに

この章では、音響的臨場感を記録・再生するためのシステムに関して考察し、その基礎的な技術を実験において確認する。そして、視覚的な臨場感と音響的な臨場感を組み合わせたシステムとしての動作を確認する。

2.4.3で議論したように、音響的な臨場感を記録・再生するためには図4.1のように音響的な情報を音源の位置情報(Position Data)と音源の音情報(Sound Data)に分離して記録しておいて、再生する段階で立体音響ディスプレイで音響的な臨場感を再合成する必要がある。

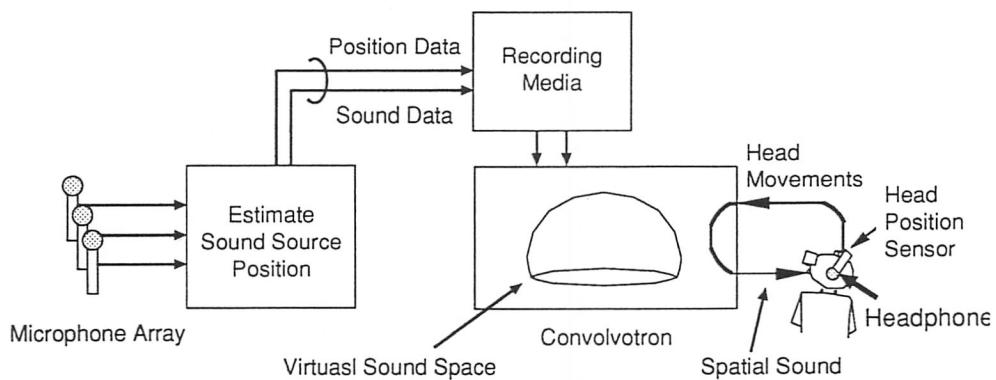


図4.1: 音響的臨場感記録・再生システムの概念図

次に、このシステムを構成する構成要素に関して考察することにする。

## 4.2 Convolvertron

今回研究に使用した立体音響ディスプレイ Convolvertron(3.4.2) の立体音響ディスプレイとしての性能に関して実験を行なった。

### 4.2.1 頭部音響伝達関数(HRTF)

Convolvertronで使用している頭部音響伝達関数(sdo\_mp50.dat)の周波数成分を解析した結果を次に示す。以下の伝達関数はすべて右耳の伝達関数を解析した結果である。

図4.2は、方位角方向の自由空間伝達関数の振幅成分を描いた図であり、図4.3は、その自由空間伝達関数を正面方向の伝達関数 (Azimuth 0[deg.], Elevation 0[deg.]) で正規化した単耳伝達関数を表している。

また、図4.4は、Azimuth 90[deg.]における仰角方向の自由空間伝達関数を描いた図であり、図4.5は、同様に単耳伝達関数表現したものである。

Convolvotronに採用されている頭部音響伝達関数は、実験の結果個人差の現れにくい伝達関数だと思われるが、それでも頭部音響伝達関数の個人差の影響は現れるようで、Convolvotronの立体音響を試聴した人の感想をまとめてみると、

- 全体的に音像が前方に定位しにくい。
- 全体的に音像の定位が水平面よりも上に生じる傾向がある。
- 頭の回りに音像を周回させた場合、音像が後方から前方へ移動する時に、頭内定位のような知覚を覚えたり、反対方向に音像が周回してしまったりする。

といった事が報告された。

個人の頭部音響伝達関数がどの程度、個人差があるのかを表した図が図4.6である。この図を見れば明らかのように、かなりはっきりした個人差が頭部音響伝達関数に認められる事が分かる。

#### 4.2.2 Convolvotronの定位精度

Convolvotronの音像の定位精度を実音源と比較した研究には、文献[27]がある。図4.7は、その文献から引用した定位精度の表であるが、この結果からは、自分の頭部音響伝達関数を使用している限り無響室での実音源と同程度の(少々悪いが)定位精度が得られている。

この研究の追試をする意味で、音像定位の実験を行なってみた。

#### 実験方法

ヘッドフォンを装着し椅子に座った被験者に対して、Convolvotronにより音像を水平面内 (Elevation 0[deg.]) に提示する。音源方向の指示方法

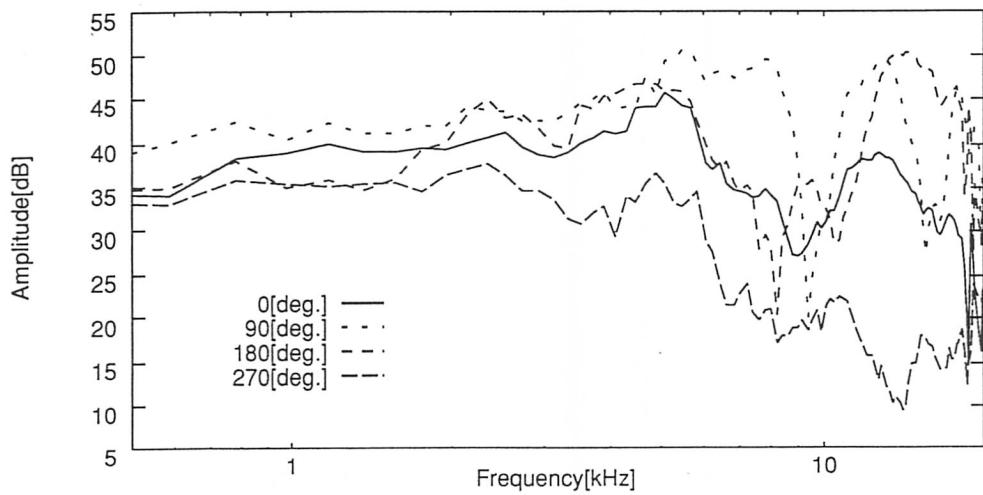


図 4.2: Azimuth 方向の自由空間伝達関数(右耳,Elevation 0[deg.])

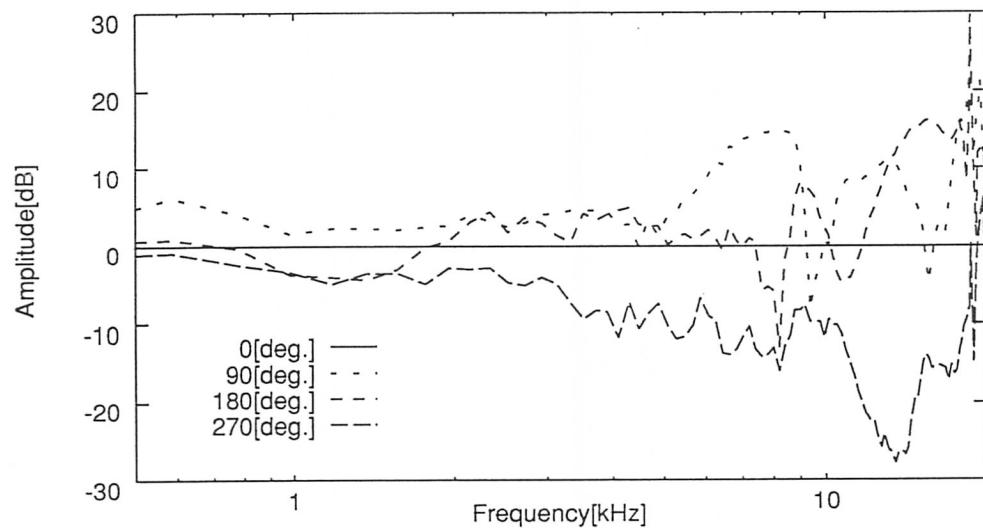


図 4.3: Azimuth 方向の単耳伝達関数(右耳,Elevation 0[deg.])

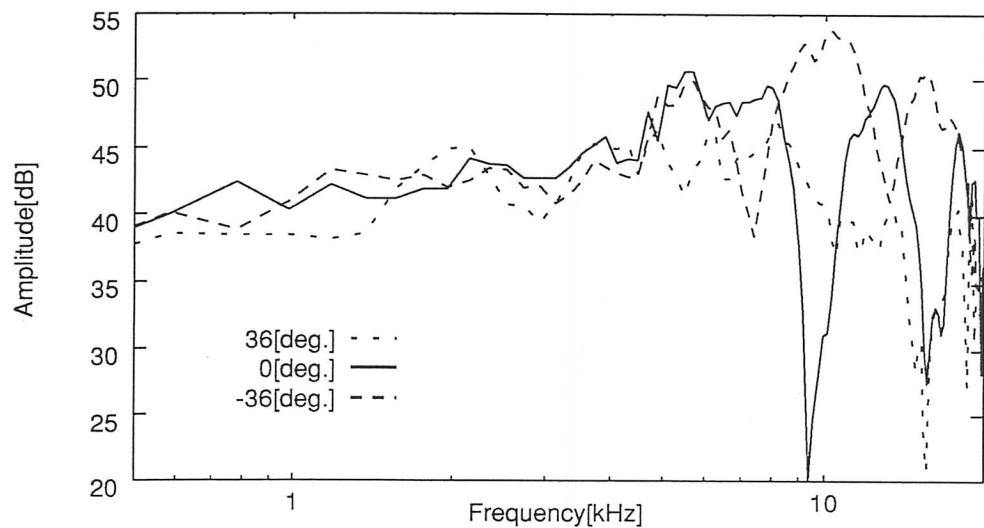


図 4.4: Elevation 方向の自由空間伝達関数(右耳,Azimuth 90[deg.])

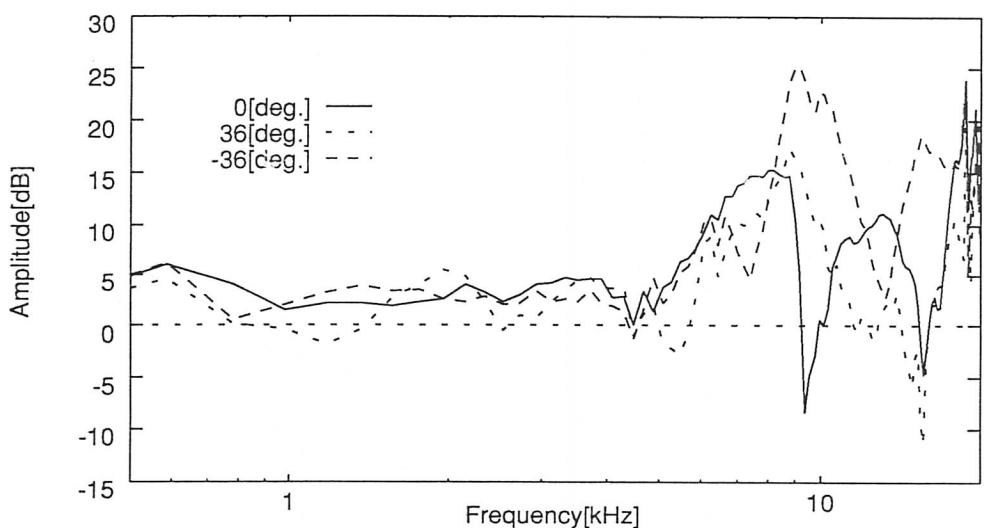


図 4.5: Elevation 方向の単耳伝達関数(右耳,Azimuth 90[deg.])

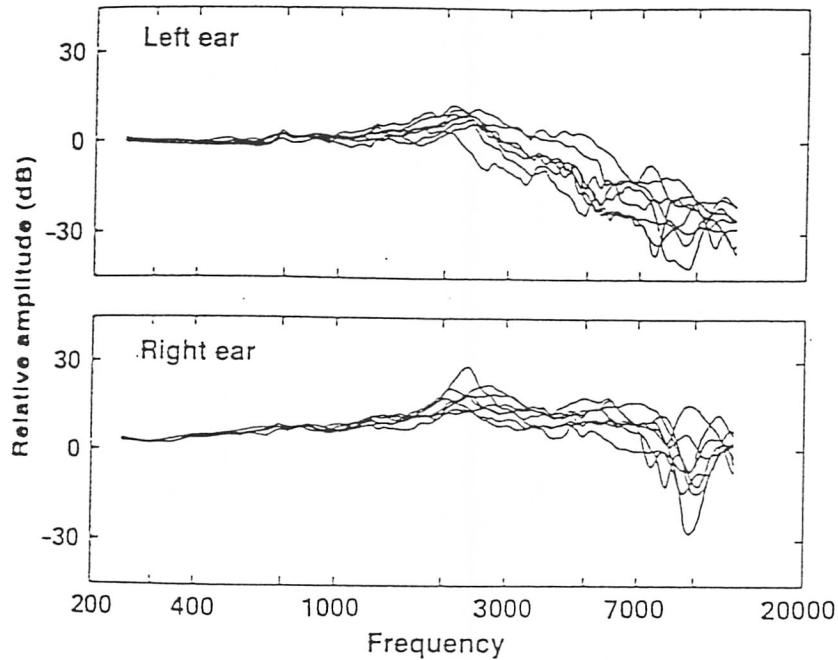


図 4.6: 頭部音響伝達関数の個人差 (文献 [27] から引用)

は、プラスチック板に Fastrak 空間位置センサを取り付け、膝の上でその板を回転させることによって音像方向を指示してもらった。

実験は、1 音源の場合と 2 音源の場合について行ない、1 音源の場合は音響信号としてシンセサイザによる合成音を、2 音源の場合はシンセサイザによる合成音と CD による音楽を提示した。

被験者は 4 人である。実験結果の表の内修正ありとは、被験者が音像を前後間違えて回答したとみなして音像の角度を補正した場合の結果である。実験結果の図に関しては、すべて修正した実験結果である。

### 実験結果

まず、1 音源の場合について考察する。表 4.1 から、前後誤りを修正した結果を見ると、4 人ともほぼ同じ定位精度 (約 20[deg.]) で、標準偏差も

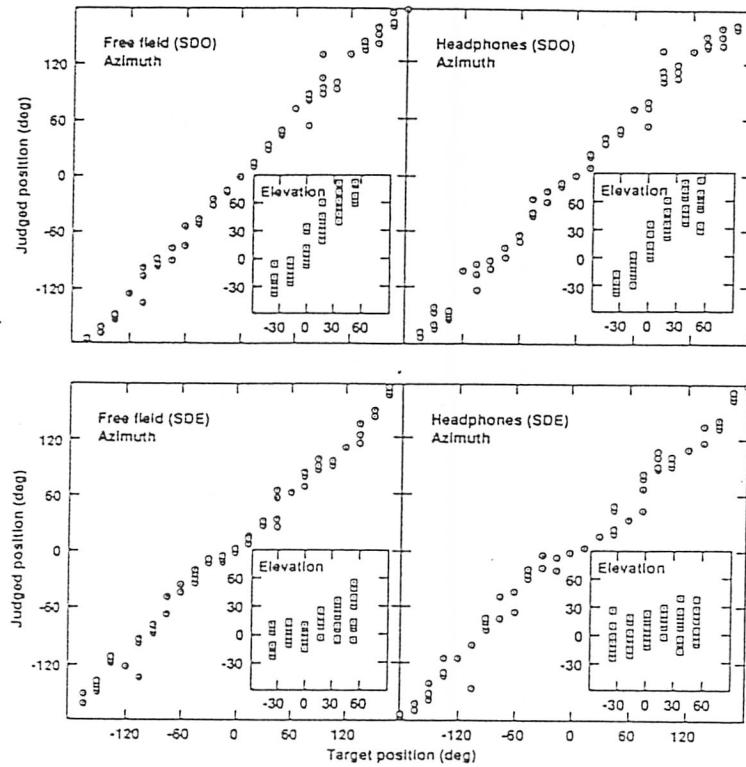


図 4.7: Convoltotron と実音源の音像定位精度 (文献 [27] から引用)

ほぼ同程度であり、明瞭な個人差は認められなかった。文献 [27] での値と比較すると、2倍程度であるが、これは文献 [27] での測定値が自分自身の頭部音響伝達関数を用いているのに対して、この実験では自分のものではない頭部音響伝達関数を使っているためと思われる。

次に 2 音源の場合について考察する。表 4.2 と表 4.3 を比較してみると、2 つの音源による定位精度の違いは見られない。

1 音源の場合と 2 音源の場合では Convoltotron が FIR として使用できるタップ数が少なくなるので音像の定位精度が悪くなることが予想されたが、今回の実験結果からは少し定位精度が悪化したが、数 % であり明瞭な差は認められなかった。

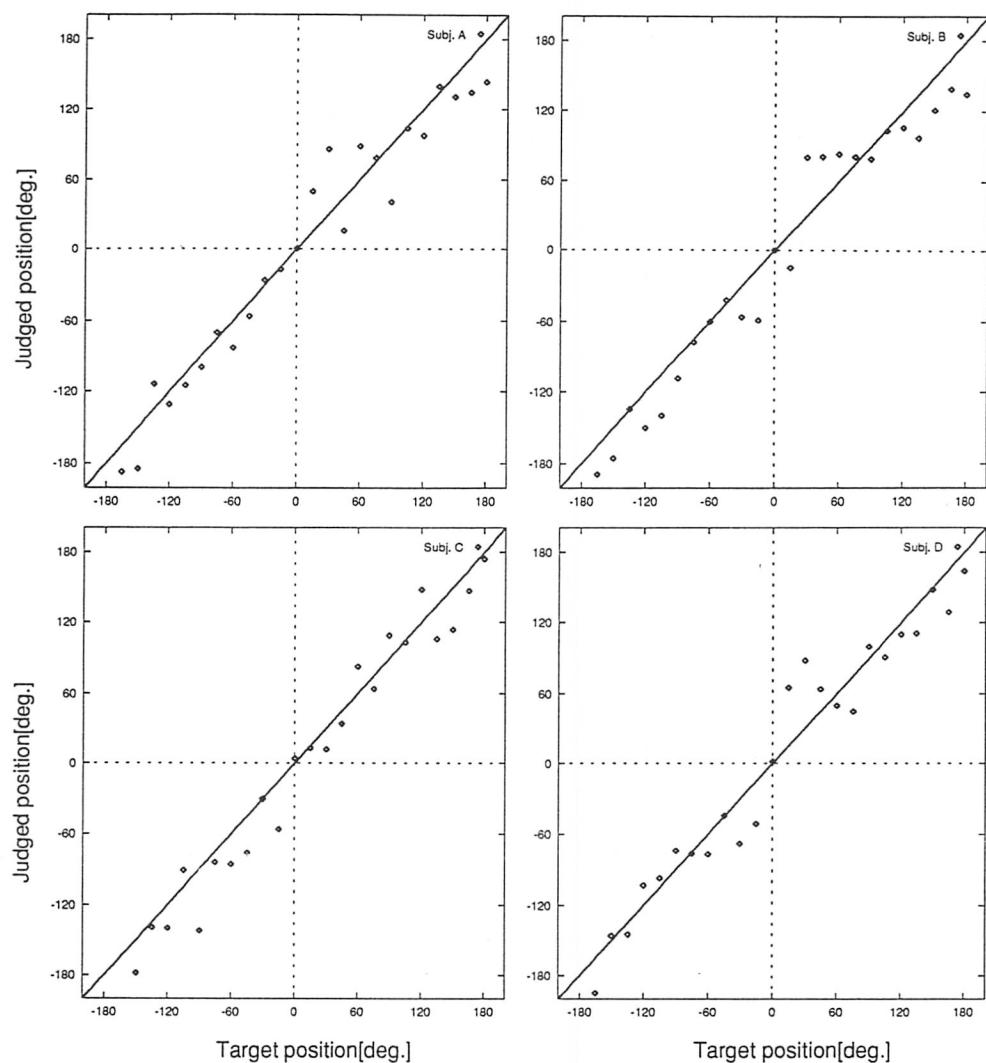


図 4.8: Convolvertron の定位精度 (1 音源)

表 4.1: Convoltotron の定位精度 (1 音源)

	平均 [deg.]		標準偏差 [deg.]		前後誤り率 [%]
	修正なし	修正あり	修正なし	修正あり	
Subject A	29.8	19.5	40.9	24.6	21
Subject B	35.2	21.7	45.5	26.5	25
Subject C	52.2	19.6	72.2	23.9	38
Subject D	39.9	19.1	53.9	24.5	38
平均	39.3	20.0	53.1	24.9	30

表 4.2: Convoltotron の定位精度 (2 音源 (1))

	平均 [deg.]		標準偏差 [deg.]		前後誤り率 [%]
	修正なし	修正あり	修正なし	修正あり	
Subject A	28.7	15.9	46.8	21.2	21
Subject B	33.4	24.9	42.6	30.9	29
Subject C	56.1	23.8	71.5	29.9	38
Subject D	46.1	22.5	65.3	28.3	38
平均	41.1	21.8	56.6	27.6	31

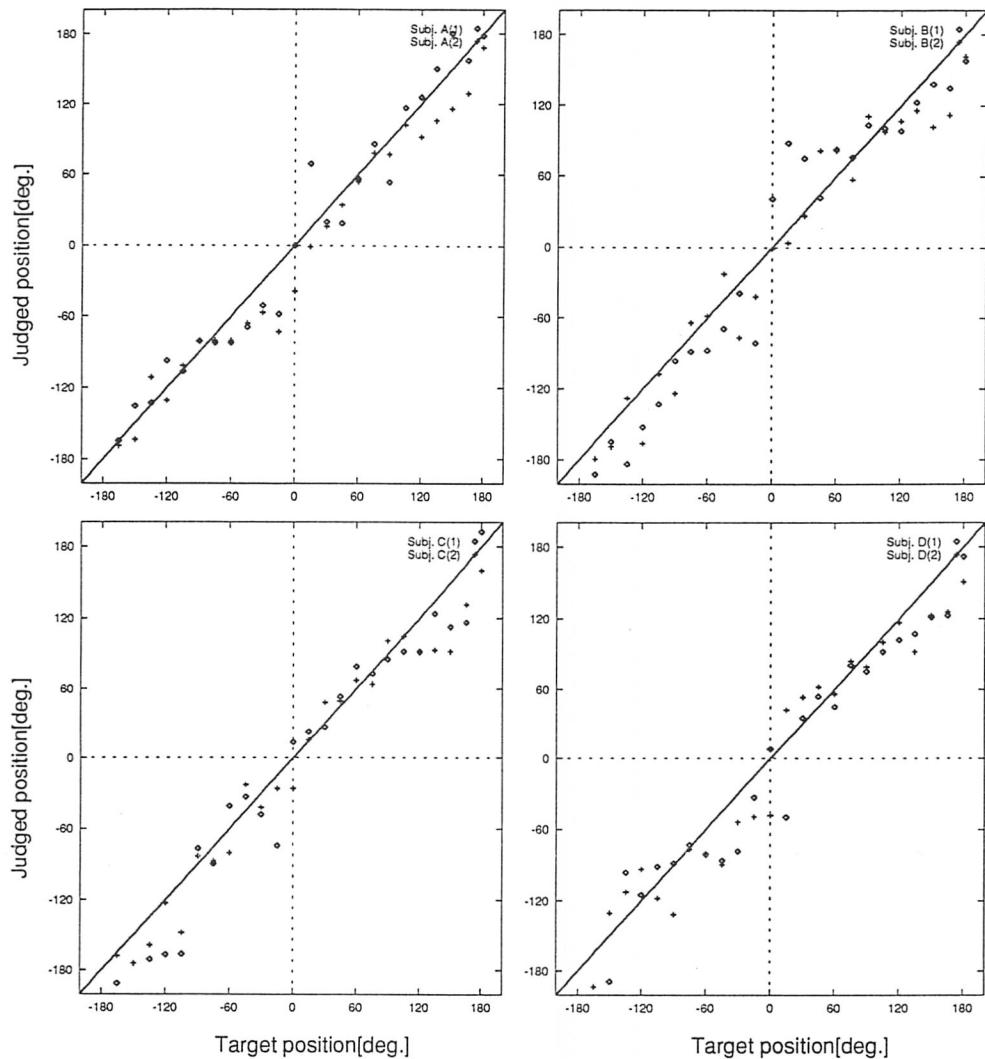


図 4.9: Convolvertron の定位精度 (2 音源)

表 4.3: Convoltotron の定位精度 (2 音源 (2))

	平均 [deg.]		標準偏差 [deg.]		前後誤り率 [%]
	修正なし	修正あり	修正なし	修正あり	
Subject A	27.6	19.3	34.4	23.4	25
Subject B	50.9	21.0	69.7	26.0	46
Subject C	52.6	18.4	71.3	23.4	50
Subject D	40.0	23.3	52.2	26.9	33
平均	42.8	20.5	56.9	24.9	39

### 4.3 音響的臨場感の記録

音像定位の重要なキーとして ITD(Interaural Time Difference) があることは、すでに述べたが、音源の位置を推定するのに複数のマイクロフォンの間の音響信号の到達時間差を利用することが可能である。

ここで、図 4.10 のように X 軸上にマイクロフォンを 2 つ配置したと考えてみる。座標は両マイクロフォン間の長さを 1 として正規化してあるとする。音源からの音響情報の到達時間差を  $t_D$  とすると、 $t_D$  は次のような式で与えられる。

$$\begin{aligned}
 t_D &= \sqrt{\left(x - \frac{1}{2}\right)^2 + y^2} - \sqrt{\left(x + \frac{1}{2}\right)^2 + y^2} \\
 &= \sqrt{r^2 - r \cos \theta + \frac{1}{4}} - \sqrt{r^2 + r \cos \theta + \frac{1}{4}} \\
 &= r \left( \sqrt{1 - \frac{1}{r} \cos \theta + \frac{1}{4r^2}} - \sqrt{1 + \frac{1}{r} \cos \theta + \frac{1}{4r^2}} \right)
 \end{aligned} \tag{4.1}$$

$r$  が十分大きく、根号の中が 1 次近似できるとすれば、

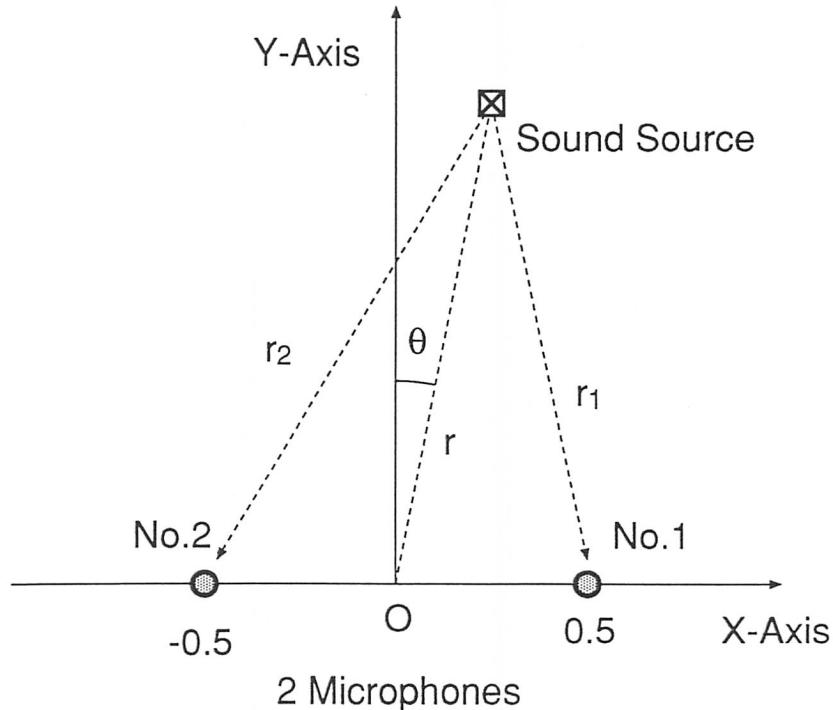


図 4.10: Microphone Array(2 Microphones)

$$\begin{aligned}
 t_D &\simeq r \left\{ \left( 1 - \frac{1}{2r} \cos \theta \right) - \left( 1 + \frac{1}{2r} \cos \theta \right) \right\} \\
 &= r \frac{1}{r} \cos \theta \\
 &= \cos \theta
 \end{aligned} \tag{4.2}$$

となって、 $r$  に依存しなくなる。

図 4.11は  $t_D$  の振舞いを厳密に計算した例で、図 4.12は  $\theta$  を固定して  $r$  に関して計算した例である。音源の距離が、マイクロフォン間の距離の 2 倍を越えれば  $t_D$  は、 $\theta$  のみに依存すると近似して良いことが分かる。

つまり、音源の距離がマイクロフォンの距離の 2 倍を越えれば、(前後の区別はつかなくとも) 音源の方向を感知することは可能である。

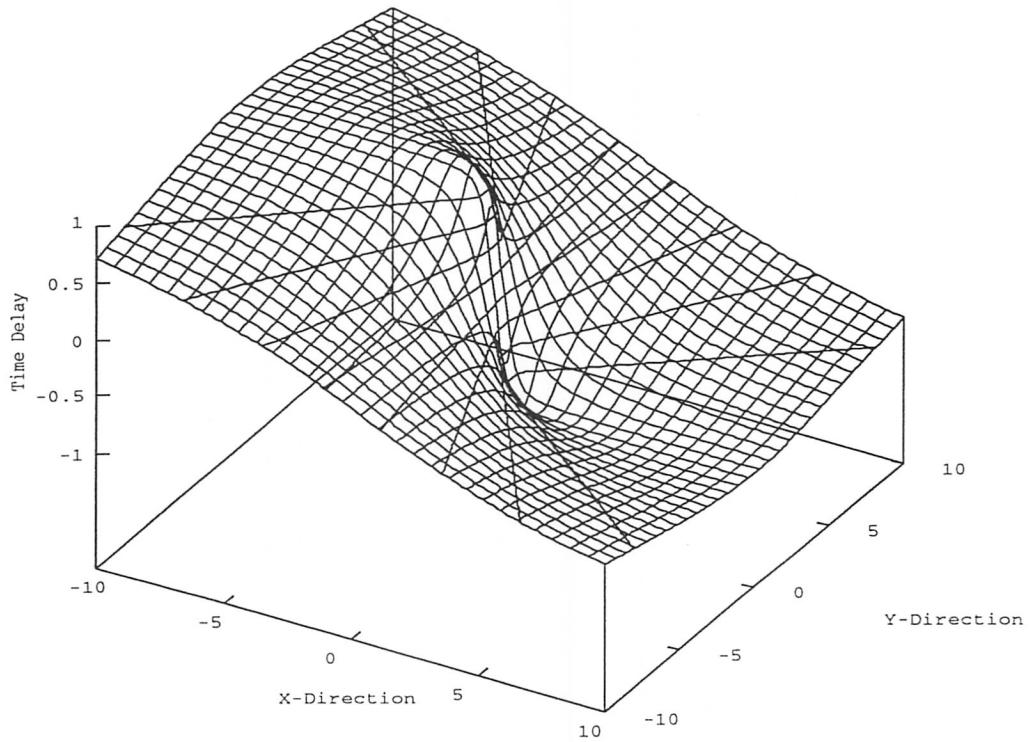


図 4.11: Time Delay

#### 4.3.1 2点法による音源位置推定

ここでは、前節で述べた理論を基に2点で音響信号を測定してそれをもとに音源位置を推定した実験として一昨年の卒業研究[20]を考察する。

#### 実験方法

音をサンプリングする装置としては、実際にサンプリング処理を行なうコンピュータにはApple社製のMacintosh II(以下Macと省略する)を採用し、外界からの音をMacに取り込む、マイクの役目をするハードウェアとして、Farallon Computing社のMacRecorderを2台、またMac上でサンプリングを行なうソフトウェアとして、MacRecorderに付属しているSoundEditというツールを使用した。

MacRecorderとSoundEditを併用した時のサンプリングレートは、

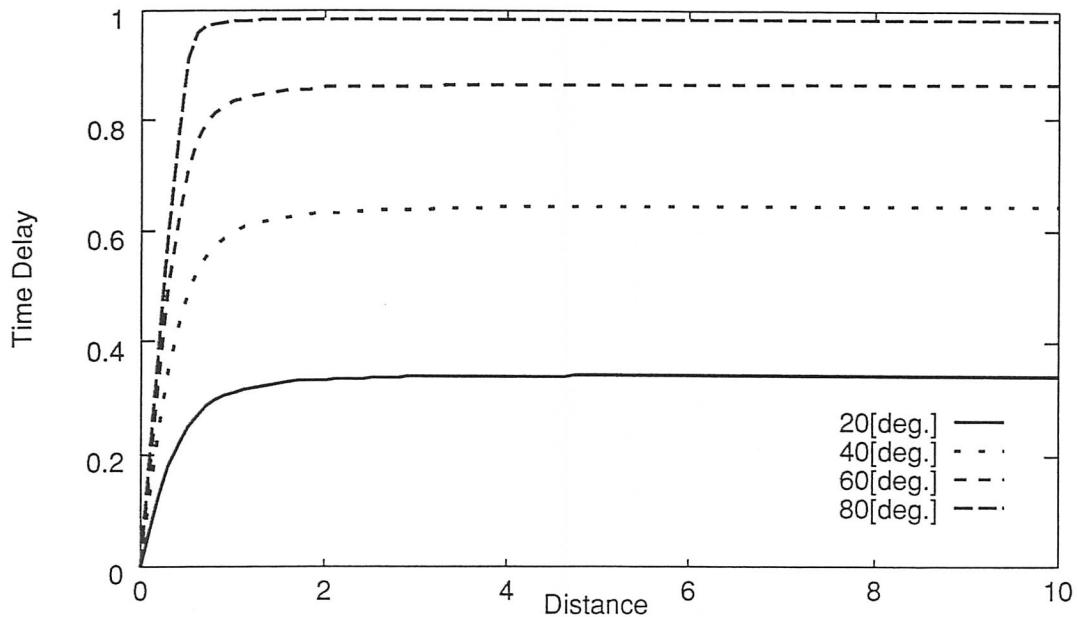


図 4.12: Time Delay

5[kHz], 7[kHz], 11[kHz], 22[kHz] の 4 通りが選べ、MacRecorder を 2 台接続すればモノラル・ステレオの両方のサンプリングが可能である。なお、サンプリングの量子化数は 8 ビットである。

この実験は、22[kHz] ステレオサンプリングで行なった。2 個の MacRecorder の間隔は約 20[cm] 離して配置した。これは、左右の MacRecorder に入力される入力信号に違いをつけ、これをもとに両耳間差などを調べるためにある。この間隔 20cm という設定の元は、音像の耳間変移が、音源から両耳への経路差が約 21[cm](位相差にして約 630[μs]) までは線形に増加することから、この数値に近い条件に設定したものである。

2 台の MacRecorder は、図 4.13 の座標系でいうと、 $(x, y) = (0.1, 0)$  と  $(x, y) = (-0.1, 0)$  の 2箇所に配置した。

今回の実験を実施した場所の、座標系つきの簡単な見取図を図 4.13 に示す。この部屋は、無響室ではなく、また図 4.13 を見れば分かるように、コンピュータ・ワークステーション・ページプリンタといった、常時ノイズを発生させている情報機器が存在し、また各机の上にも、実は様々なもの

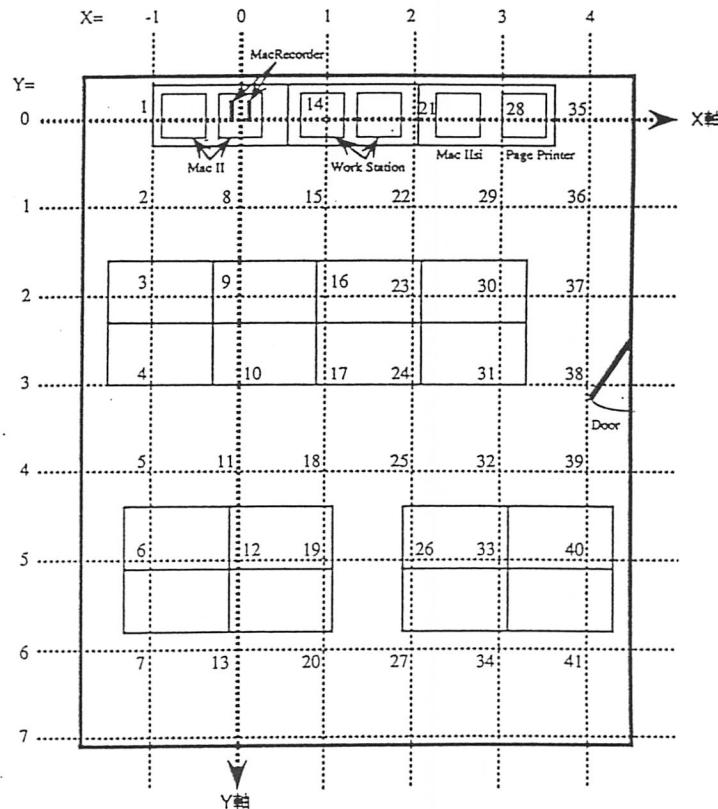


図4.13: 今回のサンプリング実験を行なった部屋の見取図

が置かれている。

音響信号の提示方法であるが、音響信号にはあらかじめ作成された合成音響信号を用い、図4.13のy軸に平行な直線 $x = -2 \sim 4$ と、x軸に平行な直線 $y = 0 \sim 6$ の交差した点（ただし原点を除く）あわせて41点において提示した（音響信号提示の対象となった41点には、図のように通し番号をつけてある）。

### 実験結果

実験結果を次に示す。図4.14は、両時間時間差としての情報と音源の距離を既知として音源位置を推定した結果である。

その次の図4.15は、音源の距離を音の音圧レベルから推定した実験結果

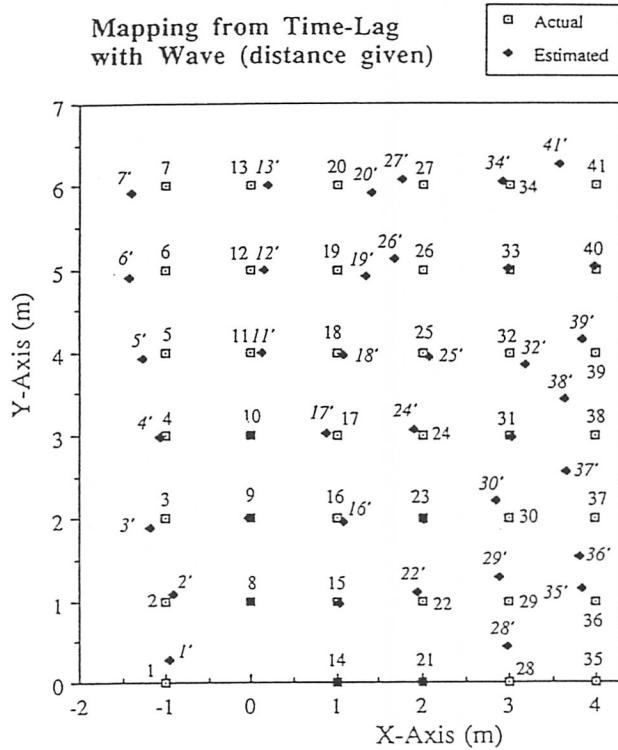


図 4.14: 両時間耳間差 + 既知の音源距離による推定

である。

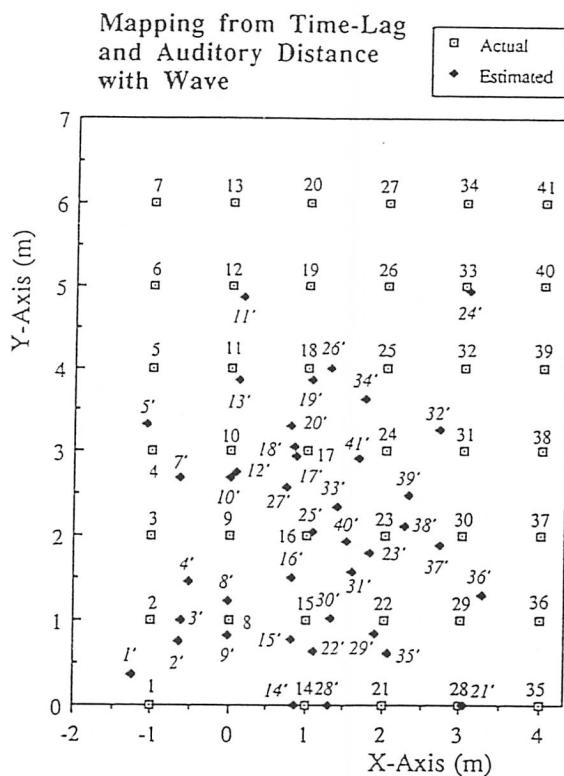
音圧レベルによる音源距離の推定のモデルとして、次のようなモデルを採用した。

左右の MacRecorder に入力される時間変化波形を、各々 FFT により周波数領域のパワースペクトルに変換して、今度は左右のパワースペクトルを合成して、その最もレベルの高いピーグレベルを検出し、音源のパワー レベルと比較して、これをもとにして音像距離を算出した。

ここで、音源のパワー レベル  $L_w$  の算出は、無響室・半無響室の場合は、次の式から計算できる。

$$L_w = \bar{L}_p + 10 \log \frac{S}{S_0} - 10 \log \left[ \sqrt{\frac{293}{273 + \theta}} \times \left( \frac{P}{1000} \right) \right] \quad (4.3)$$

ただし、 $S = 4\pi r^2 [\text{m}^2]$ (無響室の場合。半無響室の場合は $S = 2\pi r^2$ )、 $S_0 = 1 [\text{m}^2]$ 、 $\theta$ は室温(摂氏温度)、 $P$ は大気圧[hPa]、 $\bar{L}_p$ は複数の同一距離点における音圧レベルの平均値である。この実験では、実験室を半無響室として、また音波を半球面波と仮定して実験を進めた。また簡単のため、室温 $\theta = 20[^{\circ}\text{C}]$ 、大気圧 $P = 1000[\text{hPa}]$ とした。この式を用いて求めた、今回使用した音源のパワーレベルは、騒音測定技術を利用して測定した結果、約83[dB]と算出された。



### 4.3.2 3点法による音源位置推定

4.3.1で述べたように、2点での音源位置推定は困難であるが、3点での音響信号をもとに音源位置推定を行なえはどうであろうか。

#### 実験方法

図4.16のような実験装置で、1,2,3[m] の距離で鳴らした音源の位置を推定する。

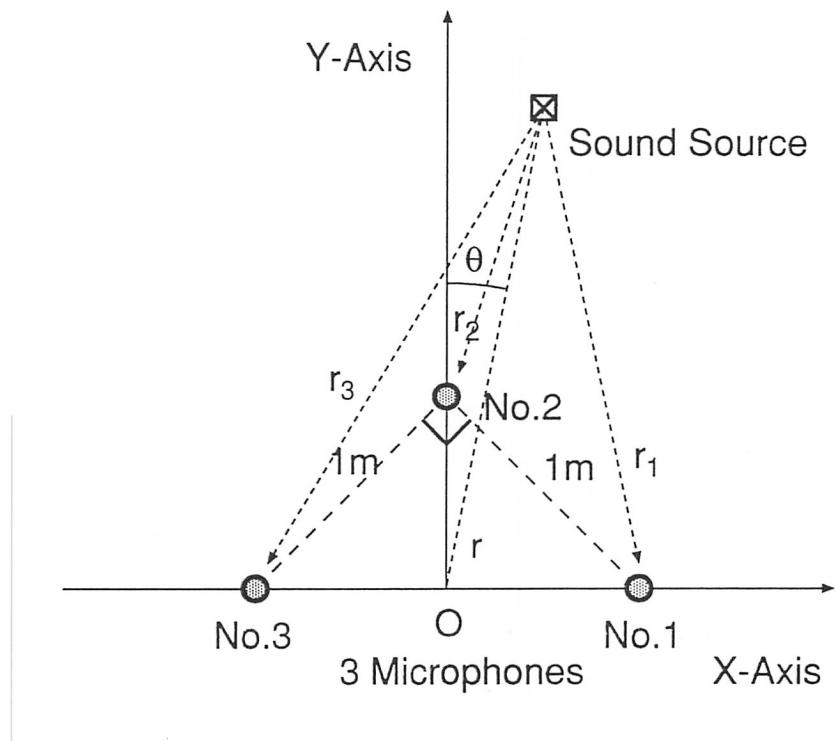


図4.16: Microphone Array(3 Microphones)

1番と2番のマイクロフォン間の音響信号の時間差  $t_{12}$  と、2番と3番のマイクロフォン間の音響信号の時間差  $t_{32}$  は次のような式で与えられる。

$$t_{12} = \sqrt{\left(x - \sqrt{\frac{1}{2}}\right)^2 + y^2} - \sqrt{x^2 + \left(y - \sqrt{\frac{1}{2}}\right)^2} \quad (4.4)$$

$$t_{32} = \sqrt{\left(x + \sqrt{\frac{1}{2}}\right)^2 + y^2} - \sqrt{x^2 + \left(y - \sqrt{\frac{1}{2}}\right)^2} \quad (4.5)$$

この  $t_{12}$  と  $t_{32}$  を図示したのが、図 4.17 である。この図から分かるように、 $t_{12}$  と  $t_{32}$  の 2 つを測定することによって、音源の位置を 1 箇所に限定して推定することが可能である。

音源としては、“あ”という音声を用いた。マイクロフォンからの音響信号は、一度アナログ 4 チャンネル MTR(Multi-Track Recorder) (図 4.18) で記録し、2 チャンネル毎に IBM-PC/AT の ISA バスに接続してある Sound Blaster 16ASP でサンプリング (44.1[kHz], 8bit) し、データをワークステーション上に転送し解析を行なった。

### 実験結果

マイクロフォンで測定された波形の一例を図 4.19 に示す。音響情報の時間差の測定方法として音の ON SET 部分(音響信号の立ち上がり部分)における時間差を、相互相関関数値によって評価した。

相互相関関数値は次のような式で計算することができる。

$$C_{mut}(\tau) = \int_{-T}^{T} R(t)L(t - \tau) dt \quad (4.6)$$

相互相関関数値を計算した一例を図 4.20 に示す。この例では、95 付近に相関関数値のピークが出ている。この図の横軸はサンプリング時間単位で量子化されているので、95 は実時間にすると  $95 \div 44100 \times 1000 = 2.15[\text{ms}]$  となる。音速を 340[m/s] とすると、1[m] の距離はサンプリング時間に換算して 130 に相当する。

このようにして、マイクロフォン間の時間差から音源位置を推定した結果を図 4.21(音源距離  $r=1[\text{m}]$ )、図 4.22( $r=2[\text{m}]$ )、図 4.23( $r=3[\text{m}]$ ) に図示する。

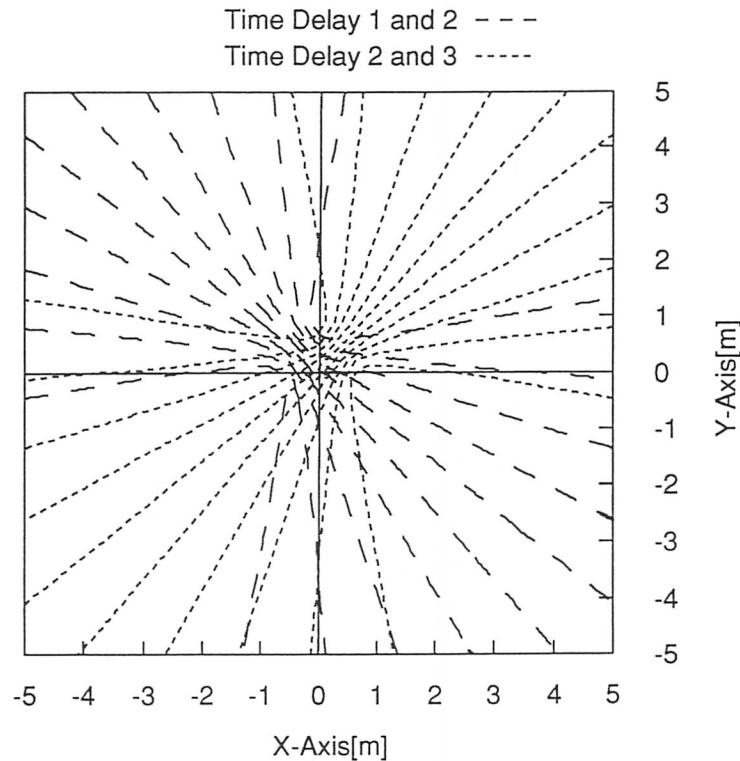


図 4.17: 時間差の分布

これらの図から分かるように音源距離に関してはかなりの誤差が生じているが、音源方向に関しては、ほぼ正確に推定することに成功している。これは音響信号の到達時間差による測定方法が方向よりは、距離の方が誤差に対してセンシティブなためである。

音源の方向の推定角度に関する誤差をまとめたのが、図 4.24 と表 4.4 である。Azimuth 方向が 90[deg.] 近辺で誤差が増加しているが、それでも最大誤差で 7[deg.]、平均誤差で 2[deg.] は、立体音響ディスプレイの提示誤差 20[deg.] に比べて十分小さいので問題はないと思われる。

また、音源の距離情報に関しては、ある意味で音源の音響情報自身に音のレベルや反響音の情報として含まれているので、今回の実験のようなシステムで音源の方向情報を測定することによって音響的臨場感記録・再生システムを組むことができると考えられる。



図 4.18: Analog 4channel MTR(Multi-Track Recorder)

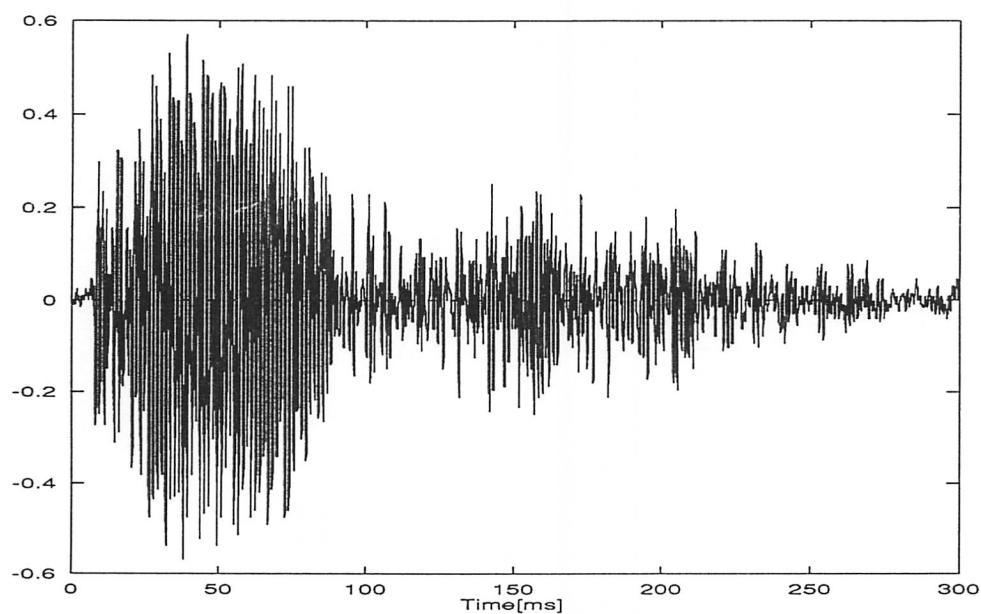


図 4.19: 測定した波形の例

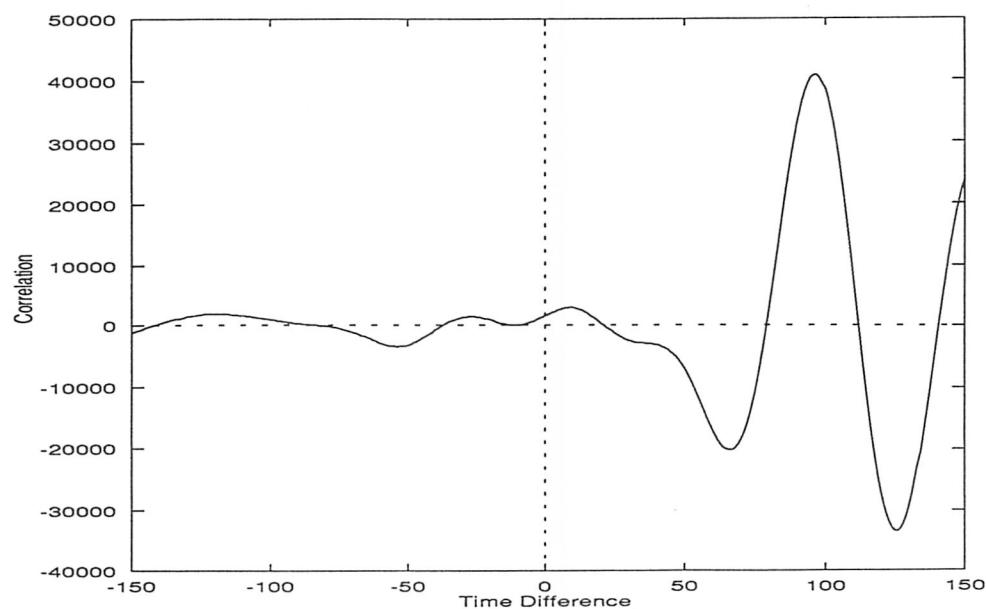


図 4.20: 相互相関関数の計算例

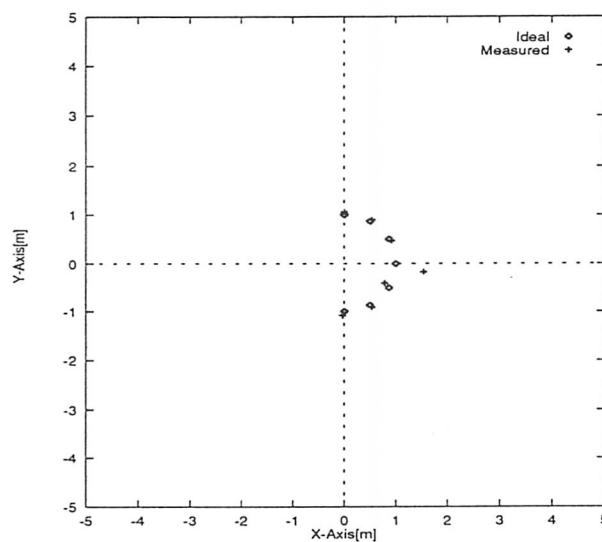


図 4.21: 音源位置の推定結果 (1[m])

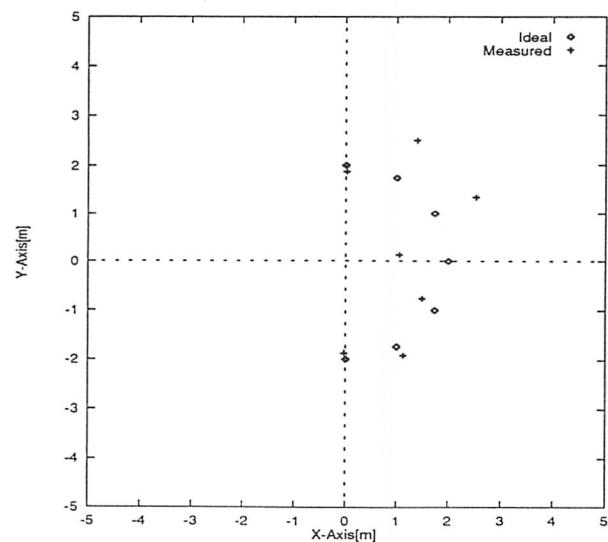


図 4.22: 音源位置の推定結果 (2[m])

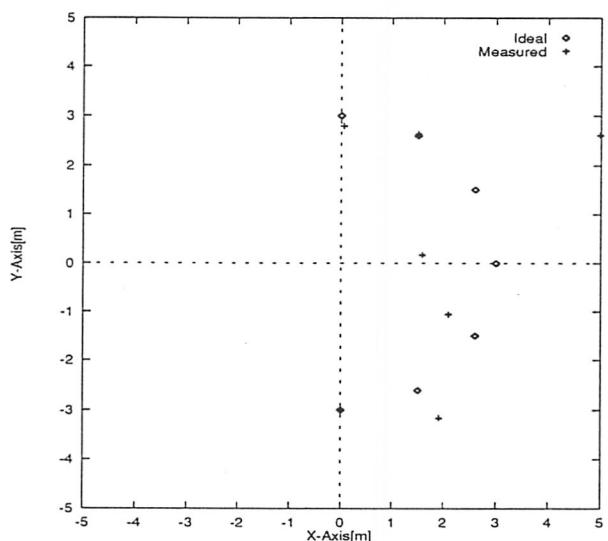


図 4.23: 音源位置の推定結果 (3[m])

また、音源分離研究の方面でも音源位置推定に関する研究 [29, 28] が行なわれているので、その技術を利用することも考えることができる。

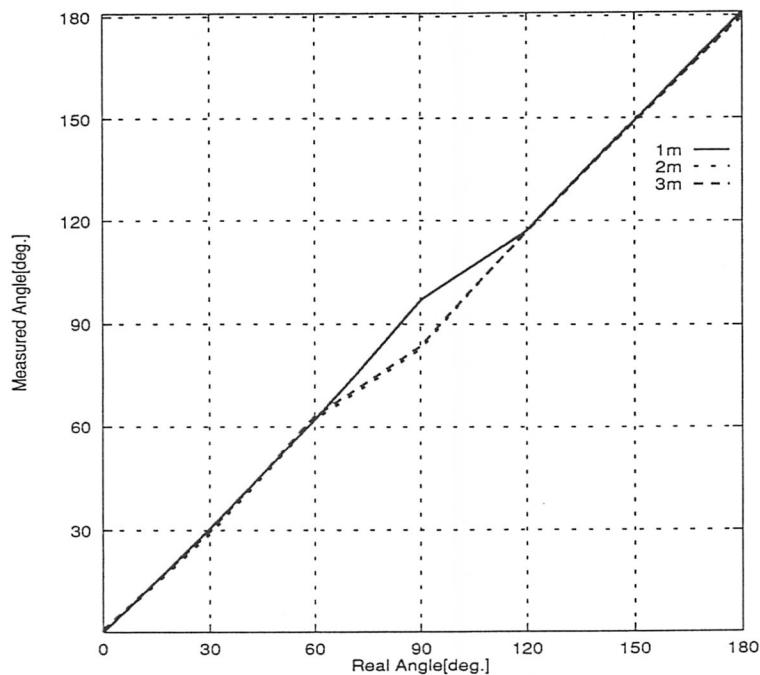


図 4.24: 音源位置の推定 (角度)

表 4.4: 音源位置の推定 (角度)

	音源位置の角度誤差		
	平均値 [deg.]	最大値 [deg.]	標準偏差 [deg.]
1[m]	2.1	6.7	2.2
2[m]	2.2	7.1	2.3
3[m]	2.2	6.2	2.1
Total	2.2	7.1	—

## 4.4 Virtual Dome

ここでは、聴覚的臨場感と組み合わせて視覚的臨場感を記録するためのシステムとして次のようなシステムを紹介する。

通信系に時間おくれが存在する場合に生じる見回し動作の問題(2.4.2を参照)は臨場感を伝達する点から考えて重要な問題になる。

例えば、テレロボットが衛星軌道上にあれば、地球からテレロボットまで片道の通信に1秒程度が必要となる。その場合、時間遅れはすでにスムーズな見回し動作の限界にある。

そこで、当研究室ではこの問題を解決するために仮想空間を用いることによって、通信系の時間遅れの問題を克服できるシステムとしてVirtual Domeを提唱、研究を行なってきた。

ここでは、その概略について説明する[49]。

### 4.4.1 Virtual Dome の構成

時間遅れの問題を解決するために、図4.25に示すような新しい原理に基づいた臨場感通信システムを試作し、このシステムをVirtual Dome(VD)と名付けた。VDがこれまでのシステムと根本的に異なる点は、カメラヘッドの動きと人間の頭部の動きが基本的に同期していないということである。それゆえに通信回線における時間遅れがHMDを用いた見回し動作に直接影響せず、同時に通信により生ずる時間遅れを最小とすることができる。

VDは

- (1) カメラヘッドシステム
- (2) 通信回線
- (3) 画像提示システム

の3つの要素から構成される。

カメラヘッドシステムは遠隔地に設置され、周囲の完全な情報を獲得するため全空間を撮影する。これらの画像は通信回線を通じて利用者のサイトにある画像提示システムに送られる。画像提示システムは、グラフィクスワークステーション、HMD、頭部位置センサといった人工現実感のハー

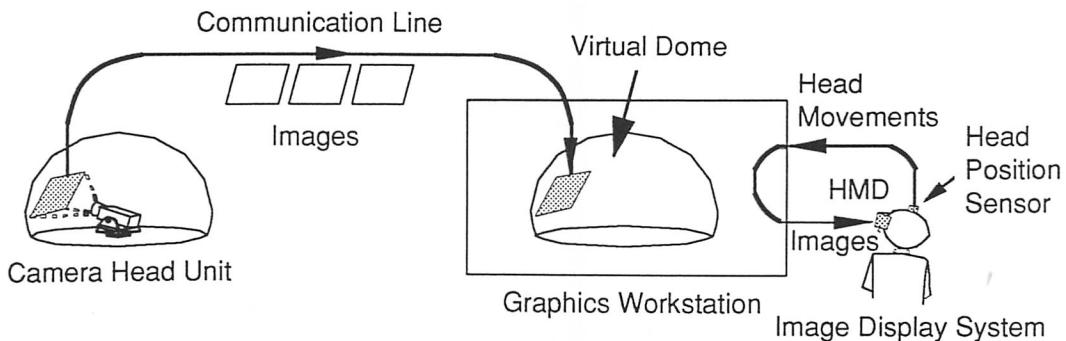


図 4.25: Virtaul Dome システムの概念図

ドウェアから構成されている。グラフィクスワークステーションでは、平面により構成される球形の物体が、仮想的なスクリーンとして描画される。カメラヘッドシステムにより撮影された画像は、この球形のスクリーンの対応する平面にテクスチャマッピングされる。こうすることにより、遠隔地の周囲の情報が利用者の側では仮想空間内のドーム状のモデルの上にコピーされることになる。HMD と空間位置センサを用いることにより、利用者の向いている方向の視野に対応する範囲の画像が HMD の液晶スクリーン上に表示される。つまり、観察者は OMNIMAX スクリーンのように、仮想空間内のドームを内側から見ることになる。全ての情報がグラフィクスワークステーションの中に存在しているので、見回し動作により生ずる時間遅れは全く存在しなくなるというわけである。

試作したシステムの構成を図 4.26 に示す。以下では、3 つの構成要素についてそれぞれ順に説明する。

#### 4.4.2 カメラヘッドシステムおよび通信回線

カメラヘッドには周囲の全空間を撮影することができるような機構が要求される。試作したカメラヘッドは、カメラ台がインデックステーブルにより駆動され、水平方向と垂直方向の回転に対応できる(図 4.27)。ステッピングモータの制御は PC-98VM 用の拡張ボードを作製し、これを介して行っている。カメラは、CCD カラーカメラ (SONY XC-711RR) を 4[ms]

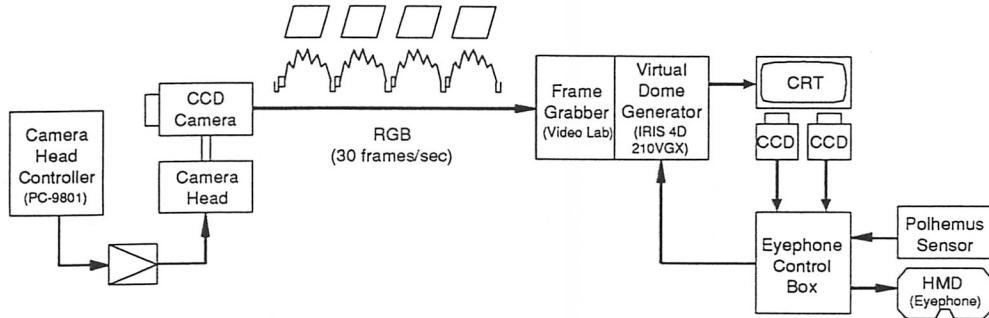


図 4.26: Virtual Dome の試作システム

のシャッタースピードで用い、水平方向は 24[deg.]、垂直方向は 18[deg.] ごとに回転を止めて画像を撮影している。この撮影条件で画像のぶれや歪みを生じずに撮影できる回転速度は毎分約 6 回転である。

画像は Frame Grabber(SGI : Video Lab) によりデジタル画像化される。実際に用いた画像のサイズは  $64 \times 48$  pixel であり 1 pixel あたりのビット数は 24bits(8bit × 3(RGB)) である。

現在、通信回線は NTSC である。将来的には、B-ISDN のような通信路が最適であろう。また、現在マルチメディアの技術と合間って注目されている MPEG のような画像の圧縮についても考える必要があるだろう。

#### 4.4.3 画像提示システム

画像提示システムには、グラフィックス・ワークステーション (SGI:IRIS 4D 210VGX)、HMD(VPL : EyePhone)、空間位置センサ (Polhemus : 3D Space) を用いている。

ドームの形状について説明する。一つの極端な例としては、コンピュータの負荷を下げるために三角柱のようにポリゴンの数を減らすことが考えられる。この場合には、それぞれのポリゴンが大きいために画像の歪みに注意を払う必要がある。別の方針は、ドームを球形にすることである。この場合には、画像の歪みをそれほど気にする必要はないが、多数のポリゴンを用いる必要がある。

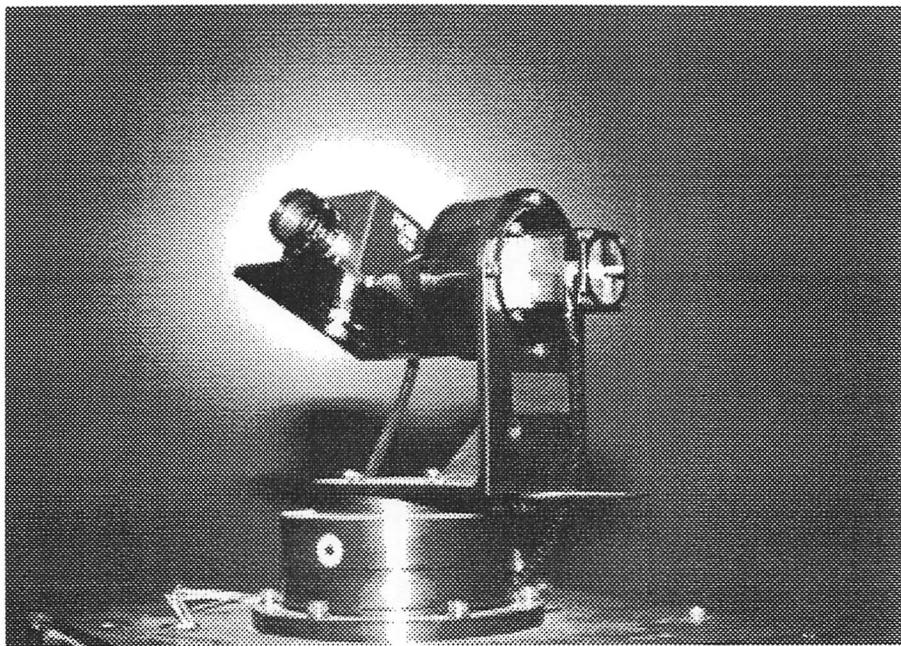


図 4.27: Virtual Dome のカメラヘッド部

実際に用いたドームの形状を図 4.28に示す。これは、カメラヘッドの構造及び動きをもとにして決定した。ドームは水平方向に 15 枚、垂直方向に 5 枚の合計 75 個のポリゴンにより構成されている。

#### 4.5 視覚的臨場感と音響的臨場感の融合実験

前述の 4.3.2 の方法を用いれば音源の位置を推定することができる。また、4.2.2で立体音響ディスプレイ Convoltotron によって立体音響を受聴者に提示できることが示された。つまり、この 2 つの事実を組み合わせれば音響的臨場感の記録・再生システムが基本的には組み立てる事が出来るという事になる。

ここでは、視覚的臨場感を提示できるものとして 4.4 で述べた Virtual Dome と音響的な臨場感を同時に提示して、双方が融合した形で臨場感を提示することに関して考察を行なう。

Virtual Dome と音響的臨場感を組み合わせた実験として次のような実

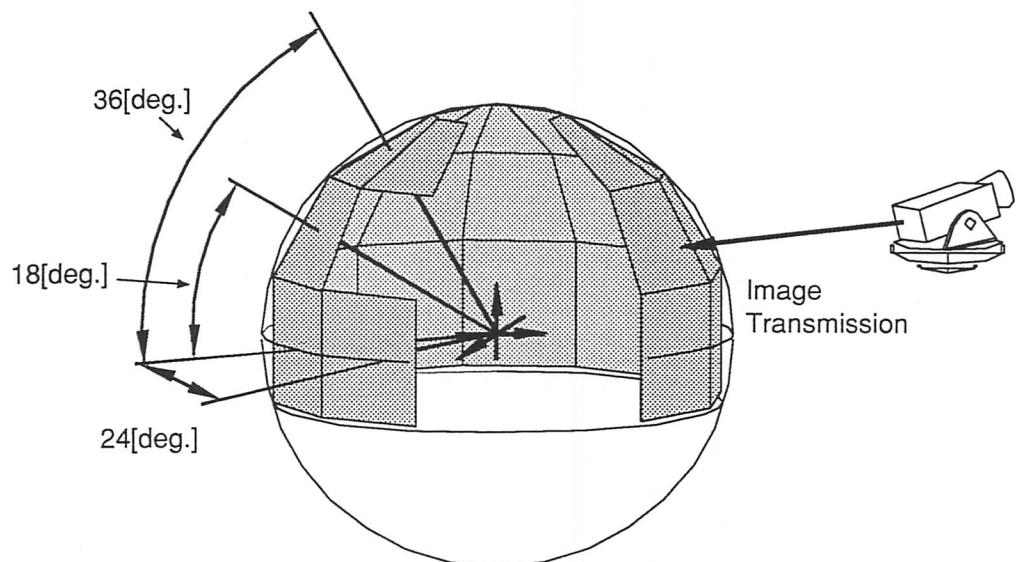


図 4.28: Virtual Dome の形状

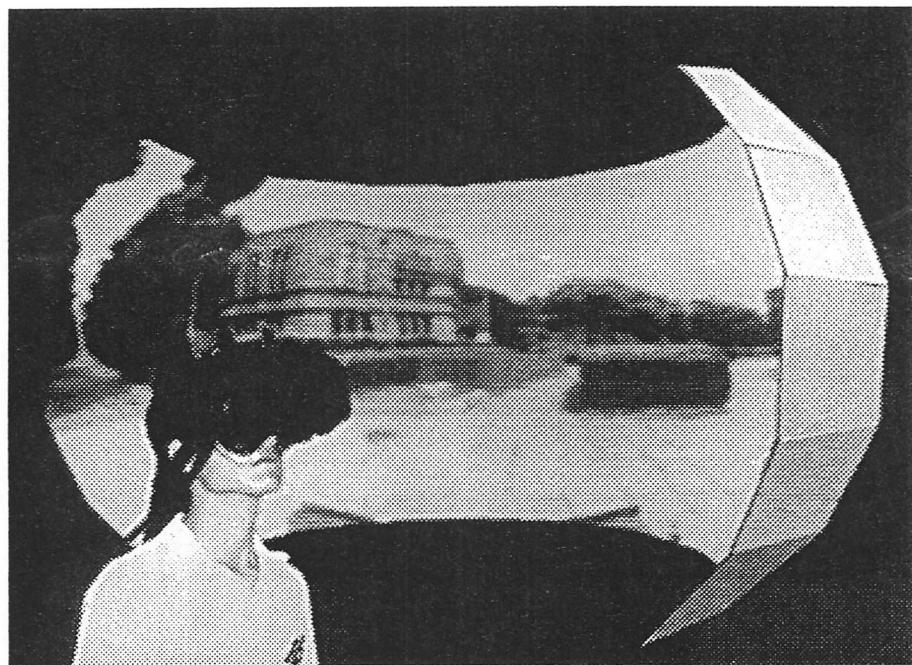


図 4.29: Virtual Dome の摸式図

験を行なった。

### 実験方法

実際のシステムに近い構成を取るために、Virtual Dome の中心部に動画を合成したシステムを基本としていくつかの条件で物体の見回し動作を被験者に行なってもらった。つまり、背景は Virtual Dome で提示し、中心部の画像だけ実画像でリアルタイムで更新されている。但し、背景は時間遅れがないのに対して、中心部は時間遅れ(半回転するのに約3秒)が存在する。音像の音としてはシンセサイザによるリズム音を提示した。

条件としては次の3つの条件で実験を行なった。

1. 基本システムのみで、見回しの目標となる物体を言葉で指示した。
2. 条件1に加えて、目標を言葉で指示すると同時に立体音響でも提示する。
3. 基本システムで、目標を立体音響のみで提示する。

目標として、条件1,2では既知の物体(Virtual Domeの背景上にも描かれている)を用いた。条件3では、目標物として判別し易いように背景にはない白い紙を用いた。すべての条件で目標として7つの物体を指示した。

被験者としては、当研究室の学生を対象に実験を行なった。

### 実験結果

条件1の時の典型的な頭部の運動を示したのが図4.30、また条件2の時の典型的な頭部の運動を示したのが図4.31である。図の横軸は実験が始まってからの時間を示しており、縦軸は視線方向とZ軸からのなす角度を表し、角度の符号は視線方向のAzimuth Angleと同じ符号を用いた。

条件1と2とを比較すると、条件2の方が全体的に見回し動作のオーバーシュートが少なくなっている。これは、立体音響で提示している音像の情報を用いて、頭部の運動に関してフィードバックをかけているためだと思われる。このことは、被験者が「音像が正中面を横切る時に音像の変化が

良く分かる」といった感想からも裏付けられる。また、条件2では条件1に比べてすみやかに次の目標に移動している。これは、音像を提示された方が直観的に頭部の回転方向を判断し易いためではないかと思われる。

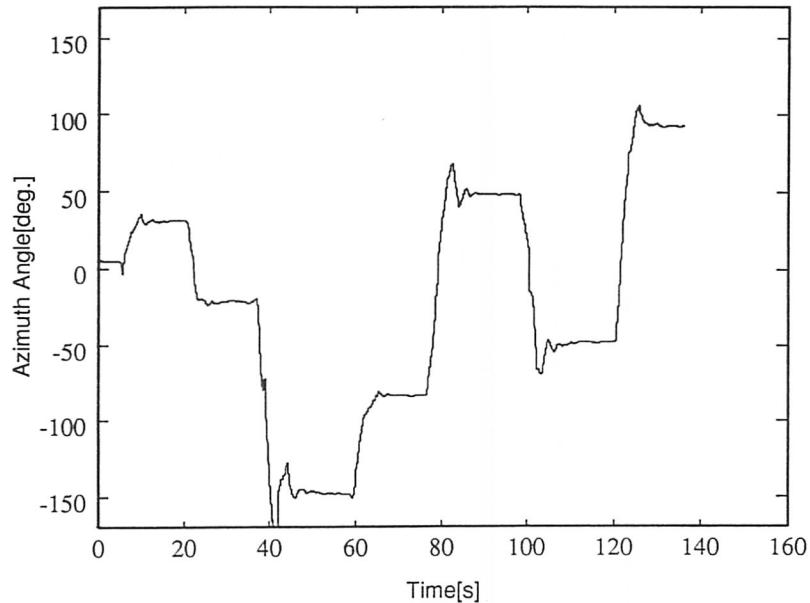


図4.30: 頭部の動きの典型例(条件1)

次に条件3の場合(図4.32)であるが、ターゲットが視覚情報からは得られない(中心の実画像でしか目標物である白い紙は見えない)が、それでも図から分かるように音像で提示した方向に向かってスムーズに頭部は運動している。

以上のことから、視覚的臨場感と音響的臨場感が仮想空間というフレームの中でうまく協調して被験者に提示されていることが確認された。

## 4.6 本章のまとめ

本章では、本研究で使用した立体音響ディスプレイの基本的な性能と考えられる音源の定位精度を測定した。

そして、音源の方向を推定する方法に関して実現可能性を実験によって

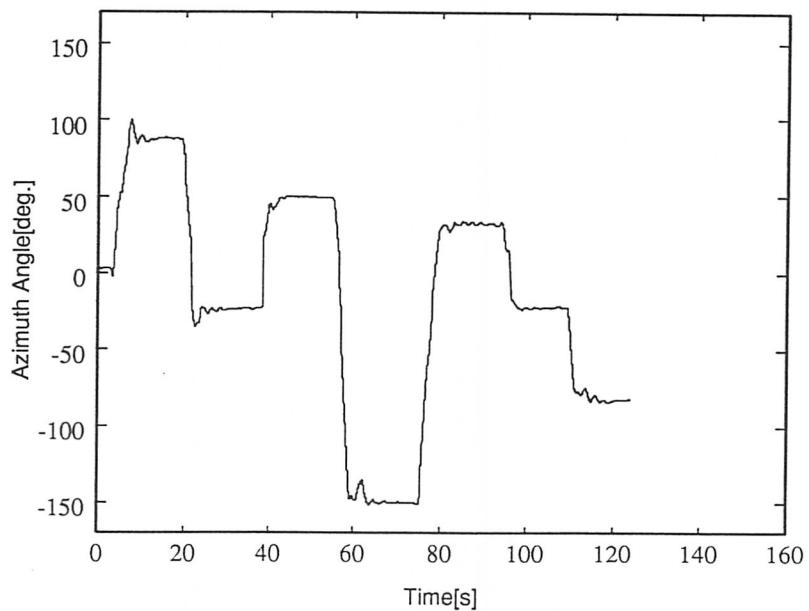


図 4.31: 頭部の動きの典型例 (条件 2)

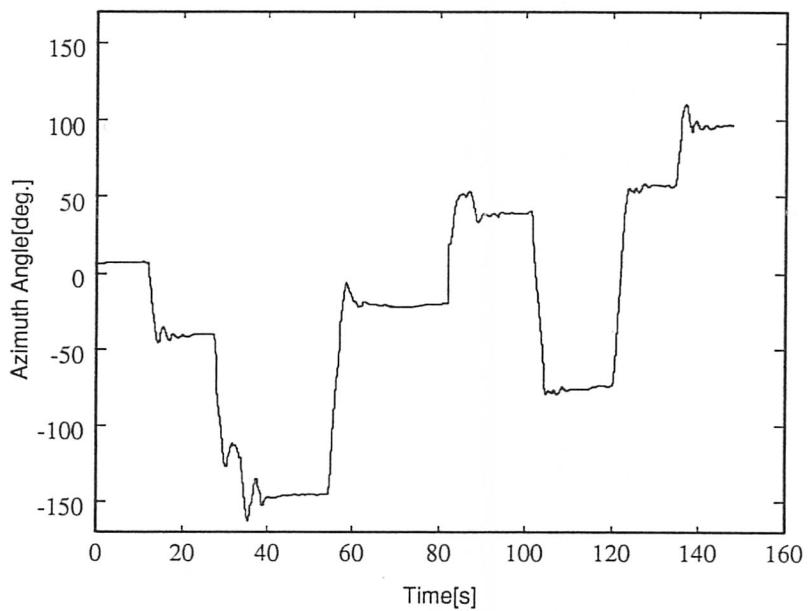


図 4.32: 頭部の動きの典型例 (条件 3)

検証した。

また、視覚的な臨場感を提示するためのシステムとしてVirtual Domeについて触れ、最後に仮想空間内の人間に対して視覚的な臨場感と聴覚的な臨場感が融合して提示し得ることを実験によって検証した。

## 第 5 章

VSS による音響的臨場感の記録・再生

## 5.1 はじめに

前章では、音響的臨場感を記録する方法として理想的な方法、つまり音響情報を音源の位置と音響信号そのものに分離して臨場感を記録する方法について検討した。

しかし、実際問題としてこの音源分離の問題は非常に困難な部分を含んでおり、今でも様々な研究が行なわれている。

そこで、この章では音響情報の分離をあえて行なわず、音響的臨場感の記録・再生をなるべく少ないチャンネルの音響信号の記録によって行なうための方法について検討を行なう。

## 5.2 原理および構成方法

受聴者が音像定位に使用するキーとしては2.3.1で述べたように、両耳間時間差、両耳間レベル差、音響信号の波形変化があるが、この中で比較的容易に測定できる両耳間時間差に注目してみる。

図5.1は、インパルス性の信号に対する両耳間時間差と耳間偏移の関係である。この図から分かるように両耳間の位相遅れを記録することによって、音像を左右に移動させることができることが分かる。

次に、両耳間時間差を記録・再生する方法に関して考察する。

### 5.2.1 2つのマイクロフォンによる方法

図5.2のようなシステムを考える。2つのマイクロフォンで収音した信号を、仮想空間に配置した2つの仮想スピーカで受聴者に音響信号を聞かせるようなシステムである。この時に頭部の垂直面方向に無限の大きさの遮音板を配置して、左右の耳へのクロストークが起きないようにする。このようなシステムでは、受聴者の両耳がマイクロフォンの位置と同じ位置に存在する場合には、ほぼ正確に両耳間時間差を再生することが可能である。

しかし、図5.2下のように頭部の方向とマイクロフォンの方向が異なる方向に向いた場合でも仮想スピーカから耳までの距離は左右で同じ距離増

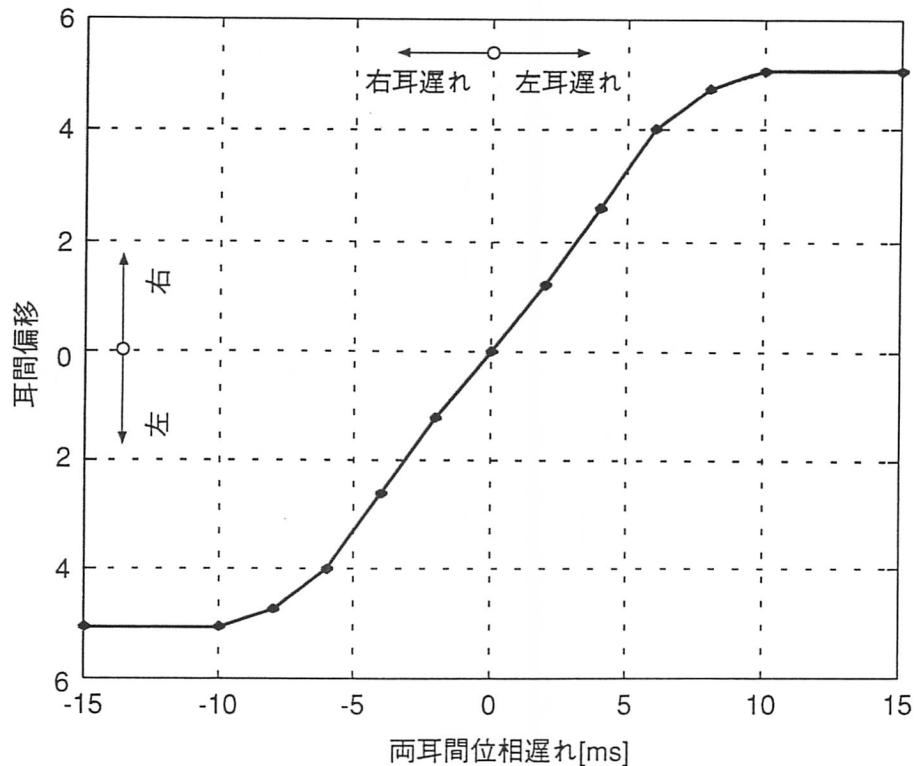


図 5.1: 両耳間時間差と耳間偏移 (文献 [3] から引用)

加するから、両耳間時間差は変化しない。つまり、このシステムで音響的臨場感を記録・再生することは不可能である。

### 5.2.2 4つのマイクロフォンによる方法

次に、図 5.3 のようにマイクロフォンと仮想スピーカを 4 つに増やしたシステムについて考察する。このシステムでも、頭部の正中面に無限大の大きさの遮音板を配置する。このシステムでは、片方の耳に到達する音響信号は 2 つの仮想スピーカからの信号が重なった形で到達するが、ここでは、2 つの信号の内より早く到達する信号によって両耳間時間差が頭内で検出されると仮定した。

図 5.4 のように頭部が仮想空間ないで回転する場合を考えてみる。この場合は、耳から仮想スピーカまでの距離が変化することによって疑似的に

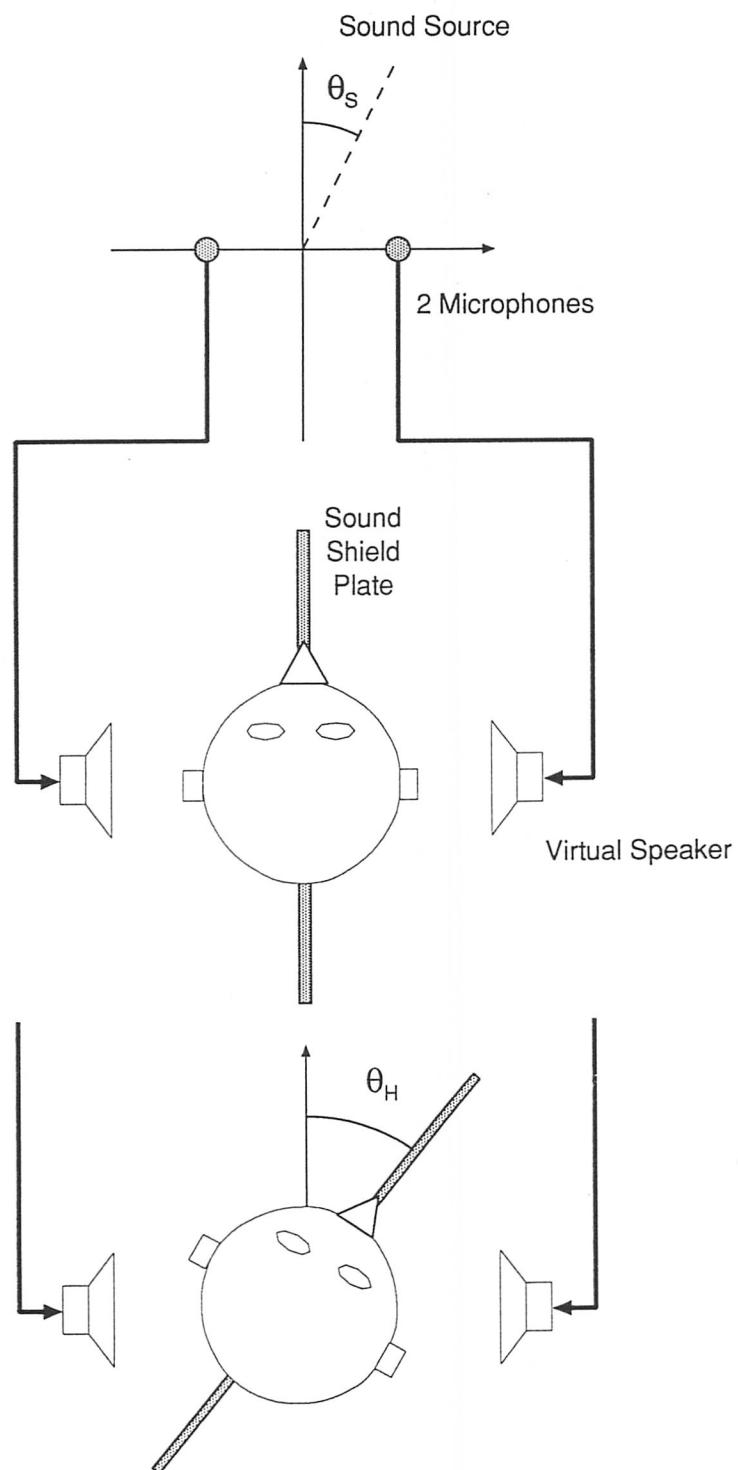


図 5.2: 2 Microphones-Virtual Speaker System

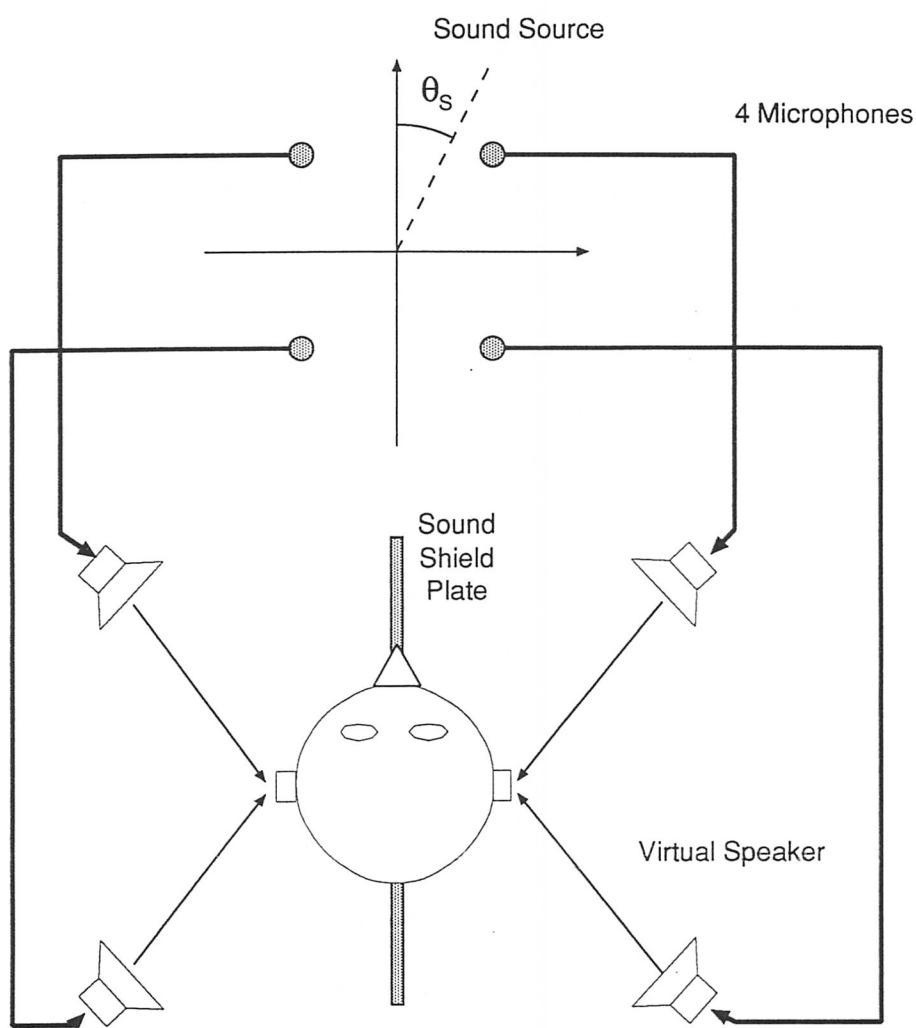


図 5.3: 4 Microphones-Virtual Speaker System

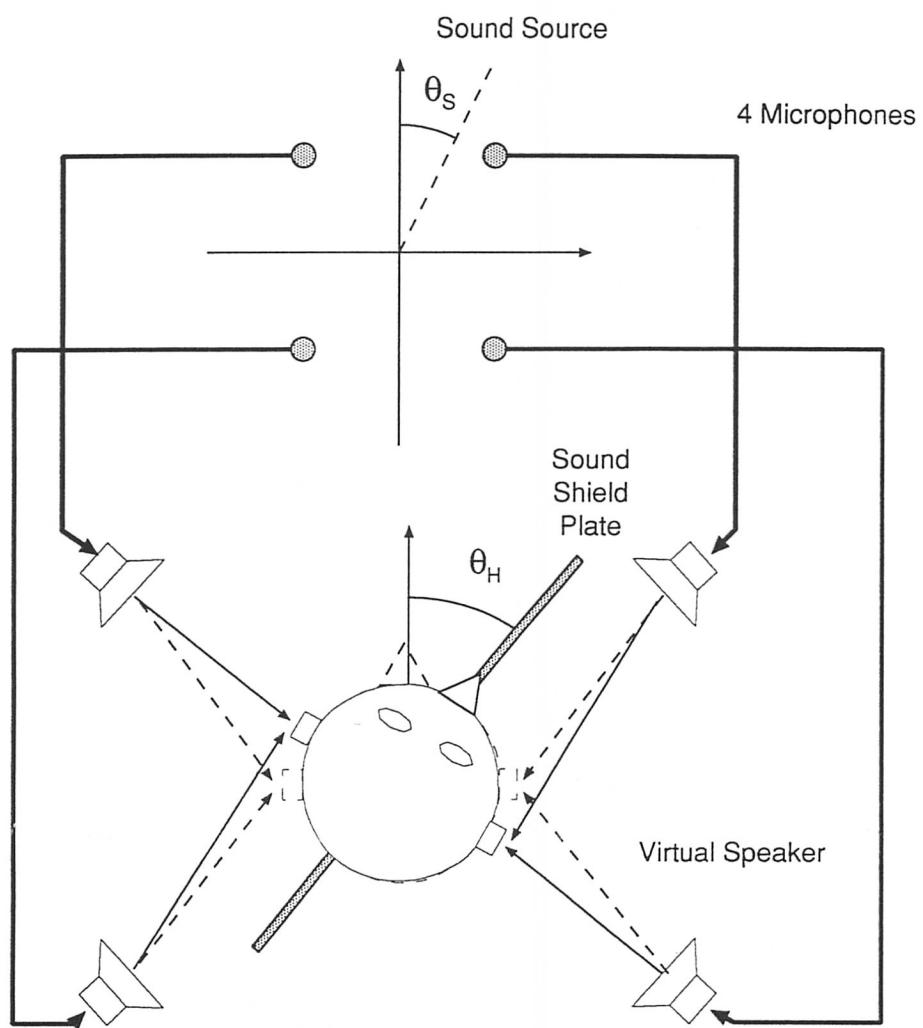


図 5.4: VSS における頭部の動き

両耳間時間差を変化させて、頭部の運動と反対方向の音像が動いたような両耳間時間差の変化を発生させることができる。

### 5.2.3 VSSのシミュレーションによる性能

水平面内だけで考えた場合、音源方向  $\theta_S$  と頭部の向いている方向  $\theta_H$  が決まれば、理想的には両耳間時間差  $t_D$  は次の式で決まる。(時間の単位は、特に断らない限り頭部の左右の耳の間の長さで正規化して表示してある。)

$$t_D = \cos(\theta_S - \theta_H) \quad (5.1)$$

(5.1) を図示したのが図 5.5 である。

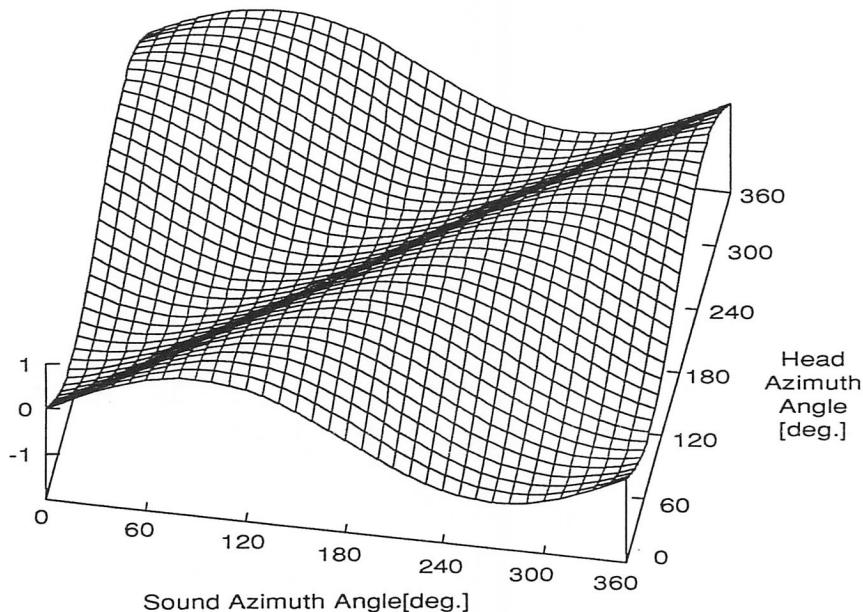


図 5.5: 理想的な両耳間時間差の値

次に、5.2.2で提案したシステム(以下、VSS: Virtual Speaker System と呼ぶことにする)の両耳間時間差をシミュレーションで計算したものを図 5.6 に示す。また、VSS で再現した両耳間時間差の理想的な値との誤差の絶対値を図に示したのが、図 5.7 である。

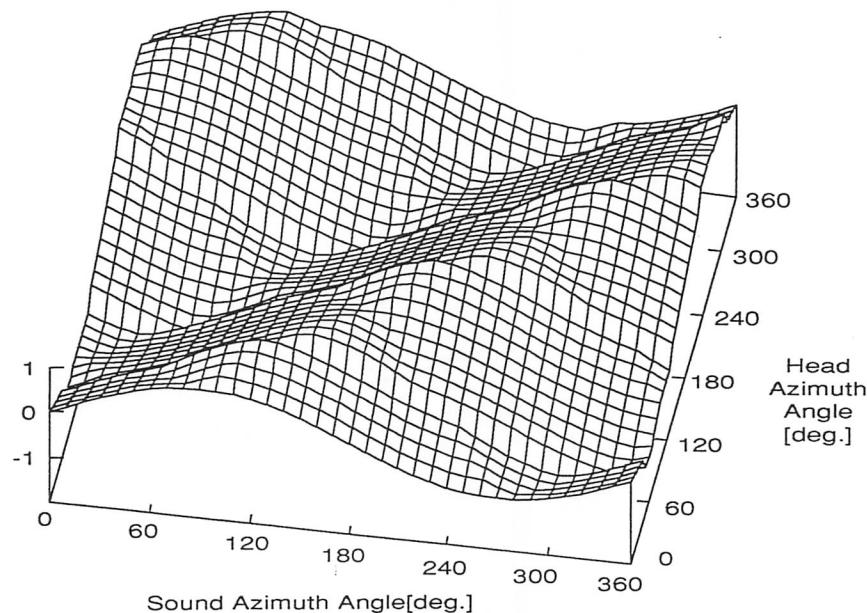


図 5.6: VSS による両耳間時間差の値

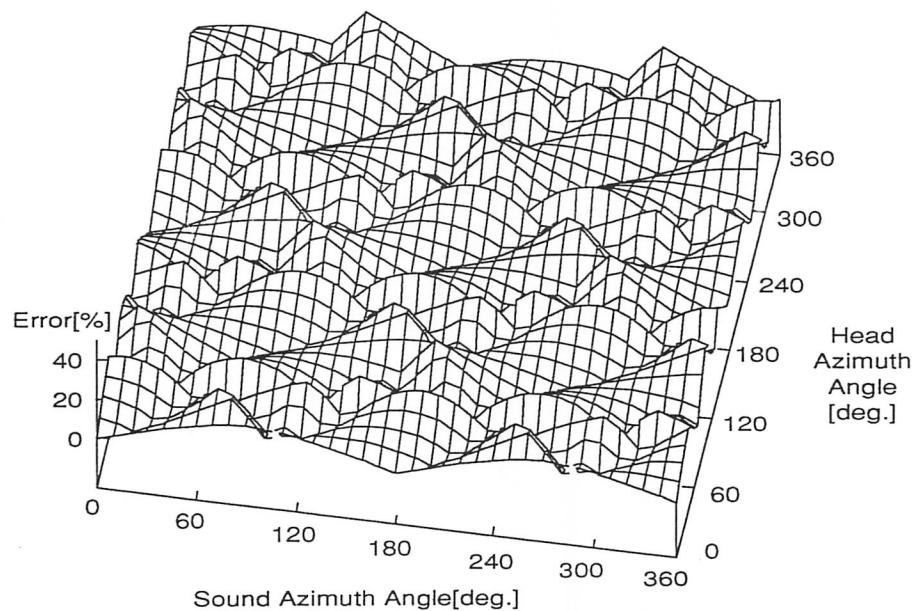


図 5.7: VSS による両耳間時間差の再現誤差

これから分かるように、VSSで両耳間時間差を再現すると最大で30[%]近い誤差が発生するが、平均では10[%]前後の誤差しか生じないことが判明した。

次にVSSのパラメータとして代表的なマイクロフォンの間隔とVirtual Speakerの仮想空間内における距離のシミュレーションでの評価を行なってみた。

図5.8は、マイクロフォンの間隔を変化させた時の両耳間時間差の誤差を最大値と平均値で評価した図である。この図から分かるように、マイクロフォンの間隔を両耳間の距離の1.2倍前後に設定するのが誤差を最小にするためには良いと思われる。

次に、図5.9は、Virtual Speakerの距離を変化させた時の両耳間時間差の誤差を最大値と平均値で評価した図である。この図からは、Virtual Speakerの距離をなるべく遠くの距離(実際には、両耳間距離の20倍程度)に設定すれば良いことを示唆している。

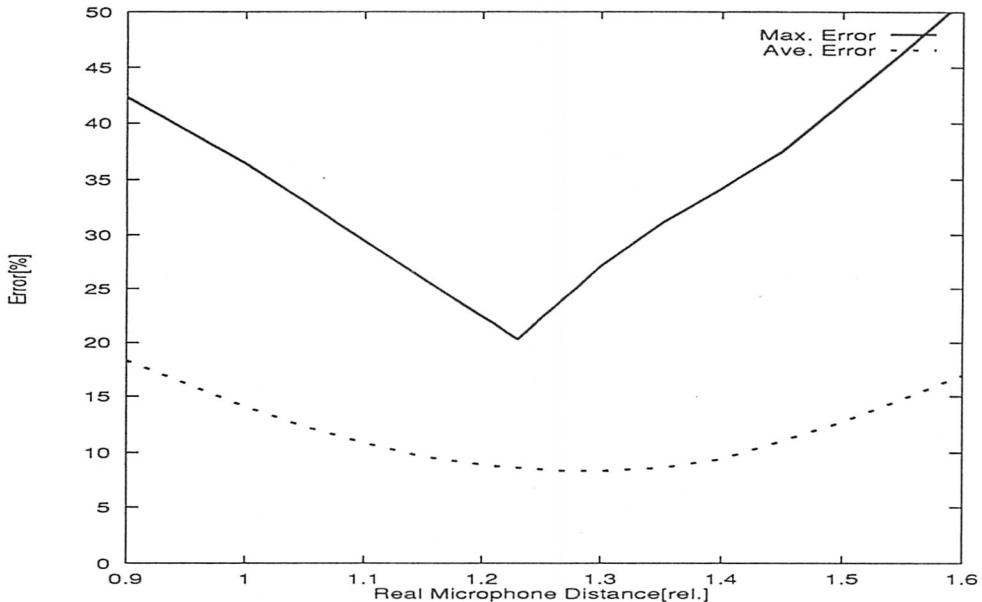


図5.8: マイクロフォンの距離と両耳間時間差の再現誤差

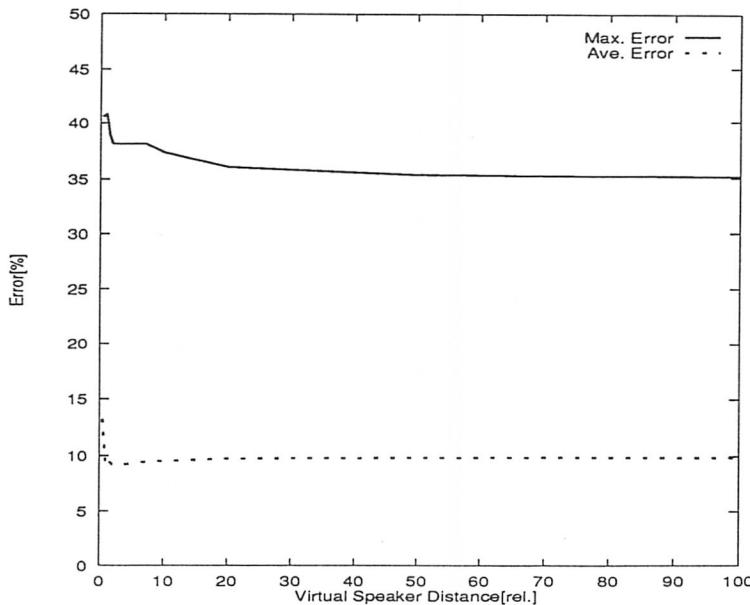


図 5.9: Virtual Speaker の距離と両耳間時間差の再現誤差

### 5.3 点音源の音像定位

VSS の基本的な性能を評価するために、Convoltotron の基礎性能を調べたのと同様に点音源に対する音像の定位精度の実験を行なった。

#### 実験方法

VSS で実音源(今回の実験では、比較的定位感が得られやすかったシンセサイザで合成したピアノの音を使用)を収音し、Convoltotron の Virutal Speaker で再生した。実音源までの距離は、マイクロフォンの中心から 2[m] である。マイクロフォンの間隔は 18[cm] に設定した。

実験は、頭部を固定した場合と、頭部を自由に回転できる状態で自分の真正面に音像が来るように頭部を回転してもらう実験の 2 つを行なった。

頭部を固定した場合には、4.2.2と同じ方法で被験者に音像の方向を提示してもらった。

被験者数は 4 人である。

実験装置では、マイクロフォンからの音響信号を録音・再生するために AKAI 製の DR4 HARD DISK RECORDER (図 5.10) を用いて 48[kHz]、16 ビットサンプリングで記録を行なった。

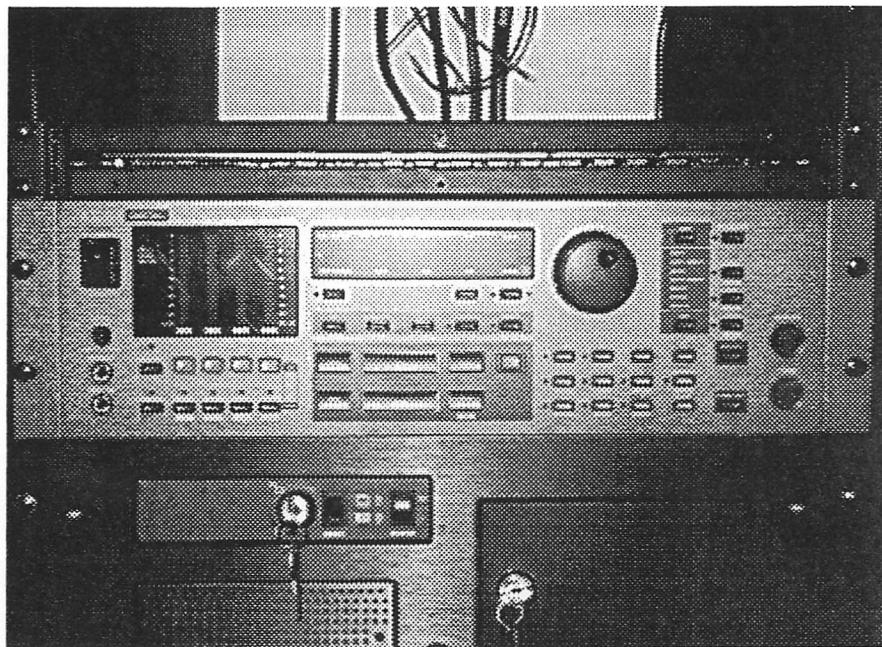


図 5.10: Digital 4-Channel MTR

### 実験結果

頭部を固定して音像を提示してもらった場合の実験結果が図 5.11 と表 5.1 に示す。平均の誤り率や前後誤り率は、Convolutron 単体の音源定位精度実験の時とほとんど変わらない値を示している。

VSS では、両耳間時間差に注目してそれをなるべく正確に再現しているだけにも関わらず、思ったよりもいい定位精度であることが確認された。

次に、頭部の回転動作を可能にした場合の実験結果に関して図 5.12 と表 5.2 に示す。本来頭部の回転動作を可能にした場合には、頭部の回転に対する音源の移動方向から前後誤りをする可能性はないと思われたが、この実験結果からは時々前後の誤りをしてしまう被験者が存在した。このことに

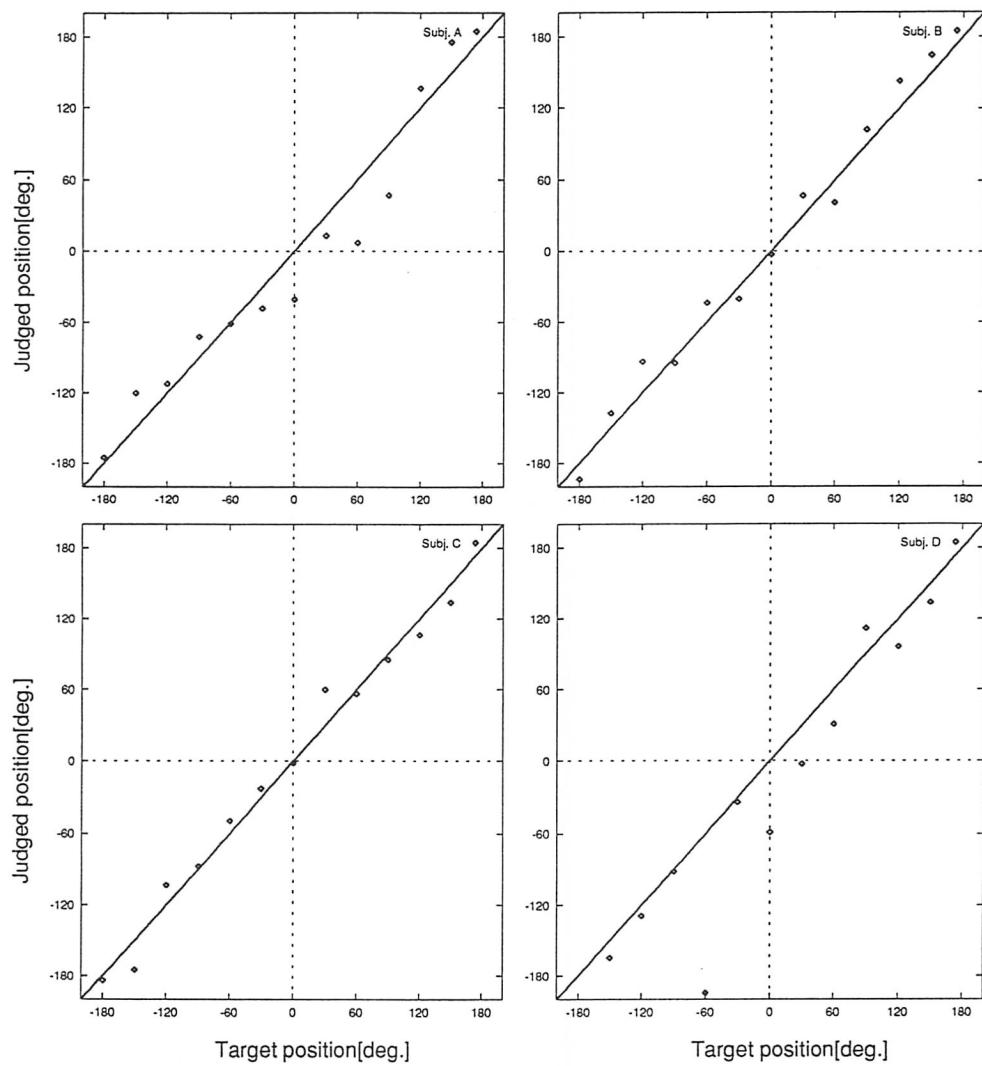


図 5.11: VSS の点音源定位精度 (頭部固定)

表 5.1: VSS の点音源定位精度 (頭部固定)

	平均 [deg.]		標準偏差 [deg.]		前後誤り率 [%]
	修正なし	修正あり	修正なし	修正あり	
Subject A	39.8	23.5	44.6	16.7	17
Subject B	42.6	14.3	54.2	6.7	25
Subject C	46.3	11.3	58.0	9.2	33
Subject D	62.6	31.7	52.1	35.9	42
平均	47.8	20.2	52.2	17.1	29

関して、実験を行なっている時の状況から言えることは、前後誤りをする場合は頭部をあまり回転させずに音源位置を確認した場合、自分が向いている方向と同じ方向に音源が提示された場合に起こる場合が多かった。

頭部の回転動作を許した場合の定位精度が一番良かったことから、VSS を用いても頭部の回転による音像の移動感を提示することが可能であるといえる。

表 5.2: VSS の点音源定位精度 (頭部回転可)

	平均 [deg.]		標準偏差 [deg.]		前後誤り率 [%]
	修正なし	修正あり	修正なし	修正あり	
Subject A	11.5	11.5	8.9	8.9	0
Subject B	19.3	19.0	15.8	15.5	8
Subject C	40.5	24.0	32.6	17.8	25
Subject D	11.9	11.9	8.8	8.8	0
平均	20.8	16.6	16.5	12.8	8

この実験から、点音源の定位精度に関してはほぼ満足のいく値が得られ

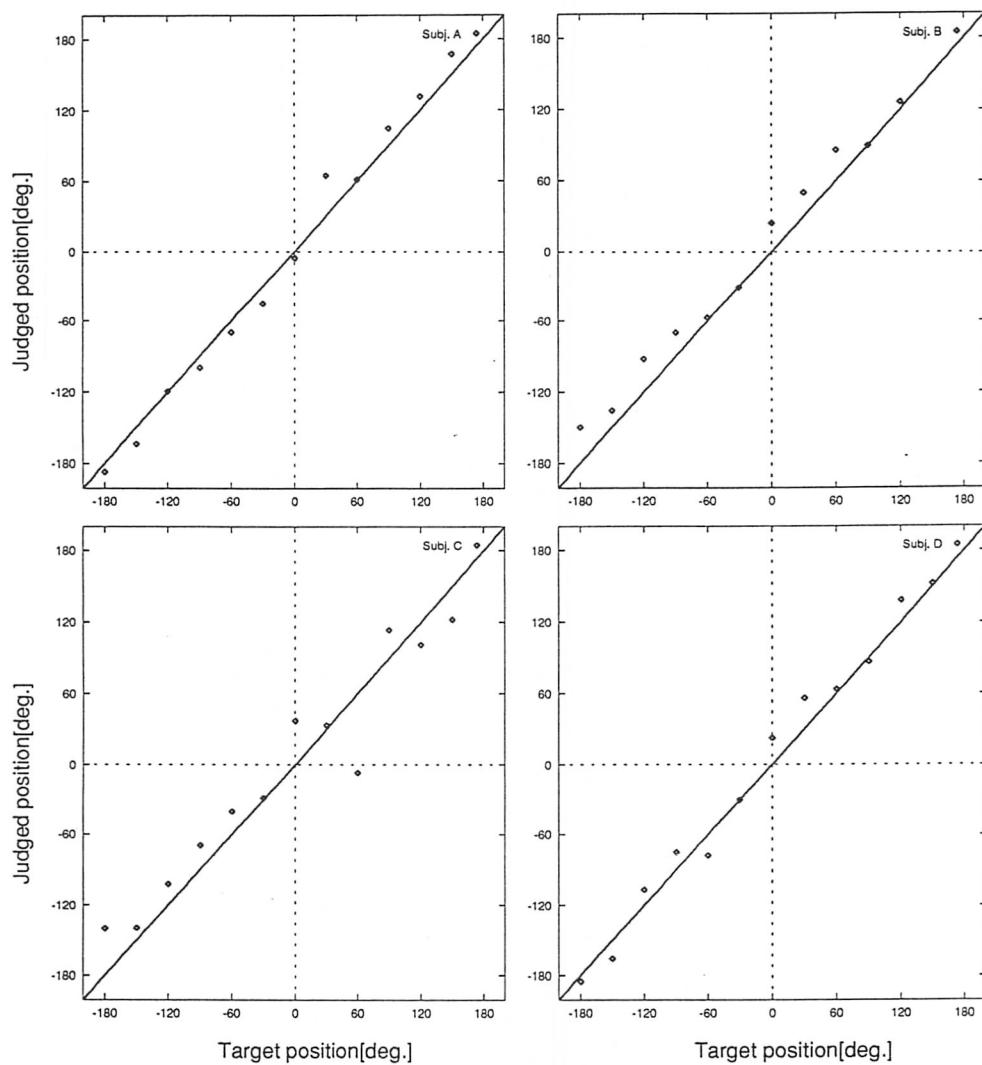


図 5.12: VSS の点音源定位精度 (頭部回転可)

たと考えられる。

#### 5.4 VSSによる状況認識実験

次の実験として、VSSが視覚的な臨場感とも融合していることを確認するためにVirtual DomeとVSSを組み合わせて状況認識の実験を行なった。

##### 実験方法

実験方法は、Virtual DomeでCamera Headの周りに半円上に4人の人が椅子に座っている場面(図5.13)を被験者に提示する。そして、VSSで4人の人間の座っている場所から音源(人間の音声/拍手の音)を15秒間提示し、どの人の場所から音が鳴っているように認識できるかを番号で指示してもらった。

##### 実験結果

実験結果を図5.3に示す。

音声の方が拍手の音よりも誤り率が高いのは、音声が時々途切れる部分が存在したためと拍手の音はインパルス的な音であるので特に定位し易い音だったためだと思われる。

音声の場合には平均誤り率42[%]と高いように感じるが、音源の位置を実際に鳴っている人間の隣の人と勘違いしている場合が多く、ある程度の音像定位が得られているものと思われる。

また拍手の音の場合には、「自信を持って音の鳴っている人を特定できた」との意見も聞かれた。

#### 5.5 本章のまとめ

本章では、音響的臨場感を記録・再生する方法として、音源の分離問題を回避できる方法としてVSSを提案した。



図 5.13: 状況認識実験

	音声	拍手
Subject A	33	0
Subject B	50	0
Subject C	58	33
Subject D	17	0
Subject E	42	33
Subject F	50	25
平均	42	15
標準偏差	15	17

表 5.3: 誤り率(状況認識実験)

VSSによる実音像の提示でも、Convoltotronの性能と同程度の定位精度を得ることができることを示した。

また、視覚的な臨場感とVSSを用いた音響的な臨場感を同時に提示した場合に、この2つの臨場感が融合できていることを実験により示した。

この章の実験を通して、VSSで実際に音響的臨場感を記録して、それを仮想空間を通して提示できることを示した。

## 第 6 章

### 今後の展望

## 6.1 はじめに

本当の意味での音響的臨場感を記録・再生するには、その空間の音響的特性をもシミュレートする必要があるのは明らかである。今後の展望として、この音響空間の特性を仮想空間の一部としてシミュレートする方法に関して考察をする。

また、視覚的臨場感を考慮したシステムに関して概略し、そのシステムに対して音響的臨場感を付加する方法について考察し、今後の研究指針に関して考察する。

## 6.2 空間情報を用いた音響空間の再構成

厳密に音響的臨場感を記録・再生するには、その音響空間の特性を測定する必要がある。また、すでに記録されてある情報に何らかの新しいバーチャル音源を合成する場合にも、その音響空間の特性を知ることが重要なファクタになる。

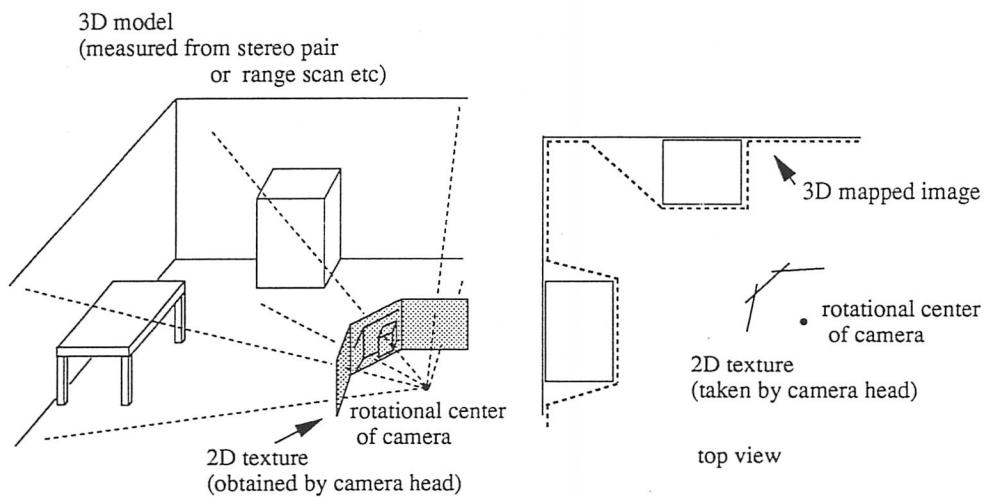


図 6.1: Virtual Dome による三次元形状情報の取得

現在、Virtual Dome の関連研究として仮想空間の形状情報も取得するような研究が行なわれている。図 6.1はその概念図を示したものである。ま

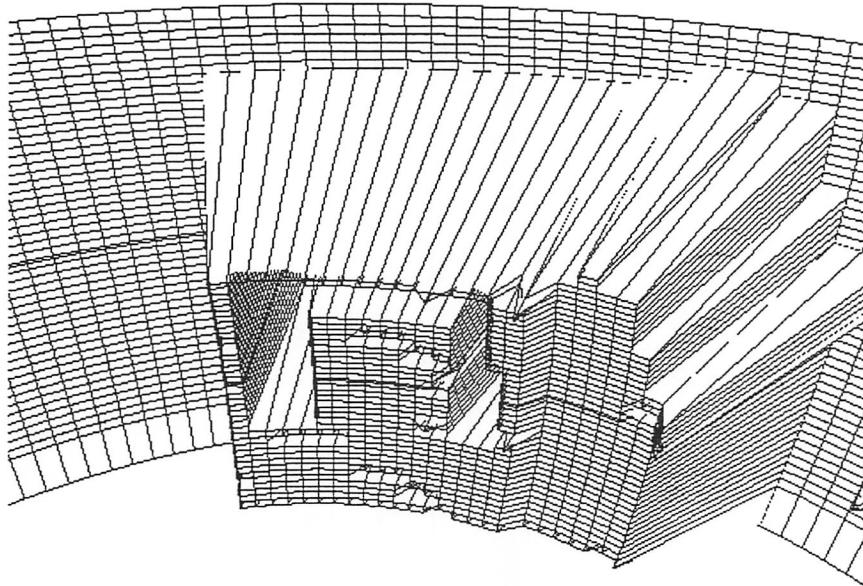


図 6.2: Virtual Dome による三次元形状情報の一例

た、Virtual Dome に対して奥行き情報を付加したものの例が図 6.2である。このように、その仮想空間の形状情報が分かっている場合にその空間の音響的特性をシミュレーションによって計算することが可能である。

音源の反射音を計算するには図 6.3のように鏡像法を利用するのが一般的である。

このような空間の音響特性を計算している例としては音響ホールの空間特性コンピュータシミュレーションの例がある [30]。

現在の Convoltotron では、リアルタイムの反射音の計算として 7 個までの鏡像の計算を行なえるようになっているがより多くの反射音の計算ができるようになれば、より臨場感のある音響仮想空間を楽しめるようになるだろう。

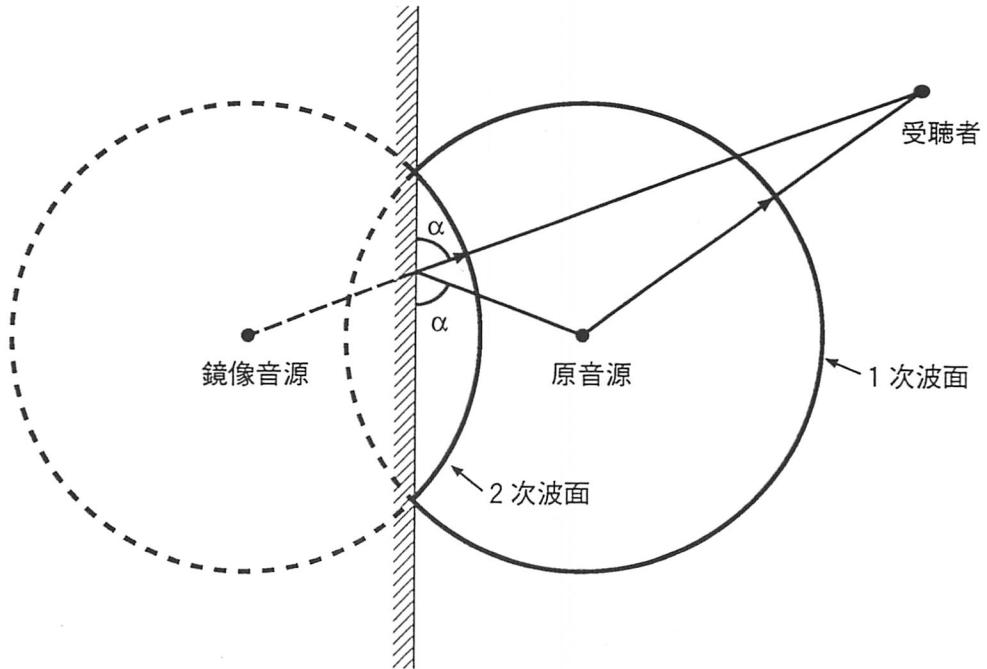


図 6.3: 鏡像音源による反射波面

### 6.3 CAVE System

頭部搭載型ディスプレイ (HMD) と対象的な視覚的臨場感の提示方法として、物理的に画像で人間の周りの空間を覆ってしまう方法がある。この方法の代表的なものには OMNIMAX があるが、装置自体が大がかりなものになってしまい人工現実感に手軽に利用するわけには行かない。

その問題点を解決する方法として、最近提案されている方法に CAVE (Cave Automatic Virtual Environment) がある [8, 6]。

CAVE の概念図を示したのが図 6.4 である。原理的にはいたって簡単で、人間のいる空間の四方をスクリーンで囲み、背面投射型のプロジェクタで画像を提示しようというものである。

この CAVE に対して、音響を付加することを考えてみると、CAVE 自身にある程度の空間的スペースを必要とするので、ヘッドフォンを用いた立体音響ディスプレイよりは、むしろスピーカシステムを用いた立体音響

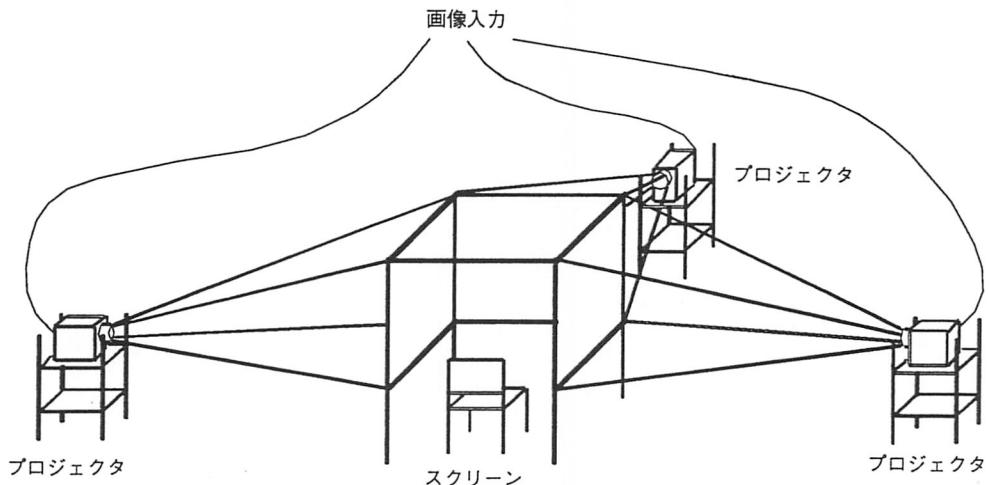


図 6.4: CAVE System

ディスプレイの方が適切かも知れない。

このCAVEと同様のシステムが当研究室でも研究されており、このシステムのデモンストレーションとして撮影された画像に対して、音響を附加したところ飛躍的に臨場感が増したとの研究結果が出ている。このことからも、音響的情報が臨場感に対していかに大切かということが分かる。

#### 6.4 KK System

より広い領域の仮想空間を簡易に生成する方法として、次のようなKK Systemが当研究室で提案・研究されている。

現在、家庭用のビデオムービーが普及しているが、それを用いて遠隔地の風景をランダムに撮影してきたとする。そのビデオテープをただ再生するだけでは、撮影者の体験を追体験するにすぎない。しかし、考え方によつては1本のビデオテープから遠隔地のコピーである仮想空間を作り出すことができる。

カメラに写る映像も人間の目に映る映像も、視点の位置と視線の方向が一致すれば基本的には同じ映像を提示するはずである。視点や視線の変化量が少なければ、変化する前の映像に対して拡大・縮小・回転などの操作

をおこなうことによって、変化後の映像を擬似的に提示することができる。つまり、画像とそれを撮影した時点でのカメラの位置・姿勢の6自由度のパラメータ $(x, y, z, \theta, \phi, \psi)$ を対応させて、視点・視線のパラメータが $(X, Y, Z, \Theta, \Phi, \Psi)$ のときの映像を合成しようというものである。

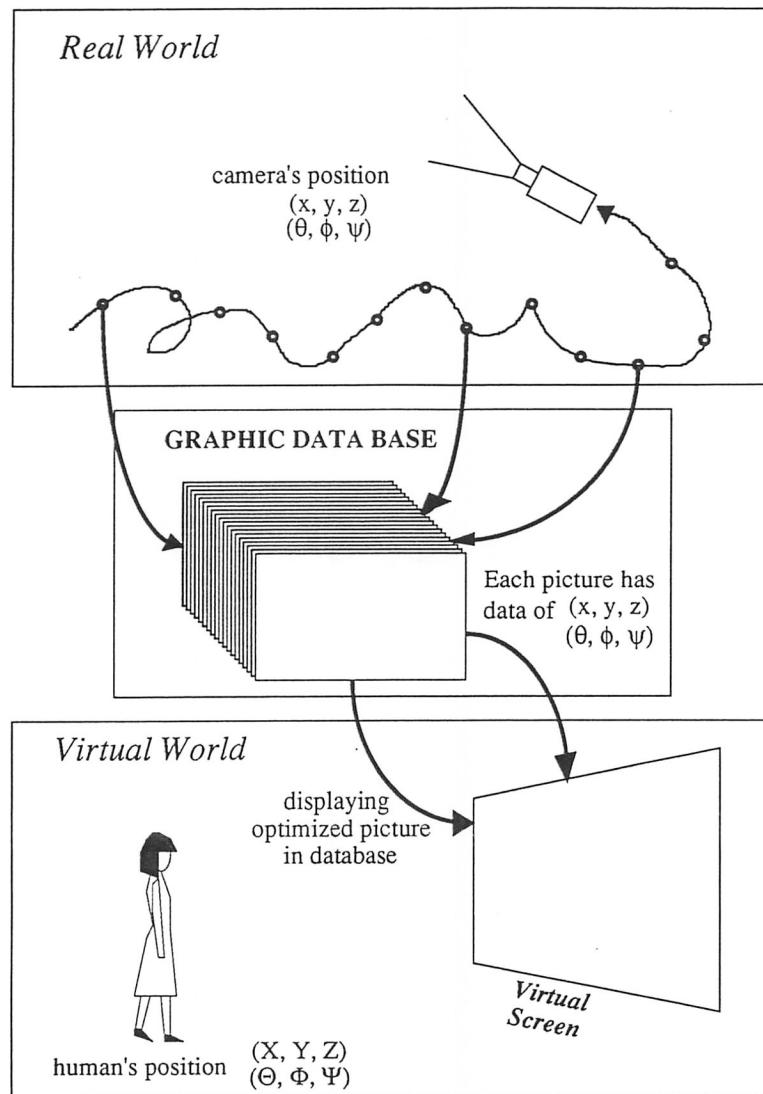


図 6.5: KK System

以上の考えに基づいて、ランダムに撮影した画像の中から適切なものを選び出して視点と視線に応じた映像を合成すれば、任意の位

置・方向での映像を作り出すことができるはずである。このことは、ビデオテープの撮影者とは違った体験を可能にし、仮想遠隔地をいわば臨場的に体験することを可能にすることができる。図6.5は、このKK Systemの概念を表したものである。

このような視覚的臨場感記録・再生システムについて音響情報をどのように付加するかは困難な問題を含んでいるが、定常的な音響空間（川辺のせせらぎのような）ならば、VSSで記録しておいて再生できると考えられる。

## 第 7 章

### 結論

本論文では音響的な臨場感を主として、それを記録・再生する方法に関して論じた。本論文では音響的臨場感を「第一人称的な体験が出来る感覺」と規定し、音響的な臨場感では特に重要と考えられる音像を頭外定位させるための立体音響技術に注目した。

この立体音響技術を人工現実感というフレームの中で利用することにより、音響的臨場感記録・再生システムを構築し得ることを示し、視覚的臨場感と組み合わせて、臨場感記録・再生システムとしての評価を行なった。

また、実際に非常に困難な問題を含んでいる音源分離の問題を回避しつつ、音響的臨場感を記録・再生するシステムとしてVSSを提案、システムとして動作することを確認し評価を行なった。

本研究で使用した立体音響提示装置、つまり高速なDSPを用い頭部音響伝達関数を利用した実時間の立体音響提示装置が利用可能になってきたのはここ数年である。この装置を人工現実感システムに立体音響ディスプレイとして組み込んだ研究が始まっている。しかし、完全な立体音響ディスプレイとして完結させるためには、立体音響提示装置の研究に加えて提示すべき音源の研究に関するもう少し真剣に取り組む必要があるのではないかだろうか。

本研究で取り上げた研究結果は臨場感の記録・再生以外に限らず、臨場感の伝送の分野においても利用できる。通信に時間遅れが存在する場合にVirtual Domeが有効なように、VSSのような技術をVirtual Domeと組み合わせれば、時間遅れのある場合であっても音響的臨場感を含めた形での臨場感伝送が可能である。これによって、Virtual Dome単独に比べて飛躍的に高い臨場感を伝送できることが期待される。

## 謝辞

修士論文を完成させるにあたり、多くの方々のご支援をいただきました。

廣瀬通孝助教授には、研究面はもちろん生活面についても多大なる御指導、御助言や興味深いお話をしていただきました。深く感謝いたします。

技官の中垣好之氏、田中雅行氏には、研究面はもちろん研究室での生活全般においても大変気をつかっていただきました。そのため、研究室では非常に快適に生活できました。本当にどうもありがとうございました。

大学院博士課程の広田光一氏、小木哲朗氏、木島竜吾氏には、研究面に関するいろいろなアドバイスをいただきました。

大学院修士2年の佐藤慎一氏、由上勤氏には短かった大学院2年間を共に過ごし、研究や大学院生活に関して多大なる援助をして下さった事に深く感謝しています。

大学院修士課程1年の小森信也氏には、同じ分野の研究を行なうにあたって多くの示唆をいただき感謝に絶えません。また、大学院修士課程1年の矢部博明氏、谷口雅昭氏には様々な面で手伝っていただきありがとうございました。

学部4年生の斎藤淳氏、高橋和久氏、腰塚智己氏、杉田雄一郎氏、遠藤隆明氏にも、お世話になりました。

また、研究生の渡辺洋一氏、鳥居龍司氏には社会人としての規範を見せてもらった思いです。

—平成6年2月里見充則—

## 参考文献

- [1] Keld Baden-Kristensen, “音響機器測定に関する応用”, 日本音響学会誌, Vol. 46, No. 8, pp. 662–677, 1990.
- [2] Durand R. Begault and Elizabeth M. Wenzel, “Headphone Localization of Speech”, *HUMAN FUCTORS*, Vol. 35, No. 2, pp. 361–376, 1993.
- [3] イエンスブラウエルト, 森本政之, 後藤敏幸編著, 「空間音響」, 鹿島出版会, 1986.
- [4] Jiashu Chen, Barry D. Van Veen, and Kurt E. Hecox, “Synthesis of 3D Virtual Auditory Space via a Spatial Feature”, In *Proc. of IEEE Virtual Reality Annual International Symposium*, pp. 188–194, 1993.
- [5] Michael Cohen, Nobuo Koizumi, and Shigeaki Aoki, “Design and Control of Shared Conferencing Environments for Audio Telecommunication”, In *Proc. of Second International Symposium on Measurement and Control in Robotics*, pp. 405–412, 1992.
- [6] Carolina Cruz-Neira, Daniel J. Sandin, and Thomas A. DeFanti, “Surround-Screen Projection-Based Virtual Reality: The Design and Implementation of the CAVE”, In *Proc. of COMPUTER GRAPHICS*, pp. 135–142, 1993.
- [7] Crystal River Engineering, Inc., “*The Convolvertron the original 3D audio convolution engine*”, 1993.
- [8] Thomas A. DeFanti, Daniel J. Sandin, and Carolina Cruz-Neira, “A ‘room’ with a ‘view’ ”, *IEEE SPECTRUM*, pp. 30–33, OCTOBER 1993.
- [9] J. L. Flanagan, J. D. Johnston, R. Zahn, and G. W. Elko, “Computer-steered microphone arrays for sound transduction in large rooms”, *J. Acoust. Soc. Am.*, Vol. 78, No. 5, pp. 1508–1518, 1985.

- [10] 藤田一郎, “耳のずれたフクロウ：聴覚空間認識の脳内機構”, 生物物理, Vol. 32, No. 3, pp. 45–51, 1992.
- [11] 畑田豊彦, “広視野動画像による臨場感の客観的測定”, テレビジョン学会技術報告, Vol. 5, No. 29, pp. 55–60, 1981.
- [12] 畑田豊彦, 坂田春夫, “視覚心理とディスプレイ”, テレビジョン学会誌, Vol. 31, No. 4, pp. 460–467, 1977.
- [13] 畑田豊彦, 坂田春夫, 日下秀夫, “画面サイズによる方向感覚誘導効果—大画面による臨場感の基礎実験—”, テレビジョン学会誌, Vol. 33, No. 5, pp. 407–413, 1979.
- [14] 早川和宏, “仮想空間における作業支援環境の開発”, 東京大学工学系研究科修士論文, 平成四年度.
- [15] Tyson R. Henry, Andrey K. Yetts, Scott E. Hudson, Brad A. Myers, and Steven Feiner, “A Nose Gesture Interface Device: Extending Virtual Realities, *PRESENCE*, Vol. 1, No. 2, pp. 258–261, Spring 1992.
- [16] 廣瀬通孝, 「バーチャル・リアリティ応用戦略」, オーム社, 1992.
- [17] 廣瀬通孝, 「バーチャル・リアリティ」, 産業図書, 1993.
- [18] 廣瀬通孝, 佐藤慎一, “人工現実感技術を用いた臨場感の合成とその伝達”, In *Proc. of Human Interface Symposium*, pp. 111–116, 1993.
- [19] Michitaka Hirose, Kensuke Yokoyama, and Shin'ichi Sato, “Transmission of Realistic Sensatio : Development of a Virtual Dome”, In *Proc. of IEEE Virtual Reality Annual International Symposium*, pp. 125–131, 1993.
- [20] 市川保, “空間音響の臨場感通信に関する研究”, 東京大学工学部卒業論文, 平成四年度.

- [21] TEXAS INSTRUMENTS, 「TMS320C25 第2世代デジタル・シグナル・プロセッサユーザーズ・マニュアル」, 日本テキサス・インスツルメンツ株式会社, 1992.
- [22] Géorges Jean, "L'écriture mémoire des hommes", Gallimard, 1987.
- [23] 菅野政明, “仮想環境における作業・動作に関する研究”, 東京大学工学系研究科修士論文, 平成四年度.
- [24] サンドラ・K・ヘルセル, ジュディス・P・ロス編, 廣瀬 通孝監訳, 「バーチャルリアリティ—理論・実践・展望—」, 海文堂出版, 1992.
- [25] Jiaxiang Liu, 阿部正人, 城戸健一, “二つのマイクロホン出力間のクロススペクトルによる音源位置推定法における移動雑音源の影響”, 日本音響学会誌, Vol. 46, No. 7, pp. 523–530, 1990.
- [26] 三浦種敏, “ダミーへッドを用いた測定”, 日本音響学会誌, Vol. 46, No. 8, pp. 633–643, 1990.
- [27] Elizabeth M. Wenzel, “Localization in Virtual Acoustic Display”, *PRESENCE*, Vol. 1, No. 1, pp. 80–107, Winter 1992.
- [28] 永田仁史, 阿部正人, “多数センサによる音源波形の推定”, 日本音響学会誌, Vol. 47, No. 4, pp. 268–273, 1991.
- [29] 永田仁史, 阿部正人, 城戸健一, “多数センサによる音源位置の推定”, 日本音響学会誌, Vol. 46, No. 7, pp. 531–540, 1990.
- [30] 難波清一郎, 後藤敏幸, 橋秀樹, 藤崎博也, 竹下信雄, 井口征士, 桑野園子, 大串健吾, 中村昭, 「音の科学」, 朝倉書店, 1989.
- [31] 及川道雅, “三次元音響ディスプレーの研究”, 東京大学工学部卒業論文, 平成三年度.
- [32] 岡部馨, “ダミーへッドを用いた音場再生”, 日本音響学会誌, Vol. 46, No. 8, pp. 650–656, 1990.

- [33] PIXEL, 「大型特集 IRIS ワークステーションのすべて」, 6 1993, 図形情報処理センター.
- [34] Stephen T. Pope and Lennart E. Fahlen, "The Use of 3-D Audio in a Synthetic Environment: An Aural Renderer for a Distributed Virtual Reality System", In *Proc. of IEEE Virtual Reality Annual International Symposium*, pp. 176–182, 1993.
- [35] 心理学実験指導研究会編, 「実験とテスト – 心理学の基礎 (解説編)」, 培風館, 1985.
- [36] 心理学実験指導研究会編, 「実験とテスト – 心理学の基礎 (実習編)」, 培風館, 1985.
- [37] Silicon Graphics, Inc., "Graphics Library Programming Guide(日本語版)", 1991.
- [38] Jon M. Speigle and Jack M. Loomis, "Auditory Distance Perception by Translating Observers", In *Proc. of IEEE Symposium on Research Frontiers in Virtual Reality*, 1993.
- [39] 館すすむ, "レイジスタンス", 計測と制御, Vol. 30, No. 6, pp. 465–471, 1991.
- [40] 館すすむ, 「人工現実感」, 日刊工業新聞社, 1992.
- [41] 館すすむ, 廣瀬 通孝 監修著, 「バーチャル・テック・ラボ」, 工業調査会, 1992.
- [42] 高田将年, "聴覚ディスプレイに関する研究", 東京大学工学部卒業論文, 平成四年度.
- [43] Tapio Takala and James Hahn, "Sound Rendering", In *Proc. SIGGRAPH'92*, volume 26, pp. 211–219. ACM Computer Graphics, 1992.

- [44] 田中良久, 「心理学的測定法」, 東京大学出版会, 1977.
- [45] Elizabeth M. Wenzel and Scott H. Foster, "PERCEPTUAL CONSEQUENCES OF INTERPOLATING HEAD-RELATED TRANSFER FUNCTIONS DURING SPATIAL SYNTHESIS", In *Proc. of the 1993 Workshop on Applications of Signal Processing to Audio and Acoustics*, 1993.
- [46] Frederic L. Wightman, "Headphone simulation of free-field listening.I:Stimulus synthesis,II:Psychophysical validation", *J. Acoustic Soc. Am.*, Vol. 85, No. 2, pp. 858–878, 1989.
- [47] Yoshio Yamasaki and Takeshi Itow, "Measurement of spatial information in sound fields by closely located four point microphone method", *J. Acoust. Soc. Jpn.*, Vol. 10, No. 2, pp. 101–110, 1989.
- [48] Hirofumi Yanagawa, "Effect of transient signal length on cross-correlation functions in a room", *J. Acoust. Soc. Am.*, Vol. 84, No. 5, pp. 1728–1733, 1988.
- [49] 横山賢介, "視覚的臨場感の合成と伝達に関する研究", 東京大学工学系研究科修士論文, 平成四年度.
- [50] Tatsuhiko Yonekura, Nobuyuki Ariyoshi, and Yoshiki Watanabe, "ASPECT:Audio SPatial Environment for Communication", In *Proc. of IEEE Virtual Reality Annual International Symposium*, pp. 183–187, 1993.
- [51] Eberhard Zwicker, 「音響心理学」, 西村書店, 1992.