

波面合成法による立体音場再生における  
情報量削減手法に関する研究

木村 敏幸

名古屋大学大学院人間情報学研究科

2005 年度



# 目次

第 1 章	序論	1
1.1	研究背景	1
1.1.1	立体音響技術	1
1.1.2	立体音場再生技術	3
1.2	本論文の目的	13
1.3	本論文の構成	14
第 2 章	波面合成シミュレーション	15
2.1	はじめに	15
2.2	波面合成実験	16
2.2.1	実験環境	16
2.2.2	実験結果及び考察	20
2.3	残響空間における波面合成の条件	24
2.4	まとめ	25
第 3 章	チャンネル信号数の削減	27
3.1	はじめに	27
3.2	主観評価実験 1 - 方向感への影響 -	28
3.2.1	チャンネル信号の作成	28
3.2.2	実験環境	29
3.2.3	実験計画	34
3.2.4	実験結果及び考察	34
3.3	主観評価実験 2 - 空間印象への影響 -	42
3.3.1	チャンネル信号の作成	42
3.3.2	実験環境	43

---

3.3.3	実験計画	47
3.3.4	実験結果及び考察	49
3.4	まとめ	49
第4章	音源抽出に基づく伝送量の削減	57
4.1	はじめに	57
4.2	アルゴリズム	58
4.2.1	音源信号の抽出	58
4.2.2	チャンネル信号の復元	62
4.3	符号化実験	62
4.3.1	室内インパルス応答の測定	63
4.3.2	チャンネル信号の作成	64
4.3.3	逆フィルタの算出	64
4.3.4	符号化及び復号化	65
4.3.5	逆フィルタによる符号化への影響	65
4.4	主観評価実験	68
4.4.1	実験環境	68
4.4.2	実験計画	68
4.4.3	実験手順	69
4.4.4	実験結果及び考察	70
4.5	まとめ	73
第5章	移動音源抽出に基づく伝送量の削減	75
5.1	はじめに	75
5.2	アルゴリズム	75
5.2.1	移動音源信号の抽出	76
5.2.2	チャンネル信号の復元	79
5.3	符号化実験	80
5.3.1	チャンネル信号の作成	80
5.3.2	室内インパルス応答データベースの算出	83
5.3.3	音源信号の抽出及びチャンネル信号の復元	83
5.4	主観評価実験	84
5.4.1	実験環境	84



---

5.4.2	実験計画	85
5.4.3	実験手順	86
5.4.4	実験結果及び考察	87
5.5	まとめ	89
第 6 章	結論	91
6.1	要約	91
6.2	今後の展望	92
	謝辞	95
	参考文献	97
付録 A	合成した波面の一例	101
	研究業績	139



# 目次

1.1	2チャンネルシステムの構成	2
1.2	5チャンネルシステムの構成	3
1.3	バイノーラルシステムの構成	4
1.4	トランスオーラルシステムの構成	5
1.5	MINT法に基づいたトランスオーラルシステムの構成	6
1.6	最小ノルム解に基づいたトランスオーラルシステムの構成	8
1.7	Bell研究所による立体音場再生システムの構成	9
1.8	Camrasによる波面合成システムの構成	10
1.9	Wave Field Synthesisシステムの構成	11
1.10	二次音源による立体音場再生システムの構成	12
2.1	本論文で検討する波面合成システムの構成	16
2.2	波面合成実験における実験環境	17
2.3	波面合成実験に用いたマイクロホンの指向特性	19
2.4	波面合成実験におけるSNRの結果(無指向性)	21
2.5	波面合成実験におけるSNRの結果(単一指向性)	22
2.6	波面合成実験におけるSNRの結果(超指向性)	23
2.7	波面合成実験における平均SNRの結果	24
3.1	方向感の主観評価における原音場	28
3.2	方向感の主観評価における実験環境(中心, 前方位置)	30
3.3	方向感の主観評価における実験環境(後方, 側方位置)	31
3.4	方向感の主観評価における実験条件(中心位置の場合)	32
3.5	方向感の主観評価における実験条件(側方位置の場合)	33
3.6	方向感の主観評価における実験流れ図	34

3.7	方向感の主観評価における実験結果 (中心位置の場合) . . . . .	35
3.8	方向感の主観評価における実験結果 (前方位置の場合) . . . . .	36
3.9	方向感の主観評価における実験結果 (後方位置の場合) . . . . .	37
3.10	方向感の主観評価における実験結果 (側方位置の場合) . . . . .	38
3.11	定位方向の偏位 (中心位置, チャネル信号の数 8, 提示方向 15° の場合) .	40
3.12	方向感の主観評価における平均二乗誤差 . . . . .	41
3.13	空間印象の主観評価における原音場 . . . . .	42
3.14	空間印象の主観評価における実験環境 . . . . .	44
3.15	空間印象の主観評価における実験条件 (中心位置の場合) . . . . .	45
3.16	空間印象の主観評価における実験条件 (側方位置の場合) . . . . .	46
3.17	方向感の主観評価における実験流れ図 . . . . .	47
3.18	見かけの音源の幅 (ASW) と音に包まれた感じ (LEV) の概念図 . . . . .	48
3.19	空間印象の主観評価における実験結果 (中心, 前方位置) . . . . .	54
3.20	空間印象の主観評価における実験結果 (後方, 側方位置) . . . . .	55
4.1	提案手法のブロック図 . . . . .	59
4.2	音源抽出のブロック図 . . . . .	59
4.3	チャネル復元のブロック図 . . . . .	62
4.4	音源用スピーカとマイクロホンアレーの配置図 . . . . .	63
4.5	符号化遅延が 320 ms (=3840 samples) の時の逆フィルタの算出結果 . . .	66
4.6	符号化遅延が 40 ms (=480 samples) の時の逆フィルタの算出結果 . . . . .	67
4.7	スピーカアレーと聴取者の位置 . . . . .	68
4.8	主観評価実験の流れ図 . . . . .	70
4.9	音場評価の結果 . . . . .	72
4.10	全体評価の結果 . . . . .	72
5.1	提案手法のブロック図 . . . . .	76
5.2	窓関数の形状 . . . . .	78
5.3	抽出及び復元のブロック図 . . . . .	79
5.4	符号化実験に用いた原音場 . . . . .	80
5.5	スピーカアレーと聴取者の位置 . . . . .	85
5.6	主観評価実験の流れ図 . . . . .	86
5.7	動き評価の結果 . . . . .	88

---

5.8	全体評価の結果 . . . . .	89
A.1	合成した波面の一例 (無指向性, $f=125$ Hz, $d=3$ m) . . . . .	102
A.2	合成した波面の一例 (無指向性, $f=125$ Hz, $d=10$ m) . . . . .	103
A.3	合成した波面の一例 (無指向性, $f=250$ Hz, $d=3$ m) . . . . .	104
A.4	合成した波面の一例 (無指向性, $f=250$ Hz, $d=10$ m) . . . . .	105
A.5	合成した波面の一例 (無指向性, $f=500$ Hz, $d=3$ m) . . . . .	106
A.6	合成した波面の一例 (無指向性, $f=500$ Hz, $d=10$ m) . . . . .	107
A.7	合成した波面の一例 (無指向性, $f=1000$ Hz, $d=3$ m) . . . . .	108
A.8	合成した波面の一例 (無指向性, $f=1000$ Hz, $d=10$ m) . . . . .	109
A.9	合成した波面の一例 (無指向性, $f=2000$ Hz, $d=3$ m) . . . . .	110
A.10	合成した波面の一例 (無指向性, $f=2000$ Hz, $d=10$ m) . . . . .	111
A.11	合成した波面の一例 (無指向性, $f=4000$ Hz, $d=3$ m) . . . . .	112
A.12	合成した波面の一例 (無指向性, $f=4000$ Hz, $d=10$ m) . . . . .	113
A.13	合成した波面の一例 (単一指向性, $f=125$ Hz, $d=3$ m) . . . . .	114
A.14	合成した波面の一例 (単一指向性, $f=125$ Hz, $d=10$ m) . . . . .	115
A.15	合成した波面の一例 (単一指向性, $f=250$ Hz, $d=3$ m) . . . . .	116
A.16	合成した波面の一例 (単一指向性, $f=250$ Hz, $d=10$ m) . . . . .	117
A.17	合成した波面の一例 (単一指向性, $f=500$ Hz, $d=3$ m) . . . . .	118
A.18	合成した波面の一例 (単一指向性, $f=500$ Hz, $d=10$ m) . . . . .	119
A.19	合成した波面の一例 (単一指向性, $f=1000$ Hz, $d=3$ m) . . . . .	120
A.20	合成した波面の一例 (単一指向性, $f=1000$ Hz, $d=10$ m) . . . . .	121
A.21	合成した波面の一例 (単一指向性, $f=2000$ Hz, $d=3$ m) . . . . .	122
A.22	合成した波面の一例 (単一指向性, $f=2000$ Hz, $d=10$ m) . . . . .	123
A.23	合成した波面の一例 (単一指向性, $f=4000$ Hz, $d=3$ m) . . . . .	124
A.24	合成した波面の一例 (単一指向性, $f=4000$ Hz, $d=10$ m) . . . . .	125
A.25	合成した波面の一例 (超指向性, $f=125$ Hz, $d=3$ m) . . . . .	126
A.26	合成した波面の一例 (超指向性, $f=125$ Hz, $d=10$ m) . . . . .	127
A.27	合成した波面の一例 (超指向性, $f=250$ Hz, $d=3$ m) . . . . .	128
A.28	合成した波面の一例 (超指向性, $f=250$ Hz, $d=10$ m) . . . . .	129
A.29	合成した波面の一例 (超指向性, $f=500$ Hz, $d=3$ m) . . . . .	130
A.30	合成した波面の一例 (超指向性, $f=500$ Hz, $d=10$ m) . . . . .	131
A.31	合成した波面の一例 (超指向性, $f=1000$ Hz, $d=3$ m) . . . . .	132

---

A.32	合成した波面の一例 (超指向性, $f=1000$ Hz, $d=10$ m) . . . . .	133
A.33	合成した波面の一例 (超指向性, $f=2000$ Hz, $d=3$ m) . . . . .	134
A.34	合成した波面の一例 (超指向性, $f=2000$ Hz, $d=10$ m) . . . . .	135
A.35	合成した波面の一例 (超指向性, $f=4000$ Hz, $d=3$ m) . . . . .	136
A.36	合成した波面の一例 (超指向性, $f=4000$ Hz, $d=10$ m) . . . . .	137

# 表目次

2.1	波面合成実験における実験条件	18
3.1	方向感の主観評価における試行条件	39
3.2	Scheffé の一対比較法の評定表	48
3.3	聴取者の識別結果 (中心位置の場合)	50
3.4	聴取者の識別結果 (前方位置の場合)	51
3.5	聴取者の識別結果 (後方位置の場合)	52
3.6	聴取者の識別結果 (側方位置の場合)	53
4.1	室内インパルス応答の測定条件	64
4.2	マルチチャネル信号の作成条件	64
4.3	逆フィルタの算出条件	65
4.4	刺激の組み合わせ	69
4.5	違いの評定表	70
4.6	二項検定の結果	71
5.1	チャンネル信号の作成条件	83
5.2	逆フィルタの算出条件	84
5.3	符号化音の種類	86
5.4	違いの評定表	87
5.5	各被験者の識別結果	87





# 第 1 章

## 序論

本論文の研究テーマは波面合成法による立体音場再生技術における情報量の削減である。1.1 節ではこの研究テーマに至るまでの研究背景を論じた上で波面合成法による立体音場再生技術について説明する。また、1.2 節ではこの技術の実用化の際に生じる問題点について説明し、その解決方法を検討する。

### 1.1 研究背景

人々は以前から演劇、オーケストラ演奏及び映画を楽しんできた。しかし、それらは主に劇場や映画館で開催するのが一般的であり、それらを楽しむためには開催されている時期に劇場や映画館にまでわざわざ足を運ぶ必要があった。

そこで、技術者たちは演劇、オーケストラ演奏及び映画をいつでも好きな時間に家庭のような場所でも楽しめるようにしたいと考えようになった。このようなことを実現するためには家庭に居ながらにしてあたかも演劇やオーケストラ演奏が行われている劇場や映画が上映されている映画館に居るような印象(この印象のことを「臨場感」という)を家庭にいる人々に与える技術を開発する必要があった。そのような臨場感を聴覚的に与える技術として立体音響技術や立体音場再生技術が開発されてきた。次節以降では現在に至るまでに開発されてきた主な立体音響技術や立体音場再生技術について概説する。

#### 1.1.1 立体音響技術

立体音響技術の開発の歴史は 1930 年代にまで遡る。その当時に開発されたなかでの代表的なシステムは図 1.1 に示すような 2 つのスピーカを用いたものである。このシステムは一般的には「ステレオホニック」[1] と呼ばれている。このシステムは 2 つのスピーカから同じような音を同時に再生すると、スピーカの間で 1 個の音が鳴っているように聞こ

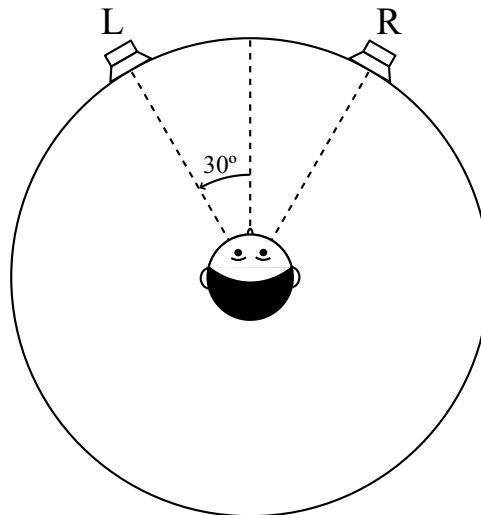


図 1.1 2 チャンネルシステムの構成

えるという現象（「合成音像定位」[2] と呼ぶ）を利用している。従って、2 つのスピーカで鳴らす信号にレベル差を設けることによって、2 つのスピーカの上に劇場の舞台のような音場を構成することができる。ゆえに、このシステムのことを「レベル差ステレオ」と呼ぶこともある。このシステムは音楽製作の分野では最も基本的な立体音響技術であり、現在でもアナログレコードや CD にまで広く応用され、普及している。

しかし、上記のシステムは構成される音場を聴取者の前方だけに限定したものである。それに対して、劇場で観賞する場合は、聴取者は劇場の壁面によって反射された音を側方や後方からも聞き、残響のような空間的な印象を感じている。そこで、このような空間印象を聴取者が感じとれるようにするための立体音響技術がその後研究されてきた。その過程で Ambisonic[3] や 3-1 方式 [4] などが開発されていったが、1990 年代に図 1.2 に示すような 5 個のスピーカを用いたシステムが開発され、ITU-R によって勧告された [5]。このシステムは図 1.1 のシステムに正面方向のスピーカと 2 個の後方のスピーカを加えたものである。その結果、前方に 3 個、後方に 2 個のスピーカが存在するので、このシステムのことを「3-2 方式」と呼ぶこともある。このシステムは映画や放送の分野では最も基本的な立体音響技術となり、音像感には寄与しない 120Hz 以下の低音を鳴らすための専用のスピーカ (0.1ch) を加えた「5.1ch」システムとして DVD-video やデジタル放送に応用され、「ホームシアター」として家庭にも広く普及している。

しかし、このシステムは聴取者が図に示されるような円の中心付近に居ることを想定している。言い換えれば聴取者の位置に対する制約がある。また、全てのスピーカを聴取者の耳と同じ高さに配置することを前提としているため、上下方向に対する音像の移動

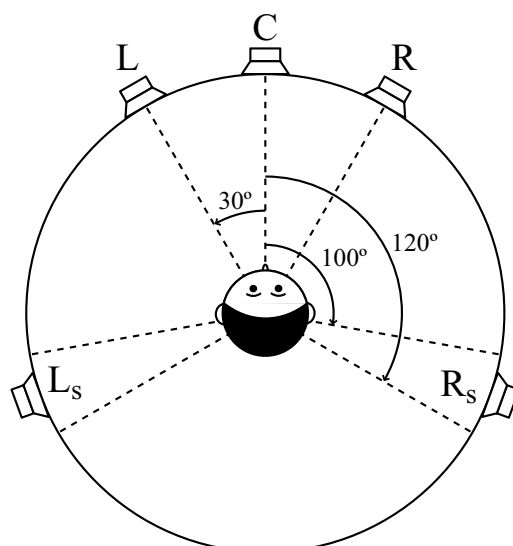


図 1.2 5チャンネルシステムの構成

感を表現することができない．そこで，最近では最適な聴取範囲を拡大したり上下の移動感を表現したりするために聴取者の上下にもスピーカを配置したシステム [6] が提案されている．このシステムは 22 個のスピーカと 2 個の低音専用のスピーカを用いるので，「22.2ch」システムと呼ばれている．

### 1.1.2 立体音場再生技術

このように 1.1.1 節で述べた立体音響技術は比較的少数のスピーカで構成されるため，現在に至るまで様々な分野に応用されている．しかし，これらの技術は音場を表現することだけに着目しているため，スピーカから再生する信号としては各音源のそばにマイクロホンを設置し，収録した音をミキシングによって合成した音が用いられることが多い．そのため，聴取者に臨場感を与えられるかどうかはミキシングを担当する技術者の力量に依存する部分が大きくなり，真の臨場感を再現しているとは言い難いのが現状である．そこで，音場を忠実に再現することによって聴取者が先に述べた技術よりも一層臨場感を感じとれるようにするために，音場の収録と再生を一体化した立体音場再生技術が盛んに研究されている．

立体音場再生技術は主に 2 種類の手法に分類される．一つのアプローチとしては聴取者の両耳の音圧を制御する手法が挙げられる．中でも，最も基本的な手法は 1933 年に Bell 研究所の Fletcher によって開発された手法で「バイノーラルシステム」[7] と呼ばれている．バイノーラルシステムの構成を図 1.3 に示す．まず，原音場に聴取者と同じ形状をし

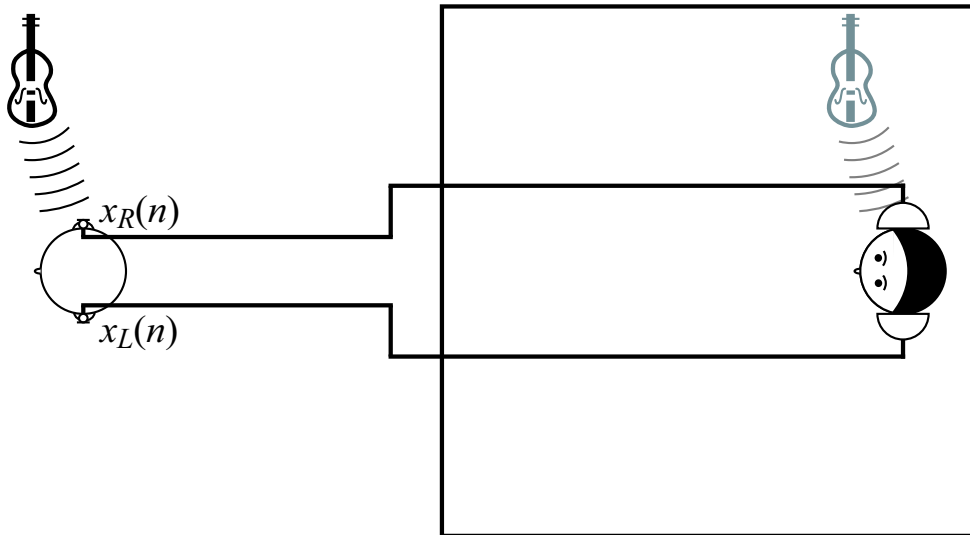


図 1.3 バイノーラルシステムの構成

たダミーヘッドを配置し，ダミーヘッドの両耳位置にマイクロホンを取り付ける．そして，マイクロホンによって原音場の音  $x_L(n)$ ， $x_R(n)$  を収録する．次に，聴取者にヘッドホンを取り付け，ヘッドホンから先ほど収録した音を再生する．すると，聴取者の両耳に到達するまでの音響的な影響（頭部による音の反射や回折など）が収録された音にあらかじめ加味されているため，聴取者はあたかも原音場で音を聞いているような臨場感を感じるようになる．そのため，聴取者はヘッドホンをしているにもかかわらず，頭の外で音が鳴っているように聞こえる（「頭外定位」と呼ぶ）．この現象は CD などの音をヘッドホンで聞いたときに頭の中で音が鳴っているように感じる現象（「頭内定位」と呼ぶ）とは明らかに異なるものである．

しかし，ダミーヘッドは一般的に汎用のものを用いるため，用いたダミーヘッドの形状が聴取者と異なっている場合には頭内定位となってしまうことが多い．また，仮にダミーヘッドの形状が聴取者と同じであったとしても，正面方向を含む正中面に位置する音像に対しては頭内定位が起きやすい．そこで，頭内定位を起こさないようにするために，ヘッドホンの代わりに 2 個のスピーカを用いて再生を行うシステムが提案された．このシステムはバイノーラルシステムと区別するために一般的に「トランスオーラルシステム」[8]と呼ばれている．トランスオーラルシステムの構成を図 1.4 に示す．まず，バイノーラルシステムと同じように原音場の音  $x_L(n)$ ， $x_R(n)$  を収録する．一方，再生音場において各スピーカから聴取者の両耳までのインパルス応答  $g_{L1}(n)$ ， $g_{L2}(n)$ ， $g_{R1}(n)$ ， $g_{R2}(n)$  をあらかじめ測定しておき，得られた室内インパルス応答から逆フィルタ  $h_{1L}(n)$ ， $h_{1R}(n)$ ， $h_{2L}(n)$ ， $h_{2R}(n)$  を算出する．そして，収録した音に逆フィルタを畳み込んで作成した音  $y_1(n)$ ， $y_2(n)$

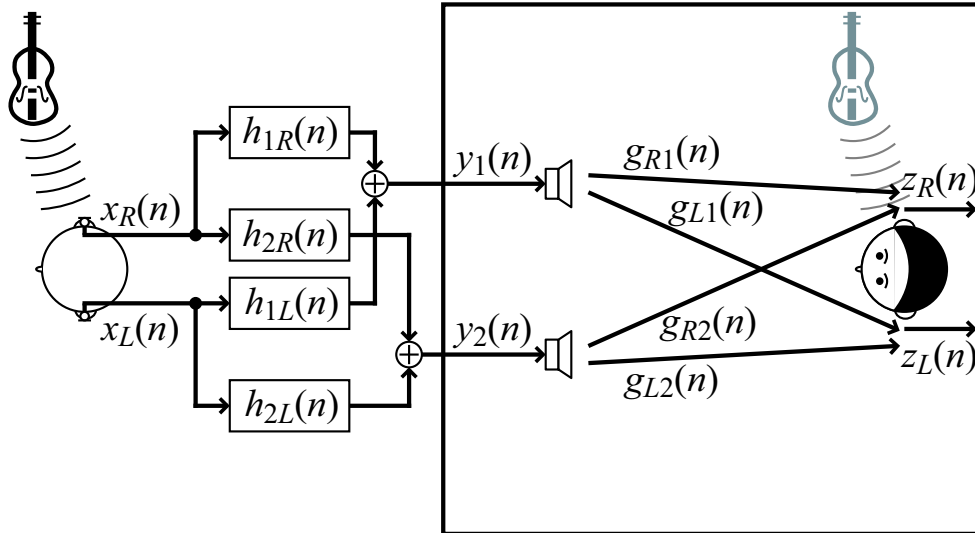


図 1.4 トランスオーラルシステムの構成

を各スピーカから再生する．すると，バイノーラルシステムと同じように聴取者はあたかも原音場で音を聞いているような臨場感を感じるようになる．

逆フィルタの算出は以下のようにして行う．まず，各スピーカから再生される音  $y_1(n)$ ， $y_2(n)$  は以下のようにして作成される．

$$\begin{aligned} y_1(n) &= h_{1L}(n) * x_L(n) + h_{1R}(n) * x_R(n) \\ y_2(n) &= h_{2L}(n) * x_L(n) + h_{2R}(n) * x_R(n) \end{aligned} \quad (1.1)$$

但し， $*$  は畳み込み演算を表す．上記の式を離散 Fourier 変換すると，以下ようになる．

$$\begin{aligned} Y_1(\omega) &= H_{1L}(\omega)X_L(\omega) + H_{1R}(\omega)X_R(\omega) \\ Y_2(\omega) &= H_{2L}(\omega)X_L(\omega) + H_{2R}(\omega)X_R(\omega) \end{aligned} \quad (1.2)$$

一方，聴取者が聞く音  $z_L(n)$ ， $z_R(n)$  は  $y_1(n)$ ， $y_2(n)$  から以下のように表される．

$$\begin{aligned} Z_L(\omega) &= G_{L1}(\omega)Y_1(\omega) + G_{L2}(\omega)Y_2(\omega) \\ Z_R(\omega) &= G_{R1}(\omega)Y_1(\omega) + G_{R2}(\omega)Y_2(\omega) \end{aligned} \quad (1.3)$$

但し， $G_{L1}(\omega)$ ， $G_{L2}(\omega)$ ， $G_{R1}(\omega)$ ， $G_{R2}(\omega)$  は  $g_{L1}(n)$ ， $g_{L2}(n)$ ， $g_{R1}(n)$ ， $g_{R2}(n)$  の離散 Fourier 変換を表す．(1.2) 式と (1.3) 式から以下のような行列関係式が導かれる．

$$\begin{aligned} \begin{pmatrix} Z_L(\omega) \\ Z_R(\omega) \end{pmatrix} &= \begin{pmatrix} G_{L1}(\omega) & G_{L2}(\omega) \\ G_{R1}(\omega) & G_{R2}(\omega) \end{pmatrix} \begin{pmatrix} H_{1L}(\omega) & H_{1R}(\omega) \\ H_{2L}(\omega) & H_{2R}(\omega) \end{pmatrix} \begin{pmatrix} X_L(\omega) \\ X_R(\omega) \end{pmatrix} \\ \mathbf{Z}(\omega) &= \mathbf{G}(\omega)\mathbf{H}(\omega)\mathbf{X}(\omega) \end{aligned} \quad (1.4)$$

ここで  $\mathbf{Z}(\omega)=\mathbf{X}(\omega)$  となるためには， $\mathbf{H}(\omega)$  が  $\mathbf{G}(\omega)$  の逆行列

$$\begin{aligned} \mathbf{G}^{-1}(\omega) &= \frac{1}{G_{00}(\omega)} \begin{pmatrix} G_{R2}(\omega) & -G_{L2}(\omega) \\ -G_{R1}(\omega) & G_{L1}(\omega) \end{pmatrix} \\ G_{00}(\omega) &= G_{L1}(\omega)G_{R2}(\omega) - G_{R1}(\omega)G_{L2}(\omega) \end{aligned} \quad (1.5)$$

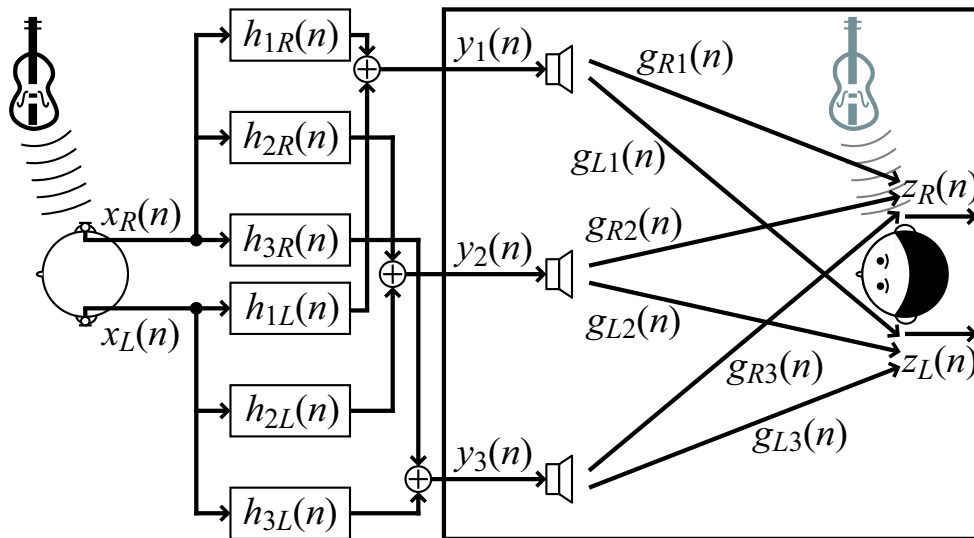


図 1.5 MINT 法に基づいたトランスオーラルシステムの構成

となればよい．従って，時間領域における逆フィルタは以下のように算出される．

$$\begin{aligned}
 h_{1L}(n) &= g_{R2}(n) * d(n) \\
 h_{1R}(n) &= -g_{L2}(n) * d(n) \\
 h_{2L}(n) &= -g_{R1}(n) * d(n) \\
 h_{2R}(n) &= g_{L1}(n) * d(n)
 \end{aligned} \tag{1.6}$$

但し， $d(n)$  は  $1/G_{00}(\omega)$  の逆離散 Fourier 変換である．

再生音場が比較的残響の少ない部屋である場合は容易に逆フィルタが算出できるため，トランスオーラルシステムはコンサートホールの比較研究 [9] などに応用されている．しかしながら，再生音場が残響の多い部屋である場合は一般に  $G_{00}(\omega) \approx 0$  となり，逆フィルタが発散してしまうことが多くなるため，逆フィルタを算出することは容易ではない．一方， $M+1$  個以上のスピーカを用いて  $M$  点の音圧を自由に制御する手法（「MINT 法」[10] と呼ばれる）が提案されており，この手法に基づいて 3 個のスピーカを用いて両耳 2 点の音圧を制御するトランスオーラルシステム [11] が提案されている．システムの構成を図 1.5 に示す．逆フィルタは以下の要件を満たすように作成する．

$$\begin{aligned}
 1 &= g_{L1}(n) * h_{1L}(n) + g_{L2}(n) * h_{2L}(n) + g_{L3}(n) * h_{3L}(n) \\
 0 &= g_{R1}(n) * h_{1L}(n) + g_{R2}(n) * h_{2L}(n) + g_{R3}(n) * h_{3L}(n) \\
 0 &= g_{L1}(n) * h_{1R}(n) + g_{L2}(n) * h_{2R}(n) + g_{L3}(n) * h_{3R}(n) \\
 1 &= g_{R1}(n) * h_{1R}(n) + g_{R2}(n) * h_{2R}(n) + g_{R3}(n) * h_{3R}(n)
 \end{aligned} \tag{1.7}$$

ここで室内インパルス応答の長さを  $L_g+1$ ，逆フィルタの長さを  $L_h+1$  とすると，(1.7) 式

は以下のような行列関係式に書き換えることができる．

$$\begin{pmatrix} \mathbf{d}_{LL} & \mathbf{d}_{LR} \\ \mathbf{d}_{RL} & \mathbf{d}_{RR} \end{pmatrix} = \begin{pmatrix} \mathbf{g}_{L1} & \mathbf{g}_{L2} & \mathbf{g}_{L3} \\ \mathbf{g}_{R1} & \mathbf{g}_{R2} & \mathbf{g}_{R3} \end{pmatrix} \begin{pmatrix} \mathbf{h}_{1L} & \mathbf{h}_{1R} \\ \mathbf{h}_{2L} & \mathbf{h}_{2R} \\ \mathbf{h}_{3L} & \mathbf{h}_{3R} \end{pmatrix} \quad (1.8)$$

$$\mathbf{d} = \mathbf{g}\mathbf{h}$$

但し， $\mathbf{g}_{ji}$ ， $\mathbf{h}_{ij}$ ， $\mathbf{d}_{jj}(i=1\dots 3, j=L, R)$  は以下のように表される．

$$\mathbf{g}_{ji} = \begin{pmatrix} g_{ji}(0) & 0 & \dots & 0 \\ g_{ji}(1) & g_{ji}(0) & \ddots & \vdots \\ \vdots & g_{ji}(1) & \ddots & 0 \\ g_{ji}(L_g) & \vdots & \ddots & g_{ji}(0) \\ 0 & g_{ji}(L_g) & \ddots & g_{ji}(1) \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & g_{ji}(L_g) \end{pmatrix} \quad (L_g + L_h + 1 \times L_h + 1) \quad (1.9)$$

$$\mathbf{h}_{ij} = \begin{pmatrix} h_{ij}(0) \\ h_{ij}(1) \\ \vdots \\ h_{ij}(L_h) \end{pmatrix} \quad (L_h + 1 \times 1)$$

$$\mathbf{d}_{LL(RR)} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad \mathbf{d}_{LR(RL)} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad (L_g + L_h + 1 \times 1)$$

ゆえに， $\mathbf{g}$  は  $2(L_g + L_h + 1)$  行  $3(L_h + 1)$  列の行列となる．ここで  $L_h = 2L_g - 1$  とすると， $\mathbf{g}$  は  $6L_g$  行  $6L_g$  列の正方行列となる．そして，全ての室内インパルス応答が共通の零点を持たなければ  $\mathbf{g}$  は正則行列となる．ゆえに，逆フィルタは以下のように算出できる．

$$\mathbf{h} = \mathbf{g}^{-1}\mathbf{d} \quad (1.10)$$

しかし，MINT 法は時間領域で直接逆フィルタを算出するので，再生音場が残響の多い部屋である場合計算量が非常に多くなる．そこで，計算量を少なくするために，複数個のスピーカを用い，周波数領域において逆フィルタを最小ノルム解を用いて算出するトランスオーラルシステム [12] が提案されている．システムの構成を図 1.6 に示す．逆フィルタは以下の式を満たすように算出する．

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} G_{L1}(\omega) & G_{L2}(\omega) & \dots & G_{LM}(\omega) \\ G_{R1}(\omega) & G_{R2}(\omega) & \dots & G_{RM}(\omega) \end{pmatrix} \begin{pmatrix} H_{1L}(\omega) & H_{1R}(\omega) \\ H_{2L}(\omega) & H_{2R}(\omega) \\ \vdots & \vdots \\ H_{ML}(\omega) & H_{MR}(\omega) \end{pmatrix} \quad (1.11)$$

$$\mathbf{D} = \mathbf{G}(\omega)\mathbf{H}(\omega)$$

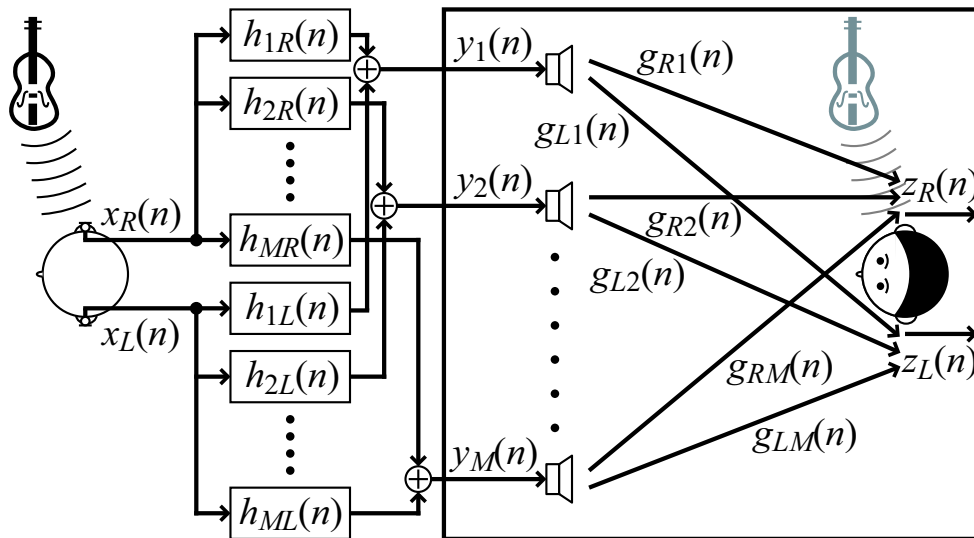


図 1.6 最小ノルム解に基づいたトランスオーラルシステムの構成

このとき、 $\mathbf{G}(\omega)$  は正則行列ではないので、最小ノルム解を用いて逆フィルタを算出する。

$$\mathbf{H}(\omega) = \mathbf{G}^H(\omega)[\mathbf{G}(\omega)\mathbf{G}^H(\omega)]^{-1}\mathbf{D} \quad (1.12)$$

但し、 $\mathbf{G}^H(\omega)$  は  $\mathbf{G}(\omega)$  の共役転置行列である。最小ノルム解を用いて逆フィルタを算出した場合、スピーカの数が多いほど逆フィルタが安定に算出されることが分かっている [12]。また、このトランスオーラルシステムにおける最適なスピーカ配置も検討されている [13]。一方で、聴取者の前方に 2 個のスピーカを近接して配置すれば逆フィルタが安定に算出できるという利点を生かしたトランスオーラルシステム（「ステレオダイポール」 [14] と呼ばれる）も提案されている。

先程まで述べてきた両耳制御手法は聴取者は再生音場のある特定の位置に居て、頭部を動かしていないことを前提条件としている。そのため、聴取者が自分の頭を動かしたり席を移動したりすると臨場感が低下してしまったり、場合によっては音がどの方向から聞こえているのか全く分からなくなってしまうこともある。また、この手法によって多人数の聴取者に聞かせたい場合にはヘッドホンやトランスオーラルシステム用の逆フィルタを人数分用意してやる必要がある。

一方、立体音場再生技術のもう一つのアプローチとして聴取者の周りの領域の音圧を制御する手法が研究されてきた。この手法を用いると、聴取者は聴取領域の中に居る限り頭部を動かしたり席を移動したりしても臨場感を感じることができる。また、制御領域を広くしてやれば、多人数の聴取者が同時に臨場感を感じるシステムを構築することも可能になる。この手法に関する試みが最初に行われたのは 1933 年 4 月 27 日、Bell 研究所の研究者たちによるもので [15]、彼らは Philadelphia の Academy of Science Hall における



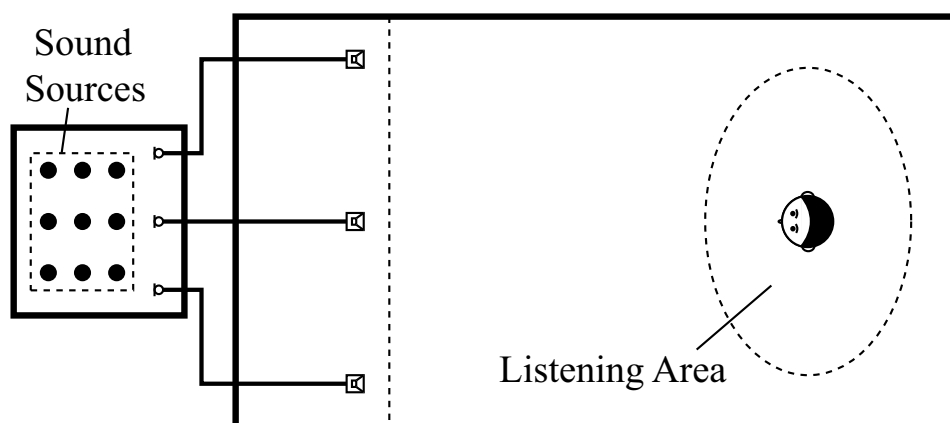


図 1.7 Bell 研究所による立体音場再生システムの構成

Philadelphia Symphony Orchestra の演奏を Washington の Constitution Hall にまで伝送し、再生することを試みた。彼らが構築したシステムの構成を図 1.7 に示す [16, 17]。まず、Philadelphia の Academy of Science Hall のステージ前方に 3 本のマイクロホンを設置し、オーケストラの音を収録する。そして、電話回線で伝送し、Washington の Constitution Hall のステージ前方に客席に向かって配置した 3 個のスピーカで再生する。その結果、聴取者は Constitution Hall で再生された各音源の位置を横方向や奥行方向においても識別することができた [16]。

その後、マイクロホンとスピーカの位置を同一にすれば Huygens の原理に従って波面が合成されることが伊藤によって示され [18]、波面合成法による立体音場再生システムが Camras によって提案された [19]。Camras による波面合成システムの構成を図 1.8 に示す。まず、原音場となる部屋に制御領域を想定し、制御領域の境界面上にマイクロホンを配置する。その際、カージオイド型の指向性マイクロホンを制御領域の外側に向けて配置する。一方、再生音場内に制御領域と同じ形をした聴取領域を想定し、マイクロホンと同じ位置にスピーカを配置する。そして、マイクロホンで収録した音をスピーカから再生すると、Huygens の原理に従って聴取領域内に制御領域の波面が合成される。Camras は原音場にコンサートホールのような非常に残響の多い空間を用い、会議室や屋外のような比較的残響の少ない空間にスピーカを配置して聴取者に音を聞かせたところ、聴取領域内にいる聴取者は比較的残響の少ない空間にいるにもかかわらず、あたかもコンサートホールにいるような印象を受けた [19]。

Huygens の原理は Kirchhoff-Helmholtz 積分方程式によって定式化されている。全ての音源が閉空間  $V$  の外側にあり、角速度  $\omega$  で振動する正弦波を放射している場合、閉空間

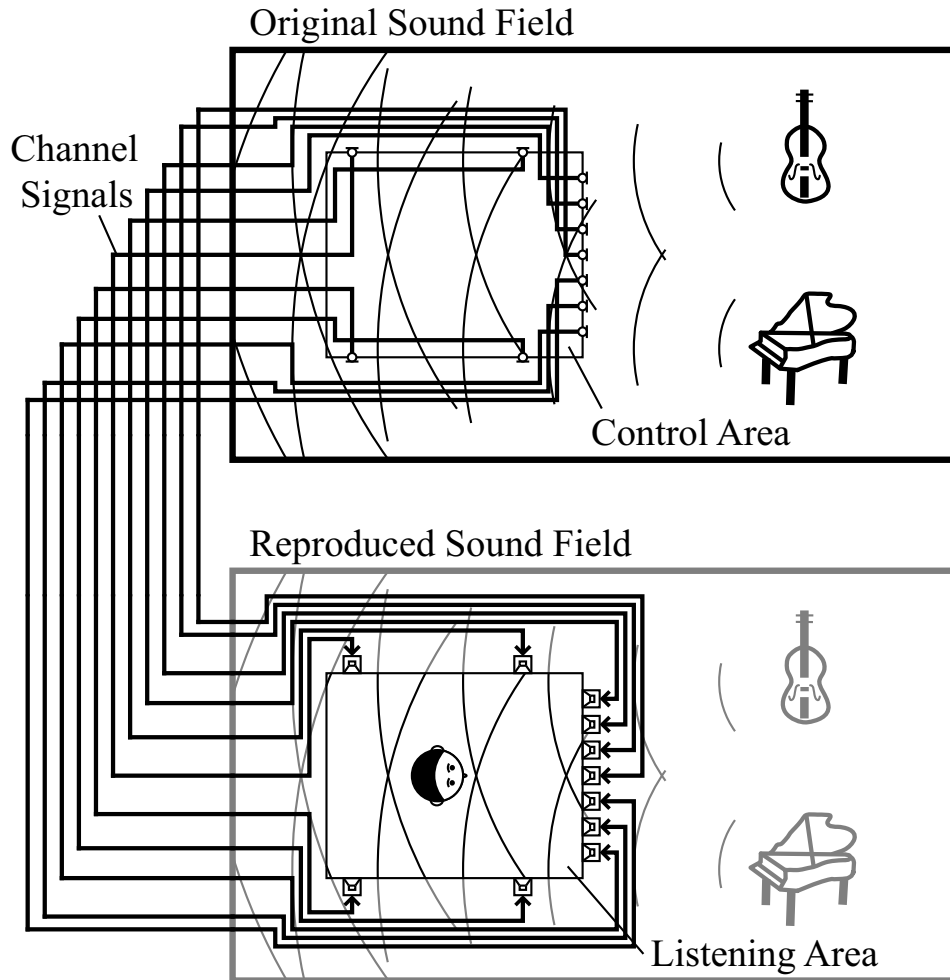


図 1.8 Camras による波面合成システムの構成

$V$  内の音圧  $P(\mathbf{r}, \omega)$  は以下の式によって表わされる .

$$P(\mathbf{r}, \omega) = \frac{1}{4\pi} \oint_S \left[ \frac{\partial P(\mathbf{r}_S, \omega)}{\partial n} \frac{e^{-jk|\mathbf{r}-\mathbf{r}_S|}}{|\mathbf{r}-\mathbf{r}_S|} - P(\mathbf{r}_S, \omega) \frac{\partial}{\partial n} \left( \frac{e^{-jk|\mathbf{r}-\mathbf{r}_S|}}{|\mathbf{r}-\mathbf{r}_S|} \right) \right] dS \quad (1.13)$$

ただし,  $k = \omega/c$  は波数,  $c$  は音速,  $\mathbf{r}$  は閉空間  $V$  内の任意の点の位置ベクトル,  $S$  は閉空間  $V$  の表面,  $\mathbf{r}_S$  は  $S$  面上の点の位置ベクトル,  $\partial n$  は  $S$  面上における外向きの法線方向,  $P(\mathbf{r}_S, \omega)$  は  $\mathbf{r}_S$  における音圧を表わす. この式は  $S$  面上における音圧  $P(\mathbf{r}_S, \omega)$  と音圧傾度  $\frac{\partial P(\mathbf{r}_S, \omega)}{\partial n}$  が定まれば,  $\mathbf{r}$  における音圧  $P(\mathbf{r}, \omega)$  が求まり, 閉空間  $V$  内の音圧が制御されるということを示している.

ここで,  $S$  を無限大の大きさを持つ平面 ( $z = z_1$ ) とし, 音源の存在する空間と再生する空間を完全に分断すると, (1.13) 式は第 2 種 Rayleigh 積分と呼ばれる以下の式に変形される [20].

$$P(\mathbf{r}, \omega) = |z - z_1| \int_S P(\mathbf{r}_S, \omega) \frac{1 + jk|\mathbf{r} - \mathbf{r}_S|}{2\pi|\mathbf{r} - \mathbf{r}_S|^3} e^{-jk|\mathbf{r}-\mathbf{r}_S|} dS \quad (1.14)$$

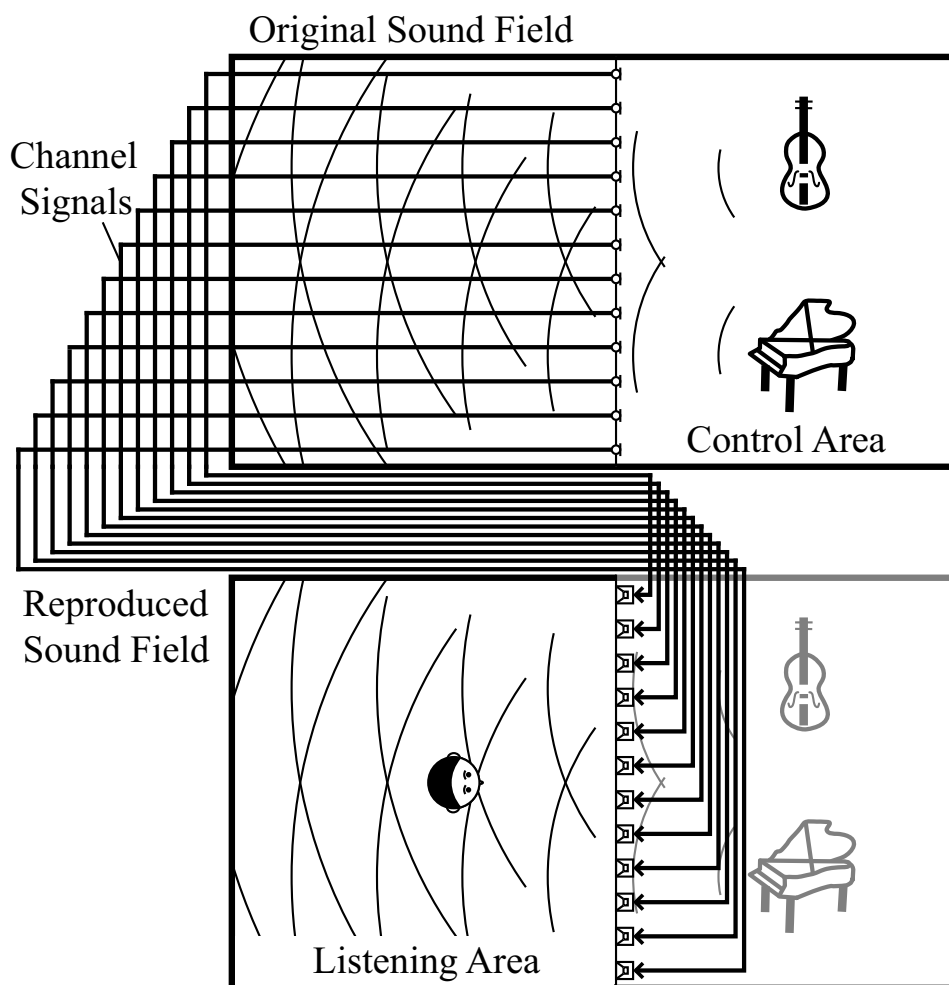


図 1.9 Wave Field Synthesis システムの構成

この式は  $S$  面上の音圧  $P(\mathbf{r}_S, \omega)$  だけが定まれば，再生する空間の音圧  $P(\mathbf{r}, \omega)$  が求まるということを示している．Berkhout らはこの式に基づき，境界面上の音圧だけを制御して波面を合成するシステムを提案した．Berkhout らによって提案されたシステムは一般に「Wave Field Synthesis」[20] と呼ばれている．システムの構成を図 1.9 に示す．まず，境界面上に無指向性マイクロホン配置し，音を収録する．そして，収録した音をマイクロホンと同じ位置に配置したスピーカから再生する．その際，ダイポール型の指向性スピーカを境界平面の反対側に向かって配置する．すると，境界平面の反対側に波面が合成される．Wave Field Synthesis システムは最初は既存のコンサートホールの反射成分を強調するために提案されたが，やがて線状のアレーを用いた立体音場再生システム [21] にも応用されるようになった．また，国内でも同様の研究が行われ [22]，遠隔音楽レッスンを行うための立体音場再生システム [23] も提案されている．

ところで，(1.13) 式に従って空間  $V$  内の音圧  $P(\mathbf{r}, \omega)$  を再生するためには， $\mathbf{r}_S$  の位置

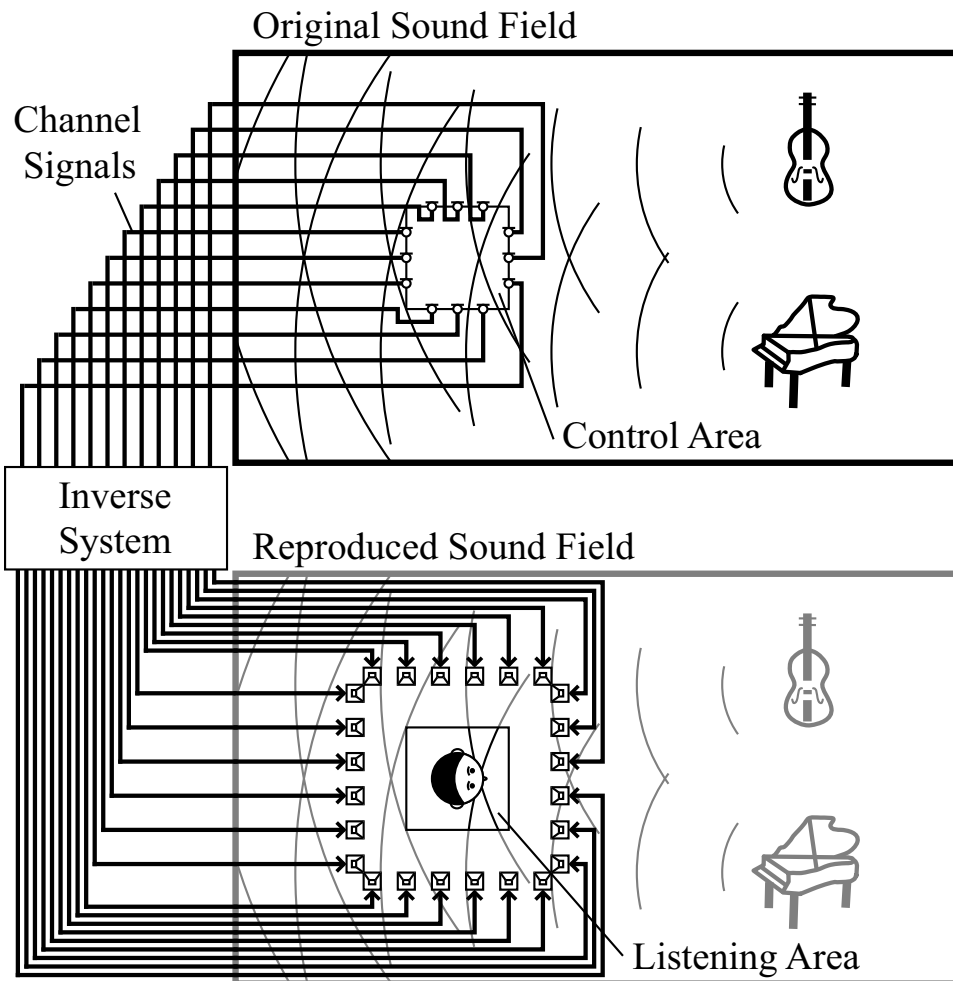


図 1.10 二次音源による立体音場再生システムの構成

に無指向性の音源とダイポール指向性の音源を配置し，無指向性の音源から音圧傾度  $\frac{\partial P(\mathbf{r}_S, \omega)}{\partial n}$ ，ダイポール指向性の音源から音圧  $-P(\mathbf{r}_S, \omega)$  を再生する必要があるが，古家らは境界値問題における解の一意性に着目すると境界面上の音圧を制御するだけで空間内の音圧が制御されることを示し [24]，神沼らはその原理に従って制御領域の周辺に配置した二次音源によって境界面の音圧を制御する立体音場再生システム [25] を提案した．システムの構成を図 1.10 に示す．まず，原音場内に制御領域を想定し，制御領域の境界面に無指向性マイクロホン配置し，音を収録する．一方，再生音場における聴取領域の周辺に二次音源を配置し，二次音源からマイクロホンまでの室内インパルス応答を測定し，逆フィルタを算出する．逆フィルタの算出手法には MINT 法や最小ノルム解法を用いる．そして，収録した音に逆フィルタを畳み込み，二次音源から再生すると，聴取領域の境界面の音圧が制御され，制御領域の波面が聴取領域内に合成される．神沼らはこのシステムを構築し，実際に聴取者に音を聞かせて評価したところ，音の方向が正しく知覚されるこ

とを報告している [25]。しかし，特定の周波数では逆フィルタが安定に算出できないことがあるので，マイクロホン新たに追加して逆フィルタを常に安定に算出できるようにする試みも行われている [26]。

一方，Ise は二次音源によって境界面上の音圧と音圧傾度を制御すれば空間内の音圧が制御されると考え，「境界音場制御」[27] と呼ばれる立体音場再生システムを提案した。このシステムは境界面上に想定した制御点の外側と内側にマイクロホン対を配置し，収録したマイクロホン対の音圧から境界面上の音圧と音圧傾度を近似するという手法に基づいている。このシステムは境界面上の音圧のみを制御する手法に比べて2倍のマイクロホンを必要とするが，逆フィルタが常に安定に算出できる。また，このシステムの応用例として，聴取者の頭部の動きにも頑健なバイノーラルシステム（「ADVISE」[28] と呼ばれている）も提案されている。このシステムはマイクロホン対で収録した音圧にマイクロホン位置から聴取者の両耳までのインパルス応答を畳み込むことによって構成されている。

## 1.2 本論文の目的

1.1.2 節で述べたように，波面合成法による立体音場再生システムは制御領域内の音場を聴取領域内に忠実に再現するものである。しかも，聴取者は聴取領域の中に居る限り頭部を動かしたり席を移動したりすることができるし，多人数の聴取者が同時に聴取することも可能である。しかし，Huygens の原理に従って波面を忠実に合成するためにはマイクロホンやスピーカを境界面上に緻密に配置する必要があるため，マイクロホンやスピーカを緻密に配置することができないという問題点がある。また，システムを遠隔で実現する際にはマイクロホンで収録した音（以後これを「チャンネル信号」と呼ぶ）を送る必要があるが，あまりにもチャンネル信号の数が多すぎるため，現在用いられている通信回線では通信容量が足りなくなる可能性がある。特に，立体映像展示技術と組み合わせたシステムを実現しようとする場合にはこの問題が顕著になる。このように，波面合成法による立体音場再生システムには必要とする情報量が多すぎるという問題点がある。そこで，本論文では立体音場再生システムにおける情報量削減手法の開発を目的とする。

本論文では波面合成法による立体音場再生システムにおいて必要とする情報量を削減するために，次の2つの試みを行う。一つは必要とするマイクロホンやスピーカの個数を減らすことを検討する。もし，マイクロホンやスピーカの個数が減っても聴取者が同じように臨場感を感じるのであれば，システムに必要なマイクロホンやスピーカの個数はより少なく済むので，ある程度の大きさを持った一般のマイクロホンやスピーカでもシステムを構築することができるようになると思われる。もう一つはチャンネル信号を直接処理す

ることによって情報量を圧縮する手法を検討する。もし、圧縮手法によって伝送すべき情報量が削減されても聴取者が同じように臨場感を感じるのであれば、現在の通信回線でも十分に情報を伝送することができるようになると考えられる。

以上の研究によって情報量の削減が達成されれば、波面合成法による立体音場再生システムを実際に構築できるようになるため、通信やバーチャルリアリティの分野に応用できるようになる。例えば、通信の分野に応用すると、遠隔通信会議のように実際は離れた所にいる人々があたかも同じ会議室内で会議をしているように思ったり、遠隔アンサンブルのように離れた所にいる人々が合奏を楽しんだりすることができるようになる。また、バーチャルリアリティの分野に応用すれば、遠隔介護のようにその場にいなくても介護ができるようになったり、遠隔手術のようにその場にいなくても医療行為が行えるようになったりする。このように立体音場再生システムが応用されれば、今までの社会では成し遂げることができなかつたことができるようになるため、人々がより便利に高度な情報化社会を利用することができるようになっていくと期待される。

### 1.3 本論文の構成

本論文の構成は以下の通りである。まず、第 2 章では本論文で検討する立体音場再生システムについて解説し、波面が忠実に合成される条件を計算機シミュレーションによって検討する。

次に、第 3 章ではより少ない個数のマイクロホン及びスピーカで立体音場再生システムを実現できるようにすることを検討する。特に、方向感と空間印象に着目し、マイクロホン及びスピーカの個数の減少に対する聴取者の臨場感の変化を主観評価実験によって検討し、第 2 章での結果によれば波面が忠実に合成されていないような場合でも聴取者は同じように臨場感を感じていることを示す。

第 4 章では第 3 章の結果を踏まえて構築された 24 個のマイクロホン及びスピーカによる立体音場再生システムの伝送量を更に削減するために、伝送量削減手法を提案する。そして、聴取実験を行い、提案手法が有効であることを示す。

第 5 章では第 4 章で提案した音源抽出に基づく伝送量削減手法を更に移動音源に対応できるように改良する。そして、聴取実験によって改良した手法の有効性を示す。

最後に、第 6 章で本論文における結論と今後の展望について述べる。

## 第 2 章

# 波面合成シミュレーション

### 2.1 はじめに

本論文で検討する波面合成法による立体音場再生システムの構成を図 2.1 に示す。まず、原音場内に半径 2 m の円状の制御領域を想定し、制御領域の境界面上にマイクロホンを配置する。一方、再生音場内に半径 2 m の円状の聴取領域を想定し、マイクロホンと同じ位置にスピーカを配置する。そして、マイクロホンで収録した音をスピーカで再生すると、原音場の波面が再生音場で合成され、聴取領域内にいる聴取者はあたかも原音場で音を聞いているような臨場感を感じる。

このシステムを構築する場合、どのような条件で波面が忠実に合成されるのかを検討することは非常に重要である。これまでも波面の再現精度に関する先行研究例はいくつか存在する。例えば、Berkhout *et al.* は図 1.9 に示すような Wave Field Synthesis システムを提案する際に、再現できる周波数帯域が制御点の間隔によってどう変わるか議論している [20]。また、Ise は境界音場制御システムを提案する際に、再現できる周波数帯域は制御点間隔に依存することを示している [27]。また、Ward *et al.* は球面上もしくは円周上に配置したスピーカアレーによって平面波を合成する際のスピーカの数による影響を球面調和解析によって検討し、波面を忠実に合成するためにはスピーカ間隔を波長の  $1/2$  以下に設定することが必要であると報告している [29]。

一方、本論文で検討する立体音場再生システムは図 1.8 に示すような Camras が提案したシステムに基づいており、先行研究例とはシステムの構成が異なるので、先行研究例をそのまま当てはめることはできない。さらに、Camras はこのシステムを提案する際になぜカージオイド型の指向性マイクロホンをういたのかについて説明していないし、システムを評価するに当たっては再生音場の音圧分布を測定するだけで、実際に波面が合成されているかについては検討していない。従って、本論文で扱うシステムにおいて波面が忠実

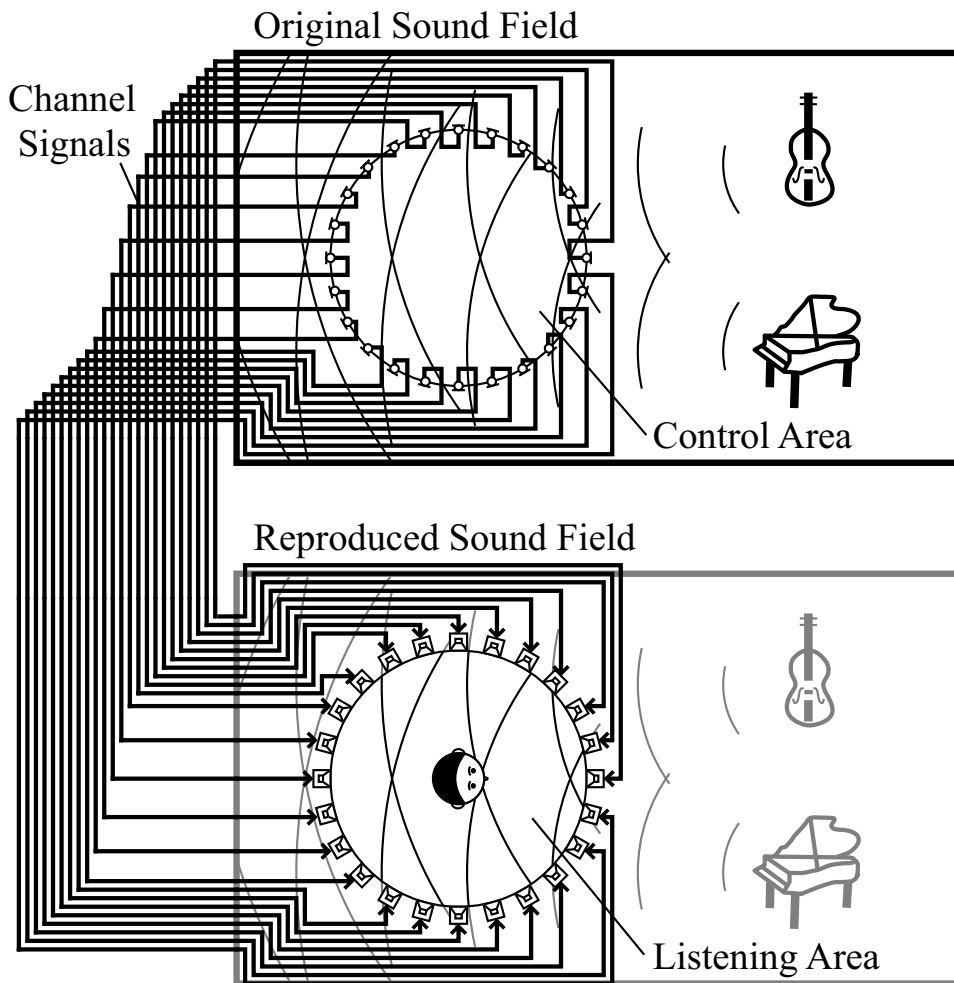


図 2.1 本論文で検討する波面合成システムの構成

に合成される条件は十分に検討されていないといえるので，本章では波面が忠実に合成される条件を検討するために計算機シミュレーションによる波面合成実験を行う．

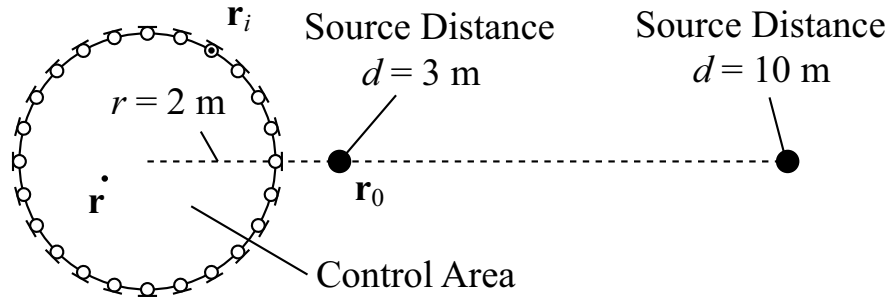
## 2.2 波面合成実験

### 2.2.1 実験環境

原音場は反射音のない自由空間とした．実験における原音場及び再生音場を図 2.2 に示す．制御領域及び聴取領域は半径 2 m の円とした．音源は 1 個とし，図 2.2 に示すように円の中心から 3 m と 10 m の地点の 2 種類を設定した．制御領域の境界線上にマイクロホンを配置する際には隣り合うマイクロホン間の距離が一定になるように円上に均等に配置した．一方，再生音場における聴取領域の境界線上にスピーカを配置した．その際，スピーカの位置はマイクロホンの位置と同じにした．そして，マイクロホンの信号をス



## Original Sound Field



## Reproduced Sound Field

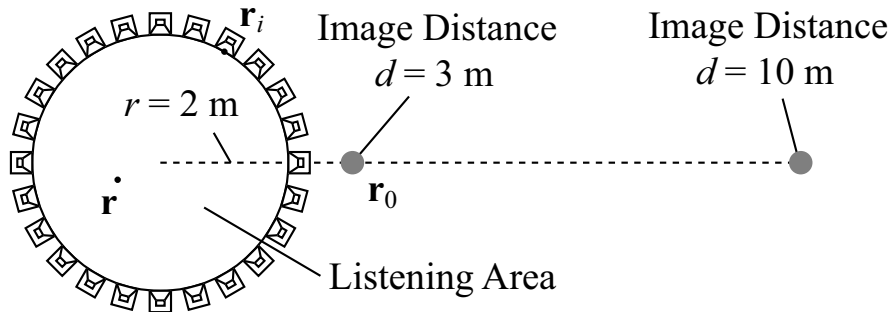


図 2.2 波面合成実験における実験環境

ピーカから再生し，図 2.2 中の灰色の円で示される音像からの波面を聴取領域内に合成させた．本実験の場合，原音場における音源は円の中心から 3 m と 10 m の 2 種類なので，再現される音像も円の中心から 3 m と 10 m の 2 種類になる．

音源信号  $s_0(t)$  は以下の式のように周波数  $f$  の正弦波信号とした．

$$s_0(t) = \sin 2\pi f t \quad (2.1)$$

制御領域内の任意の点の位置ベクトルを  $\mathbf{r}$  とすると，原音場における  $\mathbf{r}$  での音圧  $p_0(\mathbf{r}, t)$  は以下のように表わされる．

$$\begin{aligned} p_0(\mathbf{r}, t) &= \frac{1}{|\mathbf{r} - \mathbf{r}_0|} s_0\left(t - \frac{|\mathbf{r} - \mathbf{r}_0|}{c}\right) \\ &= \frac{1}{|\mathbf{r} - \mathbf{r}_0|} \sin\left[2\pi f \left(t - \frac{|\mathbf{r} - \mathbf{r}_0|}{c}\right)\right] \end{aligned} \quad (2.2)$$

但し， $\mathbf{r}_0$  は音源の位置ベクトル， $c$  は音速である．一方，マイクロホンで収録する信号

表 2.1 波面合成実験における実験条件

マイクロホンの数 ( $M$ )	2, 3, 4, 5, 6, 7, 8, 10, 11, 13, 16, 20, 24, 27, 32, 36, 45, 54, 64, 72, 90, 108, 128, 160, 180, 216, 256, 300, 360, 432, 512
音源の周波数 ( $f$ )	125, 250, 500, 1000, 2000, 4000 Hz
音源の距離 ( $d$ )	3, 10 m
制御領域の半径 ( $r$ )	2 m
音速 ( $c$ )	340 m/s
マイクロホンの指向特性	無指向性, 単一指向性, 超指向性

$x_i(t)$  は以下のように表わされる .

$$\begin{aligned} x_i(t) &= \frac{D_i}{|\mathbf{r}_i - \mathbf{r}_0|} s_0 \left( t - \frac{|\mathbf{r}_i - \mathbf{r}_0|}{c} \right) \\ &= \frac{D_i}{|\mathbf{r}_i - \mathbf{r}_0|} \sin \left[ 2\pi f \left( t - \frac{|\mathbf{r}_i - \mathbf{r}_0|}{c} \right) \right] \end{aligned} \quad (2.3)$$

但し,  $\mathbf{r}_i$  は  $i$  番目のマイクロホンの位置ベクトル ( $i=1\dots M$ ),  $M$  はマイクロホンの総数,  $D_i$  は  $i$  番目のマイクロホンの指向特性を表わす. 従って, 再生音場における聴取領域内の任意の点の位置ベクトル  $\mathbf{r}$  での音圧  $p(\mathbf{r}, t)$  は  $x_i(t)$  から以下のように算出される .

$$\begin{aligned} p(\mathbf{r}, t) &= \sum_{i=1}^M \frac{1}{|\mathbf{r} - \mathbf{r}_i|} x_i \left( t - \frac{|\mathbf{r} - \mathbf{r}_i|}{c} \right) \\ &= \sum_{i=1}^M \frac{D_i}{|\mathbf{r} - \mathbf{r}_i| |\mathbf{r}_i - \mathbf{r}_0|} \sin \left[ 2\pi f \left( t - \frac{|\mathbf{r} - \mathbf{r}_i| + |\mathbf{r}_i - \mathbf{r}_0|}{c} \right) \right] \end{aligned} \quad (2.4)$$

実験に使用した条件を表 2.1 に示す . また, 本実験では  $\mathbf{r}_0$ ,  $\mathbf{r}_i$ ,  $\mathbf{r}$  は以下のように設定した .

$$\begin{aligned} \mathbf{r}_0 &= \begin{pmatrix} 3 \\ 0 \end{pmatrix} \text{ or } \begin{pmatrix} 10 \\ 0 \end{pmatrix} \quad \mathbf{r}_i = \begin{pmatrix} 2 \cos \frac{2\pi i}{M} \\ 2 \sin \frac{2\pi i}{M} \end{pmatrix} \\ \mathbf{r} &= \begin{pmatrix} r_x \\ r_y \end{pmatrix} \quad (\text{但し, } r_x^2 + r_y^2 < 2^2) \end{aligned} \quad (2.5)$$

実験に用いた 3 種類のマイクロホンの指向特性を図 2.3 に示す . これらの指向特性は一般に流通しているマイクロホンが持つ代表的な指向特性である (例えば, ピンマイクロホン, カラオケ用マイクロホン, ガンマイクロホン) . 本実験では  $D_i$  は指向特性ごとに以下の式

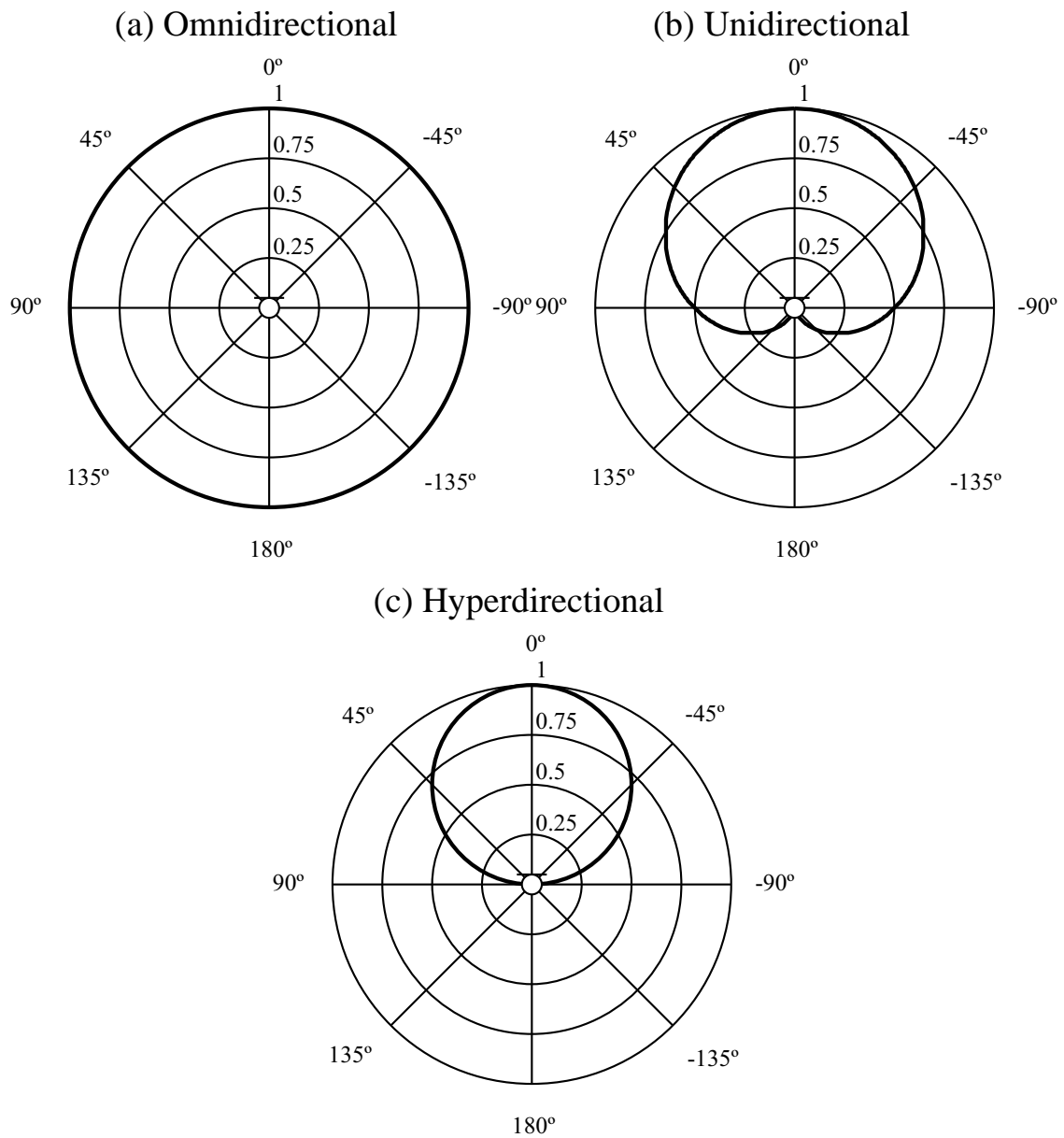


図 2.3 波面合成実験に用いたマイクロホンの指向特性

によって表わした .

$$D_i = 1 \quad (\text{無指向性}) \quad (2.6)$$

$$D_i = \frac{1}{2} \left( 1 + \frac{\mathbf{r}_i \cdot (\mathbf{r}_0 - \mathbf{r}_i)}{|\mathbf{r}_i| |\mathbf{r}_0 - \mathbf{r}_i|} \right) \quad (\text{単一指向性}) \quad (2.7)$$

$$D_i = \begin{cases} \frac{\mathbf{r}_i \cdot (\mathbf{r}_0 - \mathbf{r}_i)}{|\mathbf{r}_i| |\mathbf{r}_0 - \mathbf{r}_i|} & (\mathbf{r}_i \cdot (\mathbf{r}_0 - \mathbf{r}_i) \geq 0) \\ 0 & (\mathbf{r}_i \cdot (\mathbf{r}_0 - \mathbf{r}_i) < 0) \end{cases} \quad (\text{超指向性}) \quad (2.8)$$

### 2.2.2 実験結果及び考察

合成した波面の一例は付録 A に示す．本章では以下の式で示される SNR を算出することによって波面が忠実に合成される条件を検討する．

$$\text{SNR [dB]} = 10 \log_{10} \frac{\sum_{\mathbf{r}} \{p_0(\mathbf{r}, t)\}^2}{\sum_{\mathbf{r}} \{p(\mathbf{r}, t) - p_0(\mathbf{r}, t)\}^2} \quad (2.9)$$

但し， $\mathbf{r} = ({}^v r_x, {}^v r_y)^T$  は  $r_x^2 + r_y^2 < 1$  の範囲とした．また， $r_x^2 + r_y^2 < 1$  の範囲で  $p_0(\mathbf{r}, t)$  及び  $p(\mathbf{r}, t)$  の値を正規化してから SNR の算出を行った．音源の周波数  $f$  及びマイクロホンの数  $M$  ごとに SNR を算出した結果を図 2.4~2.6 に示す．但し，横軸は 2 を底とする対数スケールにしている．全ての図においてマイクロホンの数がある閾値を超えると，SNR はそれ以上マイクロホンの数が増えても変化しなくなる．つまり，得られた閾値が波面を忠実に合成するのに必要なマイクロホンの数であると言える．また，その閾値は合成する波面の周波数に依存しており，周波数が 2 倍になると，SNR が一定になり始めるマイクロホンの数も 2 倍になっていることが分かる．従って，以下のような条件を満たせば，波面が忠実に合成されることが出来る．

$$M \geq 0.064f \left( = \frac{256}{4000}f \right) \quad (2.10)$$

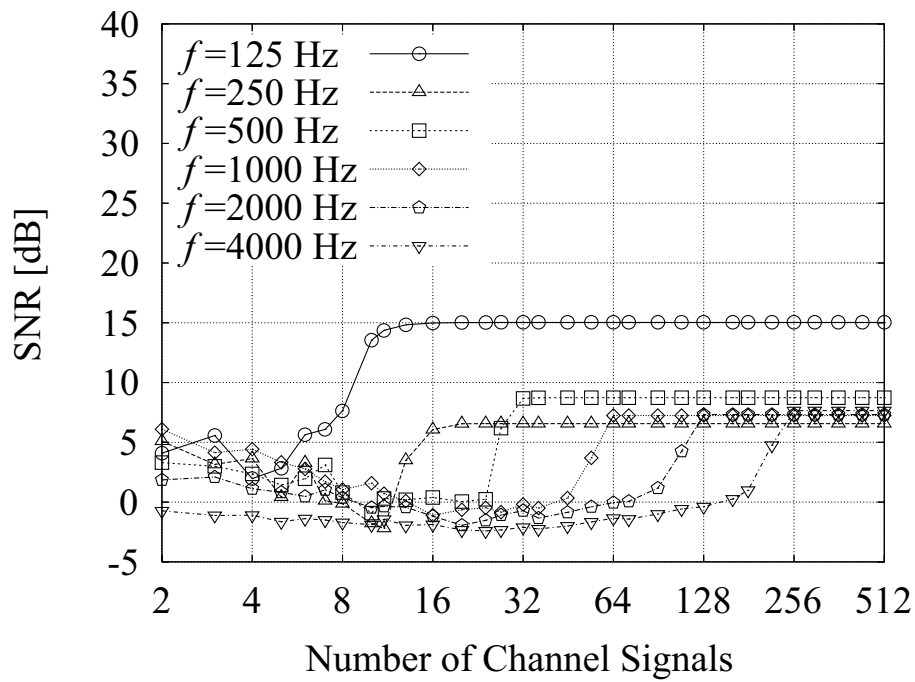
また，マイクロホンの間隔を  $\Delta$  とすると， $\Delta = 2\pi r/M$  ( $r$  は制御領域の半径) であり，波長を  $\lambda$  とすると， $f = c/\lambda$  ( $c$  は音速) であることから，(2.10) 式から以下の関係式が導かれる．

$$\Delta \leq \frac{2\pi r}{0.064c} \lambda \approx 0.57\lambda \quad (2.11)$$

ゆえに，本論文で検討する波面合成システムにおいて波面を忠実に合成するためには，マイクロホン及びスピーカ間の間隔を波長の 1/2 以下にすることが必要であると言える．

また，マイクロホンの指向特性による影響を考察するため， $M=512$  における SNR の値を音源の周波数全体で平均化したものを図 2.7 に示す．但し，エラーバーは 95% 信頼区間を表わす．どちらの音源距離においても無指向性のときの平均 SNR より単一指向性や超指向性のときの平均 SNR の方が有意に高い値を示している．これは無指向性マイクロホンを用いた場合には境界面上の音圧のみが制御されるのに対して，単一指向性や超指向性マイクロホンを用いると指向特性によって境界面上の音圧に加えて音圧傾度も制御されるようになるため，(1.13) 式で示される Kirchhoff-Helmholtz 積分方程式において波面が

Source Distance = 3 m



Source Distance = 10 m

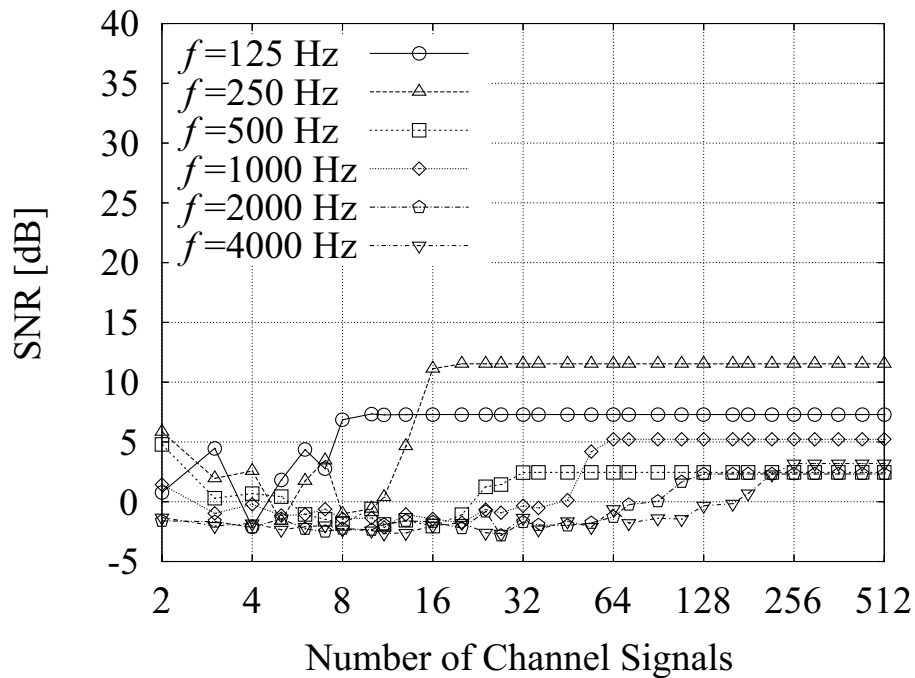
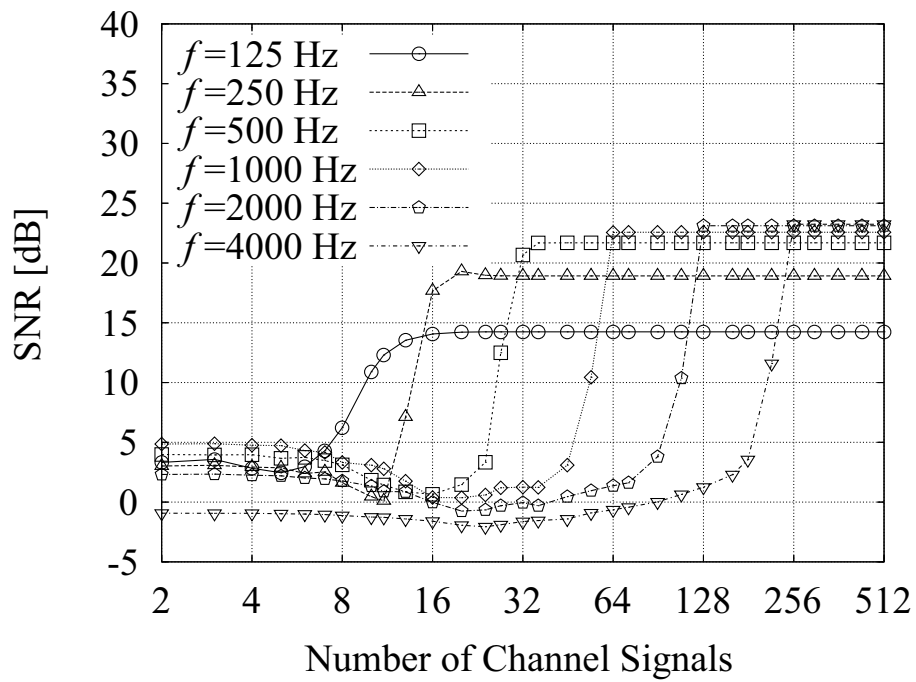


図 2.4 波面合成実験における SNR の結果 (無指向性)

Source Distance = 3 m



Source Distance = 10 m

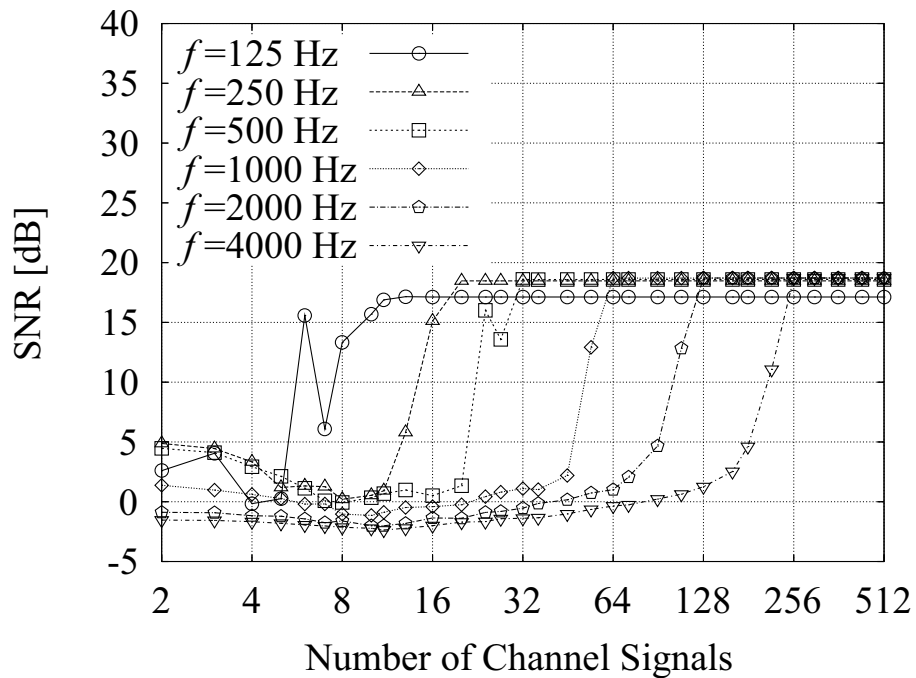
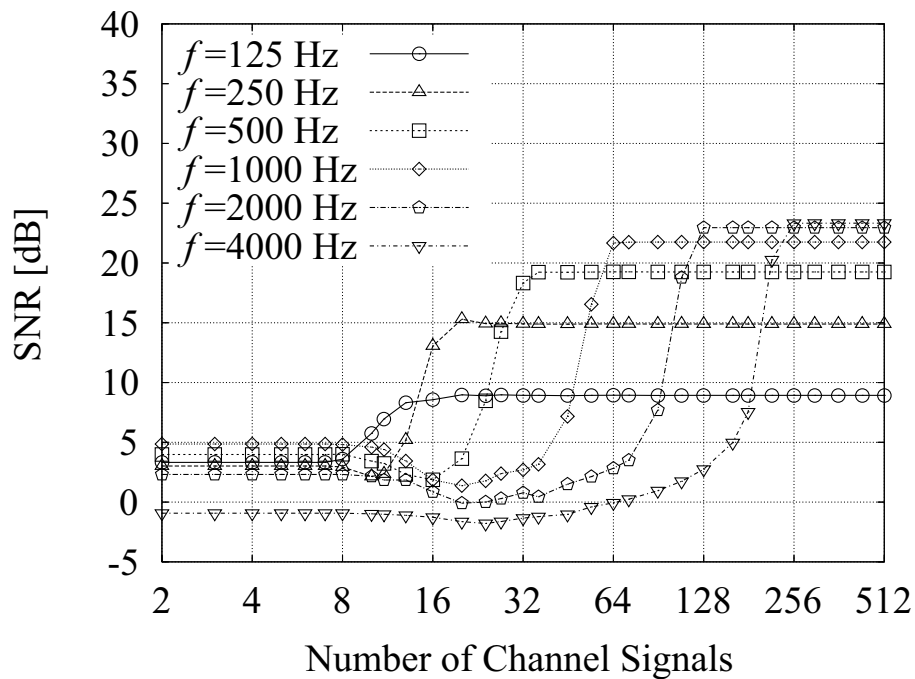


図 2.5 波面合成実験における SNR の結果 (単一指向性)

Source Distance = 3 m



Source Distance = 10 m

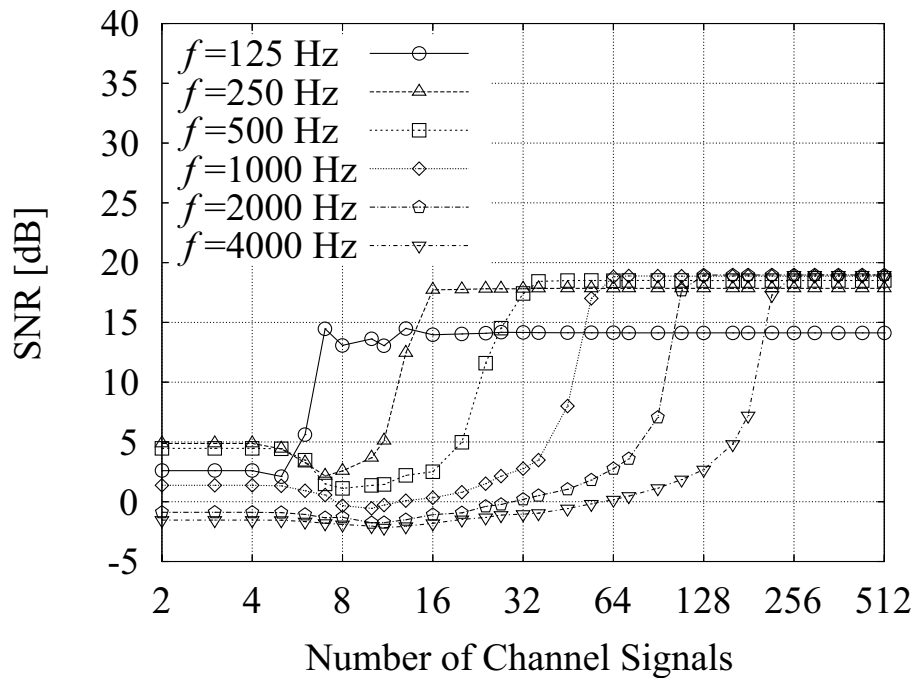


図 2.6 波面合成実験における SNR の結果 (超指向性)

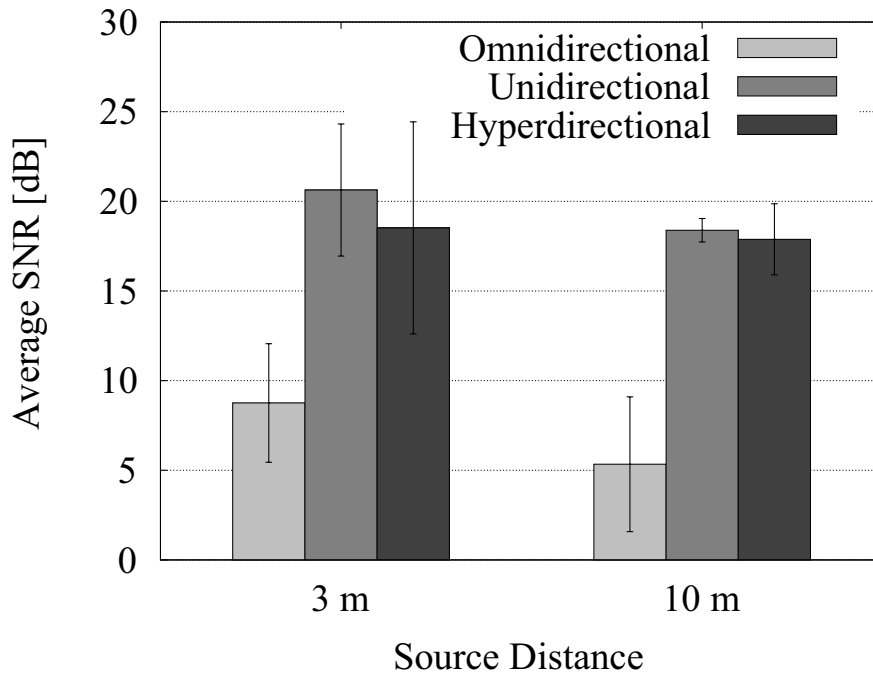


図 2.7 波面合成実験における平均 SNR の結果

忠実に合成される条件を満たすようになるためであると考えられる。従って、本論文で検討する波面合成システムを構築する際は、単一指向性マイクロホンもしくは超指向性マイクロホンをを用いるべきであると考えられる。また、2種類の音源距離においても平均 SNR の変化があまりないことから、本論文で検討するシステムは音源距離に波面合成の性能が依存しないということも言える。

## 2.3 残響空間における波面合成の条件

2.2 節で扱った原音場は反射音の存在しない自由空間であった。しかし、実際に波面合成システムを実現する際には原音場は会議室やコンサートホールなどの反射音による残響が存在する空間であることが多い。そこで、本実験による結果から残響空間における波面合成の条件を導くことができるかどうか検討する。

一般に残響空間における反射音は音源からの音波が側壁、床、天井といった壁面に反射したものであると考えることができる。その性質を利用して、反射音を音源の鏡面反射であるとみなして残響空間における室内インパルス応答を算出する手法も提案されている（「虚像法」と呼ばれる）[30]。従って、反射音は虚像音源の位置から壁面に反射した分だけ振幅が減衰されて放射している音であると考えられる。また、本論文で扱う波面合成システムは音源の距離に波面合成の性能が変化することはないので、本実験により



導かれた波面合成の条件をそのまま当てはめることができると考えられる。

従って、原音場が残響空間の場合でも波面が忠実に合成されるには 2.2 節における実験結果と同じようにマイクロホン及びスピーカ間の間隔を波長の  $1/2$  以下にし、単一指向性や超指向性マイクロホンを用いる必要があるということが言える。

## 2.4 まとめ

本章では本論文で検討する波面合成法による立体音場再生システムについて説明し、波面を忠実に合成する条件を検討するためのシミュレーションを計算機上において行った。その結果、波面を忠実に合成するためにはマイクロホン及びスピーカ間の間隔を波長の  $1/2$  以下にすること、無指向性マイクロホンよりも単一指向性マイクロホンもしくは超指向性マイクロホンを用いるのが望ましいということが分かった。



## 第 3 章

# チャンネル信号数の削減

### 3.1 はじめに

第 2 章で、波面を忠実に合成するためにはマイクロホン及びスピーカ間の間隔を波長の  $1/2$  以下に設定することが必要であるということを確認した。この場合、システムを構築する際に必要なチャンネル信号の数は対象とする音の周波数帯域に依存することになる。しかしながら、コンサートホールのように帯域が少なくとも 16 kHz であるような楽音を対象としたシステムを構築する場合には、設定すべきスピーカ間の間隔が約 1 cm になるため、実際に用いるスピーカの大きさを考慮すると、システムの構築は非現実的である。

一方、今までの議論は波面が忠実に合成されることにのみ着目しており、聴取者が音を聞いて臨場感を感じることにについてはあまり着目してこなかった。もし、マイクロホンやスピーカの数が減っても聴取者が同じように臨場感を感じるなら、一般のマイクロホンやスピーカでも十分にシステムを構築できるようになると考えられる。

そこで本章では波面合成法による立体音場再生システムにおいてチャンネル信号の数が臨場感に及ぼす影響を検討するための主観評価実験を行う。Morimoto によれば、臨場感を構成する要素の一つである音場知覚は次の 3 つに大別されるとしている [31]。

1. 方向感
2. 空間印象
3. 距離感

そこで、本論文では特に方向感と空間印象に着目し、2 種類の主観評価実験によってチャンネル信号の数による方向感及び空間印象への影響を検討する。音源が 3 m ~ 15 m 離れている場合、距離感は音の大きさに依存する [32]。従って、上記の音源距離を想定している立体音場再生システムにおいては音の大きさを制御するだけでチャンネル信号の数に関係な

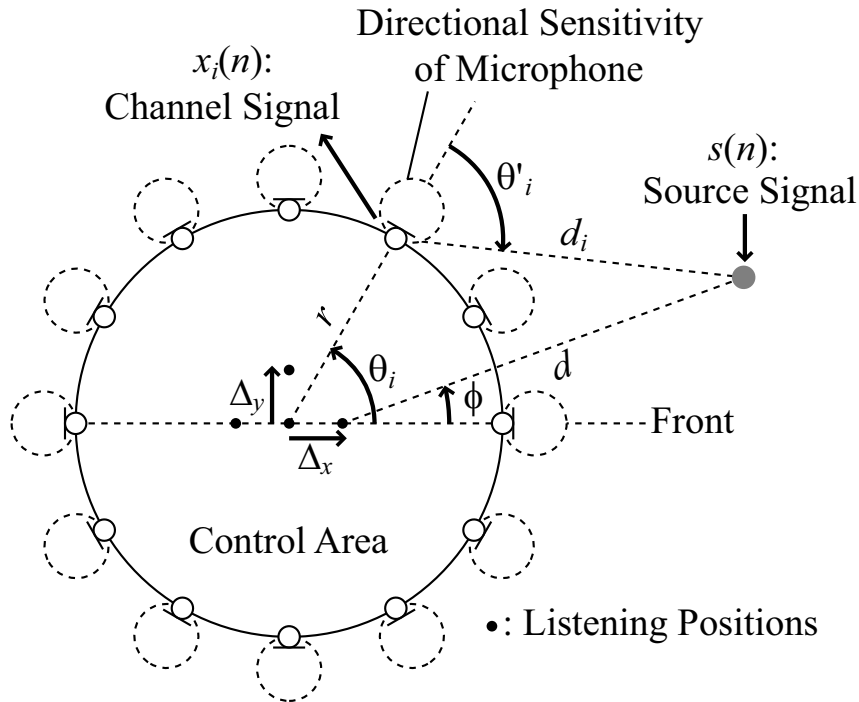


図 3.1 方向感の主観評価における原音場

く距離感が十分に再現されると考えたので、本論文では距離感については検討しない。また、複数の聴取位置を設定し、聴取位置による影響の有無についても合わせて検討する。これらの結果を踏まえて、システムに必要なチャンネル信号の数を削減できるかどうかを検討する。

## 3.2 主観評価実験 1 - 方向感への影響 -

### 3.2.1 チャネル信号の作成

方向感は主に音源からの直接音が作用している。従って、原音場は図 3.1 に示すような自由空間とし、その中に半径  $r (=2 \text{ m})$  の制御領域を想定した。各マイクロホンの指向性は制御領域の外側に向いている。聴取位置による影響も検討するために、本論文では 4 箇所の聴取位置 (中心, 前方, 後方, 側方) を設定した。このとき、中心から正面方向への聴取位置の移動距離を  $\Delta_x$ 、左側方向への移動距離を  $\Delta_y$  とすると、4 箇所の聴取位置の座標は以下のように表わされる。但し、単位は m である。

$$(\Delta_x, \Delta_y) = \begin{cases} ((0, 0) & \text{(中心)} \\ (0.5, 0) & \text{(前方)} \\ (-0.5, 0) & \text{(後方)} \\ (0, 0.5) & \text{(側方)} \end{cases} \quad (3.1)$$

灰色の円は仮想音源を表す． $\phi$  は聴取位置における音源の方位角， $d$  ( $=3, 4$  m) は音源から聴取位置までの距離である． $\theta_i$  ( $i=1\dots M$ ) は  $i$  番目のマイクロホンの方位角で， $M$  はマイクロホンの総数である．音源から  $i$  番目のマイクロホンまでの距離  $d_i$  は (3.2) 式によって算出される．

$$d_i = \sqrt{(d\cos\phi + \Delta_x - r\cos\theta_i)^2 + (d\sin\phi + \Delta_y - r\sin\theta_i)^2} \quad (3.2)$$

従って，音源から  $i$  番目のマイクロホンまでの室内インパルス応答  $g_i(n)$  は (3.3) 式のように表される．

$$g_i(n) = \frac{D_i}{d_i} \delta[n - \text{round}\left(\frac{d_i F_s}{c}\right)] \quad (3.3)$$

但し， $\delta(n)$  は Dirac のデルタ関数， $F_s$  ( $=48$  kHz) はサンプリング周波数， $c$  ( $=340$  m/s) は音速を表す．また， $D_i$  は  $i$  番目のマイクロホンの指向性関数であり，第 2 章の結果より (3.4) 式のように超指向性とした．

$$D_i = \begin{cases} \cos\theta'_i & (|\theta'_i| \leq 90^\circ) \\ 0 & (|\theta'_i| > 90^\circ) \end{cases} \quad (3.4)$$

但し， $\theta'_i$  は図 3.1 に示すように  $i$  番目のマイクロホンにおける音源の入射角を表す．

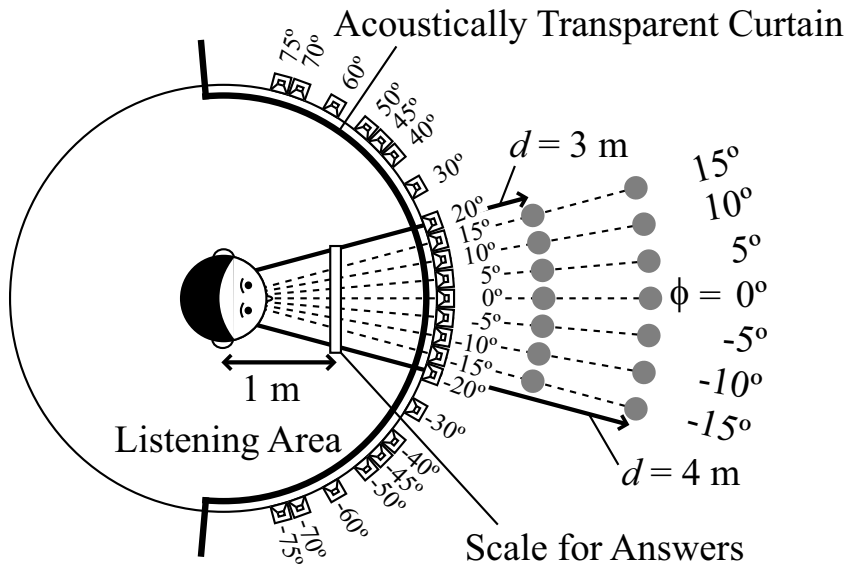
ドライソース  $s(n)$  にはサンプリング周波数 48 kHz，長さ 1 s の白色雑音と音声を用いた．白色雑音は MATLAB[33] 上で作成し，音声は SQAM-CD [34] の音声サンプルから抜粋した．そして， $s(n)$  と  $g_i(n)$  から (3.5) 式に従って  $i$  番目のマイクロホンのチャンネル信号  $x_i(n)$  を作成した．但し， $*$  は畳み込み演算を表す．

$$\begin{aligned} x_i(n) &= g_i(n) * s(n) \\ &= \frac{D_i}{d_i} s[n - \text{round}\left(\frac{d_i F_s}{c}\right)] \end{aligned} \quad (3.5)$$

### 3.2.2 実験環境

主観評価実験は本来なら無響室で行う必要があるが，実験設備の都合上，名古屋大学人間情報学研究科棟 416 号室の低残響室にて行った．この部屋の残響時間は約 80 ms である．図 3.2～3.3 に示すように聴取者の正面方向の半径 2 m の円周上に計 23 個のスピーカ (Emic: Soundevice) を配置し，さらに 4 箇所の聴取位置を設定した．灰色の円はスピーカアレーによって再生される音像を表す．音像を正面方向に配置した場合，マイクロホンの指向性関数によって後方のチャンネル信号は零信号になるので，後方にスピーカを配置する必要はなくなる．部屋の暗騒音レベルは 25.0 dB(A) で，音圧レベルは円の中心において約 70 dB(A) に設定した．また，聴取者にスピーカアレーを見せないようにするため，照明を薄暗くし，スピーカアレーの前方には音響透過カーテンを設置した．

## Center Position



## Front Position

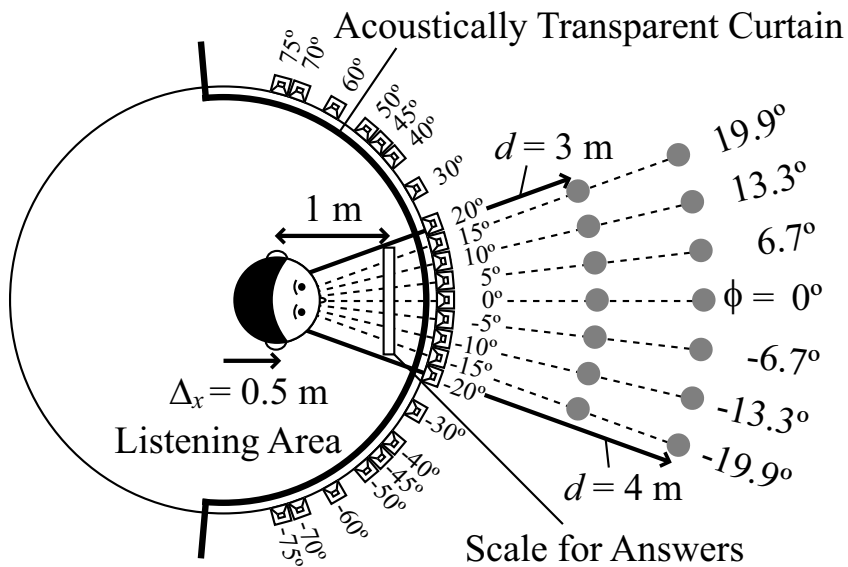
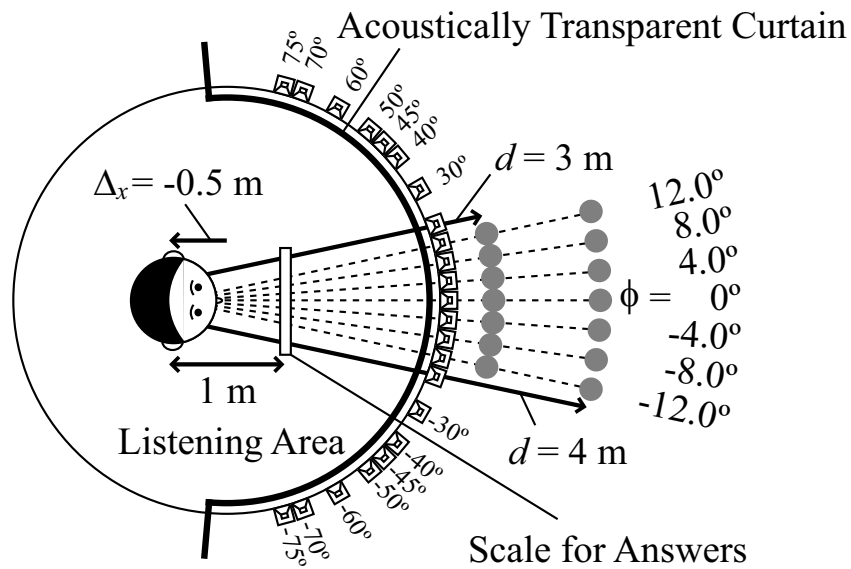


図 3.2 方向感の主観評価における実験環境 (中心, 前方位置)

本実験ではチャネル信号の数に関する条件を 5 種類設定した。実験条件を図 3.4~3.5 に示す。条件 (a)~(e) における灰色のスピーカは先に述べた理由により配置しなかったスピーカを表す。これらの条件においては 23 個のスピーカの中から選ばれた 3, 5, 7, 11, 15 個のスピーカから (3.5) 式で算出されたチャネル信号を同時に再生させた。その際、選ばれなかったスピーカからは音を再生させなかった。その結果、聴取者は図 3.4~3.5 中の灰色の円で示された位置に音像を感じるようになる。(f) に示す統制条件は従来から行

Behind Position



Lateral Position

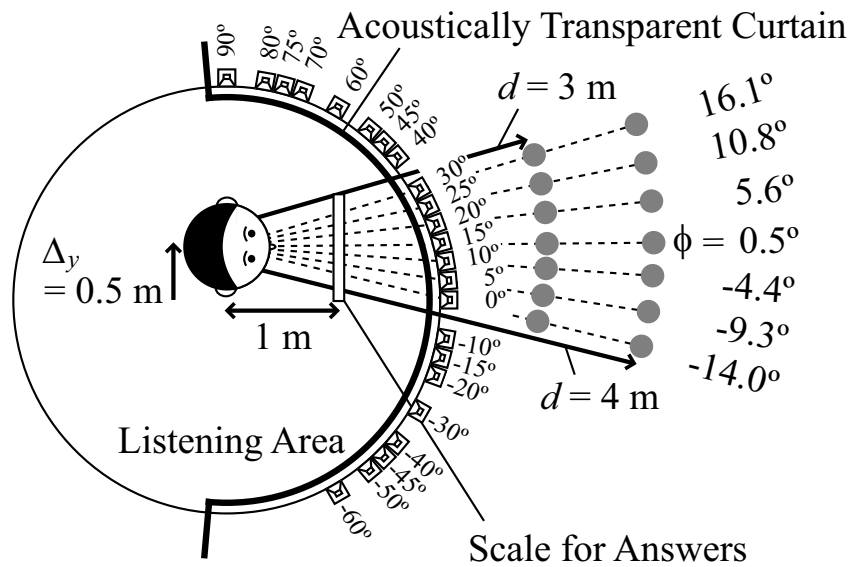


図 3.3 方向感の主観評価における実験環境 (後方, 側方位置)

われている音源定位の実験条件と同じである．ゆえに条件 (a) ~ (e) の実験結果を統制条件と比較してチャンネル信号の数が方向感に及ぼす影響を検討する．但し，聴取位置が変わると統制条件における提示方向が変わるので，その際には合成音像の提示方向  $\phi$  を (3.6) 式によって変化させた．

$$\phi_{\text{front,behind,lateral}} = \tan^{-1} \left[ \frac{r \sin \phi_{\text{center}} - \Delta_y}{r \cos \phi_{\text{center}} - \Delta_x} \right] \quad (3.6)$$

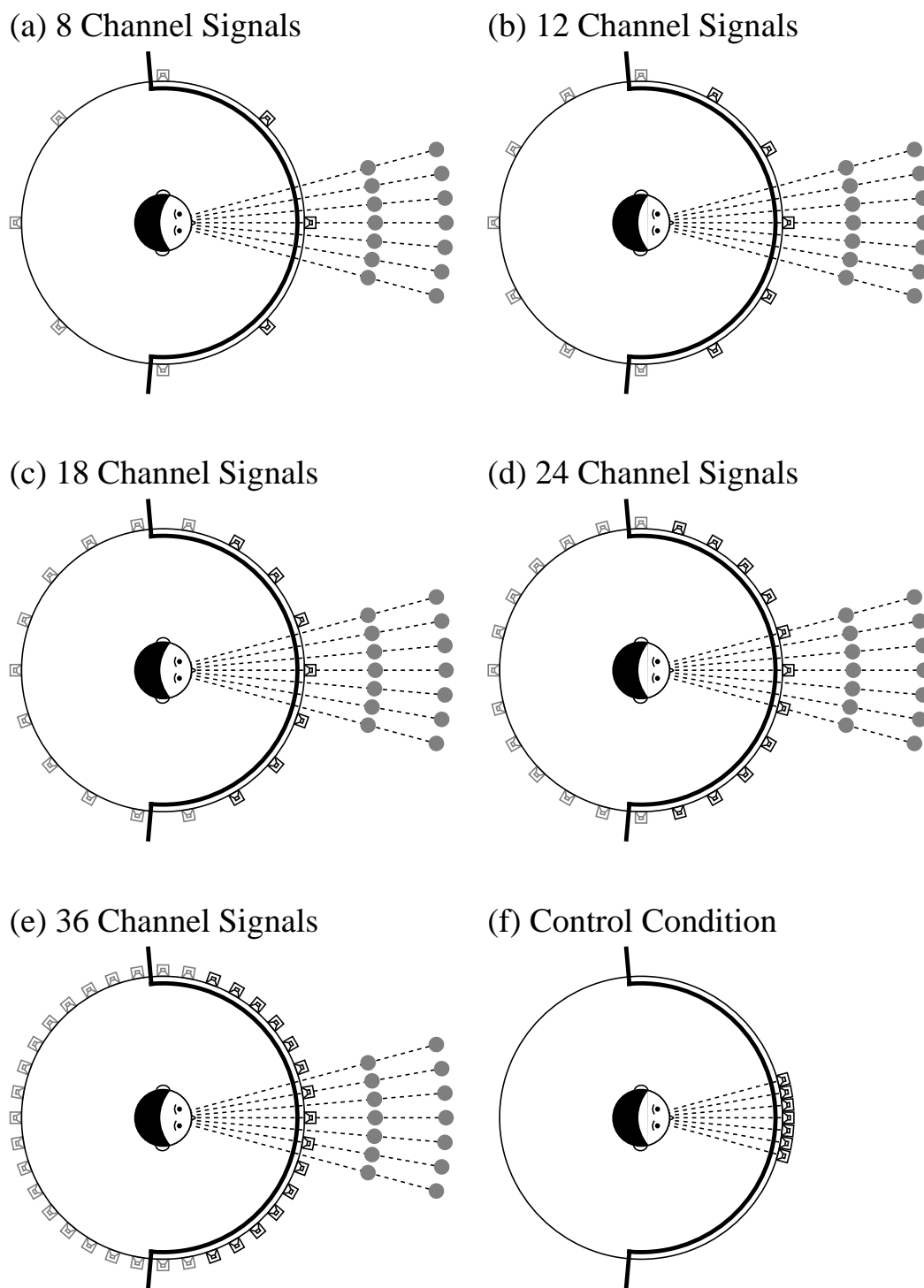


図 3.4 方向感の主観評価における実験条件 (中心位置の場合)



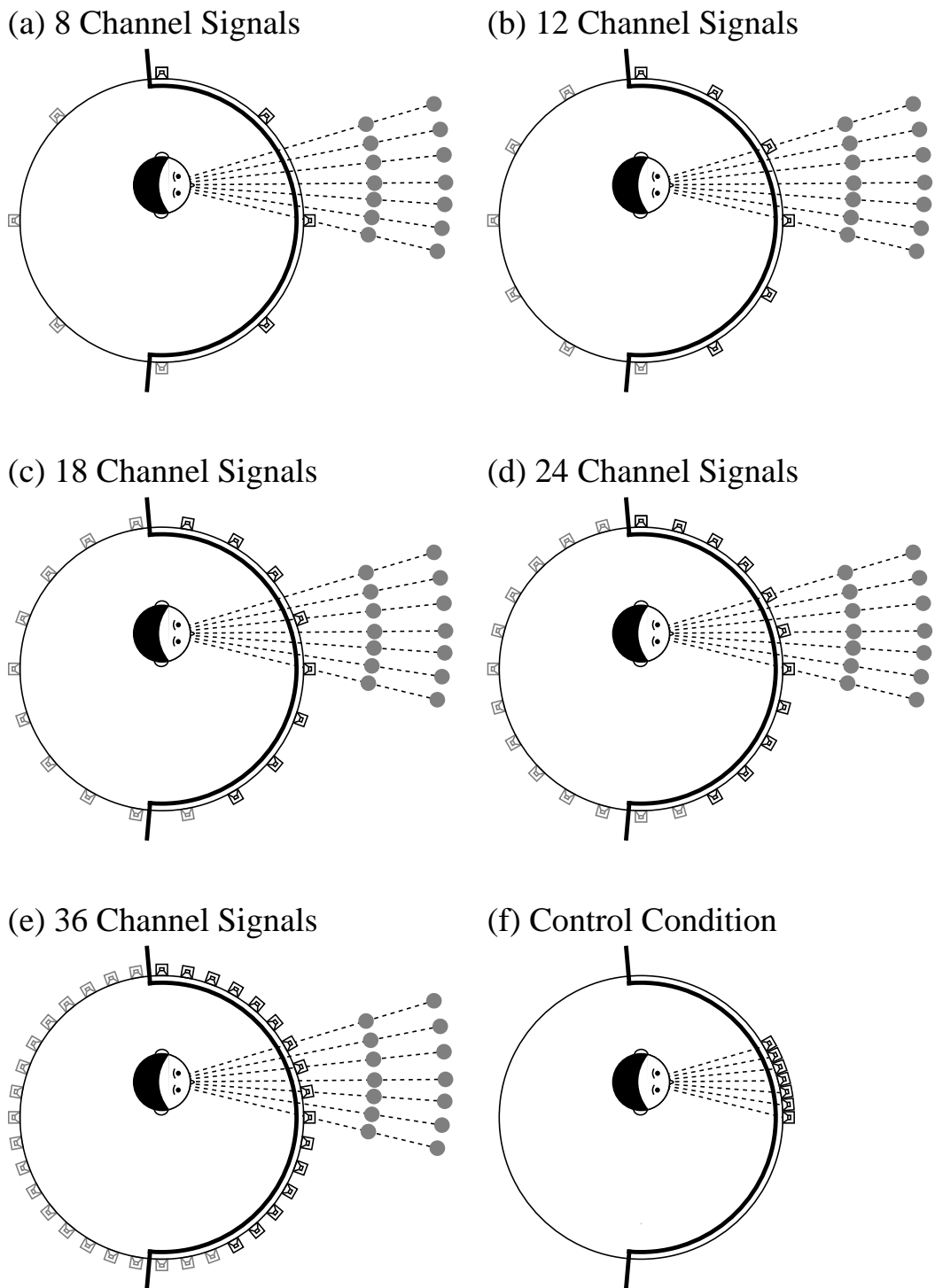


図 3.5 方向感の主観評価における実験条件 (側方位置の場合)

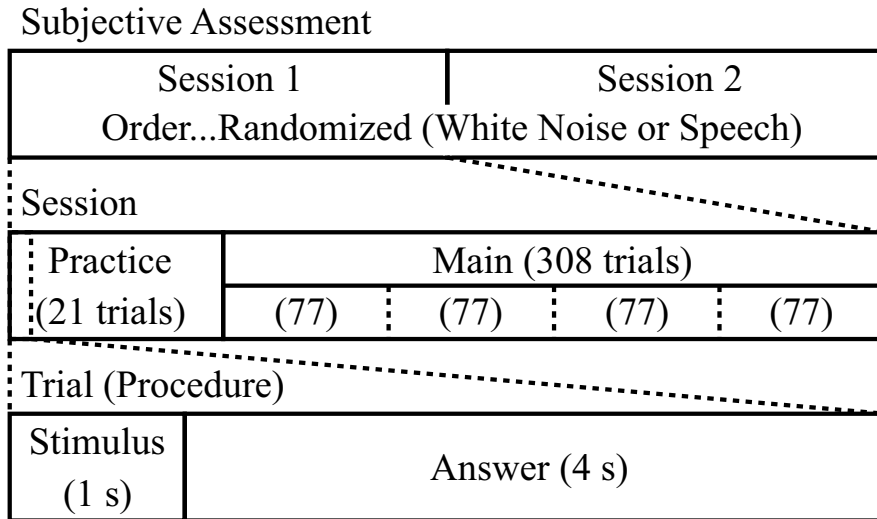


図 3.6 方向感の主観評価における実験流れ図

ここで  $\phi_{\text{center}}$ ,  $\phi_{\text{front}}$ ,  $\phi_{\text{behind}}$ ,  $\phi_{\text{lateral}}$  は中心, 前方, 後方, 側方位置における提示方向を表す. 本実験では  $\phi_{\text{center}}=0, \pm 5, \pm 10, \pm 15^\circ$ ,  $\phi_{\text{front}}=0, \pm 6.7, \pm 13.3, \pm 19.9^\circ$ ,  $\phi_{\text{behind}}=0, \pm 4.0, \pm 8.0, \pm 12.0^\circ$ ,  $\phi_{\text{lateral}}=-14.0, -9.3, -4.4, 0.5, 5.6, 10.8, 16.1^\circ$  に設定した. 但し, 側方位置の場合, 提示方向の算出に  $\phi_{\text{center}}=0, 5, 10, 15, 20, 25, 30^\circ$  を用いている.

### 3.2.3 実験計画

聴取者は日常生活において聴力に異常の見られない 12 名の学生 (うち男性 10 名, 女性 2 名) である. 1 箇所の聴取位置あたり 3 名ずつの被験者で実験を行った. 主観評価実験の流れ図を図 3.6 に示す. まず, 実験全体をドライソースごとに 2 つのセッションに分割した. その際, ドライソースの提示順序は被験者ごとにランダム化した. 各セッションにおいて 21 回の練習試行の後, 308 回の本試行を行った. その際, 77 試行ごとに休憩時間を設けた. 練習試行及び本試行の条件を表 3.1 に示す.

聴取者には音を聞いた後に音の聞こえた方向を回答するように教示した. その際, 図 3.2~3.3 に示すように聴取者の 1 m 前方に配置した目盛りに従って回答させた. この目盛りは  $-25^\circ$  から  $25^\circ$  まで  $2.5^\circ$  間隔で刻まれている. また, 回答の際には聴取者は頭部を自由に動かすことができた.

### 3.2.4 実験結果及び考察

聴取位置ごとの実験結果を図 3.7~3.10 に示す. 統制条件の場合は全ての聴取位置において回答方向が提示方向とほぼ同じである. すなわち, 聴取者は音源の方向を正しく定

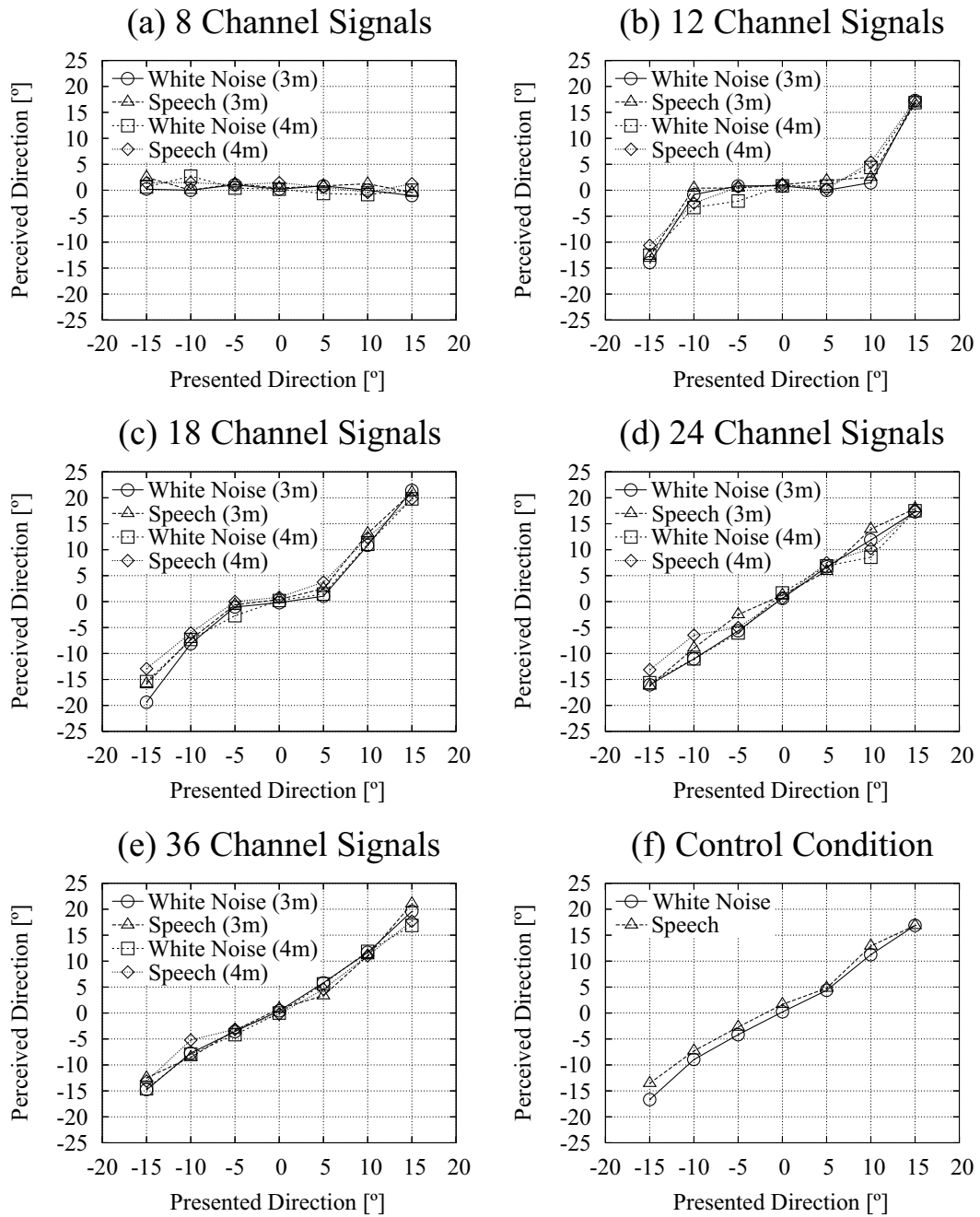


図 3.7 方向感の主観評価における実験結果 (中心位置の場合)

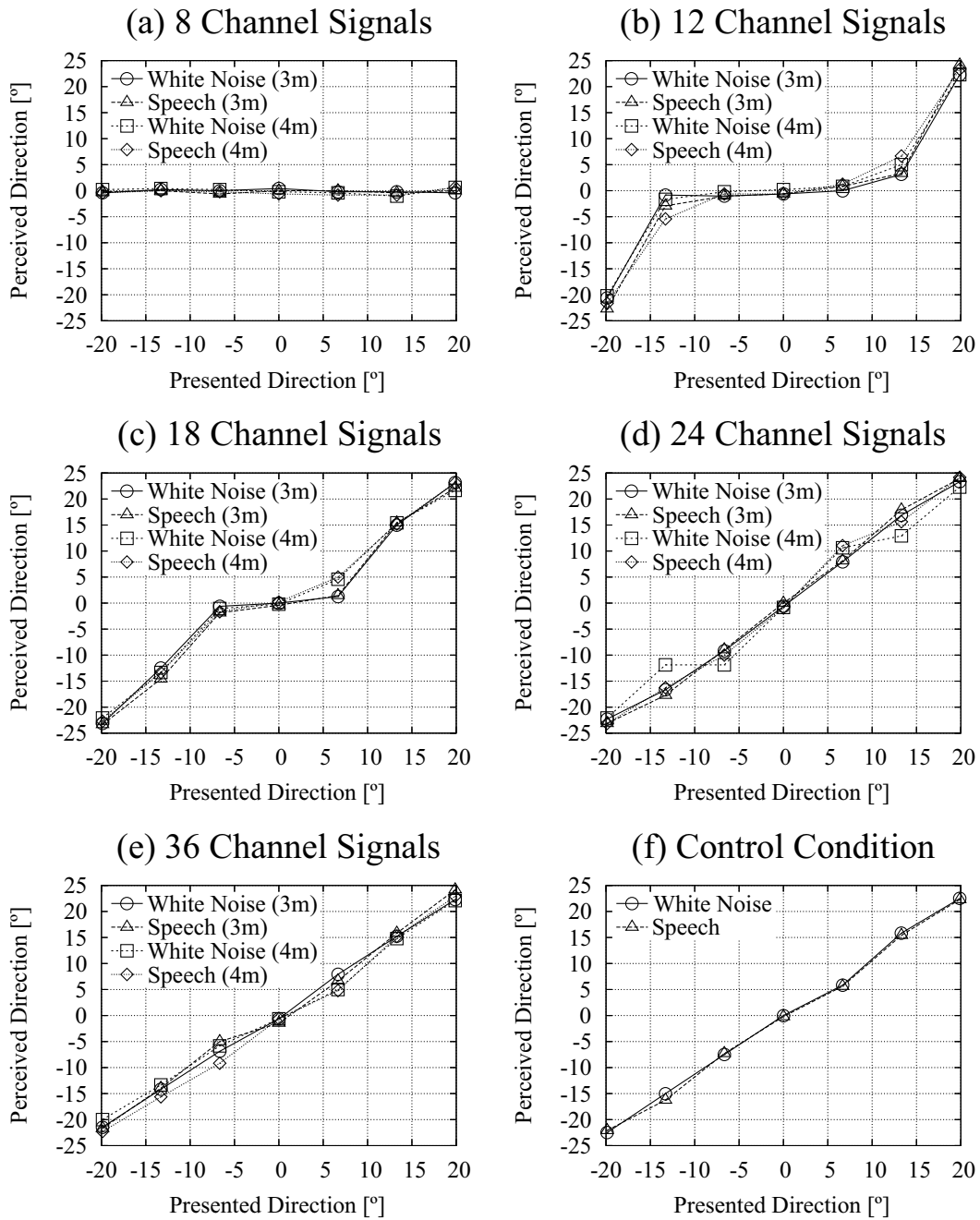


図 3.8 方向感の主観評価における実験結果 (前方位置の場合)

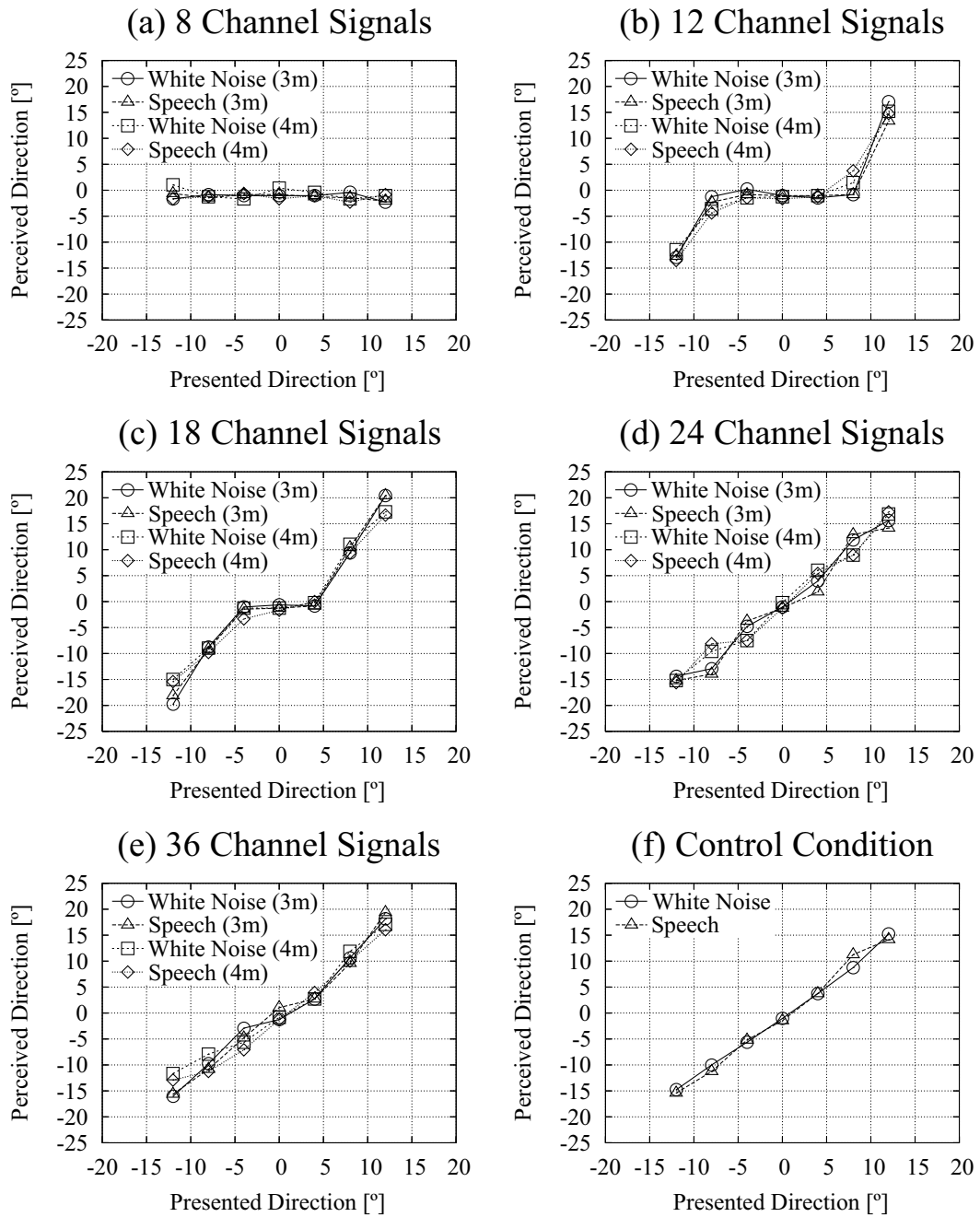


図 3.9 方向感の主観評価における実験結果 (後方位置の場合)

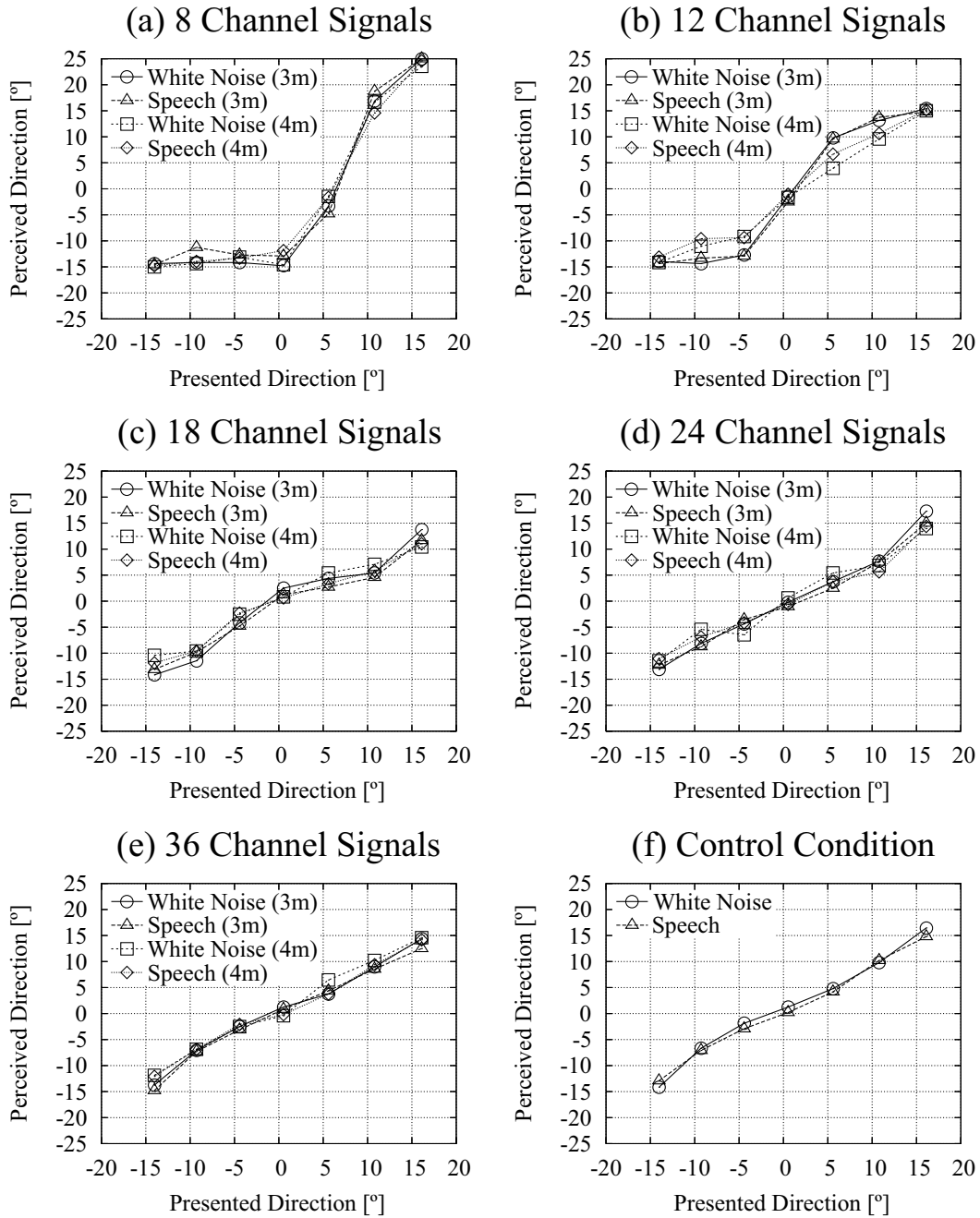


図 3.10 方向感の主観評価における実験結果 (側方位置の場合)

表 3.1 方向感の主観評価における試行条件

	要因	備考
練習試行 (21)	= 提示方向 (7) ×[チャンネル信号の数 (1) × 距離 (2) + 統制条件]	チャンネル信号の数...36 距離...3, 4 m
本試行 (308)	= 提示方向 (7) ×[チャンネル信号の数 (5) × 距離 (2) + 統制条件] × 繰り返し (4)	チャンネル信号の数...8, 12, 18, 24, 36 距離...3, 4 m

位できているので、その他の条件における回答方向のずれはチャンネル信号の数の変化によるものとみなすことができる。

チャンネル信号の数が 8, 12, 18 の場合には全ての聴取位置において回答方向が提示方向と異なる条件が存在する。これはチャンネル信号の数が少なすぎると、所望の方向へ音像を定位させることができなくなることを示している。この理由を図 3.11 を用いて説明する。

図 3.11 は中心位置に聴取者を配置し、チャンネル信号数が 8、提示方向が  $15^\circ$  の場合である。この場合、図 3.11(a) に示すように聴取者は音源信号に距離  $d_1, d_2, d_3$  に相当した遅延を設けた 3 つのチャンネル信号  $x_1(n), x_2(n), x_3(n)$  により合成された音像を定位することになる。但し、チャンネル信号のインデックスと各スピーカの方角は以下のように割り当てている： $\theta_1=0^\circ, \theta_2=45^\circ, \theta_3=-45^\circ$ 。言い換えると、聴取者は音源信号に距離  $d_1+d'_1, d_2+d'_2, d_3+d'_3$  に相当した遅延を設けた 3 つの音を聞くことになる。ここで、 $d'_1, d'_2, d'_3$  は各スピーカから聴取者までの距離である。このうち、 $d_1+d'_1$  が最も短い距離であるので、聴取者が最初に聞く音は  $x_1(n)$  である。一方、 $d_2+d'_2, d_3+d'_3$  が  $d_1+d'_1$  に比べて  $0.001 \times c$  m ( $c$  は音速) 以上長くなると、図 3.11(b) に示すように聴取者は  $x_1(n)$  から 1 ms 以上遅れて  $x_2(n), x_3(n)$  を聞くことになる。すると、 $x_1(n)$  に対する先行音効果 [35] が生じるので、 $x_2(n)$  と  $x_3(n)$  は音像の定位に全く寄与しなくなる。その結果、図 3.11(c) に示すように  $x_1(n)$  のみを聞いた場合と同じ方向 (図 3.11 の場合は  $0^\circ$ ) に音像が定位することになる。このことは聴取者が中心位置以外にいる場合でも同様に導き出される。従って、与えられた音源、聴取者、スピーカアレーの位置関係から聴取者が最初に聞く音に比べて他の音が 1 ms 以上遅れているかどうかを判定することによってチャンネル信号の数による方向感への影響を検

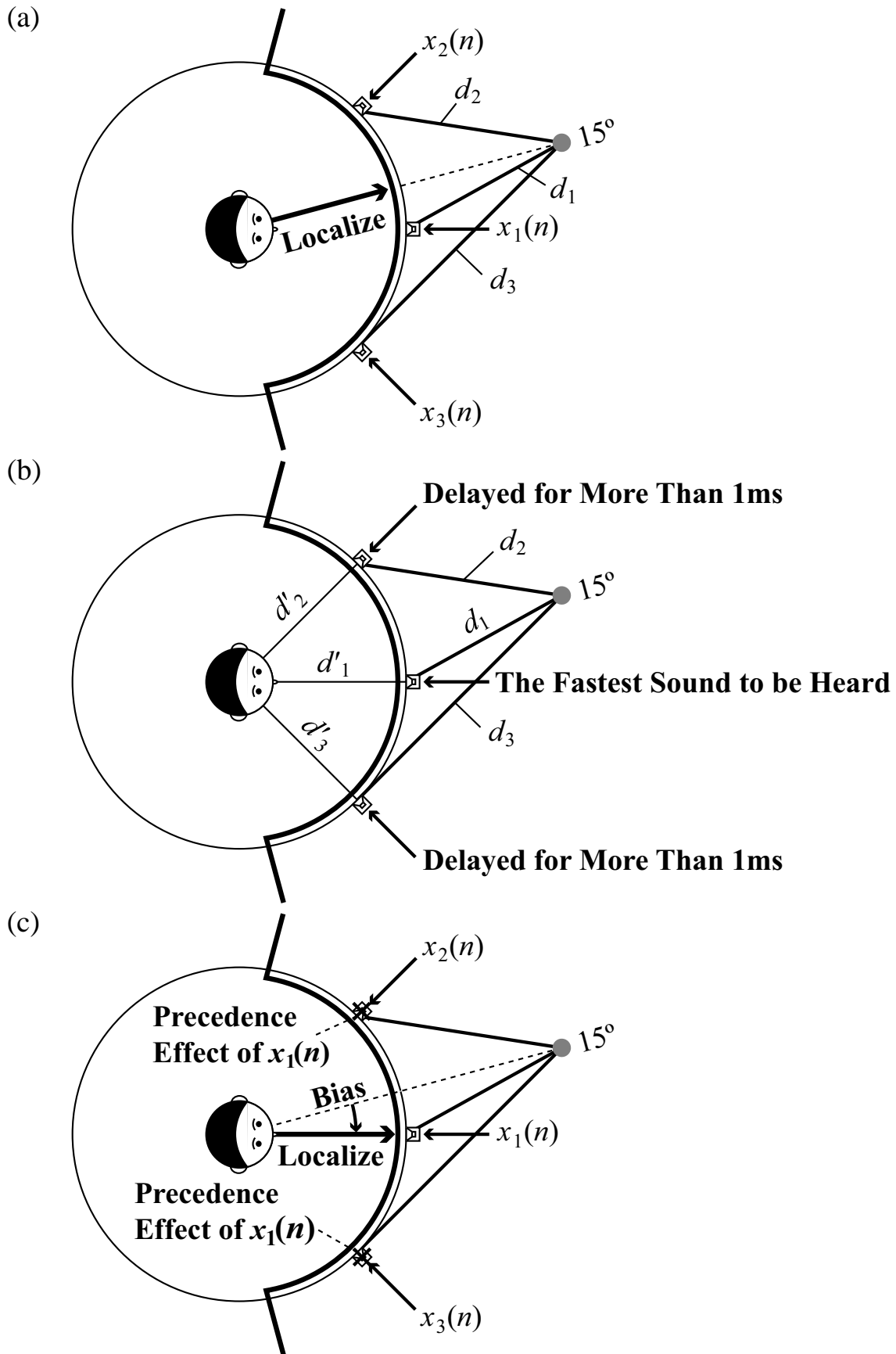


図 3.11 定位方向の偏位 (中心位置, チャネル信号の数 8, 提示方向  $15^\circ$  の場合)



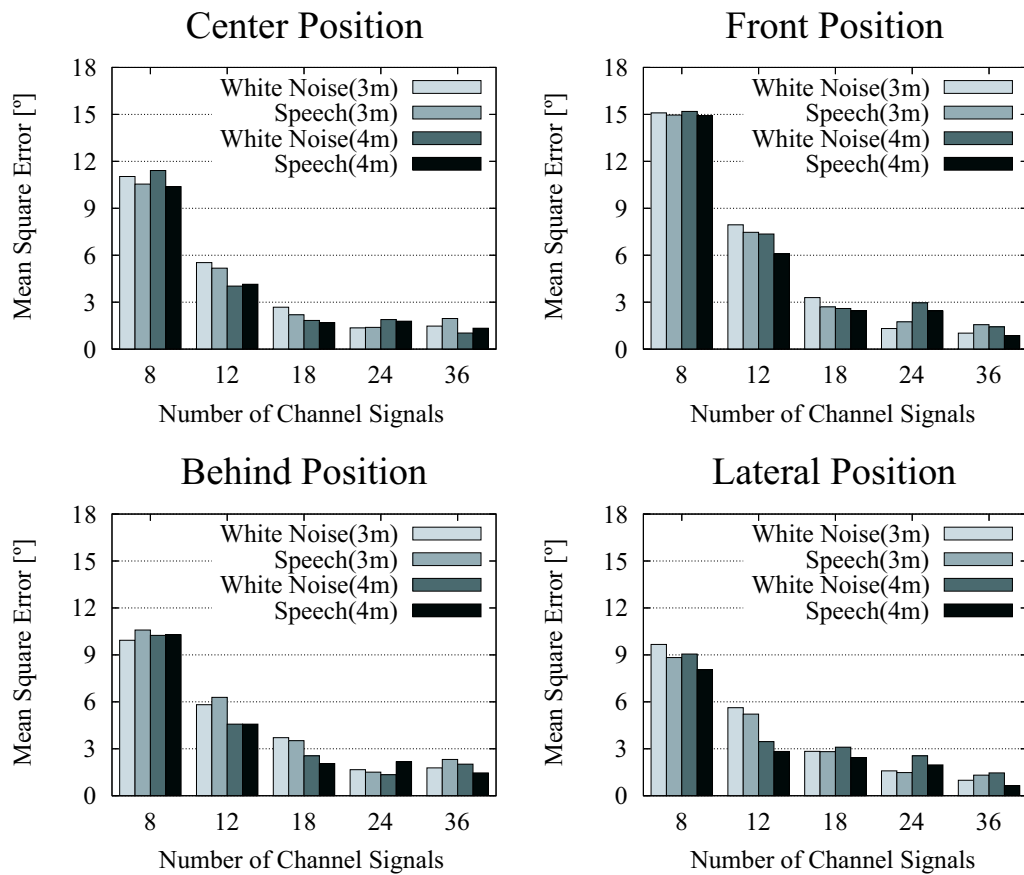


図 3.12 方向感の主観評価における平均二乗誤差

討することができると思われる。

一方，チャンネル信号の数が 24, 36 の場合には上記のような現象は生じないので，全ての聴取位置において回答方向は全て提示方向とほぼ同じである．そこで，より定量的な評価を行うために，(3.7) 式に示すような平均二乗誤差 (MSE) を聴取位置ごとに算出する．

$$\text{MSE}[\text{°}] = \sqrt{\frac{1}{7} \sum_{\phi} \{\psi(\phi) - \psi_{\text{control}}(\phi)\}^2} \quad (3.7)$$

ここで， $\phi$  は提示方向， $\psi_{\text{control}}(\phi)$  は統制条件における  $\phi$  に対する回答方向， $\psi(\phi)$  はその他の条件における  $\phi$  に対する回答方向を表す．チャンネル信号の数に関する条件ごとに算出した MSE を図 3.12 に示す．チャンネル信号の数が 24, 36 の場合，MSE の値はどの聴取位置においても  $3^{\circ}$  以下である．これは正面方向における広帯域雑音の方向弁別限 [36] よりも小さい．よって，方向感に及ばず影響という観点に基づいて波面合成法による立体音場再生システムを構築する場合，第 2 章での結果に基づけば約 300 ~ 400Hz 以下の範囲でしか波面が忠実に合成されていないような場合においても方向感十分に再現することができると思われる．

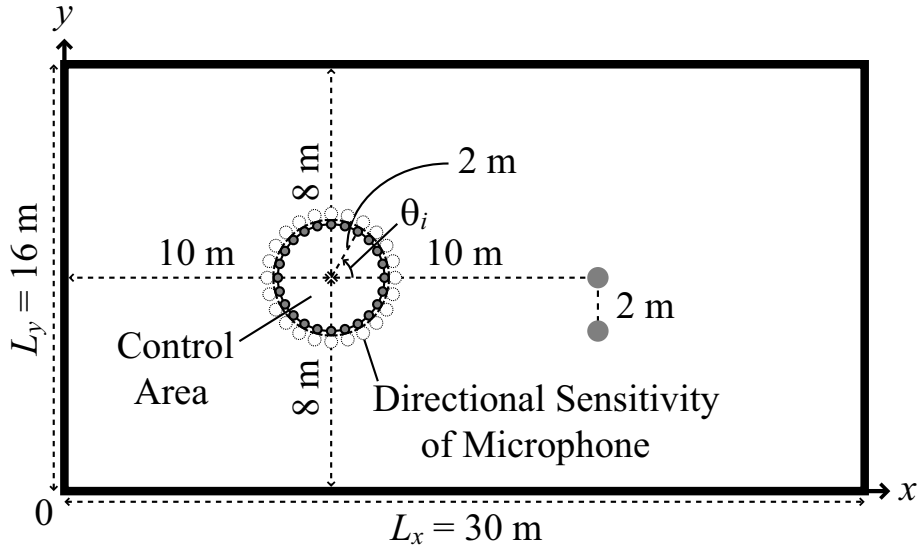


図 3.13 空間印象の主観評価における原音場

### 3.3 主観評価実験 2 - 空間印象への影響 -

#### 3.3.1 チャネル信号の作成

空間印象は主に音源からの反射音が作用している．従って，原音場は残響空間とし，虚像法 [30] を用いて 1 個の音源からマイクロホンまでの室内インパルス応答を計算機上で作成した．原音場の様子を図 3.13 に示す．各マイクロホンの指向性は制御領域の外側に向けている．虚像法によると， $x=0$ ， $x=L_x$ ， $y=0$ ， $y=L_y$  に壁面  $x_1$ ， $x_2$ ， $y_1$ ， $y_2$  を配置し， $\mathbf{s}=(s_x, s_y)^T$  に無指向性音源， $\mathbf{r}_i=(r_{ix}, r_{iy})^T$  に  $i$  番目の指向性マイクロホンを配置した場合の室内インパルス応答  $g_i(n)$  は (3.8) 式のように表わされる．

$$g_i(n) = \sum_{p_x=0}^1 \sum_{p_y=0}^1 \sum_{q_x=-\infty}^{\infty} \sum_{q_y=-\infty}^{\infty} D_i(\mathbf{p}, \mathbf{q}) \beta_{x1}^{|q_x-p_x|} \beta_{x2}^{|q_x|} \beta_{y1}^{|q_y-p_y|} \beta_{y2}^{|q_y|} \frac{\delta[n - \text{round}(\frac{|\mathbf{d}_i(\mathbf{p}, \mathbf{q})| F_s}{c})]}{4\pi |\mathbf{d}_i(\mathbf{p}, \mathbf{q})|} \quad (3.8)$$

但し， $\beta_{x1}$ ， $\beta_{x2}$ ， $\beta_{y1}$ ， $\beta_{y2}$  は壁面  $x_1$ ， $x_2$ ， $y_1$ ， $y_2$  の反射係数， $\delta(n)$  は Dirac のデルタ関数， $F_s$  (=48 kHz) はサンプリング周波数， $c$  (=340 m/s) は音速を表わす．また，

$$\begin{aligned} \mathbf{d}_i(\mathbf{p}, \mathbf{q}) &= \mathbf{s}(\mathbf{p}, \mathbf{q}) - \mathbf{r}_i \\ &= \begin{pmatrix} (1 - 2p_x)s_x + 2q_x L_x \\ (1 - 2p_y)s_y + 2q_y L_y \end{pmatrix} - \begin{pmatrix} r_{ix} \\ r_{iy} \end{pmatrix} \end{aligned} \quad (3.9)$$

で， $\mathbf{s}(\mathbf{p}, \mathbf{q})$  は虚像音源の位置ベクトルを表わす．一方， $D_i(\mathbf{p}, \mathbf{q})$  は  $i$  番目のマイクロホンの指向性を表わす関数であり，本実験では方向感の主観評価実験と同じように (3.10) 式に

示されるような超指向性に設定した。

$$D_i(\mathbf{p}, \mathbf{q}) = \begin{cases} \frac{\mathbf{u}_i \cdot \mathbf{d}_i(\mathbf{p}, \mathbf{q})}{|\mathbf{u}_i| |\mathbf{d}_i(\mathbf{p}, \mathbf{q})|} & (\mathbf{u}_i \cdot \mathbf{d}_i(\mathbf{p}, \mathbf{q}) \geq 0) \\ 0 & (\mathbf{u}_i \cdot \mathbf{d}_i(\mathbf{p}, \mathbf{q}) < 0) \end{cases} \quad (3.10)$$

但し,  $\mathbf{u}_i$  は  $i$  番目のマイクロホンの指向ベクトルである。

本実験では  $\mathbf{s}$ ,  $\mathbf{u}_i$ ,  $\mathbf{r}_i$  は以下のように設定した。

$$\mathbf{s} = \begin{pmatrix} 20 \\ 8 \end{pmatrix}, \quad \mathbf{u}_i = \begin{pmatrix} \cos \frac{\pi(i-12)}{12} \\ \sin \frac{\pi(i-12)}{12} \end{pmatrix}, \quad \mathbf{r}_i = 2\mathbf{u}_i + \begin{pmatrix} 10 \\ 8 \end{pmatrix} \quad (3.11)$$

$\beta_{x1}, \beta_{x2}, \beta_{y1}, \beta_{y2}$  は同一の値とし, 原音場をコンサートホールと想定するため, 値を 0.75 とした。  $p_x, p_y, q_x, q_y$  の値は以下の条件を満たす全ての組み合わせを用いた。

$$|q_x - p_x| + |q_x| + |q_y - p_y| + |q_y| \leq R \quad (3.12)$$

但し,  $R (=50)$  は設定した壁面反射回数の最大値である。その結果, 残響時間を Schroeder 法 [37] によって算出したところ, 1.3 s であった。ドライソース  $s(n)$  にはサンプリング周波数 48 kHz, 長さ 4 s の音声とフルートを用いた。音声は無響室内にて録音し, フルードは MIDI 音源から作成した。そして,  $s(n)$  と  $g_i(n)$  から  $i$  番目のマイクロホンのチャンネル信号  $x_i(n)$  を (3.13) 式によって作成した。

$$\begin{aligned} x_i(n) &= g_i(n) * s(n) \\ &= \sum_{p_x=0}^1 \sum_{p_y=0}^1 \sum_{q_x=-\infty}^{\infty} \sum_{q_y=-\infty}^{\infty} D_i(\mathbf{p}, \mathbf{q}) \beta_{x1}^{|q_x-p_x|} \beta_{x2}^{|q_x|} \beta_{y1}^{|q_y-p_y|} \beta_{y2}^{|q_y|} \frac{s[n - \text{round}(\frac{|\mathbf{d}_i(\mathbf{p}, \mathbf{q})| F_s}{c})]}{4\pi |\mathbf{d}_i(\mathbf{p}, \mathbf{q})|} \end{aligned} \quad (3.13)$$

### 3.3.2 実験環境

実験は 3.2 節における実験で用いた部屋と同じ部屋 (約 80 ms の低残響室) で行った。図 3.14 に示すように半径 2 m の円周上に計 24 個のスピーカ (Emic: Soundevice) を配置し, さらに 4 箇所の聴取位置を設定した。灰色の円はスピーカアレーによって再生される音像を表す。部屋の暗騒音レベルは 25.0 dB(A) で, 音圧レベルは円の中心において約 70 dB(A) に設定した。また, 聴取者にスピーカアレーを見せないようにするため, 照明を薄暗くし, スピーカアレーの前方には音響透過カーテンを設置した。

本実験ではチャンネル信号の数に関する条件を 7 種類設定した。実験条件を図 3.15 ~ 3.16 に示す。条件 (a) は典型的なモノホニックとして設定した条件である。一方, 条件 (b) ~ (g) では 24 個のスピーカの中から選ばれたスピーカから (3.13) 式で算出されたチャンネル信号を同時に再生した。その際, 選ばれなかったスピーカには零信号を割り当てた。これらの条件では第 2 章での結果より, チャンネル信号の数が増えるほど波面を忠実に合成することができる周波数帯域が広がっていくと考えることができる。

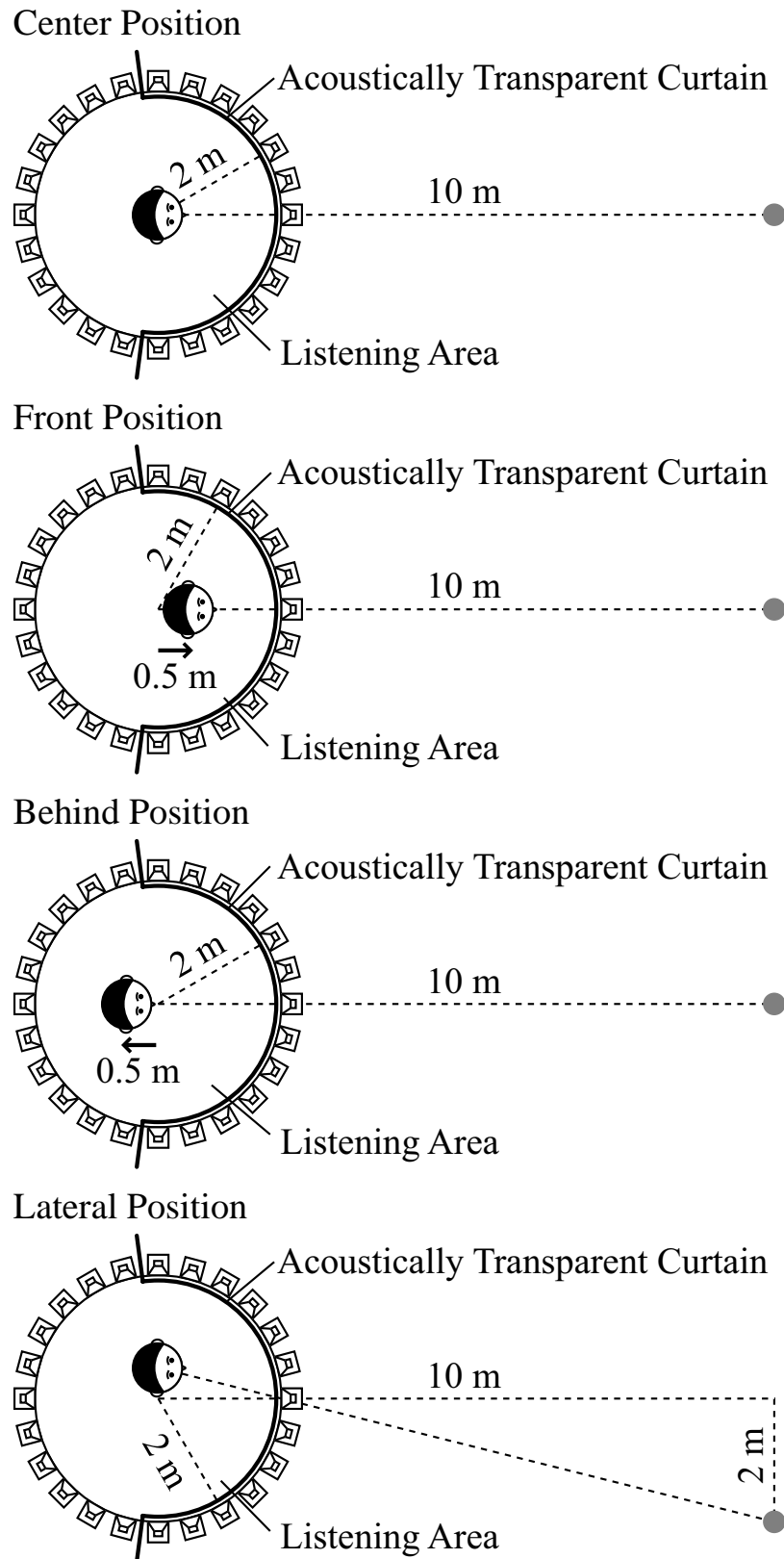


図 3.14 空間印象の主観評価における実験環境

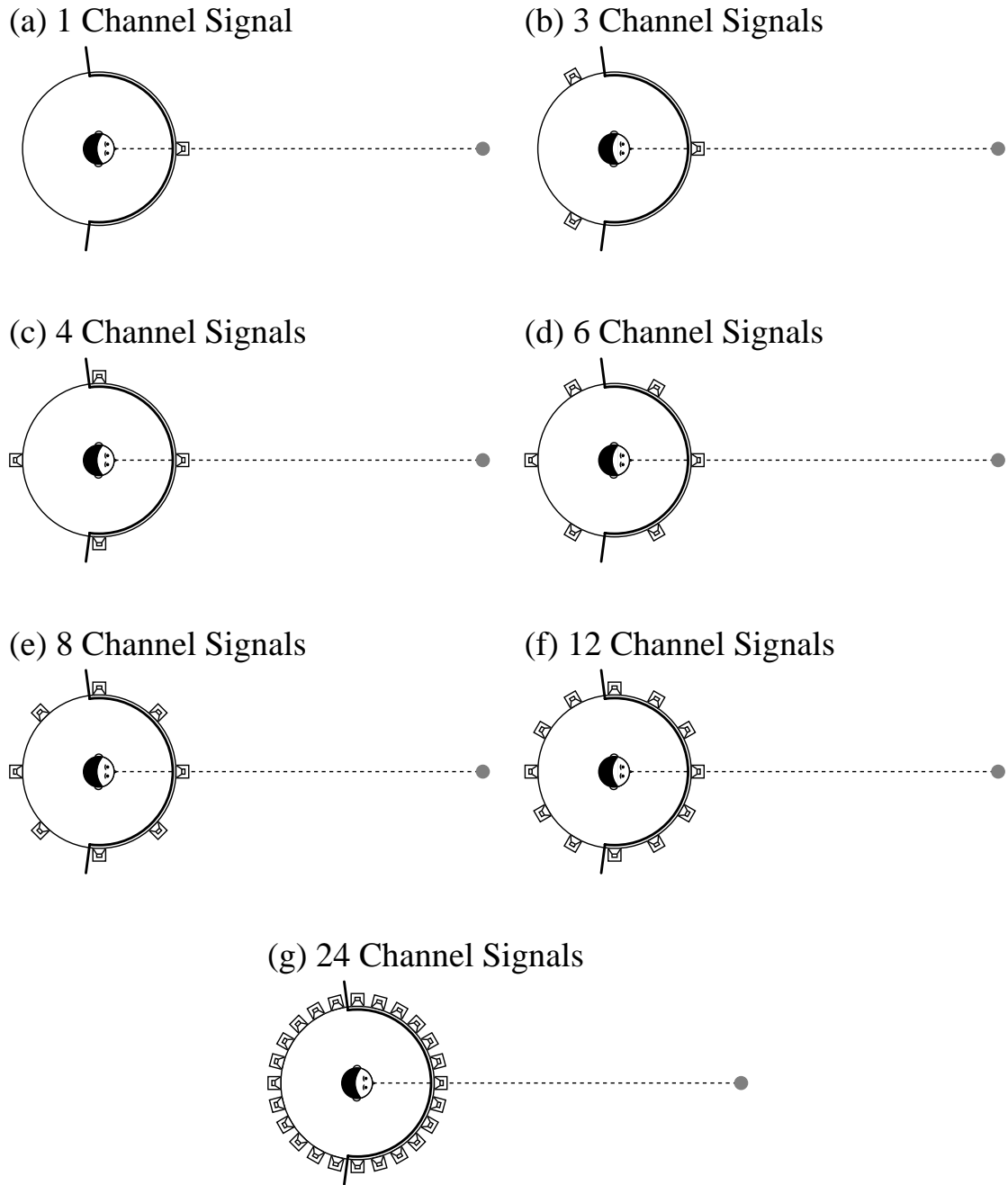


図 3.15 空間印象の主観評価における実験条件 (中心位置の場合)

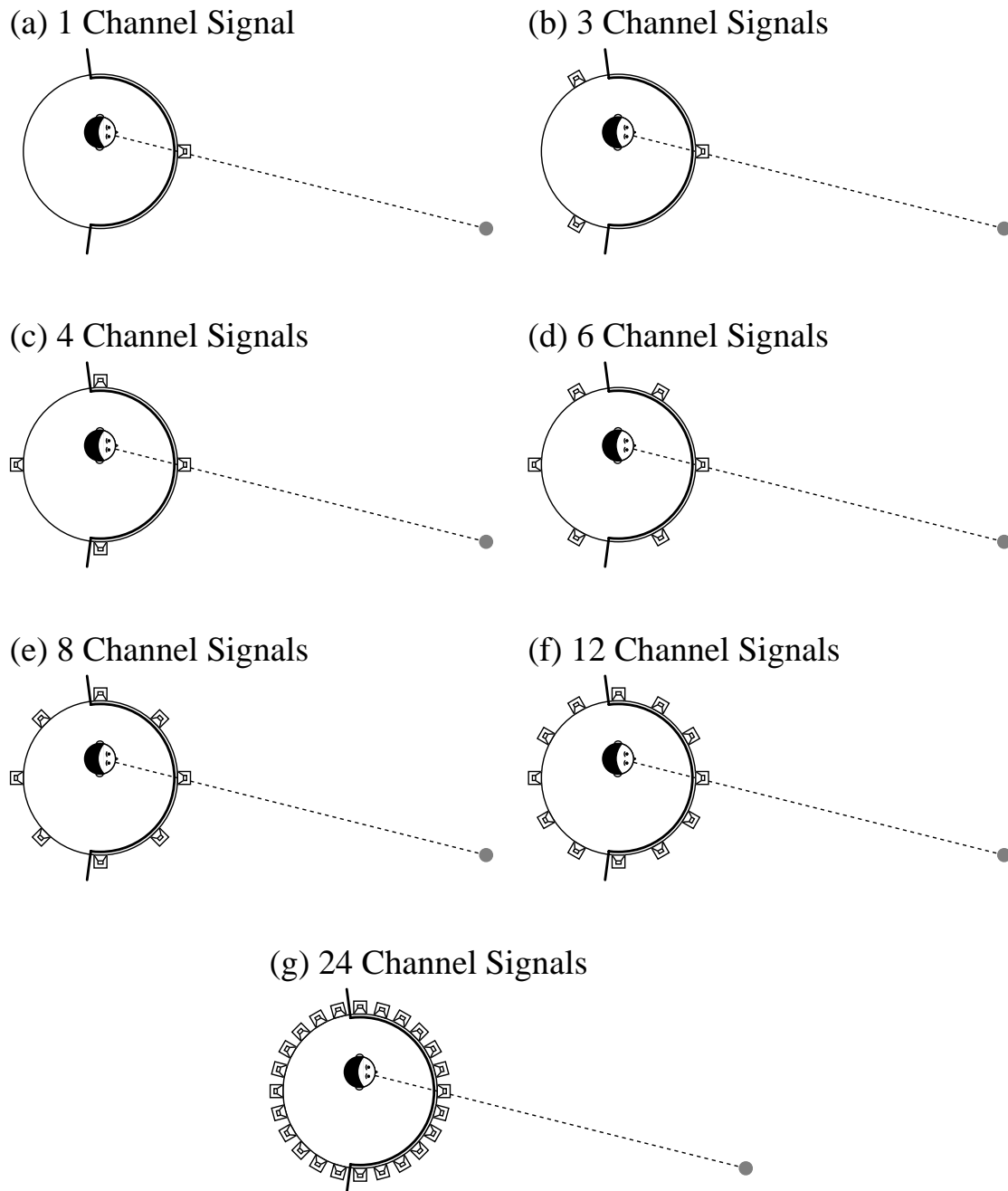


図 3.16 空間印象の主観評価における実験条件 (側方位置の場合)

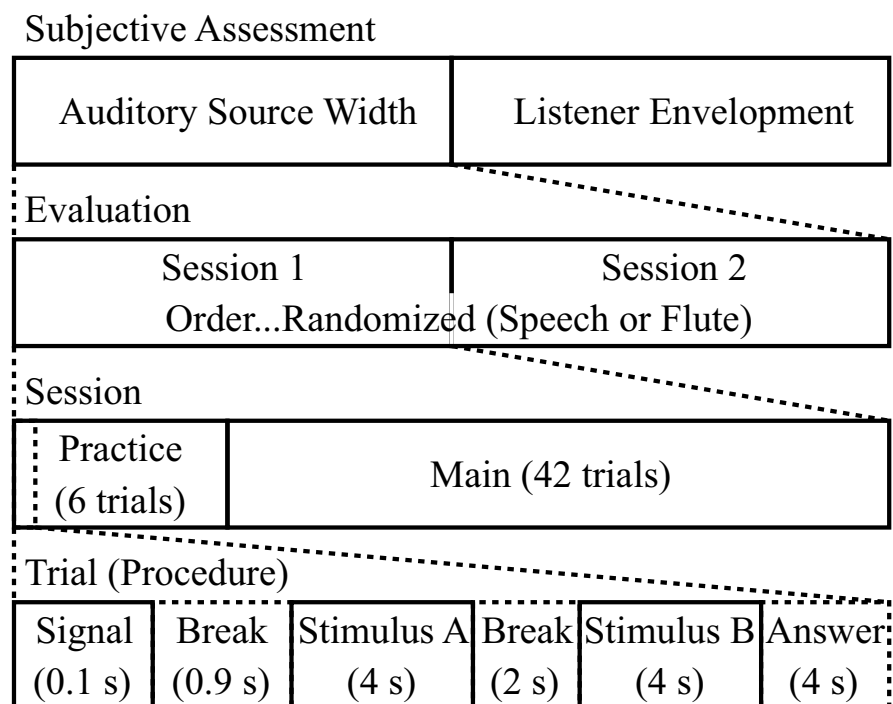


図 3.17 方向感の主観評価における実験流れ図

### 3.3.3 実験計画

聴取者は日常生活において聴力に異常の見られない 16 名の学生 (うち男性 12 名, 女性 4 名) である。1 箇所の聴取位置あたり 4 名ずつの被験者で実験を行った。評価手法には Scheffé の一対比較法 (浦の変法)[38] を用いた。主観評価実験の流れ図を図 3.17 に示す。Morimoto は空間印象は少なくとも 2 つの要素 (見かけの音源の幅 [Auditory Source Width, ASW] と音に包まれた感じ [Listener Envelopment, LEV]) に分類されるとしている [31]。そこで, 本実験においても ASW と LEV を別々に評価することで空間印象への影響を検討することにした。まず, 実験全体を評価尺度ごとに 2 つに分割し, それぞれの評価においてドライソースごとに 2 つのセッションに分割した。その際, ドライソースの提示順序は被験者ごとにランダム化した。各セッションにおいて 6 回の練習試行の後, 42 回の本試行を行った。練習試行は図 3.15 ~ 3.16 における 3 つの実験条件 [条件 (a), (b), (g)] の順列の総数 ( $3 \times 2 = 6$ ) であり, 本試行は図 3.15 ~ 3.16 における 7 つの実験条件 [条件 (a) ~ (g)] の順列の総数 ( $7 \times 6 = 42$ ) である。

主観評価実験に入る前に, 聴取者には図 3.18 を見せて ASW と LEV の定義について理解してもらった。この図は文献 [31] から引用したものである。その後, 先に提示した刺激音 (A) を基準にして後に提示した刺激音 (B) を聴取者に評定させた。その際, 評定は表

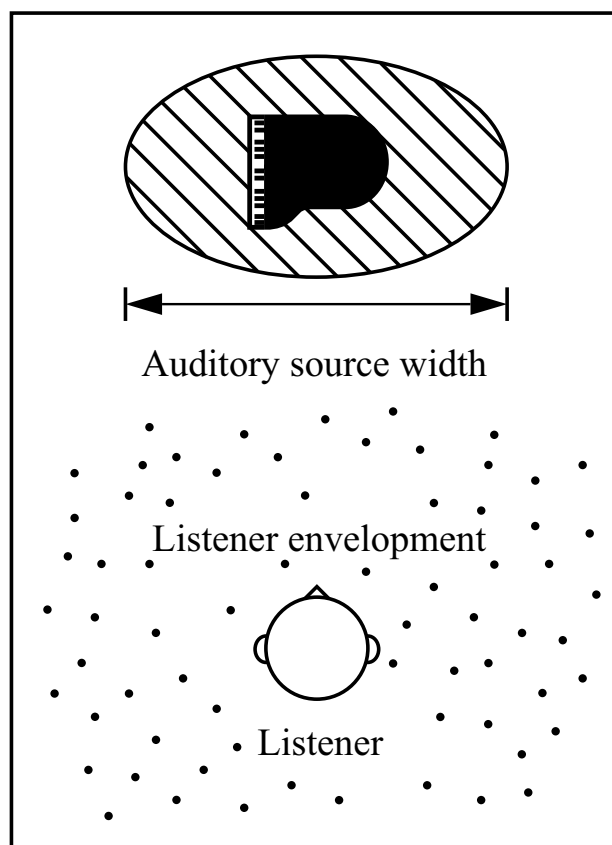


図 3.18 見かけの音源の幅 (ASW) と音に包まれた感じ (LEV) の概念図

表 3.2 Scheffé の一対比較法の評価表

評価	ASW	LEV
3	非常に広い	非常に包まれている
2	かなり広い	かなり包まれている
1	少し広い	少し包まれている
0	同じ	同じ
-1	少し狭い	少し包まれていない
-2	かなり狭い	かなり包まれていない
-3	非常に狭い	非常に包まれていない

3.2 に示すような 7 段階とした。また、評価の際には聴取者は頭部を自由に動かすことができた。



### 3.3.4 実験結果及び考察

図 3.15～3.16 で示した実験条件のうち、条件 (a) は単に聴取者の正面方向から音源信号が再生されているだけなので、ASW や LEV は非常に小さいと考えられる。一方、条件 (g) は聴取者の周りから反射音が再生されているので、ASW や LEV は非常に大きいと考えられる。ゆえに、聴取者が ASW や LEV を正しく識別しているとするならば、条件 (g) の評定値は条件 (a) の評定値よりも非常に高くなることが予測される。そこで、聴取者ごとに Scheffé の一対比較法 (浦の変法) に基づいて平均評定値を算出し、条件 (g) の評定値が条件 (a) の評定値よりも有意に高いかどうかを調べることで、聴取者の選別を行うことにした。

聴取者及びセッションごとに得られた結果を表 3.3～3.6 に示す。但し、 $\alpha_a, \alpha_g$  は条件 (a), (g) における平均評定値、 $Y_{0.01}$  は信頼度 99% のヤードスティックである。有意差水準 1% で  $\alpha_g > \alpha_a$  と判定されている場合、聴取者が ASW や LEV を正しく識別できていると判断できるため、分析には  $\alpha_g > \alpha_a$  と判定された聴取者のデータのみを用いることにする。

チャンネル信号の数に関する条件ごとの平均評定値を聴取位置ごとに図 3.19～3.20 に示す。エラーバーはヤードスティックによる 99% 信頼区間を表す。検討したチャンネル信号の数に関する条件の中で最も原音場に近いと言える 24 チャンネルを基準としてチャンネル信号の数による空間印象への影響を検討する。チャンネル信号の数が 1, 3, 4 の場合はチャンネル信号の数が 24 の場合よりも平均評定値が低いことが多い。一方、チャンネル信号の数が 6, 8, 12 の場合はチャンネル信号の数が 24 の場合と平均評定値が同じ程度であることが多い。つまり、チャンネル信号の数が 6 以上の場合は ASW や LEV に関する平均評定値に対するチャンネル信号の数による影響はないと言える。従って、空間印象への影響という観点によって波面合成法による立体音場再生システムを構築する場合、第 2 章の結果に基づけば約 100Hz 以下の範囲でしか波面が忠実に合成されていないような場合でも空間印象は十分に再現することができると考えられる。また、3.2 節で得られた結果と照らし合わせると、方向感が十分に再現されていないような場合でも空間印象は十分に再現することができると考えられる。

## 3.4 まとめ

本章では波面合成法による立体音場再生システムにおいてチャンネル信号の数が音場知覚に及ぼす影響を検討するために、方向感と空間印象に着目した 2 種類の主観評価実験を行った。その結果、正面付近に音源が存在する場合においては従来の波面の物理的な再現

表 3.3 聴取者の識別結果 (中心位置の場合)

ASW(音声)	聴取者 A	聴取者 B	聴取者 C	聴取者 D
$\alpha_a$	-1.2857	-1.4286	-0.9286	-1.5714
$\alpha_g$	0.5	0.5714	0.2143	0.3571
$Y_{0.01}$	0.8558	0.8654	2.1391	1.0282
$\alpha_g - \alpha_a - Y_{0.01}$	0.9299	1.1346	-0.9962	0.9003
$\alpha_g - \alpha_a + Y_{0.01}$	2.6415	2.8654	3.282	2.9567
判定	$\alpha_g > \alpha_a$	$\alpha_g > \alpha_a$	$\alpha_g = \alpha_a$	$\alpha_g > \alpha_a$
ASW(フルーツ)	聴取者 A	聴取者 B	聴取者 C	聴取者 D
$\alpha_a$	-1.1429	-1.7143	0.2857	-1.6429
$\alpha_g$	0.5	0.6429	0	0.9286
$Y_{0.01}$	0.8061	0.6208	1.697	1.3803
$\alpha_g - \alpha_a - Y_{0.01}$	0.8368	1.7364	-1.9827	1.1912
$\alpha_g - \alpha_a + Y_{0.01}$	2.449	2.978	1.4113	3.9518
判定	$\alpha_g > \alpha_a$	$\alpha_g > \alpha_a$	$\alpha_g = \alpha_a$	$\alpha_g > \alpha_a$
LEV(音声)	聴取者 A	聴取者 B	聴取者 C	聴取者 D
$\alpha_a$	-1.5714	-1.5714	-1.7857	-1.7143
$\alpha_g$	0.7143	0.7143	0.3571	0.2857
$Y_{0.01}$	0.9386	0.9703	0.956	1.0282
$\alpha_g - \alpha_a - Y_{0.01}$	1.3471	1.3154	1.1868	0.9718
$\alpha_g - \alpha_a + Y_{0.01}$	3.2243	3.256	3.0988	3.0282
判定	$\alpha_g > \alpha_a$	$\alpha_g > \alpha_a$	$\alpha_g > \alpha_a$	$\alpha_g > \alpha_a$
LEV(フルーツ)	聴取者 A	聴取者 B	聴取者 C	聴取者 D
$\alpha_a$	-1.2143	-1.4286	-0.3571	-1.4286
$\alpha_g$	0.5	0.5	0.1429	1
$Y_{0.01}$	0.7818	0.8525	1.268	0.9816
$\alpha_g - \alpha_a - Y_{0.01}$	0.9325	1.0761	-0.768	1.447
$\alpha_g - \alpha_a + Y_{0.01}$	2.4961	2.7811	1.768	3.4102
判定	$\alpha_g > \alpha_a$	$\alpha_g > \alpha_a$	$\alpha_g = \alpha_a$	$\alpha_g > \alpha_a$

表 3.4 聴取者の識別結果 (前方位置の場合)

ASW(音声)	聴取者 E	聴取者 F	聴取者 G	聴取者 H
$\alpha_a$	-1.6429	-0.8571	-1.2143	-0.5714
$\alpha_g$	1.3571	0.2857	0.5	0.4286
$Y_{0.01}$	0.8935	1.5441	1.8595	0.7531
$\alpha_g - \alpha_a - Y_{0.01}$	2.1065	-0.4013	-0.1452	0.2469
$\alpha_g - \alpha_a + Y_{0.01}$	3.8935	2.6869	3.5738	1.7531
判定	$\alpha_g > \alpha_a$	$\alpha_g = \alpha_a$	$\alpha_g = \alpha_a$	$\alpha_g > \alpha_a$
ASW(フルート)	聴取者 E	聴取者 F	聴取者 G	聴取者 H
$\alpha_a$	-1.5714	-0.2857	-0.9286	-0.2857
$\alpha_g$	1.0714	0.3571	0.5714	-0.1429
$Y_{0.01}$	0.9298	0.634	1.5423	0.7818
$\alpha_g - \alpha_a - Y_{0.01}$	1.713	0.0088	-0.0423	-0.639
$\alpha_g - \alpha_a + Y_{0.01}$	3.5726	1.2768	3.0423	0.9246
判定	$\alpha_g > \alpha_a$	$\alpha_g > \alpha_a$	$\alpha_g = \alpha_a$	$\alpha_g = \alpha_a$
LEV(音声)	聴取者 E	聴取者 F	聴取者 G	聴取者 H
$\alpha_a$	-2.1429	0.2857	-0.0714	-0.2857
$\alpha_g$	0.8571	0.2143	0.8571	0.2857
$Y_{0.01}$	0.8061	1.1779	1.902	0.7711
$\alpha_g - \alpha_a - Y_{0.01}$	2.1939	-1.2493	-0.9735	-0.1997
$\alpha_g - \alpha_a + Y_{0.01}$	3.8061	1.1065	2.8305	1.3425
判定	$\alpha_g > \alpha_a$	$\alpha_g = \alpha_a$	$\alpha_g = \alpha_a$	$\alpha_g = \alpha_a$
LEV(フルート)	聴取者 E	聴取者 F	聴取者 G	聴取者 H
$\alpha_a$	-1.7857	0	-0.1429	-0.1429
$\alpha_g$	1.0714	0	0	0.0714
$Y_{0.01}$	0.8329	0.5141	1.3981	0.7818
$\alpha_g - \alpha_a - Y_{0.01}$	2.0242	-0.5141	-1.2552	-0.5675
$\alpha_g - \alpha_a + Y_{0.01}$	3.69	0.5141	1.541	0.9961
判定	$\alpha_g > \alpha_a$	$\alpha_g = \alpha_a$	$\alpha_g = \alpha_a$	$\alpha_g = \alpha_a$

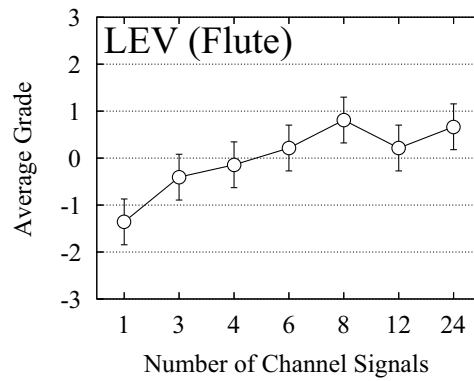
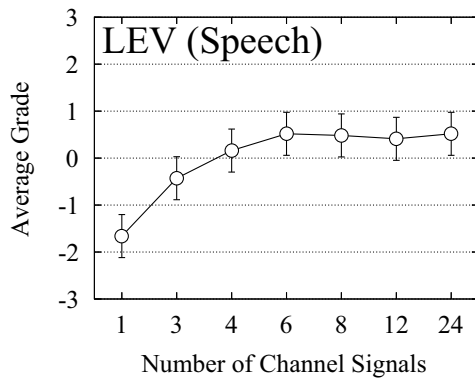
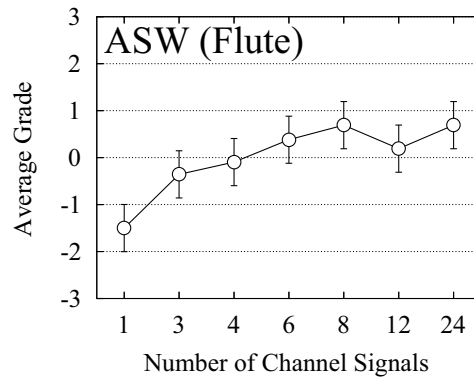
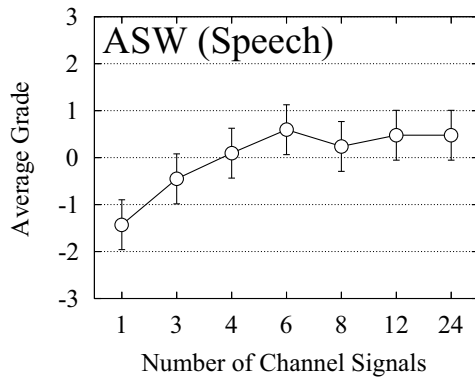
表 3.5 聴取者の識別結果 (後方位置の場合)

ASW(音声)	聴取者 I	聴取者 J	聴取者 K	聴取者 L
$\alpha_a$	-1.9286	-1.1429	-1.6429	-1.8571
$\alpha_g$	0.6429	0.5714	0.4286	0.4286
$Y_{0.01}$	1.5441	1.3803	1.3001	0.8935
$\alpha_g - \alpha_a - Y_{0.01}$	1.0274	0.334	0.7714	1.3922
$\alpha_g - \alpha_a + Y_{0.01}$	4.1156	3.0946	3.3716	3.1792
判定	$\alpha_g > \alpha_a$	$\alpha_g > \alpha_a$	$\alpha_g > \alpha_a$	$\alpha_g > \alpha_a$
ASW(フルート)	聴取者 I	聴取者 J	聴取者 K	聴取者 L
$\alpha_a$	-1.1429	-0.6429	-1.6429	-1.7143
$\alpha_g$	0.2143	0.2857	0.4286	0.3571
$Y_{0.01}$	1.3662	1.0906	1.1733	1.1423
$\alpha_g - \alpha_a - Y_{0.01}$	-0.009	-0.162	0.8982	0.9291
$\alpha_g - \alpha_a + Y_{0.01}$	2.7234	2.0192	3.2448	3.2137
判定	$\alpha_g = \alpha_a$	$\alpha_g = \alpha_a$	$\alpha_g > \alpha_a$	$\alpha_g > \alpha_a$
LEV(音声)	聴取者 I	聴取者 J	聴取者 K	聴取者 L
$\alpha_a$	-2	-1.2857	-1.5	-1.6429
$\alpha_g$	0.3571	0.6429	1	0.5714
$Y_{0.01}$	1.083	0.7957	1.1081	0.9208
$\alpha_g - \alpha_a - Y_{0.01}$	1.2741	1.1329	1.3919	1.2935
$\alpha_g - \alpha_a + Y_{0.01}$	3.4401	2.7243	3.6081	3.1351
判定	$\alpha_g > \alpha_a$	$\alpha_g > \alpha_a$	$\alpha_g > \alpha_a$	$\alpha_g > \alpha_a$
LEV(フルート)	聴取者 I	聴取者 J	聴取者 K	聴取者 L
$\alpha_a$	-1.7857	-1.1429	-1.4286	-1.7857
$\alpha_g$	0.4286	0.2143	0.5714	0.0714
$Y_{0.01}$	1.1081	0.9386	0.7853	0.6719
$\alpha_g - \alpha_a - Y_{0.01}$	1.1062	0.4186	1.2147	1.1852
$\alpha_g - \alpha_a + Y_{0.01}$	3.3224	2.2958	2.7853	2.529
判定	$\alpha_g > \alpha_a$	$\alpha_g > \alpha_a$	$\alpha_g > \alpha_a$	$\alpha_g > \alpha_a$

表 3.6 聴取者の識別結果 (側方位置の場合)

ASW(音声)	聴取者 M	聴取者 N	聴取者 O	聴取者 P
$\alpha_a$	-0.2857	-2.0714	-1.4286	-1.7143
$\alpha_g$	0.0714	0.6429	0.7857	0.7857
$Y_{0.01}$	1.0753	0.8061	1.5949	1.1006
$\alpha_g - \alpha_a - Y_{0.01}$	-0.7182	1.9082	0.6194	1.3994
$\alpha_g - \alpha_a + Y_{0.01}$	1.4324	3.5204	3.8092	3.6006
判定	$\alpha_g = \alpha_a$	$\alpha_g > \alpha_a$	$\alpha_g > \alpha_a$	$\alpha_g > \alpha_a$
ASW(フルート)	聴取者 M	聴取者 N	聴取者 O	聴取者 P
$\alpha_a$	-0.5714	-2	-1.4286	-1.6429
$\alpha_g$	0.2143	0.8571	0.7857	0.6429
$Y_{0.01}$	0.8026	0.6469	1.2527	0.8061
$\alpha_g - \alpha_a - Y_{0.01}$	-0.0169	2.2102	0.9616	1.4797
$\alpha_g - \alpha_a + Y_{0.01}$	1.5883	3.504	3.467	3.0919
判定	$\alpha_g = \alpha_a$	$\alpha_g > \alpha_a$	$\alpha_g > \alpha_a$	$\alpha_g > \alpha_a$
LEV(音声)	聴取者 M	聴取者 N	聴取者 O	聴取者 P
$\alpha_a$	-0.6429	-2	-1.2857	-1.0714
$\alpha_g$	0.5	0.4286	0.7857	0.9286
$Y_{0.01}$	1.2328	0.4809	0.9788	1.3842
$\alpha_g - \alpha_a - Y_{0.01}$	-0.0899	1.9477	1.0926	0.6158
$\alpha_g - \alpha_a + Y_{0.01}$	2.3757	2.9095	3.0502	3.3842
判定	$\alpha_g = \alpha_a$	$\alpha_g > \alpha_a$	$\alpha_g > \alpha_a$	$\alpha_g > \alpha_a$
LEV(フルート)	聴取者 M	聴取者 N	聴取者 O	聴取者 P
$\alpha_a$	-0.6429	-2.0714	-1.1429	-1.9286
$\alpha_g$	0.5	0.5	0.5	0.9286
$Y_{0.01}$	1.2328	0.5795	0.9208	1.0702
$\alpha_g - \alpha_a - Y_{0.01}$	-0.0899	1.9919	0.7221	1.787
$\alpha_g - \alpha_a + Y_{0.01}$	2.3757	3.1509	2.5637	3.9274
判定	$\alpha_g = \alpha_a$	$\alpha_g > \alpha_a$	$\alpha_g > \alpha_a$	$\alpha_g > \alpha_a$

## Center Position



## Front Position

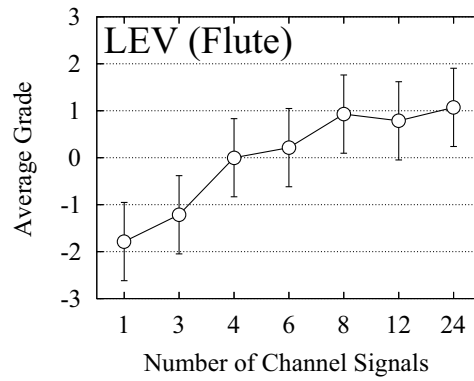
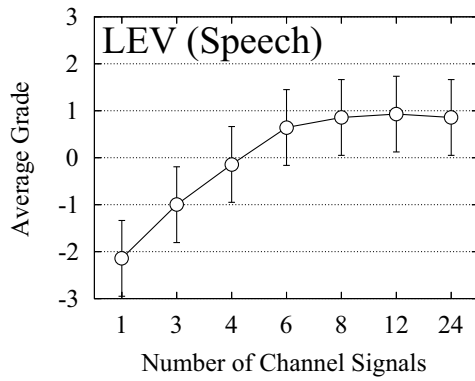
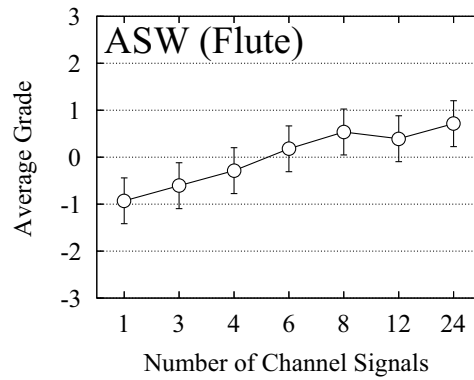
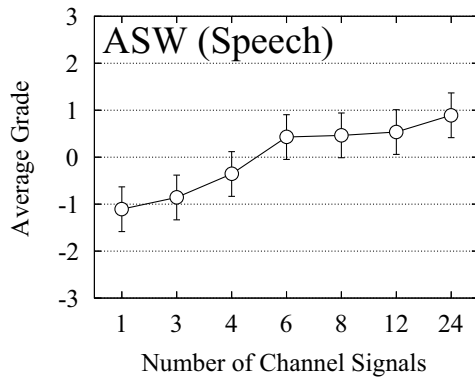
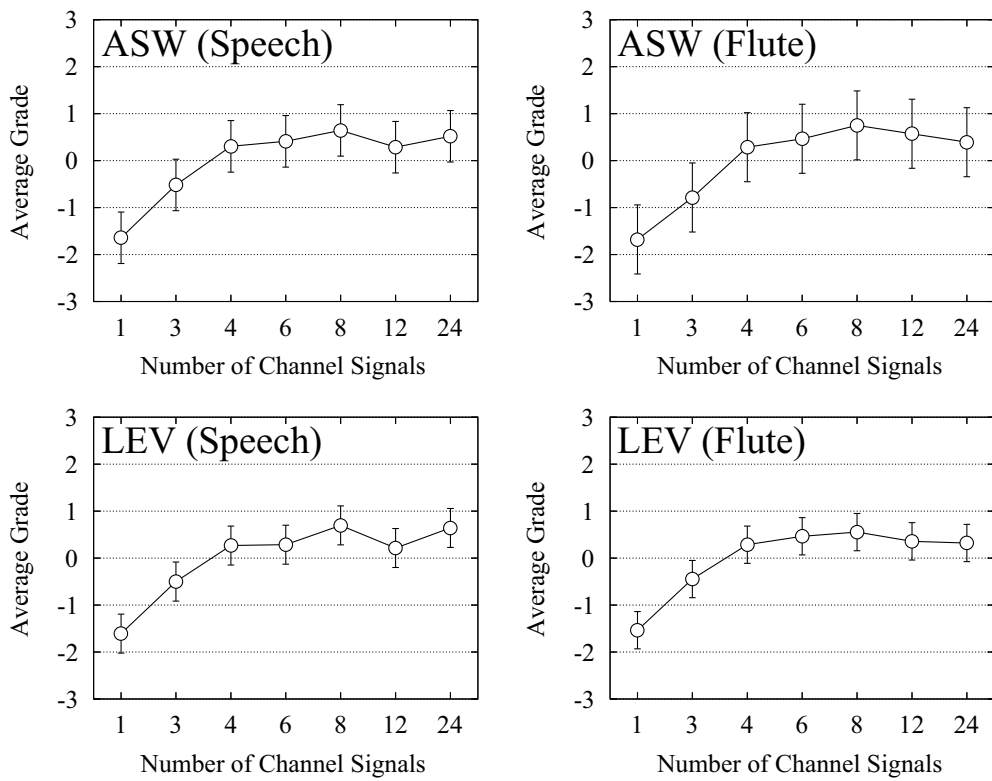


図 3.19 空間印象の主観評価における実験結果 (中心, 前方位置)

Behind Position



Lateral Position

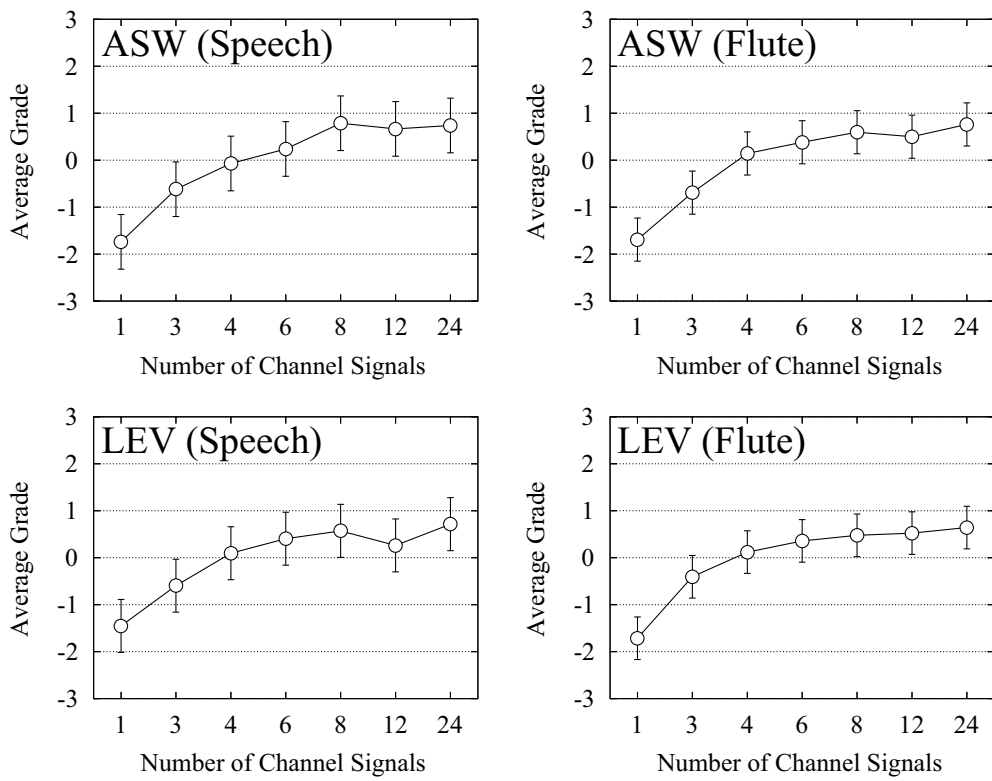


図 3.20 空間印象の主観評価における実験結果 (後方, 側方位置)

精度による観点では非常に低い周波数帯域でしか波面が忠実に合成されていなくても方向感や空間印象は十分に再現されていることが分かった。これらの主観評価実験による結果から、波面合成法による立体音場再生システムを構築する際に使用するマイクロホンやスピーカの数も24個程度に削減しても、聴取者は聴取領域の中にさえ居れば十分な臨場感を感じることができるということが分かった。



## 第 4 章

# 音源抽出に基づく伝送量の削減

### 4.1 はじめに

第 3 章での結果より，波面合成法による立体音場再生システムの構築の際にマイクロホン及びスピーカの数 を 24 個に削減しても方向感や空間印象を十分に再現できることが分かった．しかし，このシステムを遠隔地間で実現するためにチャンネル信号を伝送する際には，図 1.2 で示したような従来普及しているシステムよりも多くの伝送量を必要とするため，映像と組み合わせた通信システムを構築する際には通信容量が足りなくなる可能性がある．

このような問題を解決する一つの方法として音源のすぐそばにマイクロホンを配置し，収録した音源信号を伝送した上で音源からマイクロホンまでの室内インパルス応答を畳み込んでチャンネル信号を復元する方法が考えられる．しかし，実際にはマイクロホンが映像に映り込んでしまうなどの理由でマイクロホンを音源のすぐそばに配置できないことが多い．このような事態を避けるためにはチャンネル信号を直接処理することによって伝送量を削減する（この処理のことを以降「符号化」と呼ぶ）必要が生じてくる．

従来提案されている符号化手法としては DVD-video に応用されている AC-3[39] やデジタル放送に応用されている MPEG2 AAC[40] などが挙げられ，これらの手法を適用すれば 1 チャンネルあたり  $48 \text{ kHz} \times 16 \text{ bit} = 768 \text{ kbps}$  を必要とする伝送量を 64 kbps にまで削減することができる．しかし，達成される伝送量は 1 チャンネルあたりの値なので，波面合成法による立体音場再生システムのようにチャンネルが多数存在する場合には伝送量がチャンネル信号の数に比例して莫大な量になってしまう．一方，チャンネル信号間の冗長性を除去する空間符号化手法 [41, 42] も提案されているが，これらの手法を用いても伝送量はチャンネル信号の数に比例したままである．この問題の解決のためには従来の符号化手法とは異なる観点から伝送すべきチャンネル信号の数を削減する空間符号化手法が必要となる．

そこで、新たに空間符号化手法を構築するに当たって、次の2つの観点に着目する。一つは一般に音源の数はチャンネル信号の数よりも少ないということである。このことは会議室における話者やコンサートホールにおけるピアノ独奏を始めとする一般の室内楽など、かなりの場合に該当していると考えられる。もう一つは全てのチャンネル信号の中には音源に関する情報が必ず含まれているということである。なぜなら、各チャンネル信号は音源信号に音場の室内インパルス応答を畳み込んだ信号として表現できるからである。ゆえに、チャンネル信号から音源信号を抽出して伝送すれば、伝送量がチャンネル信号の数から音源の数に削減されることが期待できる。

本論文では上記のような観点に従って、音源抽出に基づいて伝送量を削減する空間符号化手法について検討する。第4.2節では提案する空間符号化手法について述べる。第4.3節では実際の室内での静止音源を対象とした符号化実験について述べる。提案手法は室内インパルス応答を補足情報として伝送する必要があるが、本実験においては送信側と受信側であらかじめ測定しておいた室内インパルス応答の情報が共有されているものと仮定した。第4.4節では主観評価実験によって提案手法が人間の聴覚に及ぼす影響を検討し、提案手法による知覚的な歪みは十分容認されうるものであることを示す。

## 4.2 アルゴリズム

図4.1に提案手法のブロック図を示す。まず、送信側で原音場内に配置したマイクロホンアレーによってチャンネル信号を収録する。また、音源からマイクロホンアレーまでの室内インパルス応答を測定し、音源信号を抽出するための逆フィルタを算出しておく。そして、収録したチャンネル信号に逆フィルタを畳み込んで音源信号を抽出し、得られた音源信号を伝送する。一方、受信側では、受け取った音源信号に室内インパルス応答を畳み込んでチャンネル信号を復元する。その結果、提案手法によって伝送量はチャンネル信号の数から音源の数へ削減される。これは音源の数がチャンネル信号の数よりも非常に少ない場合に大変効果的である。

### 4.2.1 音源信号の抽出

$N$  個の音源と  $M$  個のマイクロホンから成るシステム ( $N < M$ ) における音源信号抽出のブロック図を図4.2に示す。 $s_i(n)$  ( $i=1\dots N$ ) は  $i$  番目の音源信号で、 $x_j(n)$  ( $j=1\dots M$ ) は  $j$  番目のマイクロホンで収録したチャンネル信号である。 $g_{ji}(n)$  は  $i$  番目の音源信号から  $j$  番目のチャンネル信号までの室内インパルス応答で、これは送信側及び受信側においてあらかじめ知られているものとする。 $h_{ij}(n)$  は  $j$  番目のチャンネル信号から  $i$  番目の音源信号まで

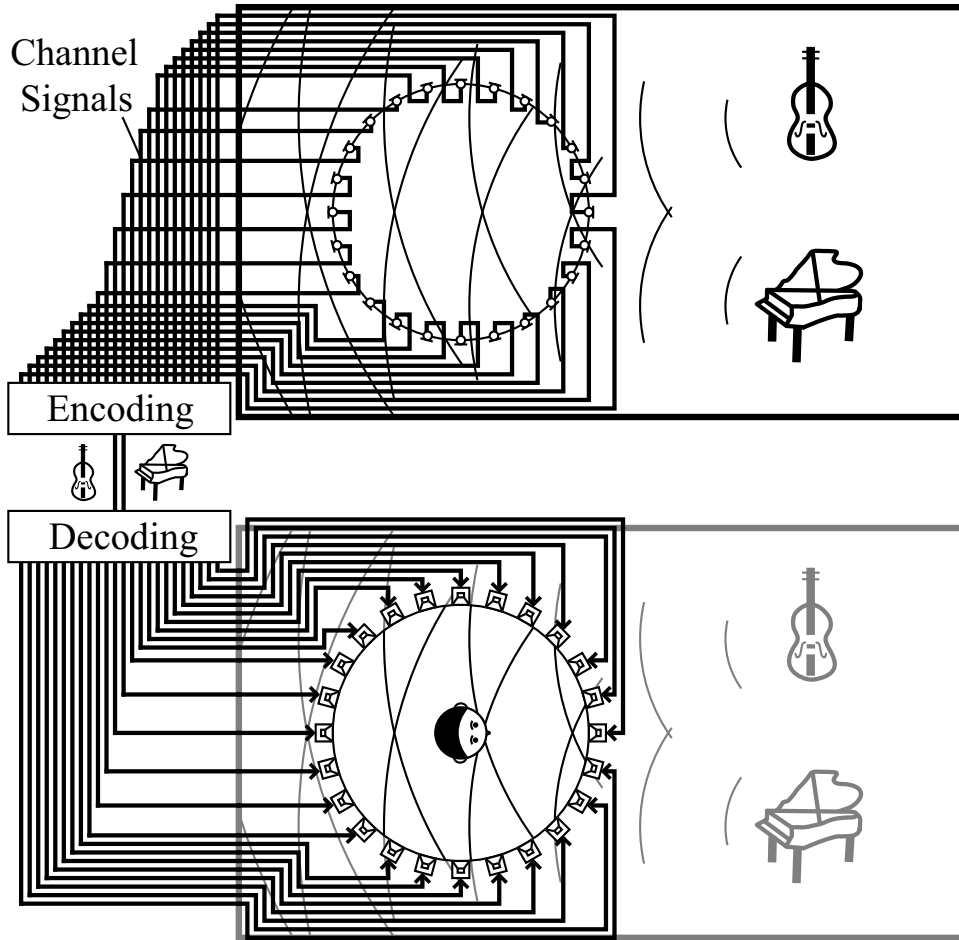


図 4.1 提案手法のブロック図

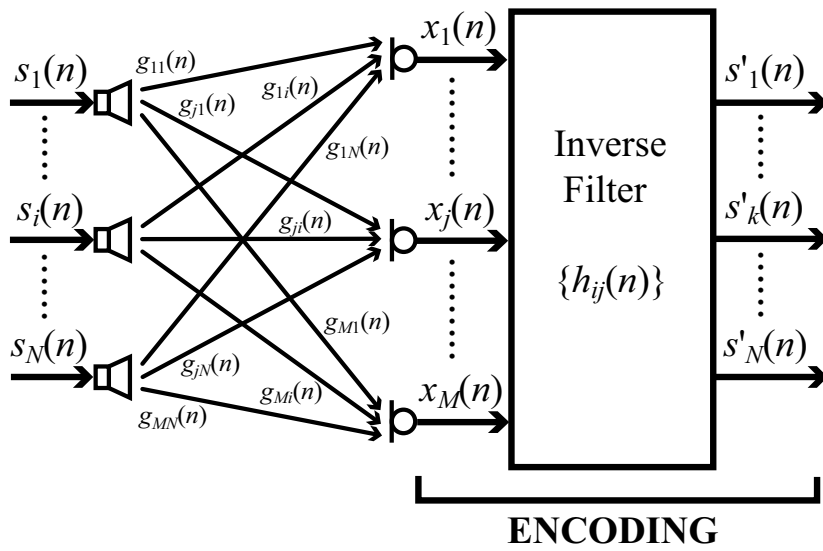


図 4.2 音源抽出のブロック図

の逆フィルタであり，チャンネル信号に逆フィルタを畳み込むことによって音源信号  $s'_k(n)$  ( $k=1\dots N$ ) が抽出される．

逆フィルタは室内インパルス応答から算出する．手順を以下に示す．まず，チャンネル信号は (4.2) 式のように音源信号に室内インパルス応答を畳み込んだ信号として表される．

$$x_j(n) = \sum_{i=1}^N g_{ji}(n) * s_i(n) \quad (4.1)$$

但し， $*$  は畳み込み演算を表わす．上記の式を離散 Fourier 変換すると，以下の式のようにになる．

$$X_j(\omega) = \sum_{i=1}^N G_{ji}(\omega) S_i(\omega) \quad (4.2)$$

また，抽出した音源信号はチャンネル信号に逆フィルタを畳み込んだ信号として表わされる．

$$s'_k(n) = \sum_{j=1}^M h_{kj}(n) * x_j(n) \quad (4.3)$$

これも同じように離散 Fourier 変換すると，以下の式のようにになる．

$$S'_k(\omega) = \sum_{j=1}^M H_{kj}(\omega) X_j(\omega) \quad (4.4)$$

(4.2) 式を (4.4) 式に代入すると，以下の関係式が得られる．

$$\sum_{j=1}^M H_{kj}(\omega) G_{ji}(\omega) = \begin{cases} S'_i(\omega)/S_i(\omega) & (i = k) \\ 0 & (\text{otherwise}) \end{cases} \quad (4.5)$$

(4.5) 式を行列形式で書き改めると，

$$\begin{pmatrix} G_{11}(\omega) & \dots & G_{M1}(\omega) \\ \vdots & \ddots & \vdots \\ G_{1N}(\omega) & \dots & G_{MN}(\omega) \end{pmatrix} \begin{pmatrix} H_{11}(\omega) & \dots & H_{N1}(\omega) \\ \vdots & \ddots & \vdots \\ H_{1M}(\omega) & \dots & H_{NM}(\omega) \end{pmatrix} = \begin{pmatrix} \frac{S'_1(\omega)}{S_1(\omega)} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \frac{S'_N(\omega)}{S_N(\omega)} \end{pmatrix} \quad (4.6)$$

元の音源信号を抽出する場合， $S'_i(\omega)/S_i(\omega)$  は 1 にしなければならないが，一般に  $G_{ji}(\omega)$  は初期遅延を持っているので，1 にしたときには  $H_{ij}(\omega)$  は因果律を満たさないフィルタになってしまう．そこで，因果律を満たす逆フィルタを算出するために  $S'_i(\omega)/S_i(\omega)$  に以下のように遅延項を設ける．

$$S'_i(\omega)/S_i(\omega) = e^{-j\omega T_0} \quad (i = 1\dots N) \quad (4.7)$$

但し,  $T_0$  は設定した遅延時間を表わす.  $\mathbf{G}(\omega)$ ,  $\mathbf{H}(\omega)$ ,  $\mathbf{D}(\omega)$  を以下の式のように定義すると,

$$\mathbf{G}(\omega) = \begin{pmatrix} G_{11}(\omega) & \dots & G_{M1}(\omega) \\ \vdots & \ddots & \vdots \\ G_{1N}(\omega) & \dots & G_{MN}(\omega) \end{pmatrix}, \quad \mathbf{H}(\omega) = \begin{pmatrix} H_{11}(\omega) & \dots & H_{N1}(\omega) \\ \vdots & \ddots & \vdots \\ H_{1M}(\omega) & \dots & H_{NM}(\omega) \end{pmatrix}, \quad (4.8)$$

$$\mathbf{D}(\omega) = \begin{pmatrix} e^{-j\omega T_0} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & e^{-j\omega T_0} \end{pmatrix}$$

(4.6) 式から (4.9) 式が導出される.

$$\mathbf{G}(\omega)\mathbf{H}(\omega) = \mathbf{D}(\omega) \quad (4.9)$$

$N < M$  の場合には,  $\mathbf{H}(\omega)$  にはいくつかの解が存在する. ここでは Moore-Penrose 擬似逆行列を用いて  $\mathbf{H}(\omega)$  を算出する [43].

$$\mathbf{H}(\omega) = \mathbf{G}^+(\omega)\mathbf{D}(\omega) \quad (4.10)$$

ここで,  $\mathbf{G}^+(\omega)$  は  $\mathbf{G}(\omega)$  の Moore-Penrose 擬似逆行列である.

Moore-Penrose 擬似逆行列は特異値分解を用いることによって算出される.  $\mathbf{G}(\omega)$  を特異値分解すると (4.11) 式のようになる.

$$\mathbf{G}(\omega) = \mathbf{U}(\omega) \left( \mathbf{A}(\omega) \mid \mathbf{O} \right) \mathbf{V}^H(\omega) \quad (4.11)$$

ここで,  $\mathbf{U}(\omega)$ ,  $\mathbf{V}(\omega)$  はそれぞれ  $N \times N$ ,  $M \times M$  次元のユニタリ行列,  $^H$  は Hermite 共役を表す.  $\mathbf{A}(\omega)$  は  $N \times N$  次元の対角行列で,  $\mathbf{O}$  は  $N \times (M - N)$  次元の零行列である.  $\mathbf{A}(\omega)$  の対角成分は  $N$  個の特異値  $\lambda_i(\omega) (i = 1 \dots N)$  で,  $\mathbf{G}(\omega)$  のランクが  $d$  の場合,  $N$  個のうち  $d$  個が 0 でない値を持つ.  $\lambda_i(\omega)$ ,  $\mathbf{U}(\omega)$ ,  $\mathbf{V}(\omega)$  を用いると, Moore-Penrose 擬似逆行列  $\mathbf{G}^+(\omega)$  は以下のように表される.

$$\mathbf{G}^+(\omega) = \mathbf{V}(\omega) \left( \frac{\mathbf{A}^+(\omega)}{\mathbf{O}^T} \right) \mathbf{U}^H(\omega) \quad (4.12)$$

但し,  $^T$  は転置を表し,  $\mathbf{A}^+(\omega)$  は以下のように定義される.

$$\mathbf{A}^+(\omega) = \begin{pmatrix} \sigma_1(\omega) & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \sigma_N(\omega) \end{pmatrix}, \quad \sigma_i(\omega) = \begin{cases} 0 & (\lambda_i = 0) \\ 1/\lambda_i & (\text{otherwise}) \end{cases} \quad (4.13)$$

ゆえに, 以下の式のように (4.10) 式で得られた  $\mathbf{H}(\omega)$  を逆離散 Fourier 変換で時間領域に戻すことによって逆フィルタが算出される.

$$h_{ij}(n) = \text{IDFT}[H_{ij}(\omega)] \quad (4.14)$$

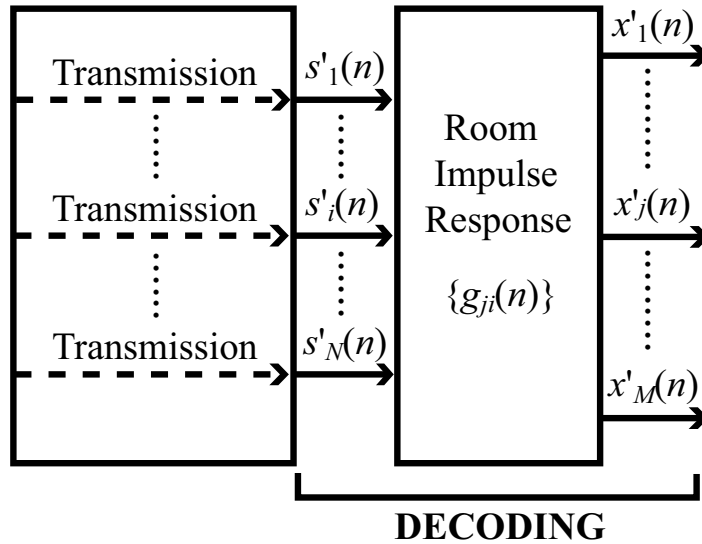


図 4.3 チャンネル復元のブロック図

#### 4.2.2 チャンネル信号の復元

チャンネル信号の復元に関するブロック図を図 4.3 に示す．以下の式のように受信した音源信号  $s'_i(n)$  に室内インパルス応答を畳み込むことによってチャンネル信号  $x'_j(n)$  が復元される．

$$x'_j(n) = \sum_{i=1}^N g_{ji}(n) * s'_i(n) \quad (4.15)$$

(4.7) 式より，音源信号を抽出する際には  $T_0 s$  の遅延が生じる．一方，チャンネル信号の復元の際には遅延は生じない．従って，提案する符号化手法における符号化全体での遅延は  $T_0 s$  となる．

また，提案手法によって取り出される信号は音源の波形信号である．ゆえに，抽出した音源信号に従来のチャンネル内符号化手法を適用することによって，伝送量を更に削減することも可能である．

### 4.3 符号化実験

この節では提案した空間符号化手法の有効性を検証するために実際に行った実験について述べる．

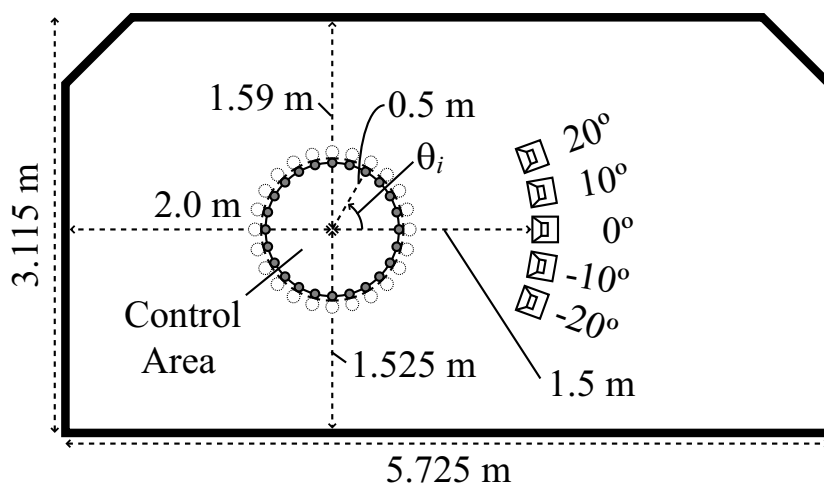


図 4.4 音源用スピーカとマイクロホンアレーの配置図

#### 4.3.1 室内インパルス応答の測定

名古屋大学工学部 7 号館 410 号室の可変残響室において室内インパルス応答を測定した。この部屋は壁面の吸音パネルを取り外すことによって、残響時間を 150 ms から 300 ms に調節することができる。部屋の天井と床を吸音状態にすることによって、2 次元の音場を想定している。音源用のスピーカとマイクロホンアレーの配置図を図 4.4 に示す。スピーカとマイクロホンアレーの高さは 0.835 m とした。制御領域は半径 0.5 m の円とし、その境界上に 15 度おきに ( $\theta_i = -165^\circ \sim 180^\circ$ ) 計 24 個のマイクロホンを設置した。円の中心から 1.5 m の距離に方位角  $-20^\circ \sim 20^\circ$  で 10 度おきにスピーカを設置した。スピーカは無指向性スピーカ (小野測器: SJ-1810)、マイクロホンは無指向性マイクロホン (Sony: ECM-77S) を用いた。アレー用の輪状アルミ製フレームの側壁にマイクロホンを取り付けることによって、各マイクロホンの指向性を制御領域の外側に向けるようにしている。なお、本実験におけるマイクロホンの指向性は単一指向性に近いものになっている。測定条件を表 4.1 に示す。音圧レベルはスピーカから 1 m 地点での値である。サンプリング周波数 48 kHz の 65536 点 TSP 信号 [44] を各スピーカから再生し、同期加算を 16 回行って FIR フィルタ形式の室内インパルス応答を測定した。

実環境における会議室やコンサートホールの残響時間に近い値での符号化実験を行うため、以降では測定した室内インパルス応答はサンプリング周波数が 12 kHz の FIR フィルタとして扱う。この操作によって部屋の残響時間は 4 倍になり、残響時間の値はそれぞれ  $150 \text{ ms} \times 4 = 0.6 \text{ s}$ 、 $300 \text{ ms} \times 4 = 1.2 \text{ s}$  となる。これらの値はそれぞれ会議室やコンサートホールにおける典型的な残響時間に相当する。この場合、音の伝達時間や伝達距離も 4

表 4.1 室内インパルス応答の測定条件

残響時間	150 ms	300 ms
室温	19.5°C	19.2°C
暗騒音レベル	20.0 dB(A)	19.4 dB(A)
音圧レベル	90.0 dB(A)	91.6 dB(A)
フィルタ次数	7200	14400

表 4.2 マルチチャンネル信号の作成条件

ドライソース	音声	木琴	木管五重奏
音源の数	1	1	5
音源の方位角	0°	0°	0°, ±10°, ±20°
残響時間	0.6 s	1.2 s	1.2 s

倍になるため、図4の部屋の大きさも等価的に4倍になり、マイクロホンアレーの半径は  $0.5\text{ m} \times 4 = 2\text{ m}$ 、中心から各スピーカまでの距離は  $1.5\text{ m} \times 4 = 6\text{ m}$  相当となる。

#### 4.3.2 チャネル信号の作成

無響室で収録したドライソースに4.3.1節で測定した室内インパルス応答を畳み込むことによってチャンネル信号を作成した。ドライソースには女性話者の音声、木琴、木管五重奏の3種類を用いた。音声は無響室で収録し、木琴と木管五重奏の各楽器音はMIDI音源から作成した。ドライソースの長さは5sで、サンプリング周波数は12kHzである。また、帯域を50–5000Hzに制限している。作成条件を表4.2に示す。音源から音が出ていない場合も考慮して、5個の音源以外の条件も設定した。その際、音源の方位角以外のスピーカには無音信号を割り当てた。

#### 4.3.3 逆フィルタの算出

逆フィルタの算出は残響時間ごとに行った。算出手順を以下に示す。まず、4.3.1節で測定した室内インパルス応答を離散 Fourier 変換で周波数領域に変換し、 $5 \times 24$  次元の  $\mathbf{G}(\omega)$  を導出した。次に、(4.10) 式によって  $24 \times 5$  次元の  $\mathbf{H}(\omega)$  を算出した。このとき、 $\mathbf{D}(\omega)$  の対角成分に FIR バンドパスフィルタの DFT 係数を用いることによって、室内インパルス応答の振幅特性が小さい周波数領域においては逆フィルタの算出を行わないようにする。



表 4.3 逆フィルタの算出条件

残響時間	0.6 s	1.2 s
DFT フレーム長	32768	65536
BPF の遅延	10ms, 20ms, 40ms, 80ms, 160ms, 320ms	
フィルタ長	14400	28800

これによって、より安定な逆フィルタを算出することができる。本研究ではサンプリング周波数 12 kHz で、帯域幅が 50–5000 Hz の FIR バンドパスフィルタを適用した。最後に、 $H(\omega)$  を逆離散 Fourier 変換によって時間領域に戻し、打ち切ることによって逆フィルタを算出した。算出条件を表 4.3 に示す。バンドパスフィルタの遅延が提案する空間符号化手法における遅延に相当する。

#### 4.3.4 符号化及び復号化

符号化は (4.3) 式で示したように 4.3.2 節で作成した 24 個のチャンネル信号に 4.3.3 節で算出した逆フィルタを畳み込んで 5 個の音源信号を抽出することによって行われた。これによって、伝送量は 24 チャンネルから 5 チャンネルに削減された。また、(4.15) 式で示したように抽出した 5 個の音源信号に室内インパルス応答を畳み込んで 24 個のチャンネル信号を復元することによって、復号化を行った。

#### 4.3.5 逆フィルタによる符号化への影響

この節では算出した逆フィルタによって符号化性能にどのような影響が及ぼされるかを考察する。まず、符号化遅延を 320 ms (=3840 samples) とした時の算出結果を図 4.5 に示す。左列が残響時間が 0.6 s、右列が残響時間が 1.2 s の時の結果である。(a) は測定した方位角  $0^\circ$  のスピーカから方位角  $\theta = 0^\circ$  のマイクロホンまでの室内インパルス応答で入力チャンネル信号に相当する。(b) は室内インパルス応答から算出した逆フィルタである。(c) は室内インパルス応答に逆フィルタを畳み込んだ波形である。この波形が逆フィルタの算出の際に用いた FIR バンドパスフィルタの波形に近ければ、音源信号の抽出精度が良いということを示している。(d) は (c) の波形に更に室内インパルス応答を畳み込んだもので、復元したチャンネル信号に相当する。

室内インパルス応答の初期遅延サンプル数と逆フィルタの初期遅延サンプル数を足し合

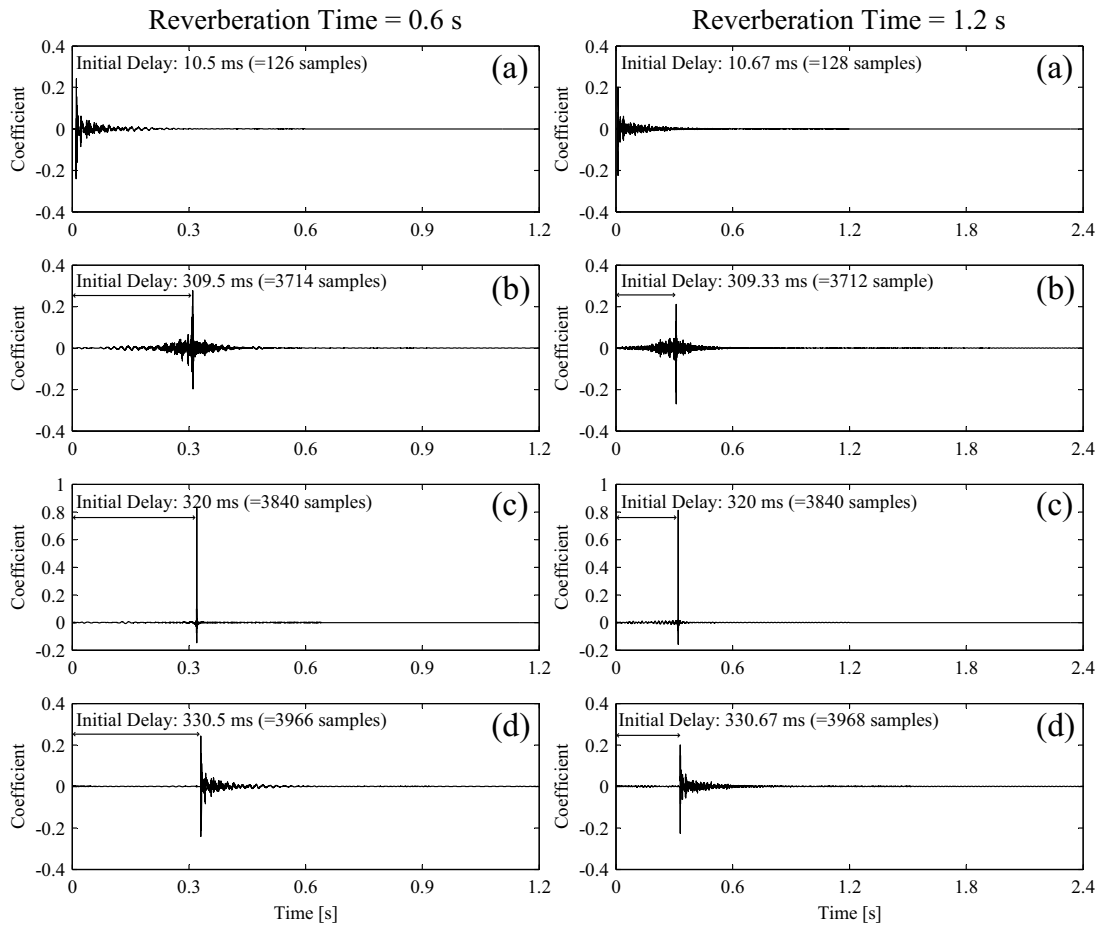


図 4.5 符号化遅延が 320 ms (=3840 samples) の時の逆フィルタの算出結果

わせると、設定した符号化遅延サンプル数  $n_0 (= T_0 F_s)$  になる。但し、 $F_s$  はサンプリング周波数を表す。このことから、逆フィルタによって音源信号を抽出することは室内インパルス応答におけるスピーカからマイクロホンまでの全ての経路 (経路に対応する遅延サンプル数によって直接音、初期反射音、残響音に分類される [20]) に対応した遅延とアレー処理 [45] に従って音源信号を抽出することに相当する。符号化遅延が 320 ms の場合は残響音まで含めて音源信号を抽出しているため、(c) に示される波形は 4.3.3 節で用いた FIR バンドパスフィルタに近い波形となっている。また、(d) の波形も (a) の波形と全く同じ形になっている。このように、完全に音源を抽出するためには直接音や初期反射音に加えて残響音の遅延サンプルまで含まれる必要があるため、符号化遅延は部屋の残響時間を考慮したサンプル数が必要となる。しかし、会議室のような双方向通信を実現しようとする場合には符号化遅延が長すぎると円滑な会話の妨げとなる恐れがある [46]。そのため、双方向通信を実現する際には符号化遅延をなるべく短く設定することが必要となる。

そこで、符号化遅延を会話の支障とされないと考えられる 40 ms (=480 samples) とした

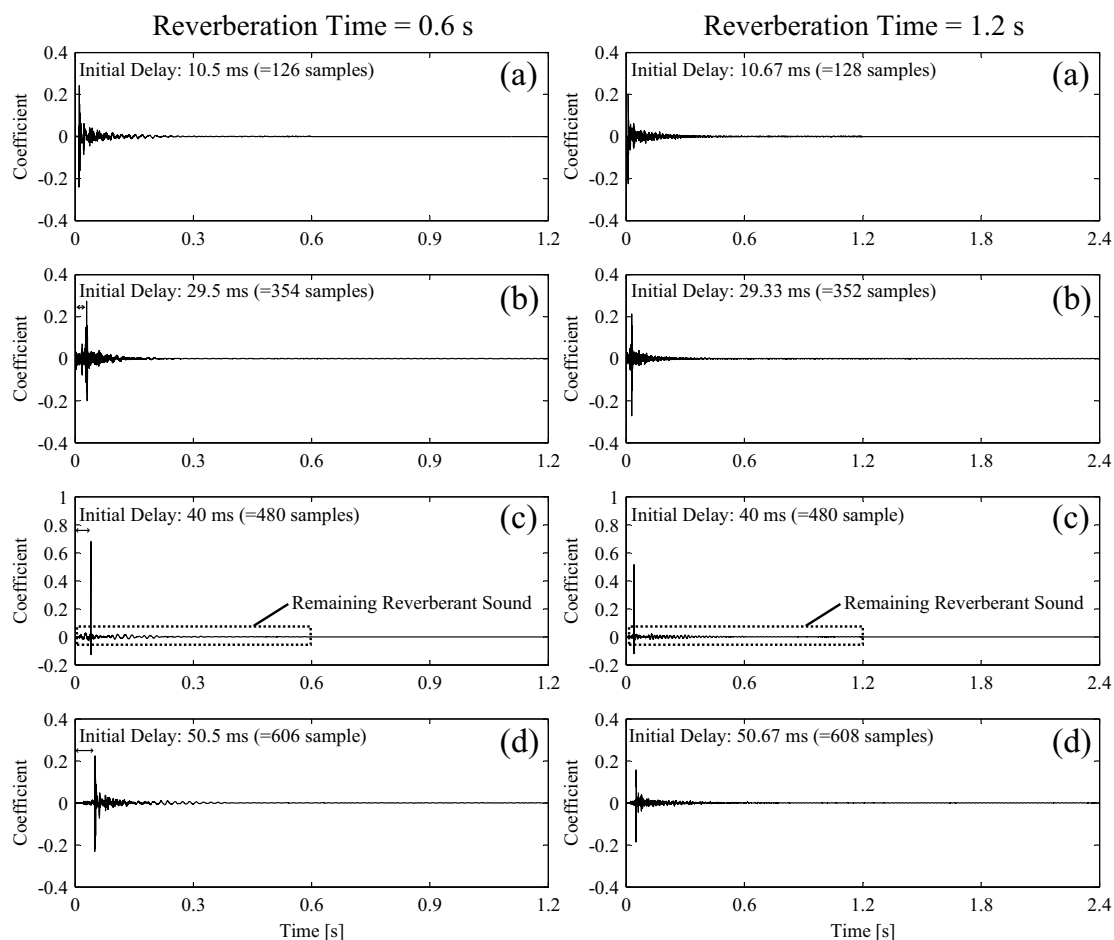


図 4.6 符号化遅延が 40 ms (=480 samples) の時の逆フィルタの算出結果

時の算出結果を図 4.6 に示す．この場合には，直接音や初期反射音の遅延サンプル数は含まれるが，残響音の遅延サンプル数は含まれないため，(c) で示されるように抽出された音源信号は残響音が残留したままの波形になる．しかし，(d) に示される復元したチャンネル信号は (a) の波形と比べると概形があまり変化していない．これは抽出された音源信号が残留した残響音に比べて十分にパワーが大きいいため，復元の際に付加された残響成分が残留した残響音をマスクしてしまうためであると考えられる．これより，符号化遅延を音源信号の抽出が完全に行われないような比較的短い時間に設定しても，チャンネル信号の聴感上の違いはほとんど出て来ないことが予想される．以降では双方向通信の応用も考慮して，符号化遅延を会話に支障のない時間に設定した条件のみで主観評価実験を行うことにする．

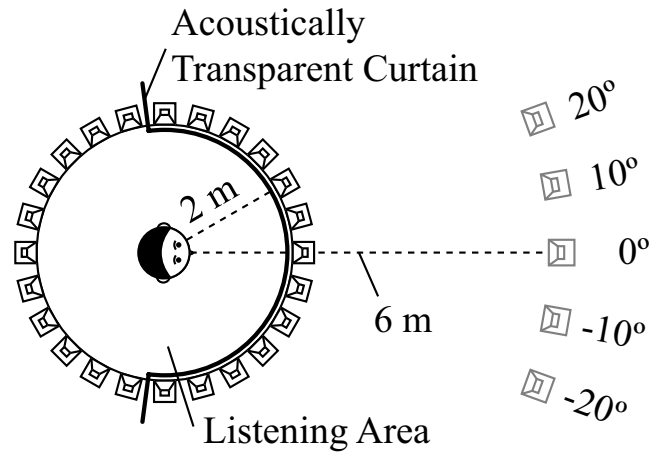


図 4.7 スピーカアレーと聴取者の位置

## 4.4 主観評価実験

提案した符号化手法が知覚的にどのような影響を及ぼすのかを主観評価実験によって検討した。

### 4.4.1 実験環境

主観評価実験は本来なら無響室で行う必要があるが、実験設備の都合上、第3章における主観評価実験で用いた部屋と同じ部屋（約 80 ms の低残響室）で行った。スピーカアレーと聴取者の位置を図 4.7 に示す。4.3.1 節で説明したように、マイクロホンアレーの半径は 2 m になるので、スピーカアレーの半径も 2 m にした。各スピーカ (Emic: Soundevice) からチャンネル信号を再生すると、スピーカアレーの中に音場が再現され、図 4.7 の灰色で示されるように中心から 6 m 離れた 5 個のスピーカの位置に音像が生じることになる。スピーカの周波数特性は 50–5000 Hz でほぼ一定であったので、特にスピーカに関する周波数特性の補正は行わなかった。暗騒音レベルは 25.0 dB(A)、音圧レベルは聴取者の位置において約 70 dB(A) とした。また、視覚の影響を低減するため、照明を薄暗くし、スピーカアレーの前方は音響カーテンで覆った。

### 4.4.2 実験計画

聴取者は日常生活において聴力的に異常の見られない 7 名の男子学生と 1 名の女子学生である。実験は ITU-R 勧告 [47] に従って三刺激二重盲験法を用いた。各試行において

表 4.4 刺激の組み合わせ

X	A	B	練習試行	本試行
原音	原音	符号化音 (遅延=10 ms)	1	2
原音	符号化音 (遅延=10 ms)	原音	1	2
原音	原音	符号化音 (遅延=20 ms)	1	2
原音	符号化音 (遅延=20 ms)	原音	1	2
原音	原音	符号化音 (遅延=40 ms)	1	2
原音	符号化音 (遅延=40 ms)	原音	1	2

3種類の刺激(以下“X”, “A”, “B”とする)が提示される。“X”には常に符号化しない原音が割り当てられ, “A”と“B”には原音もしくは符号化した音がランダムに割り当てられる。刺激の全ての組み合わせ及び試行数を表 4.4 に示す。

実験計画は以下のように設定した。まず, 符号化による影響の違いを検討するために, 実験全体を2つに分割し, それぞれを“音場評価”, “全体評価”と名付けた。次に, それぞれの評価を3つのセッションに分割した。それぞれのセッションにおいて音源の異なる3種類のチャンネル信号(“音声”, “木琴”, “木管五重奏”)を提示した。その際, 音源の提示順序は聴取者ごとにランダムイズした。各セッションにおいて6回の練習試行と12回の本試行を行い, 1回の試行において“X”, “A”, “B”の順に提示した。試行の提示順序も聴取者ごとにランダムイズした。実験計画の流れ図を図 4.8 に示す。

#### 4.4.3 実験手順

前半の音場評価では符号化による音場への影響を評価した。聴取者に以下のように教示した。「“X”は基準となる部屋で鳴っています。“A”と“B”のどちらかは必ず基準となる部屋と同じ部屋で鳴っています。“A”と“B”のうち, 基準となる部屋で鳴っていると思った方を“5.0”と評定して下さい。もう一つは表 4.5 に従って, 部屋の差がどれくらいあるかを小数点第1位単位で“1.0”から“4.9”の範囲で評定して下さい。」

後半の全体評価では先に評価した音場への影響に加え, 符号化による音質への影響も含んだ全体的な影響を検討した。聴取者に以下のように教示した。「“X”は原音です。“A”と“B”のどちらかは必ず原音と同じです。“A”と“B”のうち, 原音と思った方を“5.0”と評定して下さい。もう一つは表 4.5 に従って, 原音との差がどれくらいかを小数点第1位単位で“1.0”から“4.9”までで評定して下さい。」

音を聞いている間, 聴取者は頭部を自由に動かすことができた。また, 判断がつくまで

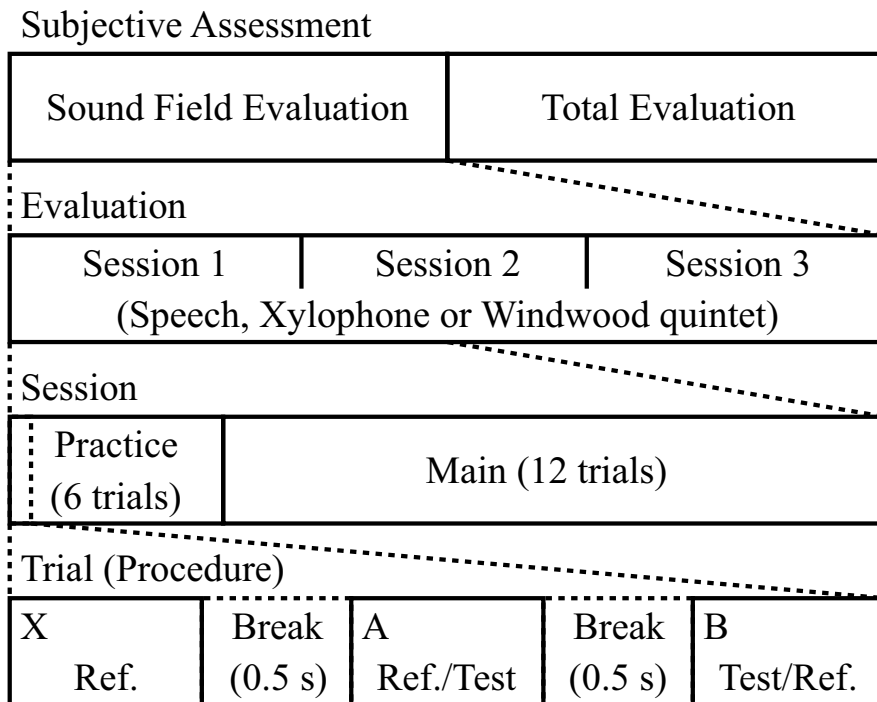


図 4.8 主観評価実験の流れ図

表 4.5 違いの評定表

差の程度	評定
差が分からない	5.0
差は分かるが、気にならない	4.0
差がわずかに気になる	3.0
差が気になる	2.0
差が非常に気になる	1.0

は何度も“X”，“A”，“B”を聞くことができた。

#### 4.4.4 実験結果及び考察

“A”と“B”のうち原音を割り当てた方に“5.0”と評定したかを確認することによって聴取の評定能力を検証する。聴取者の評定能力が十分でない場合、その聴取者は当て推量で回答しているため、正答率がチャンスレベルに落ち着くと考えられる。各聴取者の全体の正答数を二項検定した結果を表 4.6 に示す。\*\* は 1% で有意であったことを表し、\*\*\* は 0.1% で有意であったことを表す。以降の分析では 0.1% で有意であると判定された聴取

表 4.6 二項検定の結果

聴取者	標本数	音場評価		全体評価	
		正答数	危険率	正答数	危険率
A	36	33	0.0000***	36	0.0000***
B	36	34	0.0000***	35	0.0000***
C	36	32	0.0000***	35	0.0000***
D	36	31	0.0000***	32	0.0000***
E	36	32	0.0000***	31	0.0000***
F	36	25	0.0057**	28	0.0002***
G	36	25	0.0057**	27	0.0006***
H	36	26	0.0020**	25	0.0057**

者のデータのみを用いることとする。

各試行において符号化音を割り当てた音の評定値から原音を割り当てた音の評定値を引いたランク差を算出する。ランク差の平均値を図 4.9, 4.10 に示す。音場評価については 20 標本，全体評価については 28 標本のデータである。図中のエラーバーは 95% 信頼区間を表す。もし，符号化による影響がない場合には，ランク差平均は 0 に近づくことになる。

音場評価の結果は全体評価の結果とほとんど同じである。このことは，提案した空間符号化手法による影響は音場への影響がほとんどであるということを意味している。符号化遅延が 10 ms のときは被験者は少なくとも符号化による劣化を“わずかに気にしている”。これは設定した符号化遅延が室内インパルス応答の初期遅延よりも小さいために，音源抽出の際に 4.3.5 節で示したような遅延和アレー処理ができなかったためだと考えられる。一方，符号化遅延が 20 ms や 40 ms の場合には，音声や木管五重奏においてランク差が -1 より大きい。つまり，被験者は音声や木管五重奏のように持続している音に対しては符号化による劣化を“知覚できるが，気にしていない”ということが言える。これは 4.3.5 節で説明したように，直接音や初期反射音を含めた遅延和アレー処理を行い，抽出した音源信号が残留した残響音に比べて十分に大きくなるため，復元の際に付加された残響成分が抽出の際に残留してしまった残響音をマスクしてしまっているためであると考えられる。

しかし，木琴のような衝撃音に対しては多少評定値が下がっている。これは逆フィルタを有限長にしたために起こるプリエコーのような時間的な歪み [48] が衝撃音に対しては鋭敏に感じられることが原因であり，より正確な逆フィルタを算出する手法 [10, 49, 50]

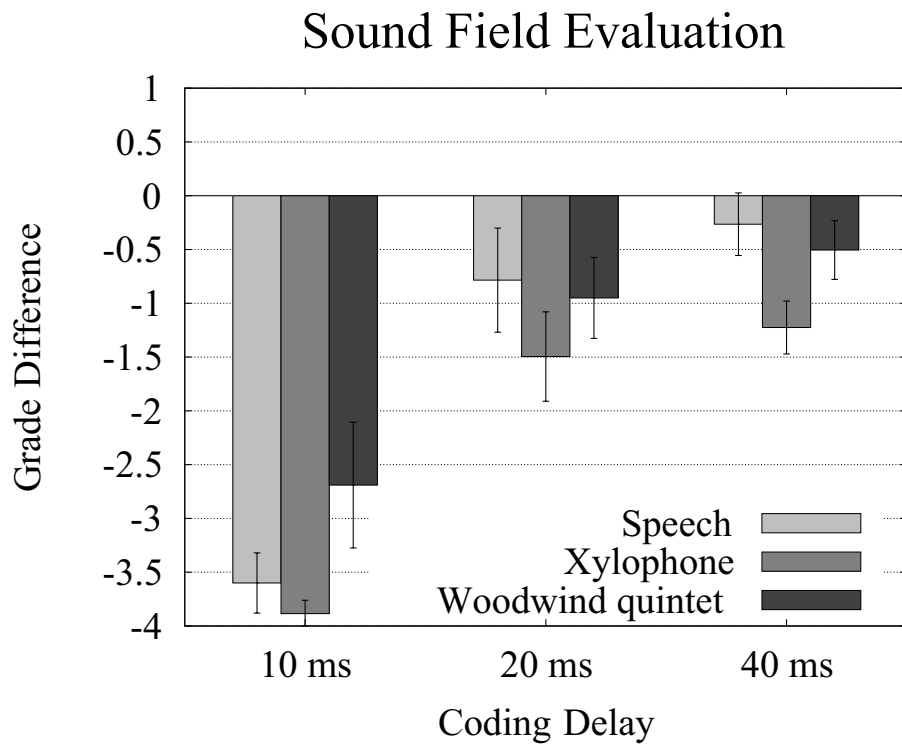


図 4.9 音場評価の結果

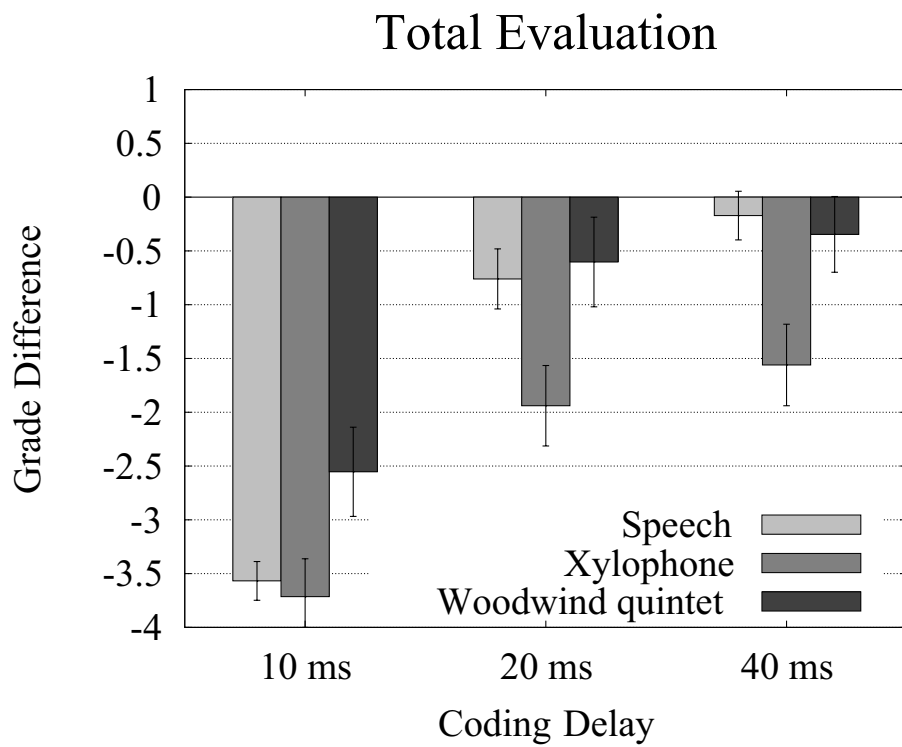


図 4.10 全体評価の結果



を適用すればこのような問題は解決されると考えられる。

## 4.5 まとめ

本章では波面合成法による立体音場再生システムを構築する際に伝送量をより削減するために音源抽出に基づく空間符号化手法を提案した。実環境にて測定された室内インパルス応答を基に伝送量を 24 個のチャンネル信号から 5 個の音源信号に削減する符号化実験を行い、知覚への影響を検討するために主観評価実験を行ったところ、提案手法による知覚への影響は多くの場合において問題にならない程度であることが示された。



## 第 5 章

# 移動音源抽出に基づく伝送量の削減

### 5.1 はじめに

第 4 章では音源からマイクロホンアレーまでの室内インパルス応答を基にチャンネル信号から音源信号を抽出することによって伝送量をチャンネル信号の数から音源の数に削減する空間符号化手法を提案し、伝送量を 24 個のチャンネル信号から 5 個の音源信号に削減してもその影響は知覚的に十分許容できるということが分かった。しかし、第 4 章では音源が静止している場合を扱っていた。波面合成法による立体音場再生システムの実際の応用例を考えてみると、演劇のように演者という音源が移動している場合も考えられる。そこで、移動音源も取り扱うことができる空間符号化手法を開発する必要があると考えられる。

そこで、本章では第 4 章で提案した空間符号化手法を改良し、移動音源信号の抽出に基づいて伝送量を削減する空間符号化手法を提案する。5.2 節では提案手法のアルゴリズムについて説明する。5.3 節では虚像法によって作成したチャンネル信号を対象とした符号化実験について述べる。5.4 節では主観評価実験によって提案手法の符号化品質について検討し、提案手法が遠隔通信システムを構築する際に有効であることを示す。

### 5.2 アルゴリズム

図 5.1 に提案した空間符号化手法のブロック図を示す。但し、図中の RIR, IF は室内インパルス応答、逆フィルタの略称を表わす。入力チャンネル信号は原音場内に配置したマイクロホンアレーによって収録された信号である。その際、位置センサによって音源の位置も同時に収録する。送信側では収録した音源の位置情報を基に適切な室内インパルス応答を求め、逆フィルタを算出する。そして、チャンネル信号に逆フィルタを畳み込んで音源信

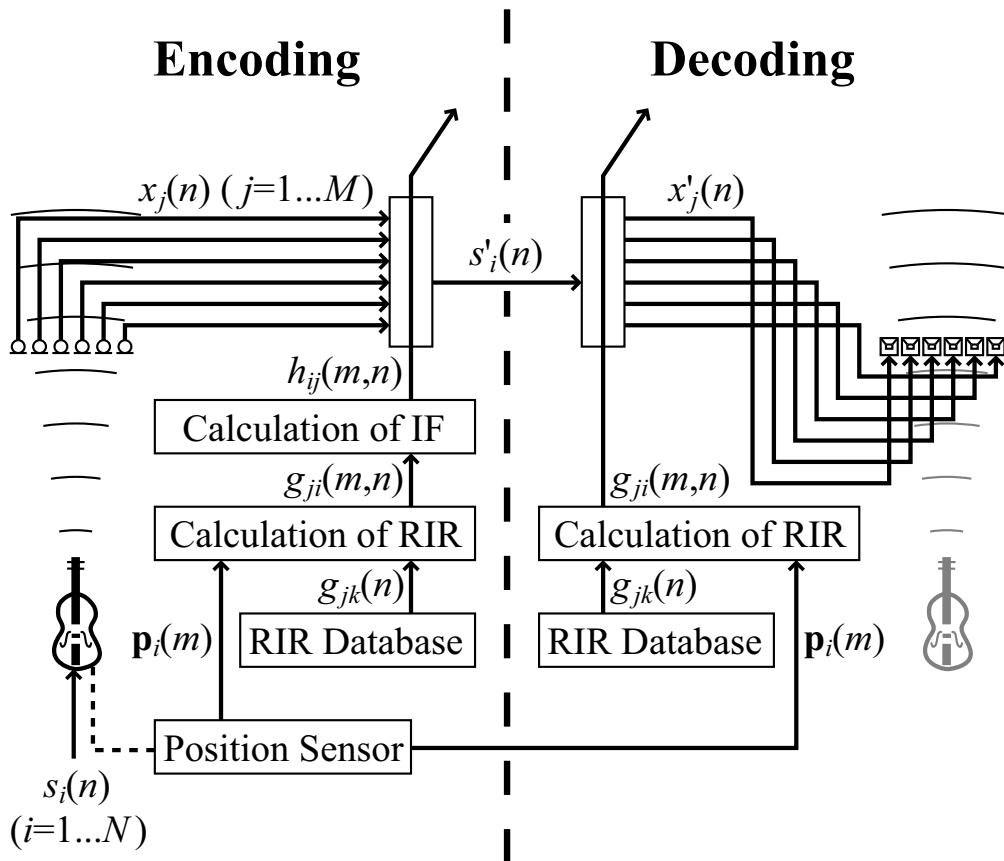


図 5.1 提案手法のブロック図

号を抽出し，音源の位置情報と共に伝送する．受信側ではまず受け取った音源の位置情報を基にそれに対応する室内インパルス応答を求める．そして，受信した音源信号に室内インパルス応答を畳み込んで出力チャンネル信号を復元する．この手法によって伝送量はチャンネル信号の数から音源の数へ削減される．これは音源の数がチャンネル信号の数よりも非常に少ない場合に大変効果的である．但し，音源の位置情報を付加情報として伝送する必要が生じてくるが，伝送すべき音源位置の情報量は音源信号の情報量と比べると非常に小さく，伝送量全体には影響を及ぼさないと考えられるので，以降では伝送すべき音源位置の情報量は考慮しないことにする．

### 5.2.1 移動音源信号の抽出

原音場中の  $N$  個の移動音源信号を  $s_i(n)$  ( $i=1\dots N$ )， $M$  個 ( $M > N$ ) のマイクロホンで収録したチャンネル信号を  $x_j(n)$  ( $j=1\dots M$ ) とする．但し，信号のサンプリング周波数は  $F_s$  [Hz] である．それと同時に，位置センサによって音源の位置情報  $\mathbf{p}_i(m)=[p_{ix}(m), p_{iy}(m), p_{iz}(m)]^T$  もサンプリング周波数  $F_p$  [Hz] で収録されている．但し， $m, n$  は離散時間を表わ

す整数で，連続時間  $t$  においてはそれぞれ  $\frac{m}{F_p}$ ， $\frac{n}{F_s}$  に換算される．

送信側ではまず得られた音源位置情報  $\mathbf{p}_i(m)$  を基に時間  $m$  における  $i$  番目の音源信号から  $j$  番目のチャンネル信号までの室内インパルス応答  $g_{ji}(m, n)$  を以下の式によって求める．

$$g_{ji}(m, n) = g_{jk}(n) \quad (5.1)$$

但し， $g_{jk}(n)$  は  $\mathbf{p}_k [= \mathbf{p}_i(m)]$  に存在する音源から  $j$  番目のチャンネル信号までの室内インパルス応答である．これらの室内インパルス応答の情報は図 5.1 に示すようにデータベースとして送信側と受信側であらかじめ保存されているものとする．室内の形状から室内インパルス応答を算出する手法 (例えば [51]) などを適用すれば，全ての  $\mathbf{p}_k$  に対する  $g_{jk}(n)$  を既知にすることは可能であるので，十分に実用的なシステムを構築できると考えられる．

得られた室内インパルス応答  $g_{ji}(m, n)$  を基に時間  $m$  における  $j$  番目のチャンネル信号から  $i$  番目の音源信号までの逆フィルタ  $h_{ij}(m, n)$  を (4.10) 式と同じように以下の式によって算出する．

$$\mathbf{H}(m, \omega) = \mathbf{G}^+(m, \omega) \mathbf{D}(\omega) \quad (5.2)$$

但し， $\mathbf{G}^+(m, \omega)$  は  $\mathbf{G}(m, \omega)$  の Moore-Penrose 擬似逆行列で， $\mathbf{H}(m, \omega)$ ， $\mathbf{G}(m, \omega)$ ， $\mathbf{D}(\omega)$  は以下のように定義される行列である．

$$\mathbf{H}(m, \omega) = \begin{pmatrix} H_{11}(m, \omega) & \dots & H_{N1}(m, \omega) \\ \vdots & \ddots & \vdots \\ H_{1M}(m, \omega) & \dots & H_{NM}(m, \omega) \end{pmatrix} \quad (5.3)$$

$$\mathbf{G}(m, \omega) = \begin{pmatrix} G_{11}(m, \omega) & \dots & G_{M1}(m, \omega) \\ \vdots & \ddots & \vdots \\ G_{1N}(m, \omega) & \dots & G_{MN}(m, \omega) \end{pmatrix}, \quad \mathbf{D}(\omega) = \begin{pmatrix} e^{-j\omega T_c} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & e^{-j\omega T_c} \end{pmatrix} \quad (5.4)$$

また， $H_{ij}(m, \omega)$ ， $G_{ji}(m, \omega)$  は  $h_{ij}(m, n)$ ， $g_{ji}(m, n)$  を離散 Fourier 変換したもので， $T_c (= \frac{P_c}{F_s})$  は因果律を満たす逆フィルタを算出するために必要な遅延時間であり，この値が符号化遅延に相当する．

得られた逆フィルタ  $h_{ij}(m, n)$  を以下の式のようにチャンネル信号  $x_j(n)$  に畳み込むことによって移動音源信号  $s'_i(n)$  を抽出する．

$$\begin{aligned} s'_i(n) &= \sum_{m=0}^{\infty} w(m, n) \sum_{j=1}^M h_{ij}(m, n) * x_j(n) \\ &= \sum_{m=0}^{\infty} w(m, n) \sum_{j=1}^M \sum_{l=0}^{L_h-1} h_{ij}(m, l) x_j(n-l) \end{aligned} \quad (5.5)$$

但し， $*$  は畳み込み演算を表わし， $L_h$  は逆フィルタの次数である．また， $w(m, n)$  は以下

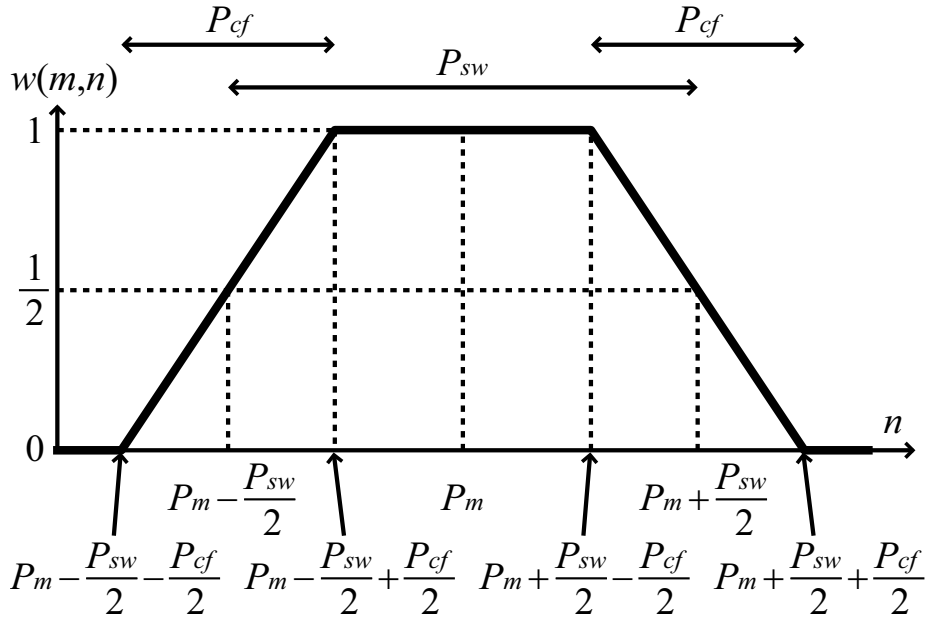


図 5.2 窓関数の形状

の式のように定義される窓関数である。

$$w(m, n) = \begin{cases} \frac{1}{P_{cf}} [n - P_m + \frac{P_{sw}}{2} + \frac{P_{cf}}{2}] & (P_m - \frac{P_{sw}}{2} - \frac{P_{cf}}{2} \leq n < P_m - \frac{P_{sw}}{2} + \frac{P_{cf}}{2}) \\ 1 & (P_m - \frac{P_{sw}}{2} + \frac{P_{cf}}{2} \leq n \leq P_m + \frac{P_{sw}}{2} - \frac{P_{cf}}{2}) \\ -\frac{1}{P_{cf}} [n - P_m - \frac{P_{sw}}{2} - \frac{P_{cf}}{2}] & (P_m + \frac{P_{sw}}{2} - \frac{P_{cf}}{2} < n \leq P_m + \frac{P_{sw}}{2} + \frac{P_{cf}}{2}) \\ 0 & (\text{otherwise}) \end{cases} \quad (5.6)$$

但し,  $P_{sw}(=\frac{F_s}{F_p})$  は音源位置の切替サンプル数である。窓関数の形状を図 5.2 に示す。この関数によって時間  $P_m(=P_c + mP_{sw})$  を中心に  $P_{sw}$  点の範囲で音源信号が抽出され, 隣接する音源位置との間で抽出した音源信号の波形が不連続にならないようにするために窓の両脇に  $P_{cf}(=F_s T_{cf})$  点のリニアなクロスフェードが設けられる。その際,  $P_c, P_{sw}, P_{cf}$  は以下の条件を満たす必要がある。

$$P_c \geq \frac{P_{sw} + P_{cf}}{2}, \quad P_{sw} \geq P_{cf} \quad (5.7)$$

音源信号が抽出される様子を図 5.3 の左側に示す。図では 1 入力 1 出力なので,  $i, j$  は省略してある。この図のように音源位置が  $\mathbf{p}(0), \mathbf{p}(1), \mathbf{p}(2), \dots$  と変化するに従って逆フィルタ  $h(0, n), h(1, n), h(2, n), \dots$  をチャネル信号  $x(n)$  に畳み込めば, 抽出される音源  $s'(n)$  は移動音源のドライソースそのものになる。

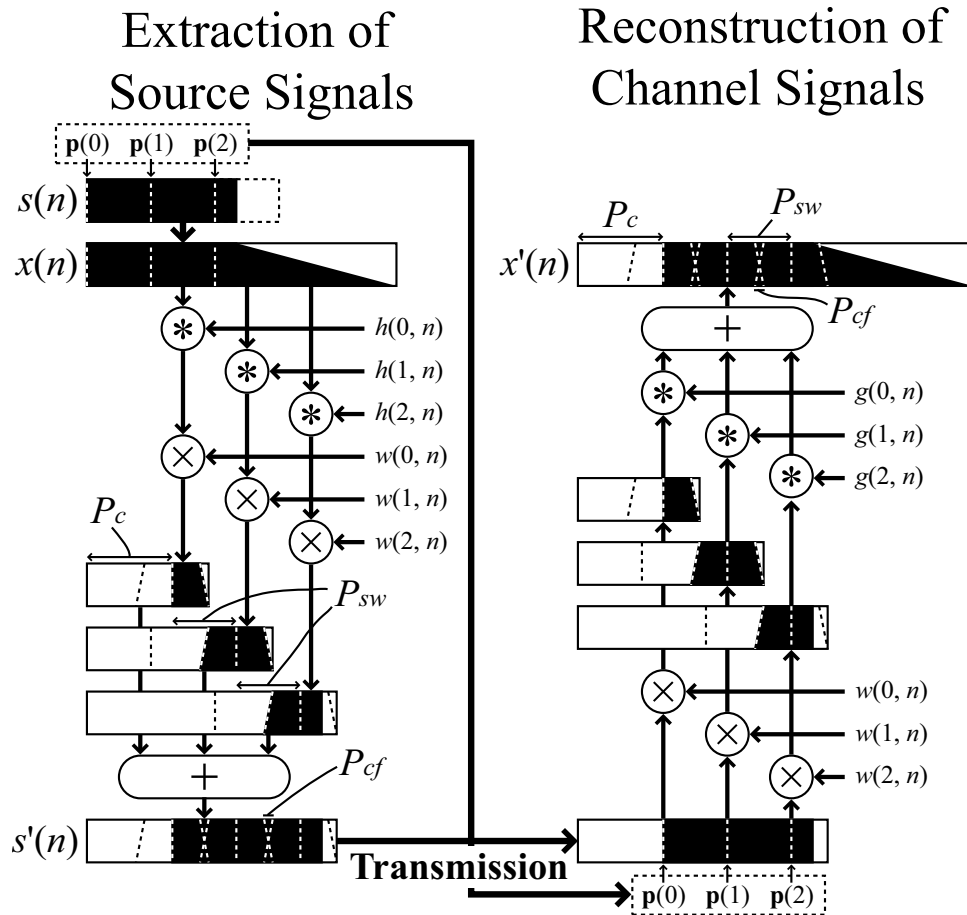


図 5.3 抽出及び復元のブロック図

### 5.2.2 チャネル信号の復元

受信側ではまず受信した音源の位置情報  $\mathbf{p}_i(m)$  を基に (5.1) 式によって室内インパルス応答  $g_{ji}(m, n)$  を求める。そして、得られた  $g_{ji}(m, n)$  を以下の式のように受信した音源信号  $s'_i(n)$  に畳み込むことによってチャネル信号  $x'_j(n)$  を復元する。

$$\begin{aligned} x'_j(n) &= \sum_{m=0}^{\infty} \sum_{i=1}^N g_{ji}(m, n) * [w(m, n)s'_i(n)] \\ &= \sum_{m=0}^{\infty} \sum_{i=1}^N \sum_{l=0}^{L_g-1} g_{ji}(m, l)w(m, n-l)s'_i(n-l) \end{aligned} \quad (5.8)$$

但し、 $L_g$  は室内インパルス応答の次数である。

チャネル信号が復元される様子を図 5.3 の右側に示す。この図のように音源位置が  $\mathbf{p}(0)$ ,  $\mathbf{p}(1)$ ,  $\mathbf{p}(2)$ , ... と変化するに従って室内インパルス応答  $g(0, n)$ ,  $g(1, n)$ ,  $g(2, n)$ , ... を音源信号  $s'(n)$  に畳み込めば、音源の動きまで再現されたチャネル信号  $x'(n)$  が復元される。元の

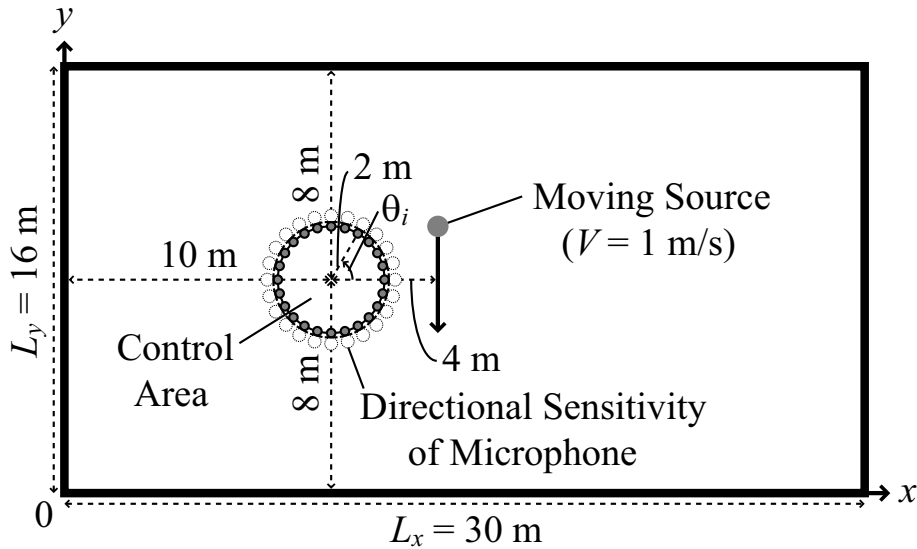


図 5.4 符号化実験に用いた原音場

音の動きが連続的であるのに対して，再現された音の動きはサンプリング周波数  $F_p$  [Hz] で離散化されている． $F_p$  が小さすぎると音の動きが元の音の動きと異なって知覚されるので， $F_p$  は音の動きが正しく知覚されるように設定する必要がある．一方，符号化遅延時間  $T_c$  が大きすぎると双方向通信システムの構築の際に円滑なコミュニケーションの妨げとなる [46]．ゆえに， $T_c$  も双方向通信システムの構築に支障がおきないように設定する必要がある．5.3 節ではこれらのことをふまえて  $F_p$  や  $T_c$  の設定を行う．

## 5.3 符号化実験

本節では提案した空間符号化手法の有効性を検証した実験について述べる．

### 5.3.1 チャネル信号の作成

実験に用いるチャネル信号は虚像法 [30] を用いて計算機上で作成した．原音場は図 5.4 に示すような残響空間とし，その中に 1 個の移動音源と 24 個のマイクロホンが配置されているとした．各マイクロホンの指向性は半径  $2\text{ m}$  の円で定義された制御領域の外側に向いている．

虚像法によると  $x=0$ ,  $x=L_x$ ,  $y=0$ ,  $y=L_y$  に壁面  $x_1$ ,  $x_2$ ,  $y_1$ ,  $y_2$  を配置し  $\mathbf{s}_i = (s_{ix}, s_{iy})^T$  (5.2 節における  $\mathbf{p}_k$  に相当する) に  $i$  番目の無指向性音源， $\mathbf{r}_j = (r_{jx}, r_{jy})^T$  に  $j$  番目の指向性受音



点を配置した場合の室内インパルス応答  $g_{ji}(t)$  は以下の式のように表わされる .

$$g_{ji}(t) = \sum_{p_x=0}^1 \sum_{p_y=0}^1 \sum_{q_x=-\infty}^{\infty} \sum_{q_y=-\infty}^{\infty} D_j(\mathbf{p}, \mathbf{q}) \beta_{x1}^{|q_x-p_x|} \beta_{x2}^{|q_x|} \beta_{y1}^{|q_y-p_y|} \beta_{y2}^{|q_y|} \frac{\delta[t - \frac{|\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q})|}{c}]}{4\pi|\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q})|} \quad (5.9)$$

但し ,  $\delta(t)$  は Dirac のデルタ関数 ,  $c$  は音速 ( $=340[\text{m/s}]$ ) ,  $\beta_{x1}$  ,  $\beta_{x2}$  ,  $\beta_{y1}$  ,  $\beta_{y2}$  は壁面  $x_1$  ,  $x_2$  ,  $y_1$  ,  $y_2$  の反射係数を表わす . 以降では  $\beta_{x1}^{|q_x-p_x|} \beta_{x2}^{|q_x|} \beta_{y1}^{|q_y-p_y|} \beta_{y2}^{|q_y|} = B(\mathbf{p}, \mathbf{q})$  ,  $\sum_{p_x=0}^1 \sum_{p_y=0}^1 = \sum_{\mathbf{p}}$  ,  $\sum_{q_x=-\infty}^{\infty} \sum_{q_y=-\infty}^{\infty} = \sum_{\mathbf{q}}$  と表記する . また ,

$$\begin{aligned} \mathbf{d}_{ji}(\mathbf{p}, \mathbf{q}) &= \mathbf{s}_i(\mathbf{p}, \mathbf{q}) - \mathbf{r}_j \\ &= \begin{pmatrix} (1 - 2p_x)s_{ix} + 2q_xL_x \\ (1 - 2p_y)s_{iy} + 2q_yL_y \end{pmatrix} - \begin{pmatrix} r_{jx} \\ r_{jy} \end{pmatrix} \end{aligned} \quad (5.10)$$

で ,  $\mathbf{s}_i(\mathbf{p}, \mathbf{q})$  は虚像音源の位置ベクトルを表わす . 一方 ,  $D_j(\mathbf{p}, \mathbf{q})$  は  $j$  番目の受音点の指向性を表わす関数であり , 本実験では第 2 章における結果から以下の式に示されるような単一指向性とした .

$$D_j(\mathbf{p}, \mathbf{q}) = \frac{1}{2} \left( 1 + \frac{\mathbf{u}_j \cdot \mathbf{d}_{ji}(\mathbf{p}, \mathbf{q})}{|\mathbf{u}_j| |\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q})|} \right) \quad (5.11)$$

但し ,  $\mathbf{u}_j = (u_{jx}, u_{jy})^T$  は  $j$  番目の受音点の指向ベクトルである . 得られた  $g_{ji}(t)$  を音源信号  $s_i(t)$  に畳み込むことによって音源が静止している場合のチャンネル信号  $x_j(t)$  が得られる .

$$\begin{aligned} x_j(t) &= \sum_{i=1}^N g_{ji}(t) * s_i(t) \\ &= \sum_{i=1}^N \sum_{\mathbf{p}} \sum_{\mathbf{q}} D_j(\mathbf{p}, \mathbf{q}) B(\mathbf{p}, \mathbf{q}) \frac{s_i[t - \frac{|\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q})|}{c}]}{4\pi|\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q})|} \end{aligned} \quad (5.12)$$

音源が移動する場合には音源や虚像音源の位置が時間  $t$  によって変化する . 本節で作成するチャンネル信号は符号化する前の信号であり , 音源の位置は全ての  $t$  において判明しているとする . 従って , (5.12) 式は以下のように変形される .

$$x_j(t) = \sum_{i=1}^N \sum_{\mathbf{p}} \sum_{\mathbf{q}} D_j(\mathbf{p}, \mathbf{q}, t) B(\mathbf{p}, \mathbf{q}) \frac{s_i[t - \frac{|\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q}, t)|}{c}]}{4\pi|\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q}, t)|} \quad (5.13)$$

計算機上で扱う音源信号やチャンネル信号は離散時間信号なので , 時間  $t$  を離散時間  $n (= t \cdot F_s)$  にすると ,

$$x_j(n) = \sum_{i=1}^N \sum_{\mathbf{p}} \sum_{\mathbf{q}} D_j(\mathbf{p}, \mathbf{q}, n) B(\mathbf{p}, \mathbf{q}) \frac{s_i[n - \frac{F_s |\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q}, n)|}{c}]}{4\pi|\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q}, n)|} . \quad (5.14)$$

(5.14) 式は時間  $n$  に音源信号を入力すると , 時間  $n + \frac{F_s |\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q}, n)|}{c}$  にチャンネル信号が出力されるということを表わしている . しかし , 一般的には  $n$  に整数を代入しても  $n + \frac{F_s |\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q}, n)|}{c}$  は

整数にならないため，(5.14) 式をそのまま適用しても出力信号は  $n$  の離散時間信号にならない．そこで，入力時間を  $n+a(0 \leq a < 1)$  とし，得られる出力時間  $n+a+\frac{F_s|\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q}, n+a)|}{c}$  を整数にすることによって，出力信号が  $n$  の離散時間信号として出力されるようにする．但し，

$$\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q}, n+a) = (1-a)\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q}, n) + a\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q}, n+1). \quad (5.15)$$

その際，出力時間を  $K=n+a+\frac{F_s|\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q}, n+a)|}{c}$  と表記すると， $a$  は  $i, j, \mathbf{p}, \mathbf{q}, n$  ごとに以下の二次方程式を解くことによって得られる．

$$\begin{aligned} \{|\Delta\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q}, n)|^2 - \frac{c^2}{F_s^2}\}a^2 + 2\{\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q}, n) \cdot \Delta\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q}, n) + \frac{c^2(K-n)}{F_s^2}\}a \\ + \{|\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q}, n)|^2 - \frac{c^2(K-n)^2}{F_s^2}\} = 0 \end{aligned} \quad (5.16)$$

但し，

$$\Delta\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q}, n) = \mathbf{d}_{ji}(\mathbf{p}, \mathbf{q}, n+1) - \mathbf{d}_{ji}(\mathbf{p}, \mathbf{q}, n). \quad (5.17)$$

また， $K$  は以下の条件を満たす整数に設定する．

$$\text{ceil}\left(n + \frac{F_s|\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q}, n)|}{c}\right) \leq K \leq \text{floor}\left(n + 1 + \frac{F_s|\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q}, n+1)|}{c}\right) \quad (5.18)$$

ゆえに，音源が移動している場合のチャンネル信号は以下のようにして得られる．

$$x_j(n) = \sum_{i=1}^N \sum_{\mathbf{p}} \sum_{\mathbf{q}} D_j(\mathbf{p}, \mathbf{q}, n+a) B(\mathbf{p}, \mathbf{q}) \frac{s_i[n+a - \frac{F_s|\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q}, n+a)|}{c}]}{4\pi|\mathbf{d}_{ji}(\mathbf{p}, \mathbf{q}, n+a)|} \quad (5.19)$$

その際の入力信号  $e_i(n+a)$  は以下の式のように線形補間を行ってから用いる．

$$s_i(n+a) = (1-a)s_i(n) + as_i(n+1) \quad (5.20)$$

チャンネル信号の作成条件を表 5.1 に示す．2 種類の作成条件は会議室で歩きながら話をしたり，コンサートホールでフルートを演奏しながら歩いたりしている場合を想定した．この場合，移動速度は音速に対して十分小さいため，ドップラー効果は発生しない．音声は無響室で収録し，フルートは MIDI 音源から作成したものをを用いた．また，4 つの壁面の反射係数は同一とし， $p_x, p_y, q_x, q_y$  の値は以下の条件を満たす全ての組み合わせを用いた．

$$|q_x - p_x| + |q_x| + |q_y - p_y| + |q_y| \leq R \quad (5.21)$$

但し， $R$  は表 5.1 に示すように設定した壁面反射回数の最大値である．また，部屋の残響時間は Schroeder 法 [37] により算出した．

表 5.1 チャネル信号の作成条件

ドライソース	音声	フルーツ
$F_s$ (サンプリング周波数)	48 kHz	
音源の長さ	4 s	
$\beta$ (壁面の反射係数)	0.5	0.7
$R$ (最大の壁面反射回数)	6	10
部屋の残響時間	0.6 s	1.0 s
$V$ (音源の移動速度)	1 m/s=3.6 km/h	

### 5.3.2 室内インパルス応答データベースの算出

室内インパルス応答の情報をデータベース化するために，図 5.4 中の直線状の音源の軌跡を 8.33 mm ごとに区切り，得られた計 481 点の音源位置から 24 点の受音点位置までの室内インパルス応答  $g_{jk}(n)$  ( $k=1\dots 481$ ,  $j=1\dots 24$ ) を以下の式によって算出した．

$$g_{jk}(n) = \sum_{\mathbf{p}} \sum_{\mathbf{q}} D_j(\mathbf{p}, \mathbf{q}) B(\mathbf{p}, \mathbf{q}) \frac{\delta[n - \text{round}(\frac{F_s |\mathbf{d}_{jk}(\mathbf{p}, \mathbf{q})|}{c})]}{4\pi |\mathbf{d}_{jk}(\mathbf{p}, \mathbf{q})|} \quad (5.22)$$

壁面の反射係数や最大の壁面反射回数は表 5.1 と同じ条件とした．また， $D(j)$  を算出するために必要な  $\mathbf{u}_j$  や  $\mathbf{s}_k$ ,  $\mathbf{r}_j$  は以下のように設定した．

$$\mathbf{u}_j = \begin{pmatrix} \cos \frac{\pi(j-12)}{12} \\ \sin \frac{\pi(j-12)}{12} \end{pmatrix}, \quad \mathbf{s}_k = \begin{pmatrix} 14 \\ 10 - \frac{k-1}{120} \end{pmatrix}, \quad \mathbf{r}_j = 2\mathbf{u}_j + \begin{pmatrix} 10 \\ 8 \end{pmatrix} \quad (5.23)$$

### 5.3.3 音源信号の抽出及びチャネル信号の復元

まず，(5.1) 式によって  $g_{jk}(n)$  から移動音源を表わす室内インパルス応答  $g_{ji}(m, n)$  を求めた．本実験では  $m = \frac{F_p(k-1)}{120}$  とし， $F_p$  には 30, 60, 120 Hz の値を用いた．

次に， $g_{ji}(m, n)$  から (5.2) 式によって  $h_{ij}(m, n)$  を算出した．算出条件を表 5.2 に示す．第 4 章での結果より符号化遅延時間  $T_c$  を 20 ms 以上にすれば符号化歪みによる品質への影響は小さくなるので， $T_c$  を 20 ms にすれば逆フィルタの算出による品質への影響は少ないと考え，本実験では符号化遅延を 20 ms に設定した．また，5.2 節でも述べたように，設定した符号化遅延時間は双方向通信システムを構築する場合においても円滑なコミュニケーションの妨げにならない値である．

表 5.2 逆フィルタの算出条件

残響時間	0.6 s	1.0 s
FFT フレーム長 [samples]	65536	131072
符号化遅延時間 $T_c$ (=符号化遅延サンプル数 $P_c$ )	20 ms (=960 samples)	
フィルタ長 [samples]	28800	48000

得られた逆フィルタ  $h_{ij}(m, n)$  を 5.3.1 節で得られたチャネル信号  $x_j(n)$  に畳み込んで音源信号  $s'_i(n)$  を抽出した。これによって伝送量は 24 個のチャネル信号から 1 個の移動音源信号に削減された。その際、窓関数  $w(m, n)$  に用いた切替サンプル数  $P_{sw}$  の値 (=1600, 800, 400 samples) は先ほど設定した  $F_p$  を換算した値で、クロスフェード時間  $T_{cf}$  には 1, 4 ms の値を用いた (クロスフェードサンプル数  $P_{cf}$  に換算すると 48, 192 samples)。  $P_{sw}$  や  $P_{cf}$  の値は (5.7) 式で与えられた条件に従っている。また、抽出した音源信号  $s'_i(n)$  に室内インパルス応答  $g_{ji}(m, n)$  を畳み込んでチャネル信号  $x'_j(n)$  を復元した。

## 5.4 主観評価実験

4.3.5 節によると、波形的な歪みが大きい場合においても音源定位や音像の大きさへの影響は必ずしも大きくなかった。そこで、本章でも主観評価実験によって提案した空間符号化手法が知覚に及ぼす影響を検討する。

### 5.4.1 実験環境

実験条件は 4.4 節の実験とほぼ同じ条件に設定した。主観評価実験は本来なら無響室で行う必要があるが、実験設備の都合上、4.4 節における主観評価実験で用いた部屋と同じ部屋 (約 80ms の低残響室) で行った。実験に用いたスピーカアレーと聴取者の位置を図 5.5 に示す。実験条件は 4.4 節の実験とほぼ同じ条件に設定した。主観評価実験は本来なら無響室で行う必要があるが、実験設備の都合上、第 3 章における主観評価実験で用いた部屋と同じ部屋 (約 80ms の低残響室) で行った。実験に用いたスピーカアレーと聴取者の位置を図 5.5 に示す。各スピーカ (Emic: Soundevice) からチャネル信号を再生すると、スピーカアレーの中に音場が再現され、アレー内にいる聴取者は図 5.5 の灰色で示されるように音像が移動しているように感じる。暗騒音レベルは 25.0 dB(A) で、音圧レベルは聴取者の位置において約 70 dB(A) とした。また、視覚の影響を低減するため、照明を薄

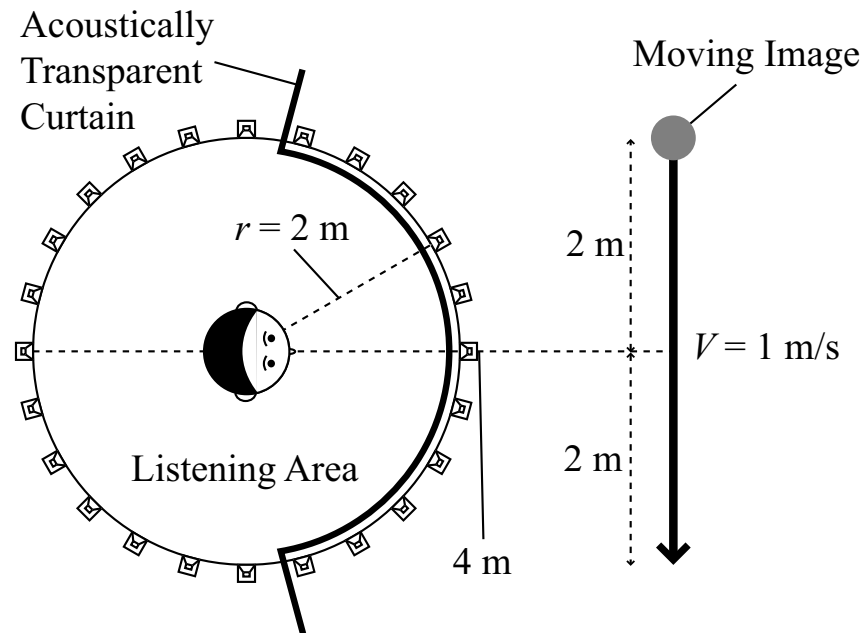


図 5.5 スピーカアレーと聴取者の位置

暗くし，スピーカアレーの前方は音響透過カーテンで覆った．

#### 5.4.2 実験計画

聴取者は日常生活において聴力に異常の見られない 8 名の男子学生である．ITU-R 勧告 [47] に従って三刺激二重盲験法を実験手法とした．主観評価実験の流れを図 5.6 に示す．まず，実験全体を 2 つに分割し，それぞれを“動き評価”，“全体評価”と名付けた．次に，それぞれの評価を 2 つのセッションに分割し，音源の異なる（音声とフルート）2 種類のチャンネル信号を提示した．その際，音源の提示順序は聴取者ごとにランダムイズした．セッションごとに練習試行と本試行を順に行い，1 回の試行において 3 種類の刺激を順に提示した．1 つの刺激の長さは 4 s である．最初に提示される刺激（以下“X”）には常に符号化しない原音を割り当て，2 番目に提示される刺激（以下“A”）と 3 番目に提示される刺激（以下“B”）には原音もしくは表 5.3 に示されるような 6 種類の符号化音を割り当てた．試行の提示順序も聴取者ごとにランダムイズした．練習試行は  $6(\text{符号化音の種類}) \times 2(\text{A or B に割り当て}) = 12$  試行を行い，本試行は  $6(\text{符号化音の種類}) \times 2(\text{A or B に割り当て}) \times 2(\text{繰り返し}) = 24$  試行を行った．

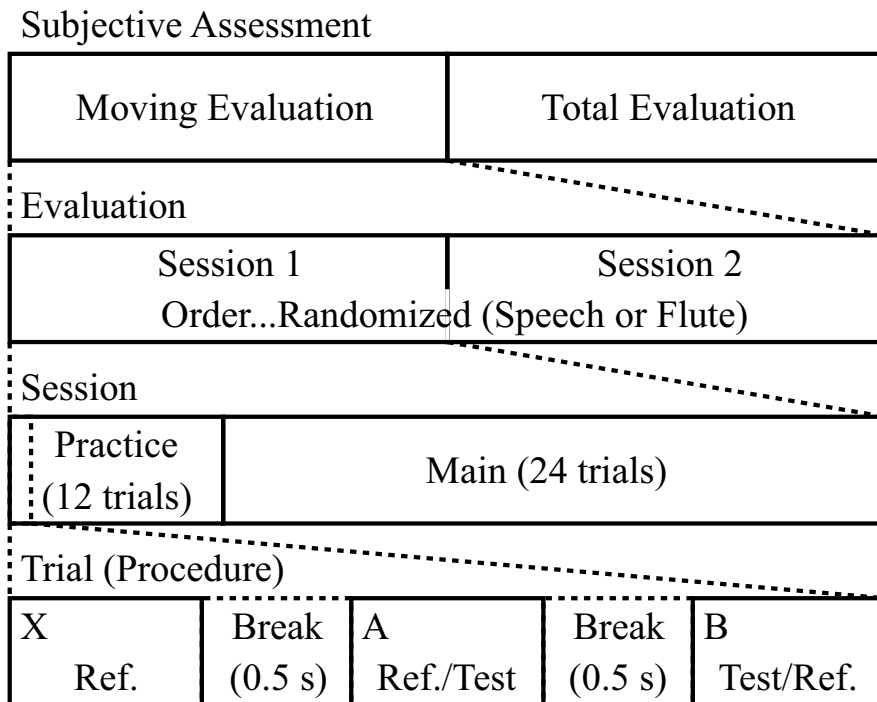


図 5.6 主観評価実験の流れ図

表 5.3 符号化音の種類

	1	2	3	4	5	6
$F_p$	30 Hz	30 Hz	60 Hz	60 Hz	120 Hz	120 Hz
$T_{cf}$	1 ms	4 ms	1 ms	4 ms	1 ms	4 ms

### 5.4.3 実験手順

前半の動き評価では符号化による音像の動き知覚への影響を評価した。聴取者に以下のように教示した。「“X”は基準となる動き方をしている音です。“A”と“B”のどちらかは必ず基準の音と同じ動き方をする音です。“A”と“B”のうち、基準の音と同じ動き方をする音だと思った方を“5.0”と評定して下さい。もう一方の音は表 5.4 に従って、音の動き方の差がどれくらいあるかを小数点第 1 位単位で“1.0”から“4.9”の範囲で評定して下さい。」

後半の全体評価では先に評価した音の動き知覚への影響に加え、符号化による音質への影響をも含んだ全体的な知覚への影響を検討した。聴取者に以下のように教示した。「“X”は原音です。“A”と“B”のどちらかは必ず原音と同じ音です。“A”と“B”のうち、

表 5.4 違いの評定表

差の程度	評定
差が分からない	5.0
差は分かるが，気にならない	4.0
差がわずかに気になる	3.0
差が気になる	2.0
差が非常に気になる	1.0

表 5.5 各被験者の識別結果

聴取者	標本数	動き評価		全体評価	
		音声	フルート	音声	フルート
A	24	<b>15</b>	<b>23</b>	<b>15</b>	<b>22</b>
B	24	7	<b>23</b>	<b>16</b>	<b>24</b>
C	24	7	<b>24</b>	14	<b>24</b>
D	24	<b>14</b>	21	11	21
E	24	<b>14</b>	14	<b>15</b>	17
F	24	11	22	9	22
G	24	11	19	9	20
H	24	9	18	10	17

原音と思った方を“5.0”と評定して下さい。もう一つは表 5.4 に従って，原音との差がどれくらいかを小数点第 1 位単位で“1.0”から“4.9”までで評定して下さい。」

主観評価実験を行っている間，聴取者は頭部を自由に動かすことができた。また，判断がつくまでは何度も“X”，“A”，“B”を繰り返し聞くことができた。

#### 5.4.4 実験結果及び考察

実験結果を分析する前に，聴取者が原音と符号化音を正しく識別し，原音を割り当てた刺激を「5.0」と評定しているかを検査する。各聴取者が原音を「5.0」と評定した回数をセッションごとに表 5.5 に示す。評定した回数が多いほど，聴取者の判断は信頼性が高いと考えられるので，分析には各セッションにおける上位 3 名(表中で太字で示した箇所)の結果を用いることにした。

符号化音の評定値から原音の評定値の差を算出した結果を図 5.7，5.8 に示す。エラー

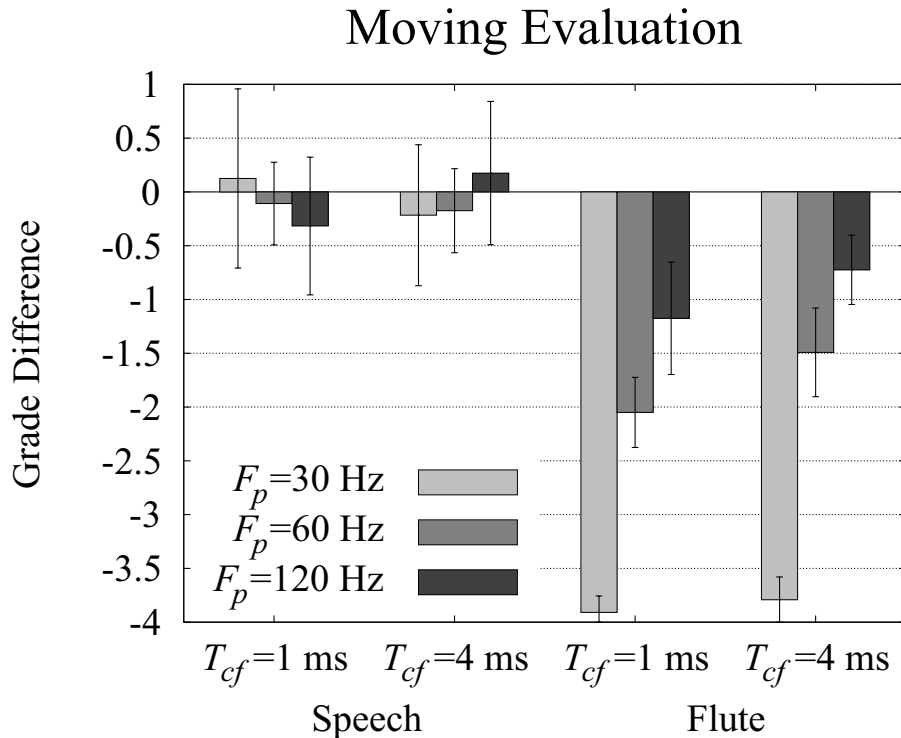


図 5.7 動き評価の結果

バーは 95% 信頼区間を表わす．動き評価の結果と全体評価の結果はほとんど同じである．これは提案した符号化手法による知覚的な歪みは 5.2 節で述べたように音の動きが離散的に再現されることで引き起こされる歪みに依存するというを示している．フルートの評定差は音声よりも全体的に低くなっている．これはフルートは調波複合音で，どちらかという純音に近い音であるため，音の動きを再現する際に窓関数を重ね合わせることによって生じるスペクトル歪みが音声の場合よりも鋭敏に感知されてしまったと考えられる．ゆえに，提案した符号化手法は音源の種類によって知覚的な影響が大きく異なるということが言える．

これらの実験結果から遠隔通信システムが実用的に構築できるかどうかを考察する．遠隔アンサンブルシステムを構築する場合，対象となる音源は主に楽音であるため，フルートの実験結果を参考にすることが必要である．そのため，符号化による知覚的な影響を低減するためには  $F_p$  をなるべく大きい値に設定する必要がある． $F_p$  が 120 Hz の場合には音の動きや全体的な知覚への影響は「差は分かるが，気にならない」程度である．この値はカメラなどを用いた位置検出システムで十分に実現できる性能であるので，映像と組み合わせた遠隔アンサンブルシステムを構築する場合，提案した空間符号化手法を導入すれば十分実用的なシステムを構築することができると考えられる．また，カメラなどを用いた位置



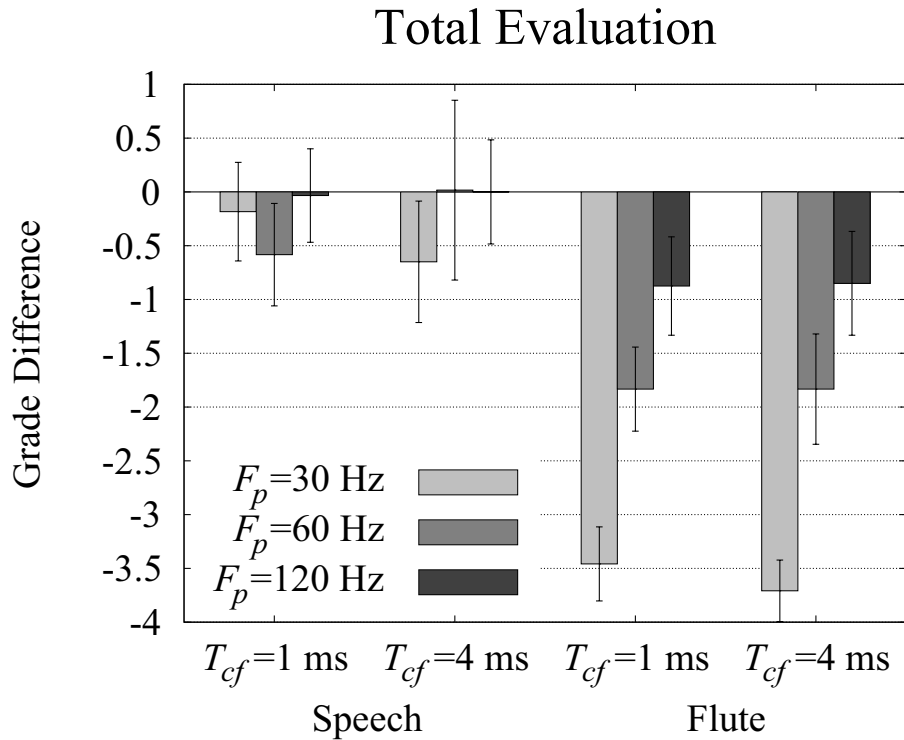


図 5.8 全体評価の結果

検出システムを用いると音源の位置を収録するための位置センサを音源に直接配置する必要はなくなるので、立体映像展示技術と組み合わせた立体音場再生システムを構築する際にも音源収録用のマイクロホンや位置センサが映像に映り込むことがないシステムを構築することも可能であると言える。

一方、遠隔通信会議システムの場合は対象とする音源は主に音声である。音声に対する実験結果を見てみると、 $F_p$  が 30 Hz の場合でも、被験者は知覚的な影響を識別できていないことが多い。ゆえに、映像と組み合わせた遠隔通信会議システムを構築する場合、遠隔アンサンブルシステムを構築する場合に比べてより簡易な位置検出システムを導入しても知覚上問題ないシステムを実現できると考えられる。

## 5.5 まとめ

本章では波面合成法による立体音場再生システムを遠隔で実現する際の伝送量の削減のために、音源の位置情報を基に移動音源を抽出することで伝送量をチャンネル信号の数から音源の数に削減する空間符号化手法を提案した。虚像法によって作成した残響空間におけるチャンネル信号を対象に符号化実験を行い、伝送量を 24 個のチャンネル信号から 1 個の移動音源信号に削減した。そして、符号化性能を評価するために主観評価実験を行ったとこ

る，システムを構築する際に対象となる音源に対して適切なパラメータを設定すれば，音の動き知覚や全体的な知覚への影響のほとんどないシステムを実用化できるということが分かった．

## 第 6 章

# 結論

### 6.1 要約

波面合成法による立体音場再生技術はステレオホニックや 5.1ch オーディオといった従来の立体音響技術よりもあたかも劇場や映画館でコンサートや映画を鑑賞しているかのような臨場感をより忠実に再現できるため、次世代の音響技術として通信やバーチャルリアリティの分野での応用が期待されている。しかし、この技術は音場再生に必要な情報量が非常に多くなるという問題点があった。そこで、本論文では立体音場再生システムに必要な情報量を削減できるかどうかを主観評価実験をベースとして検討した。

第 2 章では本論文で取り扱う立体音場再生システムについて解説した上で、波面が忠実に合成される条件を計算機上におけるシミュレーション実験によって検討した。その結果、波面を忠実に合成するためにはマイクロホン及びスピーカ間の間隔を波長の  $1/2$  以下にすること、無指向性マイクロホンよりも単一指向性マイクロホンもしくは超指向性マイクロホンをういた方がよいことを明らかにした。

第 2 章での結果に従ってマイクロホン及びスピーカ間の間隔を波長の  $1/2$  以下に設定しようとする、帯域が少なくとも 16 kHz であるような楽音を対象としたコンサートホールの音場再生システムを構築する場合には設定すべきスピーカ間の間隔が約 1 cm になるため、実際に用いるスピーカの大きさを考慮すると、システムの実現は非常に困難となる。そこで、第 3 章ではマイクロホン及びスピーカの個数が音場知覚に及ぼす影響を方向感と空間印象に着目した 2 種類の主観評価実験によって検討した。その結果、比較的低い周波数帯域で波面が忠実に合成されていれば方向感や空間印象は十分に再現されることを示した。即ち、一般に出回っているマイクロホンやスピーカを 24 個用いたような立体音場再生システムでも十分に臨場感を再現できることが分かった。

しかし、第 3 章での結果より構築した立体音場再生システムを遠隔地間でリアルタイ

ムに実現するための伝送量は従来のシステムの伝送量に比べて非常に多いため、映像と組み合わせた通信システムを構築する場合には通信容量が不足する可能性もある。そこで、第 4 章では立体音場再生システムにおいて伝送量を削減する手法を提案した。この手法は音源からマイクロホンの室内インパルス応答の情報に基づきチャンネル信号から音源信号を抽出して伝送するものである。提案手法を検討するために、実環境で測定した室内インパルス応答に基づき 24 個のチャンネル信号を 5 個の音源信号に伝送量を削減する符号化実験を行い、さらに主観評価実験によって知覚への影響を検討した。その結果、提案手法による知覚への影響は対象となる音源が音声のような持続音の場合には問題にならないということを示した。

第 5 章では演劇のように演者という音源が移動している場合に伝送量を削減することを試み、チャンネル信号から移動音源信号を抽出して伝送する伝送量削減手法を提案した。この手法は第 4 章で提案した伝送量削減手法に音源の位置センサを追加することによって構成されている。提案手法を検討するために、虚像法によって作成した残響空間におけるチャンネル信号を対象に伝送量を 24 個のチャンネル信号から 1 個の移動音源信号に削減する符号化実験を行い、主観評価実験によって提案した符号化手法の性能を評価した。その結果、システムを構築する際に対象となる音源に対して適切なパラメータを設定すれば、音の動き知覚や全体的な知覚への影響はほとんどないということを示した。

以上の結果から総合的な考察を行う。本論文では波面合成法による立体音場再生システムにおいて必要とする情報量を削減するための手法の開発を行った。情報量を削減する一つの手法として使用するマイクロホン及びスピーカの個数を減らすことを試みたところ、波面合成精度の観点からは比較的低い周波数帯域でしか波面が忠実に合成されていないような場合でも聴取者は十分に臨場感を感じるということが示された。また、もう一つの手法として伝送量削減手法を検討したところ、伝送すべき情報量が削減されても聴取者は知覚的な歪みをほとんど感じないということが示された。これらのことを利用すると、現在の通信インフラでも情報が十分に伝送できるようなシステムを実現することができるようになるので、今までは情報量が多過ぎるために実現は困難であると言われていた波面合成法による立体音場再生システムを実現できるという工学的な応用の可能性を示すことができた。

## 6.2 今後の展望

今後の展望としては以下のようなことが挙げられる。まず、本論文では従来の立体音響システムを拡張する形をベースとしたために、制御領域を半径 2 m の円とし、その境界面

上にマイクロホン及びスピーカを配置した。しかし、実際に家庭などに応用する場合にはマイクロホン及びスピーカを半径 2 m の円上に配置することは部屋の形状などの問題から困難であることが予想される。また、波面合成法による立体音場再生システムはシステム特性上、制御領域を 3 次元に設定することも可能である。そこで、制御領域を半径 2 m の円以外にした場合 (3 次元の制御領域にした場合も含む) に必要なマイクロホン及びスピーカの個数を主観評価実験などを基に検討する必要があると考えられる。また、主観評価実験の実施には多大な労力と時間を要するので、今までに得られた主観評価実験の結果から任意の制御領域を設定した場合の主観評価実験の結果を予測することができるようなモデルを構築する必要もあると考えられる。

一方、伝送量削減手法についてもまだまだ検討すべき点が多い。例えば、第 4 章において伝送量削減手法を検討した際には、測定した室内インパルス応答のサンプリング周波数を 1/4 にしたので、検討した音は 5 kHz 帯域の音であった。従って、5 kHz 以上の帯域においては品質の劣化が生じている可能性があるため、今後は実際の楽音のような 16 kHz 帯域の音を対象にした場合に本手法の有効性を検討する必要がある。また、第 5 章において音源が移動している場合の伝送量削減手法を検討した際には、室内インパルス応答は虚像法によって計算機上で算出した。そこで、今後は実環境において測定した室内インパルス応答及び収録した音を用いた場合について検討する必要がある。

また、実際の会議室やコンサートホールにおいて室内インパルス応答を測定したり、音を収録したりした上で立体音場再生システムの実現の可能性を検討する必要があると考えられる。実際にシステムを応用していくに当たってはオーケストラのようにチャンネル信号の数よりも音源の数の方が多き場合もあるので、そのような場合に音源セットに対して本伝送量削減手法を適用し際の有効性を検討していくことも課題である。これらの課題を克服した上で、将来的には実用的な波面合成法による立体音場再生システムを構築していくことを考えている。



# 謝辞

本研究を遂行するにあたって、本研究の内容のみならず研究者としての姿勢・心構えについて親切丁寧に御指導下さるだけでなく、本論文の審査委員をも勤めて下さいました中京大学情報科学部箕一彦教授(元:名古屋大学大学院人間情報学研究科・教授)に心より感謝の意を申し上げます。

本研究の内容について大変貴重な御意見を下さいました名城大学理工学部板倉文忠教授に心より感謝の意を申し上げます。また、様々な御意見を下さいました名古屋大学大学院情報科学研究科武田一哉教授に心より感謝の意を申し上げます。

名古屋大学大学院人間情報学研究科博士後期課程在籍中に講座ゼミにおいて適切な御助言・御指摘を下さるばかりでなく、本論文の審査委員をも勤めて下さいました名古屋大学大学院情報科学研究科齋藤洋典教授，三輪和久教授，川合伸幸助教授諸先生方に心より感謝の意を申し上げます。

同課程在籍中に本研究について様々な御相談に応じて下さいました名古屋大学情報メディア教育センター助手西野隆典博士，名城大学理工学部坂野秀樹講師両氏に心より感謝の意を申し上げます。

同課程在籍中に本研究について熱心な御議論に応じて下さるばかりでなく、本研究を遂行するにあたって必要な実験機器の手配をして下さいました岐阜工業高等専門学校助手西田鶴代氏に心より感謝の意を申し上げます。

日頃の研究生生活において熱心な議論に応じてくださった早稲田大学人間科学学術院助手白勢彩子博士，東海女子大学人間関係学部野村理朗講師，金沢工業大学人間情報システム研究所研究員沢田晴彦氏諸氏に心より感謝の意を申し上げます。

日頃から本研究に関する議論に応じて下さっただけでなく、本論文についても御一読し，適切な御助言を下さいました神戸大学大学院自然科学研究科研究員阿瀬見典昭博士に心より感謝の意を申し上げます。

また、本研究を遂行するにあたって主観評価実験の被験者に御協力して下さいました名古屋大学大学院人間情報学研究科(現:情報科学研究科)認知情報論講座及び名古屋大学大

学院工学研究科板倉研究室 (現:情報科学研究科武田研究室) の院生の皆様に心より感謝の意を申し上げます。

最後に、本研究生生活を陰ながら支えて下さいました家族・親類に改めて感謝の意を述べると共に、この論文を今は亡き父に捧げます。



## 参考文献

- [1] A. D. Blumlein, Improvements in and relating to sound-transmission, sound-recording and sound-reproducing systems, British Patent, No. 394325, 1931.
- [2] J. Blauert, Spatial Hearing, p. 204, MIT Press, Cambridge, 1997.
- [3] M. A. Gerzon, "Periphony: With-height sound reproduction," Journal of Audio Engineering Society, Vol. 21, No. 1, pp. 2–10, January 1973.
- [4] 濱崎公男, "放送における立体音響," JAS Journal, Vol. 44, No. 10, pp. 36–42, October 2004.
- [5] ITU-R Recommendation, Multichannel Stereophonic Sound System with and without Accompanying Picture, BS.775-1, 1992-1994.
- [6] 火山浩一郎, 濱崎公男, "走査線 4000 本級超高精細映像のための 3 次元音響システム," 日本音響学会講演論文集, No. 2-10-12, pp. 567–568, March 2004.
- [7] 岡部馨, "ダミーヘッドを用いた音場再生," 日本音響学会誌, Vol. 46, No. 8, pp. 650–656, August 1990.
- [8] 安藤四一, 設楽貞樹, 前川純一, 城戸健一, "電子計算機による室内音響設計法に関する二, 三の基礎的考察," 日本音響学会誌, Vol. 29, No. 3, pp. 151–159, March 1973.
- [9] M. R. Schroeder, D. Gottlob, and K. F. Siebrasse, "Comparative study of european concert halls: Correlation of subjective preference with geometric and acoustic parameters," Journal of Acoustical Society of America, Vol. 56, No. 4, pp. 1195–1201, October 1974.
- [10] M. Miyoshi and Y. Kaneda, "Inverse filtering of room acoustics," IEEE Transaction on Acoustics, Speech and Signal Processing, Vol. 36, No. 2, pp. 145–152, February 1988.
- [11] 三好正人, 小泉宣夫, "多数のスピーカを用いたトランスオーラル系の構成について," 日本音響学会講演論文集, No. 1-7-1, pp. 371–372, March 1991.
- [12] 神沼充伸, 伊勢史郎, 鹿野清宏, "周波数領域における最小ノルム解を利用した多チャンネル音場再現システムにおける逆フィルタの設計," 日本音響学会誌, Vol. 57, No. 3, pp. 175–183, March 2001.

- [13] 古見芳幸, 穂刈治英, 島田正治, “立体音像再生に関するスピーカ配置の検討,” 電子情報通信学会論文誌, Vol. J83-A, No. 7, pp. 920–923, July 2000.
- [14] 浜田晴夫, 得能宏則, O. Kirkeby, P. A. Nelson, “新しい音場再生方式 (Stereo Dipole) に関する研究,” 電子情報通信学会技術研究報告, No. EA96-2, pp. 9–17, April 1996.
- [15] H. Fletcher, “Symposium on wire transmission of symphonic music and its reproduction on auditory perspective: Basic requirement,” Bell System Technical Journal, Vol. 13, No. 2, pp. 239–244, April 1934.
- [16] J. C. Steinberg and W. Snow, “Symposium on wire transmission of symphonic music and its reproduction on auditory perspective: Physical factors,” Bell System Technical Journal, Vol. 13, No. 2, pp. 245–258, April 1934.
- [17] E. H. Bedell and I. Kerney, “Symposium on wire transmission of symphonic music and its reproduction on auditory perspective: System adaptation,” Bell System Technical Journal, Vol. 13, No. 2, pp. 301–308, April 1934.
- [18] 伊藤毅, 音響工学原論, pp. 587–589, コロナ社, 東京, 1973.
- [19] M. Camras, “Approach to recreating a sound field,” Journal of Acoustical Society of America, Vol. 43, No. 6, pp. 1425–1431, November 1968.
- [20] A. J. Berkhout, D. de Vries, and P. Vogel, “Acoustic control by wave field synthesis,” Journal of Acoustical Society of America, Vol. 93, No. 5, pp. 2764–2778, May 1993.
- [21] D. de Vries and J. Baan, “Auralization of sound fields by wave field synthesis,” Preprints of AES 106th Convention, No. 4927, May 1999.
- [22] 及川靖広, 馬場貞如, 山崎芳男, “人間の聴覚を考慮した波面合成法による音場制御と音場再現,” 日本バーチャルリアリティ学会論文誌, Vol. 5, No. 3, pp. 919–926, September 2000.
- [23] 平井徹, 本地由和, 池田雅弘, 清水寧, 川上福司, 東山三樹夫, “遠隔地音楽レッスン用 AVW システムとその原理,” 電子情報通信学会技術研究報告, No. EA2000-17, pp. 25–32, June 2000.
- [24] 古家賢一, 一ノ瀬裕, “境界面音圧による閉空間の音場制御,” 電子情報通信学会技術研究報告, No. EA90-15, pp. 25–32, May 1990.
- [25] 神沼充伸, 伊勢史郎, 鹿野清宏, “受聴者の頭部の動きを考慮した多チャンネル音場再現システム,” 日本バーチャルリアリティ学会論文誌, Vol. 5, No. 3, pp. 957–964, September 2000.
- [26] S. Takane, Y. Suzuki, and T. Sone, “A new method for global sound reproduction based on Kirchhoff’s integral equation,” ACUSTICA - Acta Acustica, Vol. 85, No. 2, pp. 250–

- 257, March/April 1999.
- [27] S. Ise, "A principle of sound field control based on the Kirchhoff-Helmholtz integral equation and the theory of inverse systems," *ACUSTICA - Acta Acustica*, Vol. 85, No. 1, pp. 78–87, January/February 1999.
- [28] S. Takane, Y. Suzuki, T. Miyajima, and T. Sone, "A new theory for high definition virtual acoustic display named ADVISE," *Acoustical Science & Technology*, Vol. 24, No. 5, pp. 276–283, September 2003.
- [29] D. B. Ward and T. D. Abhayapala, "Reproduction of a plane-wave sound field using an array of loudspeakers," *IEEE Transaction on Speech and Audio Processing*, Vol. 9, No. 6, pp. 697–707, September 2001.
- [30] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *Journal of Acoustical Society of America*, Vol. 65, No. 4, pp. 943–950, 1979.
- [31] M. Morimoto, "The relation between spatial impression and the precedence effect," *Proceedings of International Conference on Auditory Display*, No. SS#2-1, pp. 297–306, Kyoto, Japan, July 2002.
- [32] J. Blauert, *Spatial Hearing*, p. 118, MIT Press, Cambridge, 1997.
- [33] <http://www.mathworks.com/products/matlab/>.
- [34] Technical Centre of the European Broadcasting Union (EBU), *Sound Quality Assessment Material - Recordings for Subjective Tests*, 1988.
- [35] H. Wallach, E. B. Newman, and M. R. Rosenzweig, "The precedence effect in sound localization," *American Journal of Psychology*, Vol. 62, No. 3, pp. 315–336, July 1949.
- [36] J. Blauert, *Spatial Hearing*, p. 39, MIT Press, Cambridge, 1997.
- [37] M. R. Schroeder, "New method of measuring reverberation time," *Journal of Acoustical Society of America*, Vol. 37, No. 3, pp. 409–412, March 1965.
- [38] 官能検査ハンドブック, 日科技連官能検査委員会 (編), pp. 366–374, 日科技連出版社, 東京, 1973.
- [39] S. Vernon, "Design and implementation of AC-3 coders," *IEEE Trans. Consumer Electronics*, Vol. 41, No. 3, pp. 754–759, August 1995.
- [40] ISO/IEC, *Information Technology - Genetic Coding of Moving Pictures and Associated Audio Information - Part 7 Advanced Audio Coding*, No. 13818-7, 1997.
- [41] D. Yang, H. Ai, C. Kyriakakis, and C.-C. Jay Kuo, "High-fidelity multichannel audio coding with Karhunen-Loève transform," *IEEE Transaction on Speech and Audio Pro-*

- cessing, Vol. 11, No. 4, pp. 365–380, July 2003.
- [42] Y. Wang, M. Vilermo, M. Väänänen, and L. Yaroslavsky, “A multichannel audio coding algorithm for inter-channel redundancy removal,” Preprints of AES 110th Convention, No. 5295, May 2001.
- [43] J. Bauck and D. H. Cooper, “Generalized transaural stereo and applications,” *Journal of Audio Engineering Society*, Vol. 44, No. 9, pp. 683–705, September 1996.
- [44] Y. Suzuki, F. Asano, H. Y. Kim, and T. Sone, “An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses,” *Journal of Acoustical Society of America*, Vol. 97, No. 2, pp. 1119–1123, February 1995.
- [45] デジタル音声・オーディオ技術, 北脇信彦 (編), pp. 172–174, 電気通信協会, 東京, 1999.
- [46] 伊藤憲三, 北脇信彦, “双方向通信系における伝搬遅延の許容限に関する検討,” *日本音響学会聴覚研究会資料*, No. H84-57, 1984.
- [47] ITU-R Recommendation, *Methods for the Subjective Assessment of Small Impairments in Audio Systems Including Multichannel Sound Systems*, BS.1116-1, 1994-1997.
- [48] 中島弘史, 東山三樹夫, “周波数領域処理を繰り返して利用する逆フィルタの時間領域設計法,” *日本音響学会講演論文集*, No. 2-5-9, pp. 431–432, September 1999.
- [49] S. J. Elliott, I. M. Stothers, and P. A. Nelson, “A multiple error lms algorithm and its application to the active control of sound and vibration,” *IEEE Transaction on Acoustics, Speech and Signal Processing*, Vol. 35, No. 10, pp. 1423–1434, October 1987.
- [50] Y. Tatekuta, H. Saruwatari, and K. Shikano, “An iterative inverse filter design method for the multichannel sound field reproduction system,” *IEICE Transaction on Fundamentals of Electronics, Communications and Computer Sciences*, Vol. E84-A, No. 4, pp. 991–998, April 2001.
- [51] B.-I. Dalenbäck, “The rediscovery of diffuse reflection in room acoustics prediction,” *Journal of Acoustical Society of America*, Vol. 112, No. 5, p. 2225, November 2002. CATT-Acoustic (<http://www.catt.se/>).

## 付録 A

# 合成した波面の一例

第 2 章における波面合成シミュレーションによって合成された波面の一例を図 A.1 ~ A.36 に示す。「Original Wavefront」は原音場内における波面を表わす。 $M$  はシミュレーションで用いたマイクロホン及びスピーカの数、「Wavefront」は聴取領域内において合成された波面、「Difference」は合成された波面と原音場内の波面との差分を表わす。この差分表示の中心付近が黒いほど波面が忠実に合成されていることを示している。

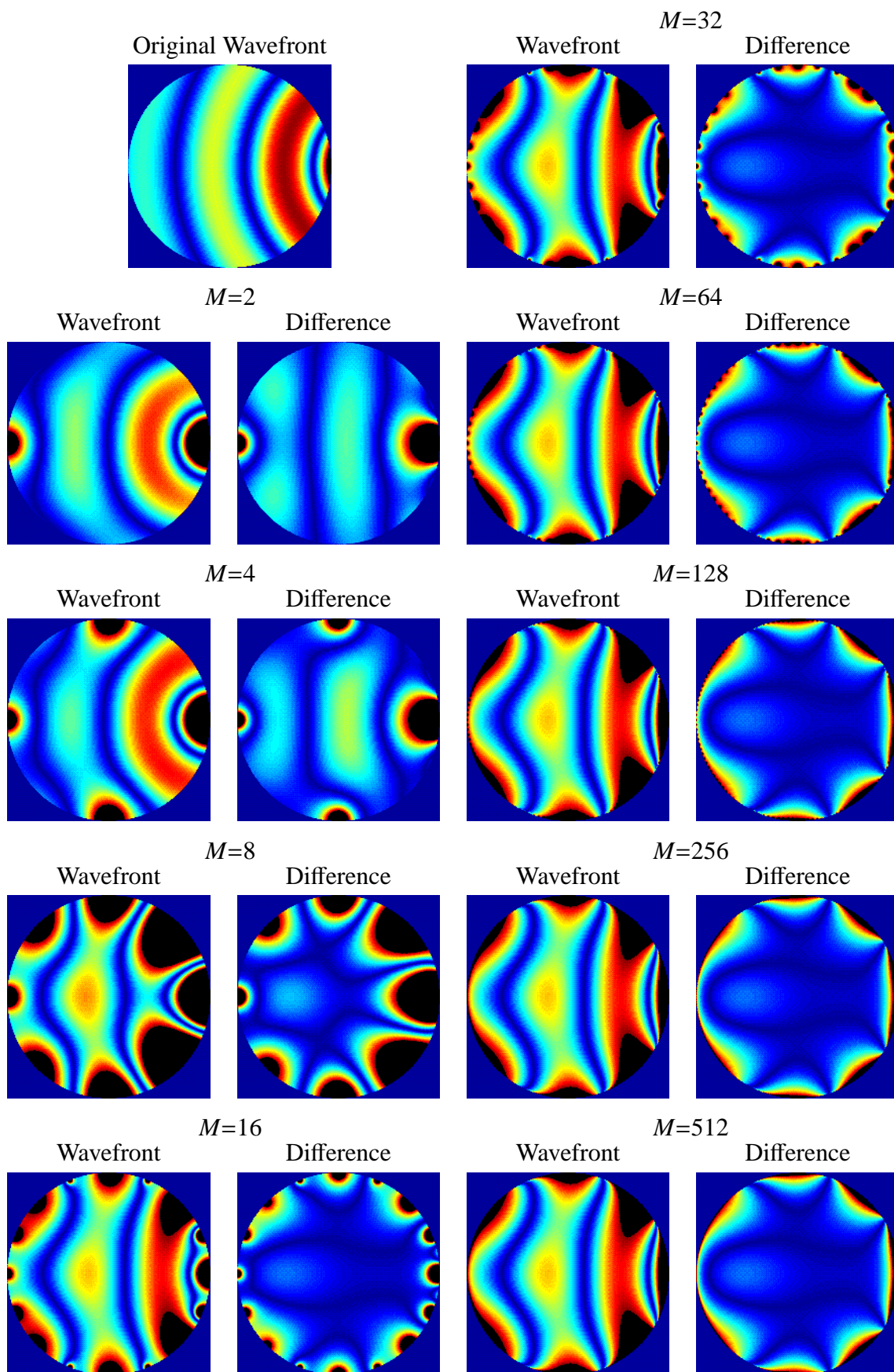


図 A.1 合成した波面の一例 (無指向性,  $f=125$  Hz,  $d=3$  m)

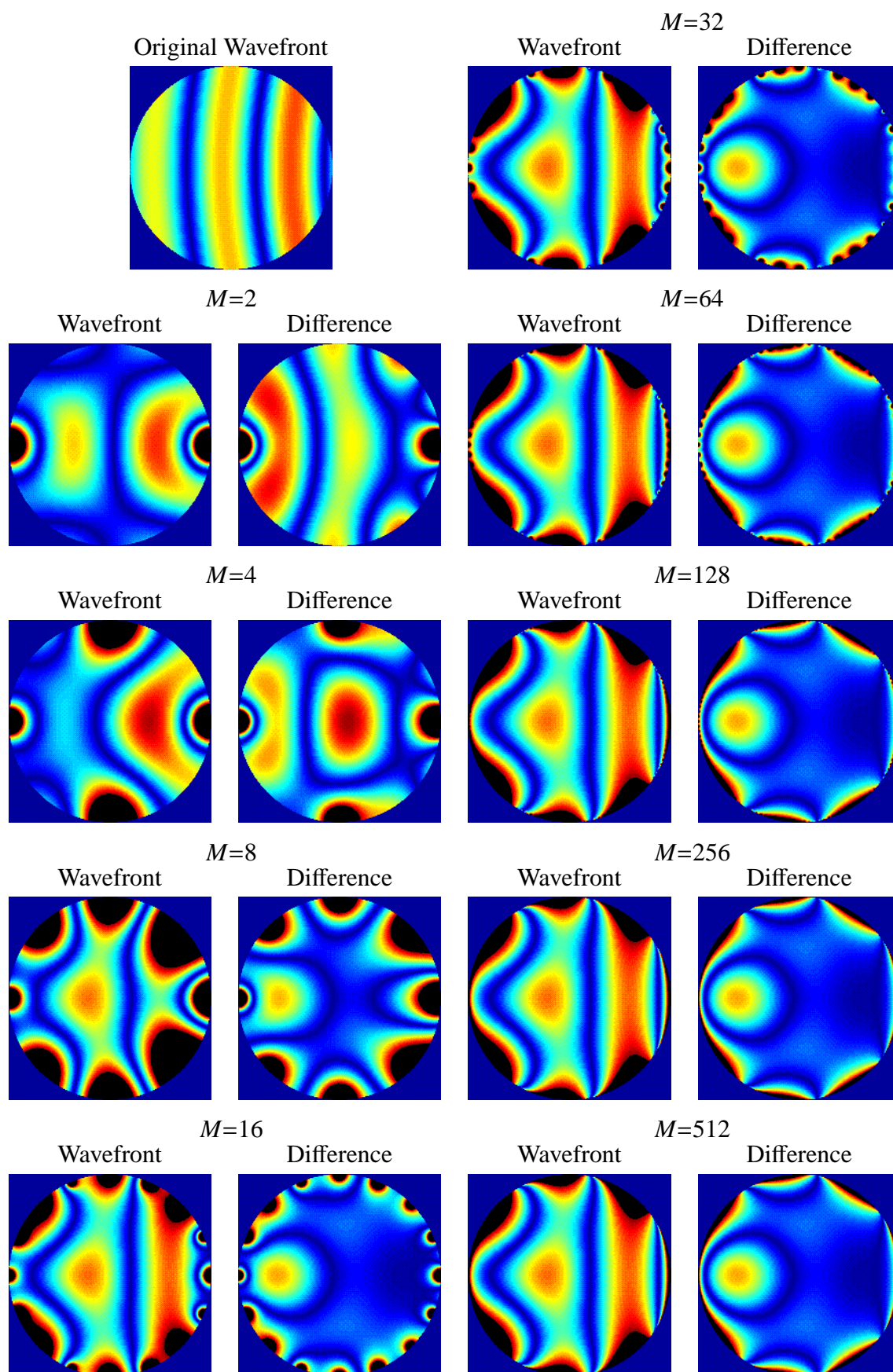


図 A.2 合成した波面の一例 (無指向性,  $f=125$  Hz,  $d=10$  m)



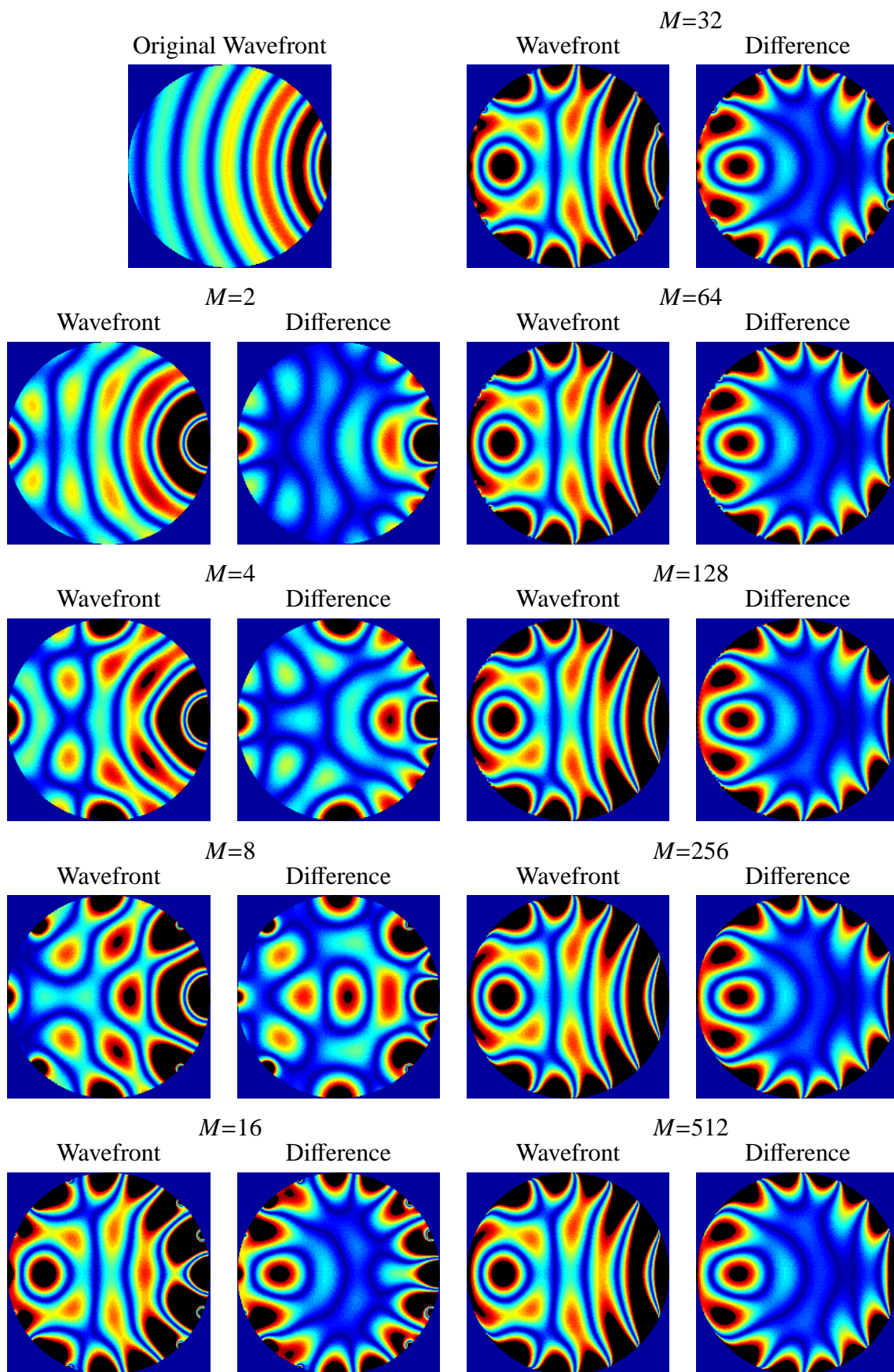


図 A.3 合成した波面の一例 (無指向性,  $f=250$  Hz,  $d=3$  m)



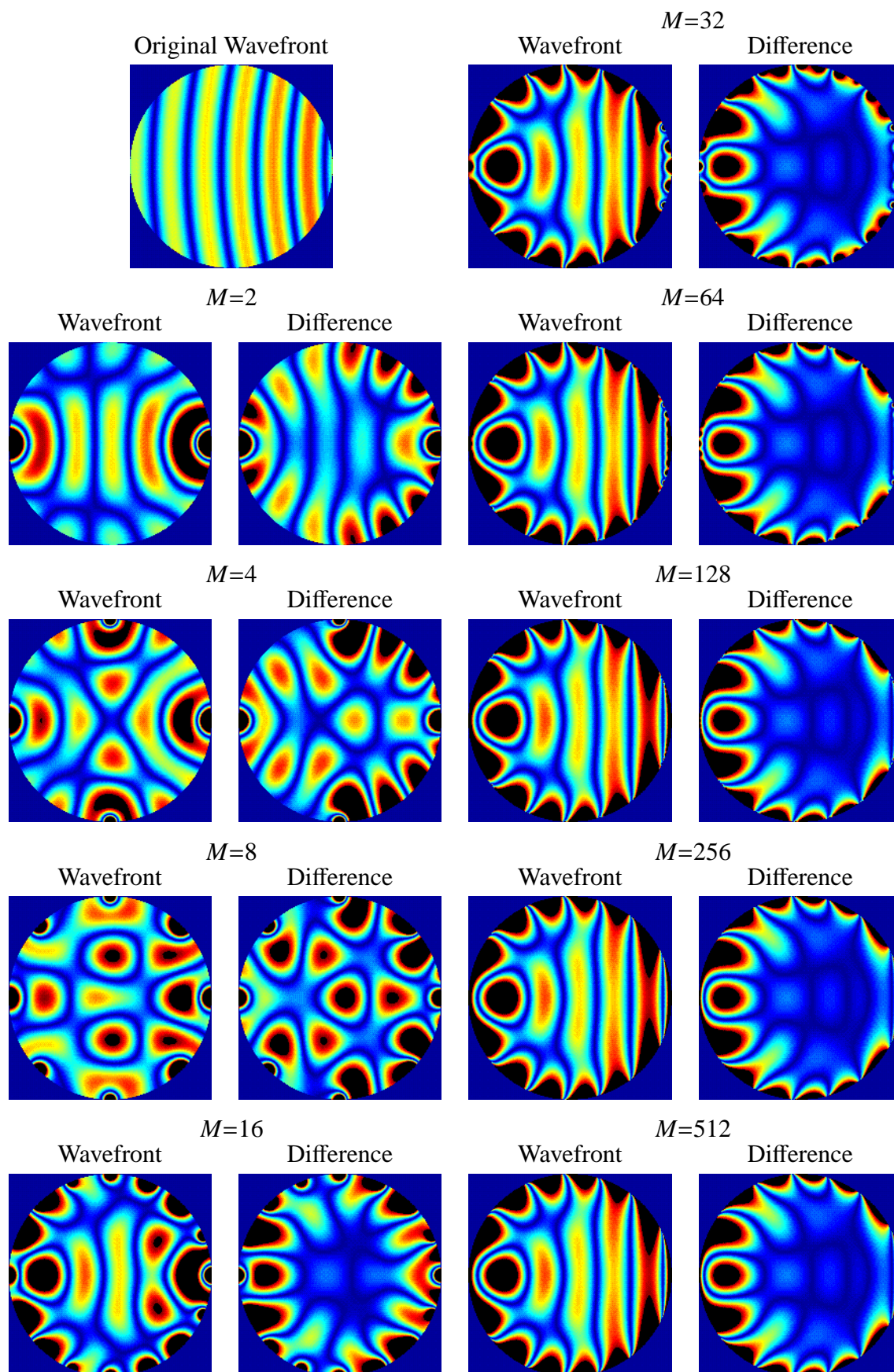


図 A.4 合成した波面の一例 (無指向性,  $f=250$  Hz,  $d=10$  m)

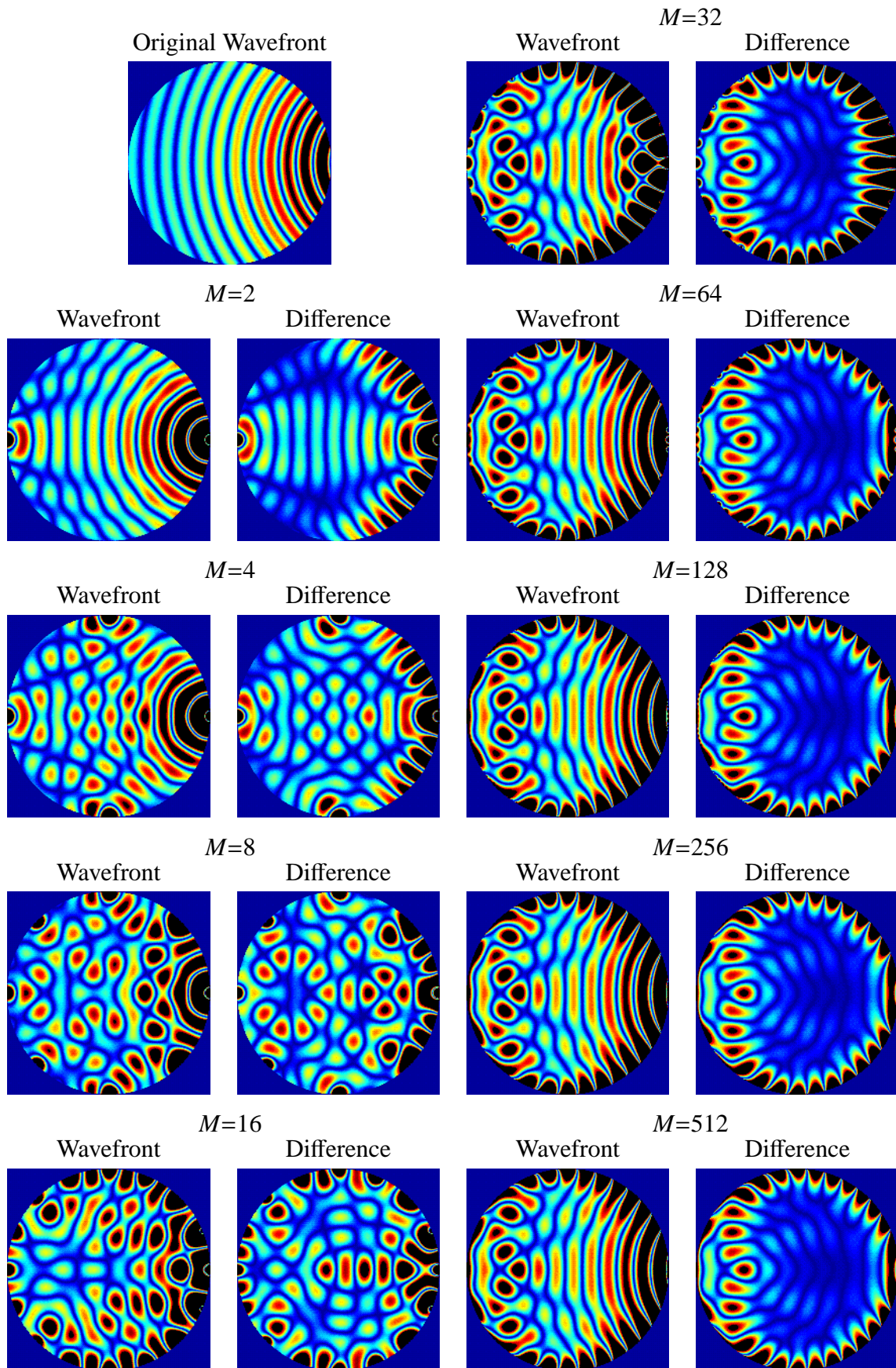


図 A.5 合成した波面の一例 (無指向性,  $f=500$  Hz,  $d=3$  m)



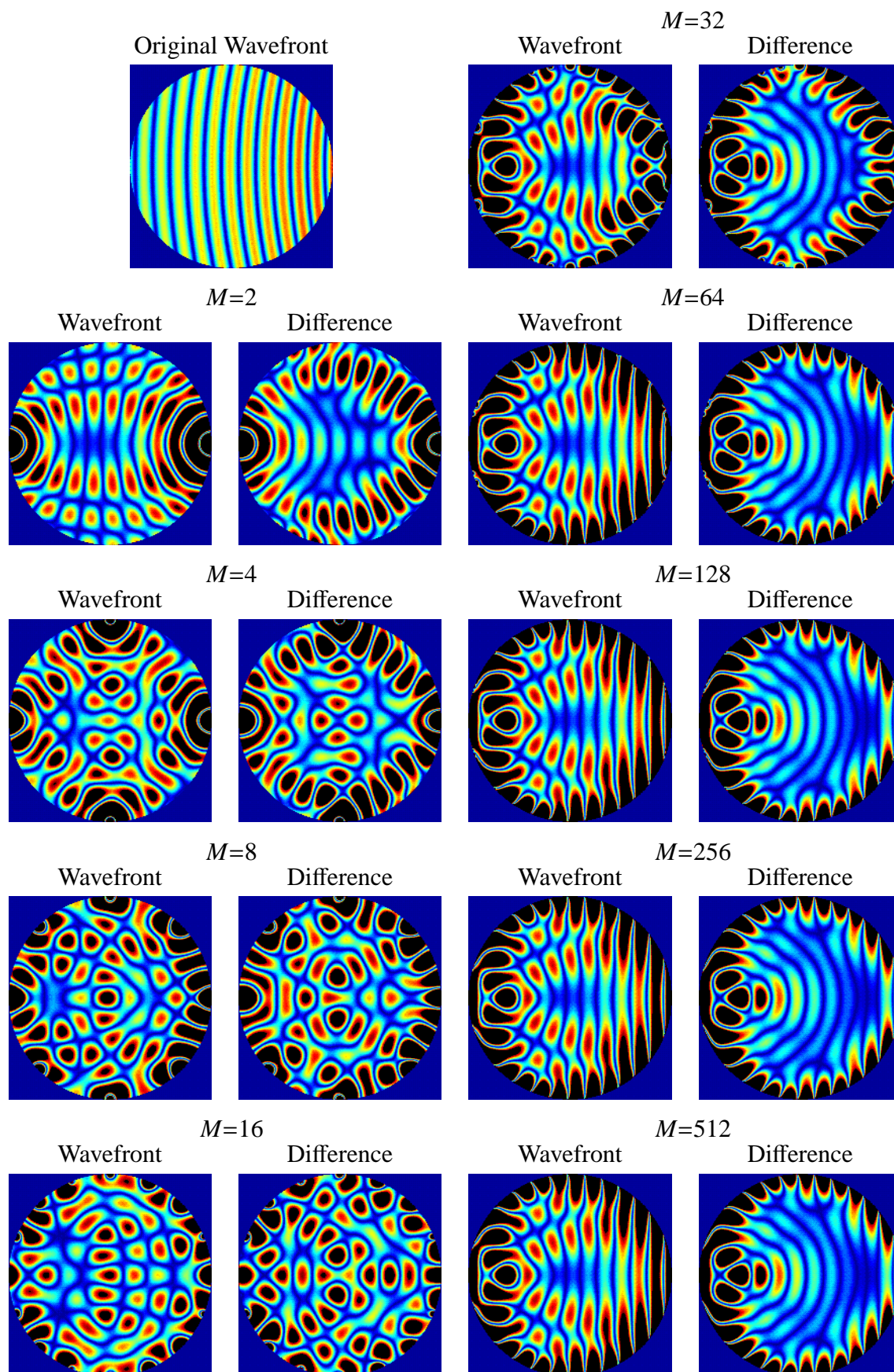


図 A.6 合成した波面の一例 (無指向性,  $f=500$  Hz,  $d=10$  m)



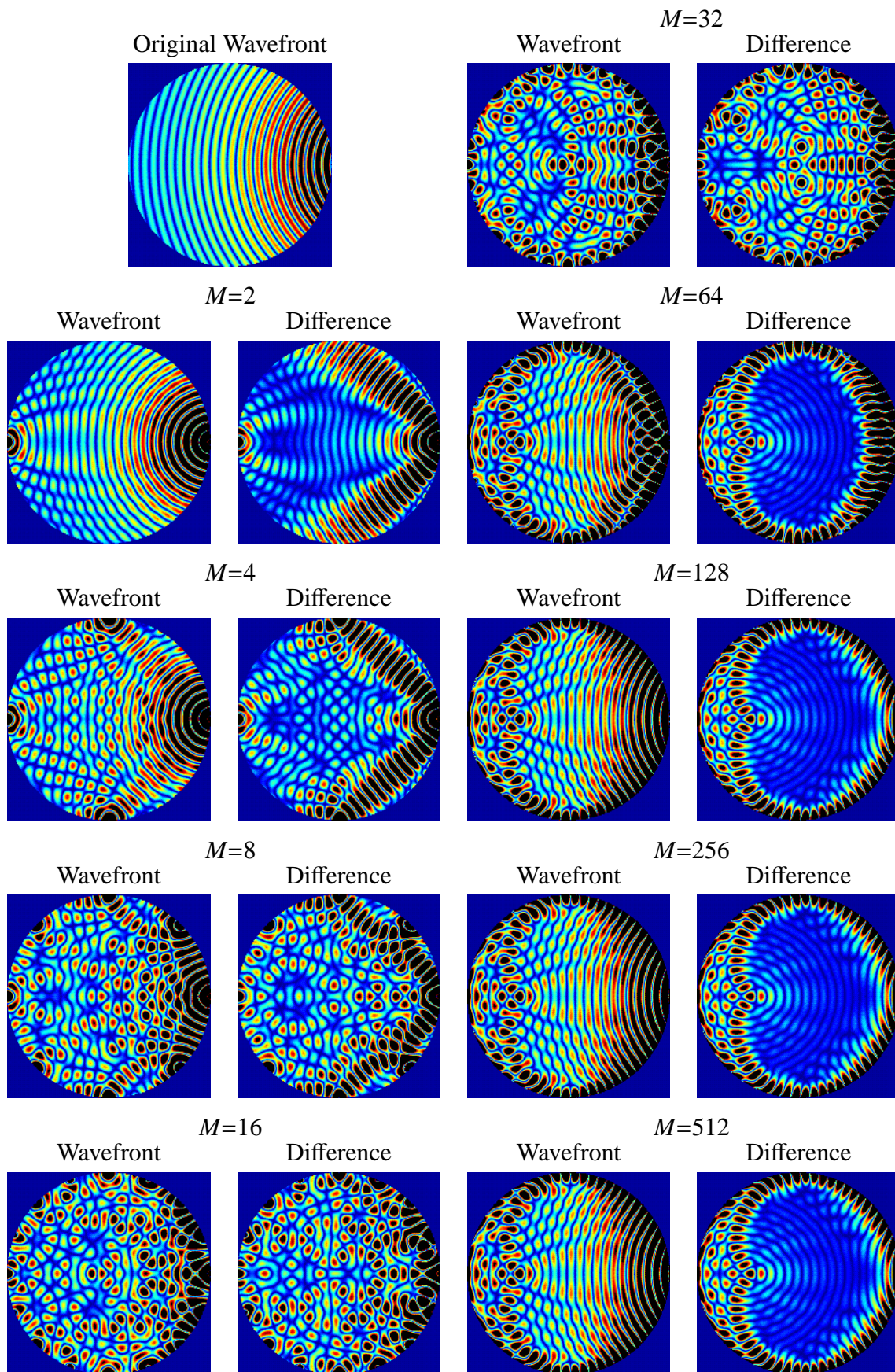


図 A.7 合成した波面の一例 (無指向性,  $f=1000$  Hz,  $d=3$  m)



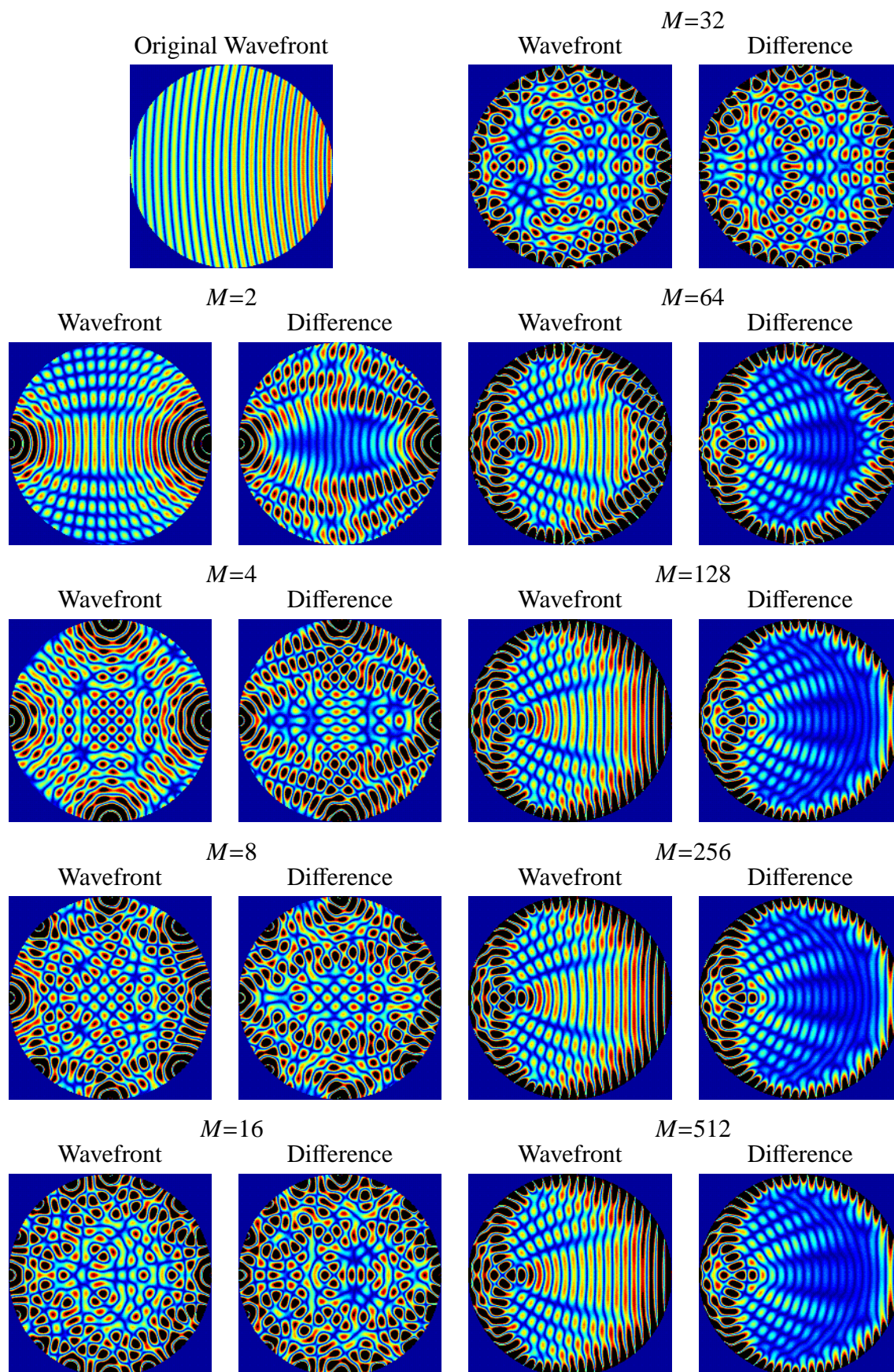


図 A.8 合成した波面の一例 (無指向性,  $f=1000$  Hz,  $d=10$  m)



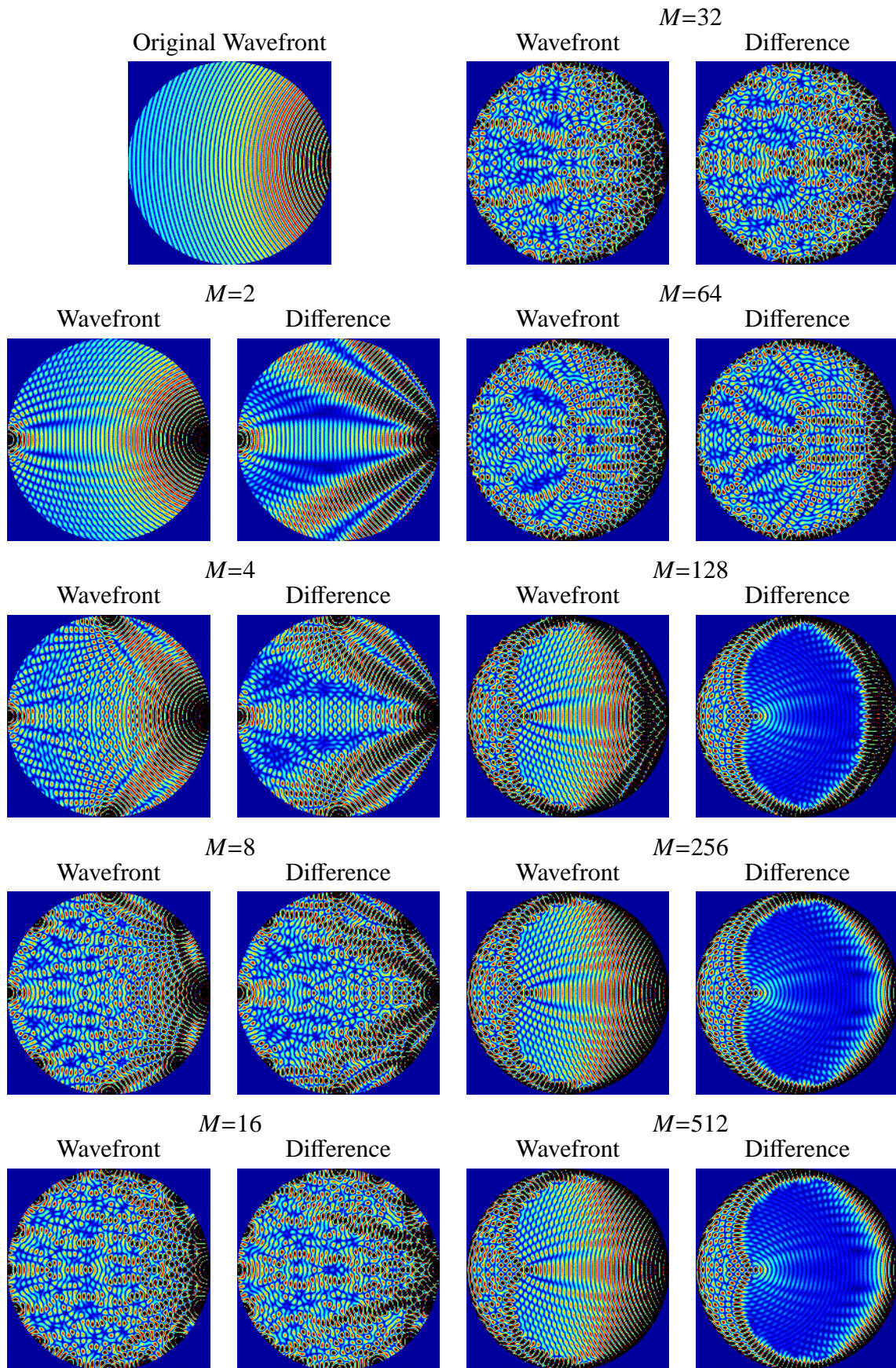


図 A.9 合成した波面の一例 (無指向性,  $f=2000$  Hz,  $d=3$  m)



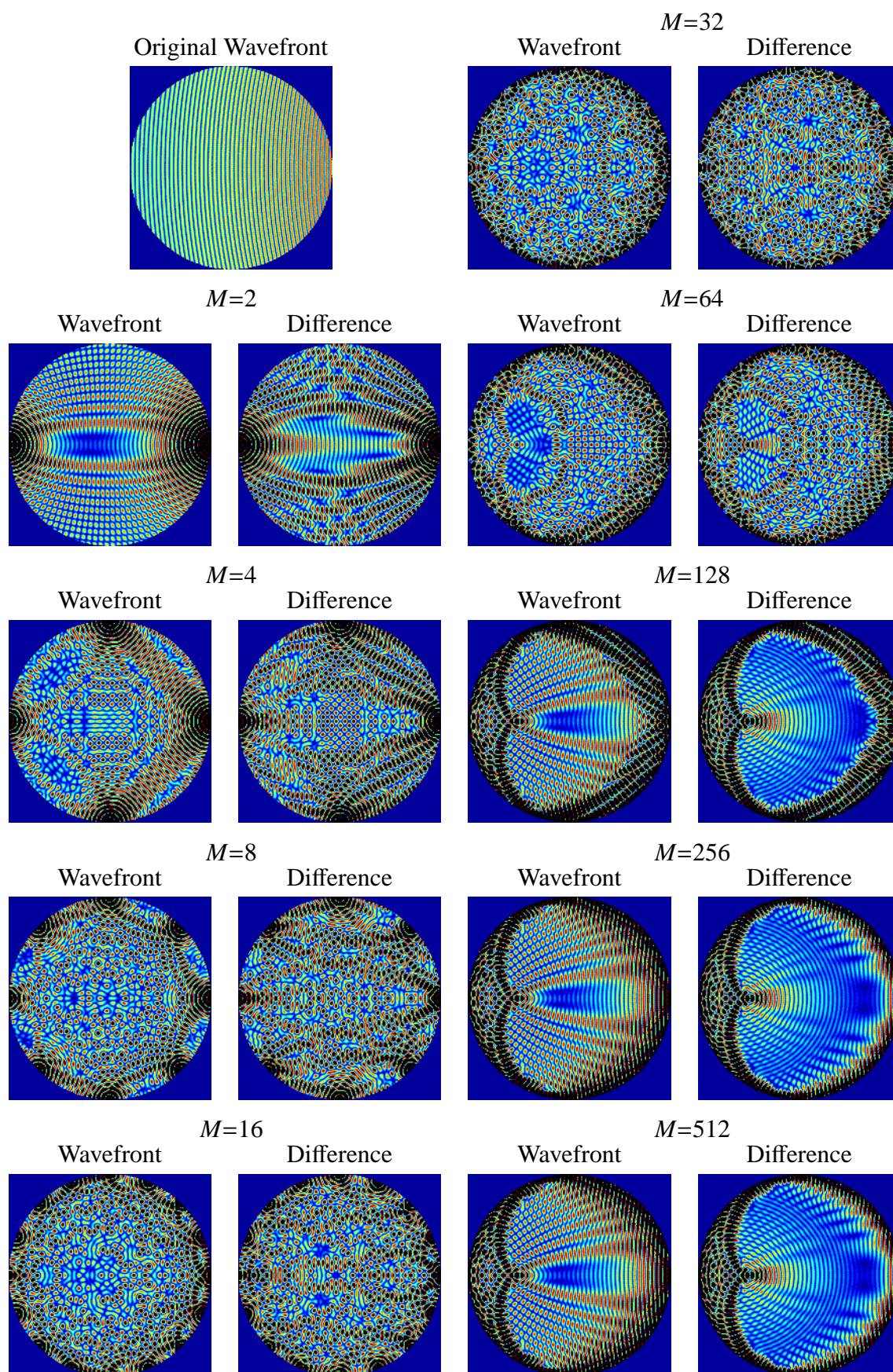


図 A.10 合成した波面の一例 (無指向性,  $f=2000$  Hz,  $d=10$  m)



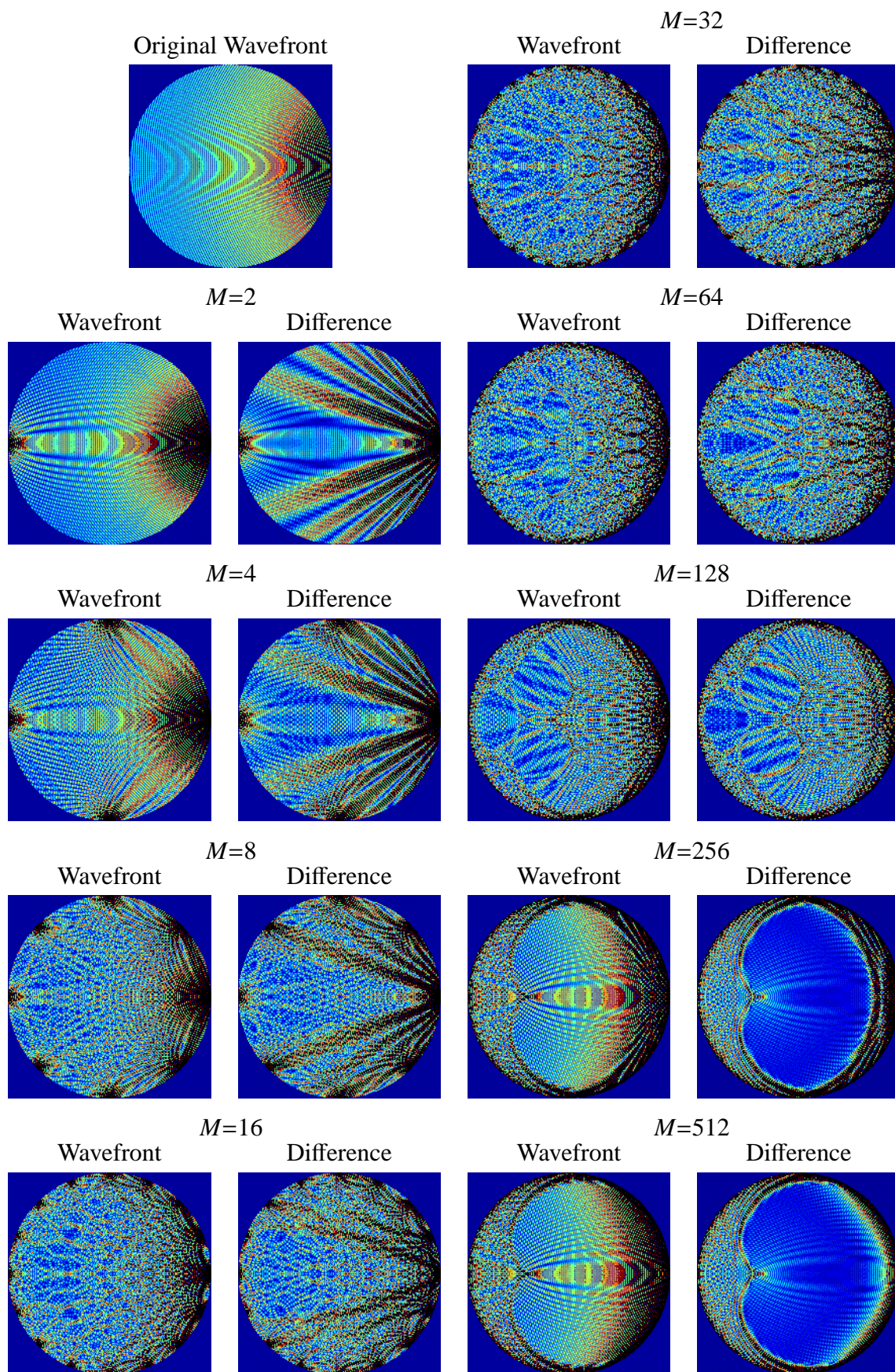


図 A.11 合成した波面の一例 (無指向性,  $f=4000$  Hz,  $d=3$  m)



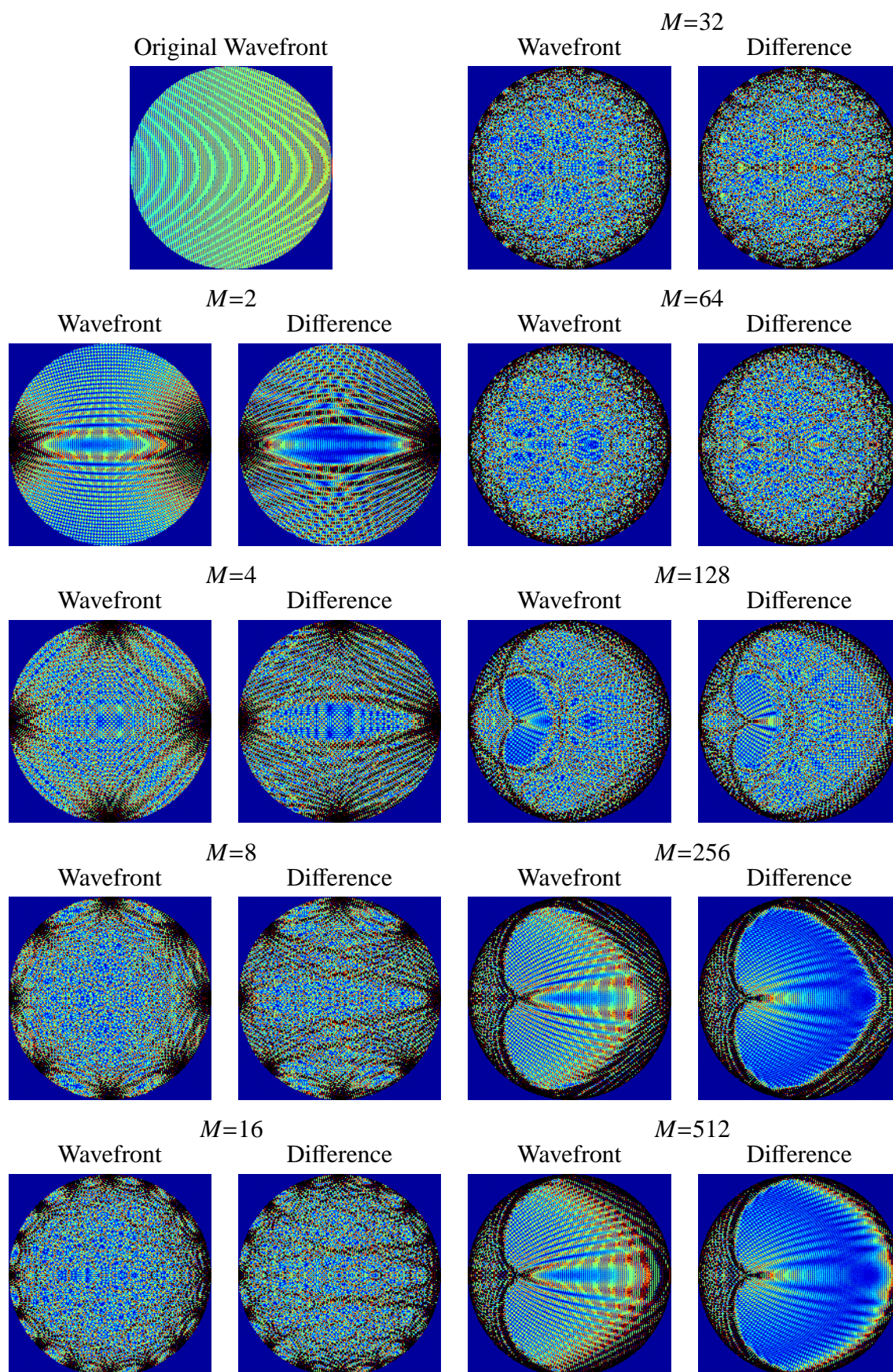


図 A.12 合成した波面の一例 (無指向性,  $f=4000$  Hz,  $d=10$  m)



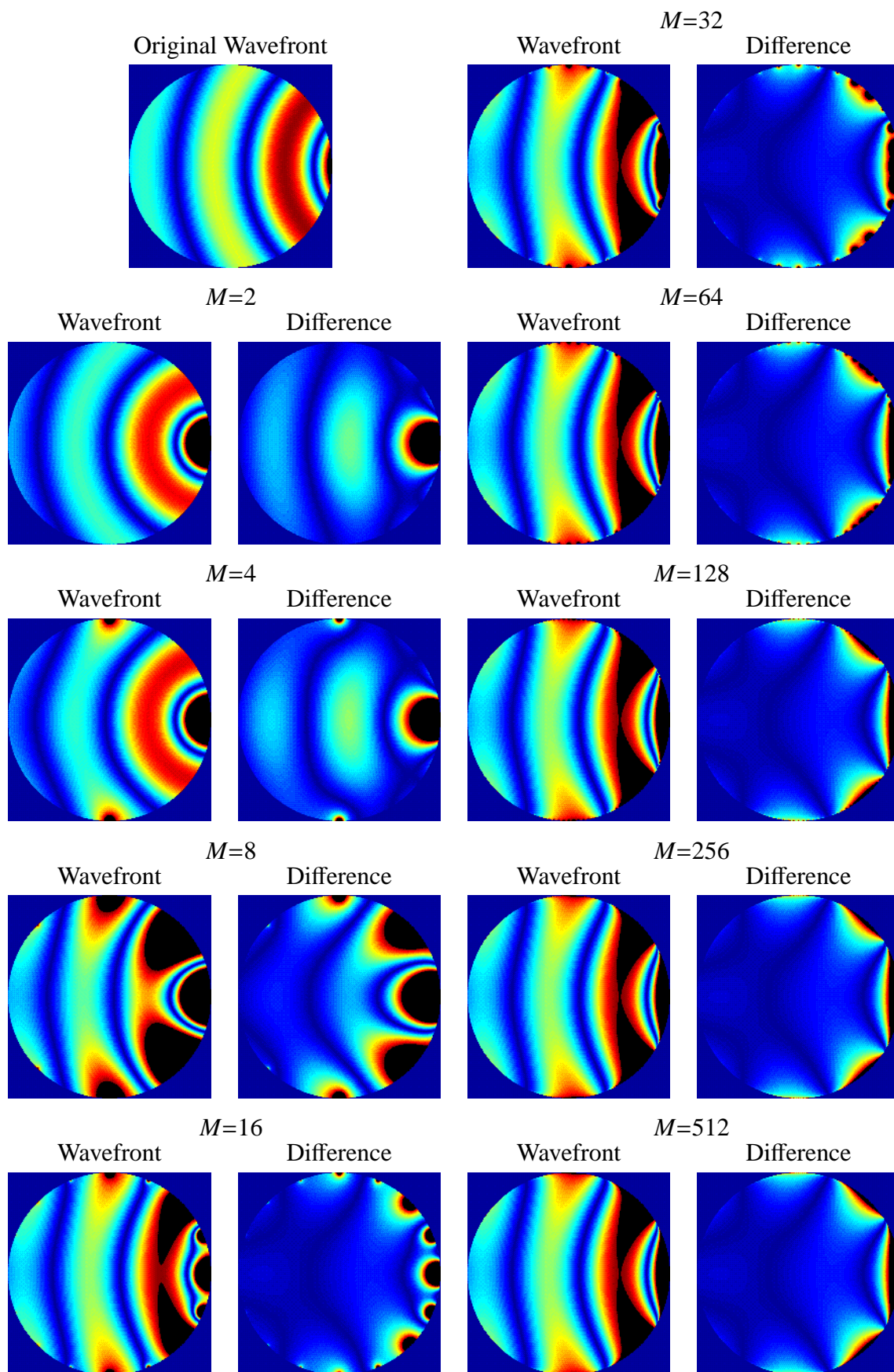


図 A.13 合成した波面の一例 (単一指向性,  $f=125$  Hz,  $d=3$  m)

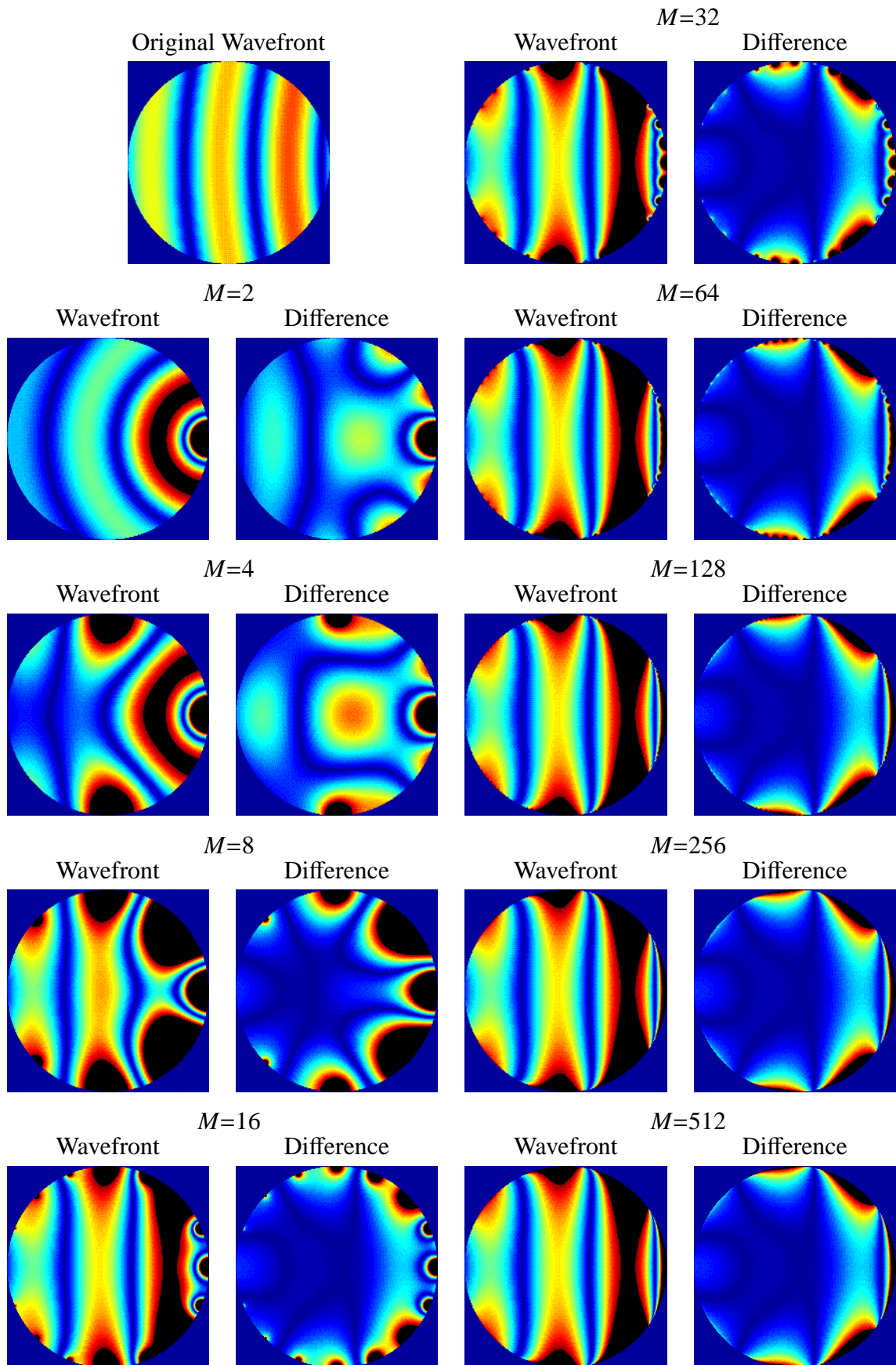


図 A.14 合成した波面の一例 (単一指向性,  $f=125$  Hz,  $d=10$  m)



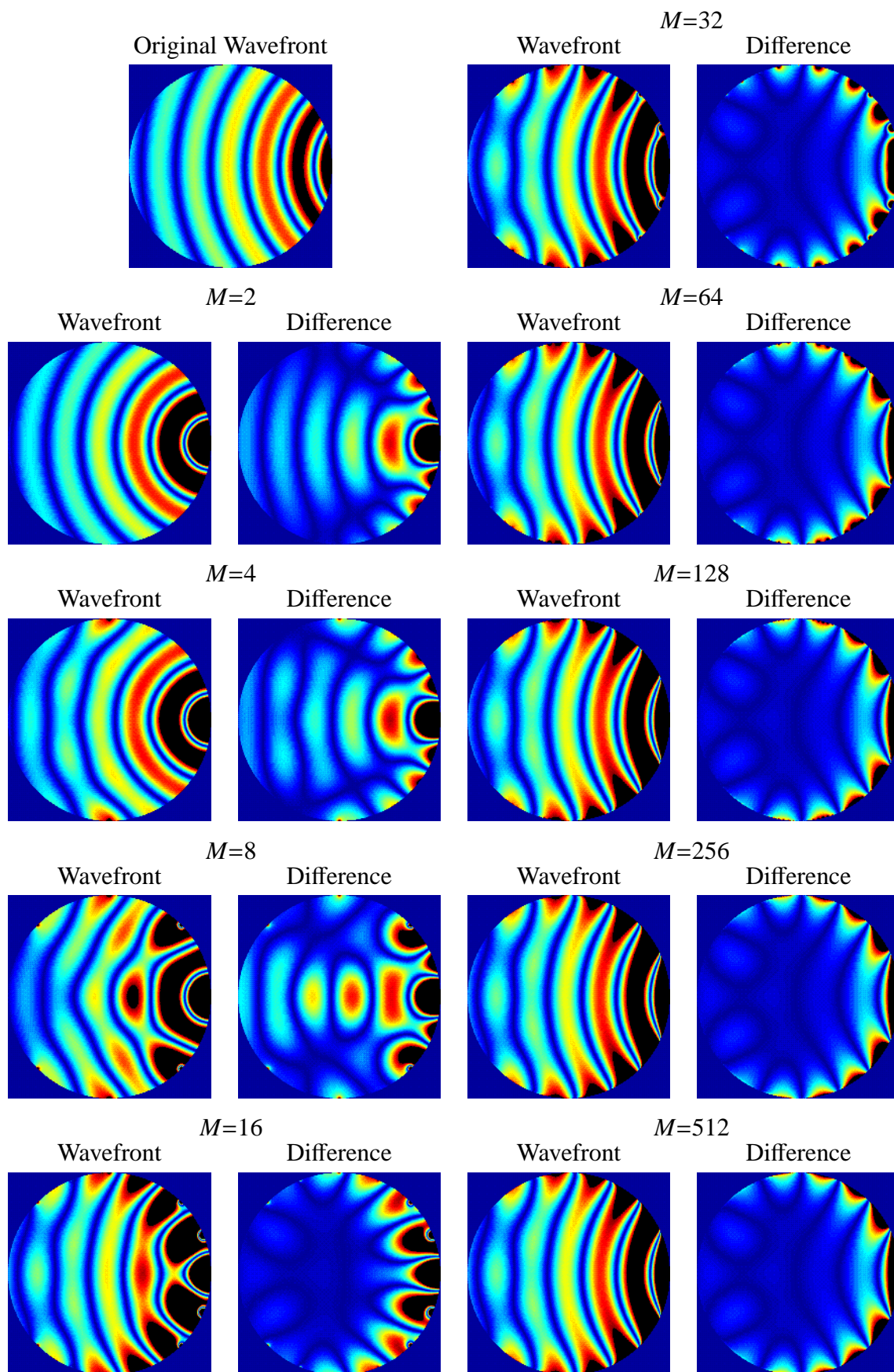


図 A.15 合成した波面の一例 (単一指向性,  $f=250$  Hz,  $d=3$  m)

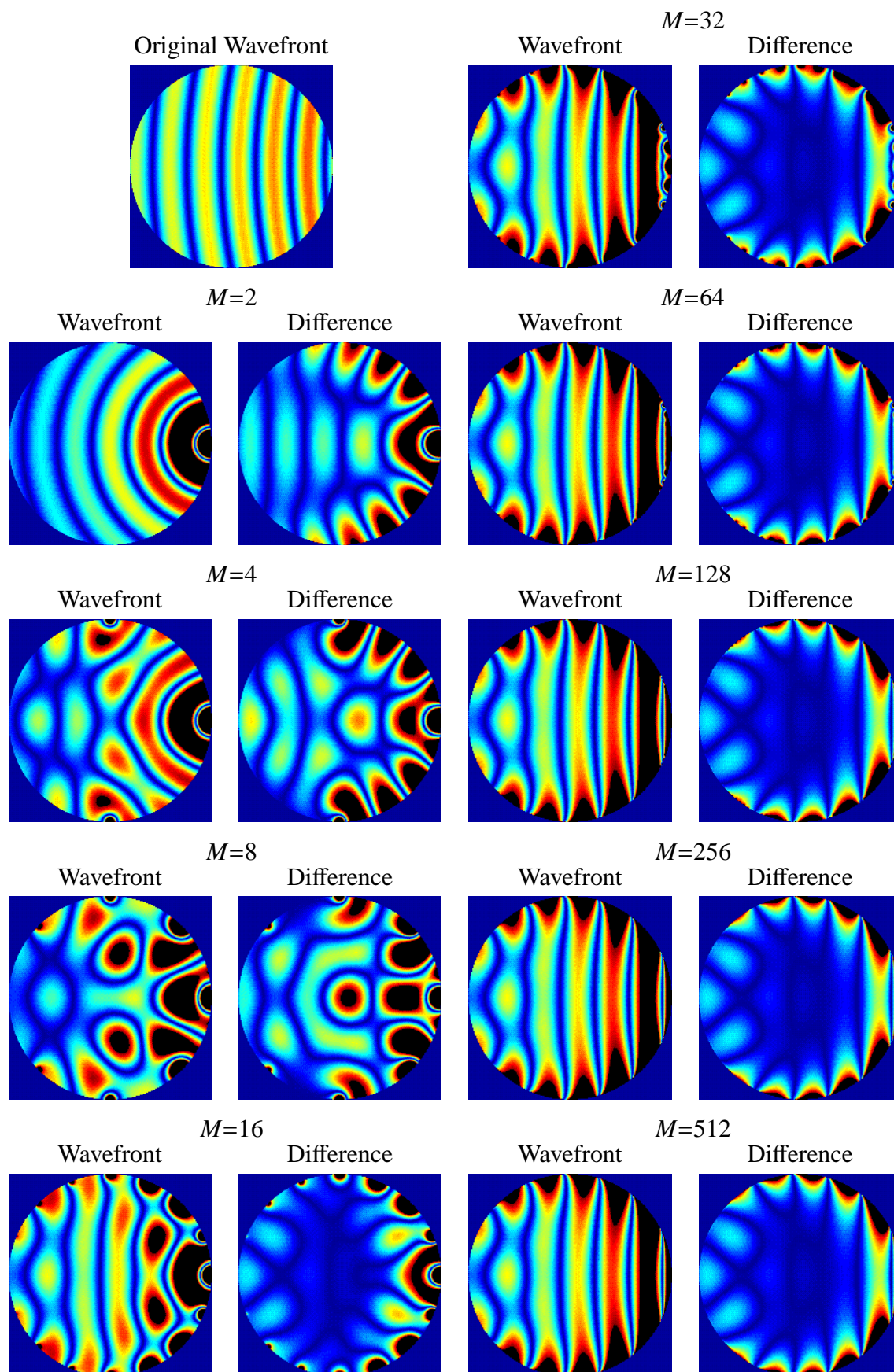


図 A.16 合成した波面の一例 (単一指向性,  $f=250$  Hz,  $d=10$  m)



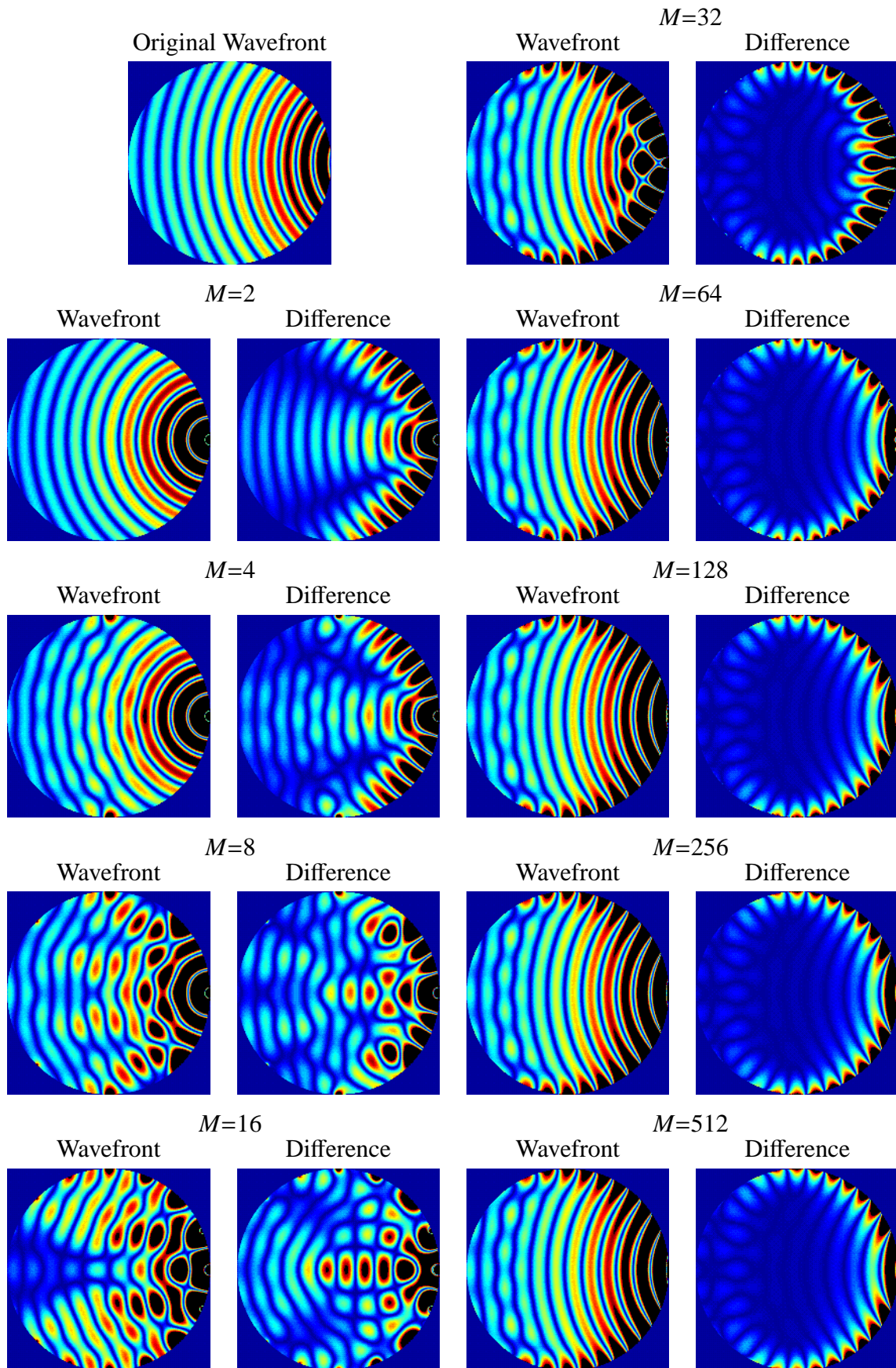


図 A.17 合成した波面の一例 (単一指向性,  $f=500$  Hz,  $d=3$  m)

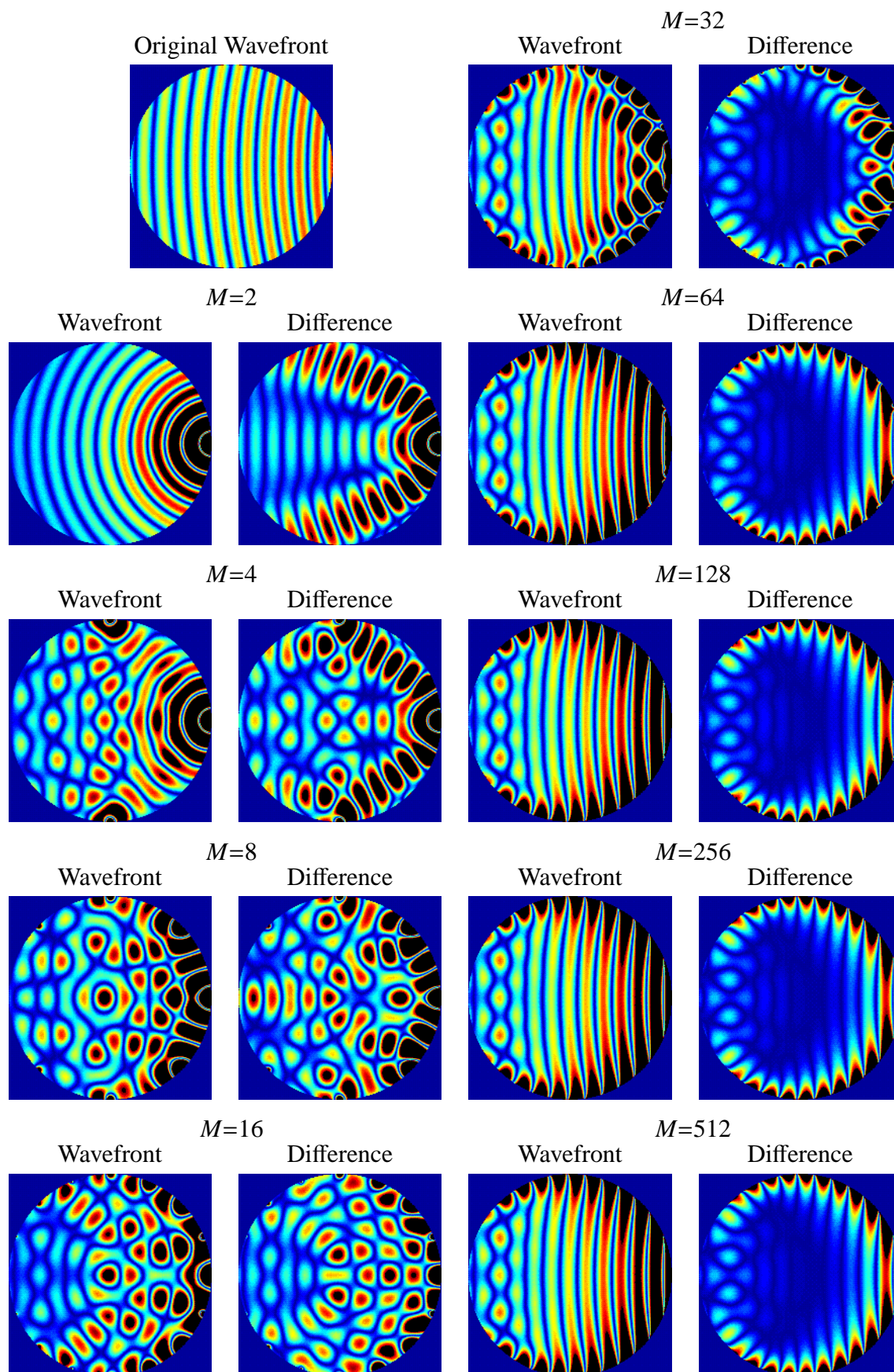


図 A.18 合成した波面の一例 (単一指向性,  $f=500$  Hz,  $d=10$  m)



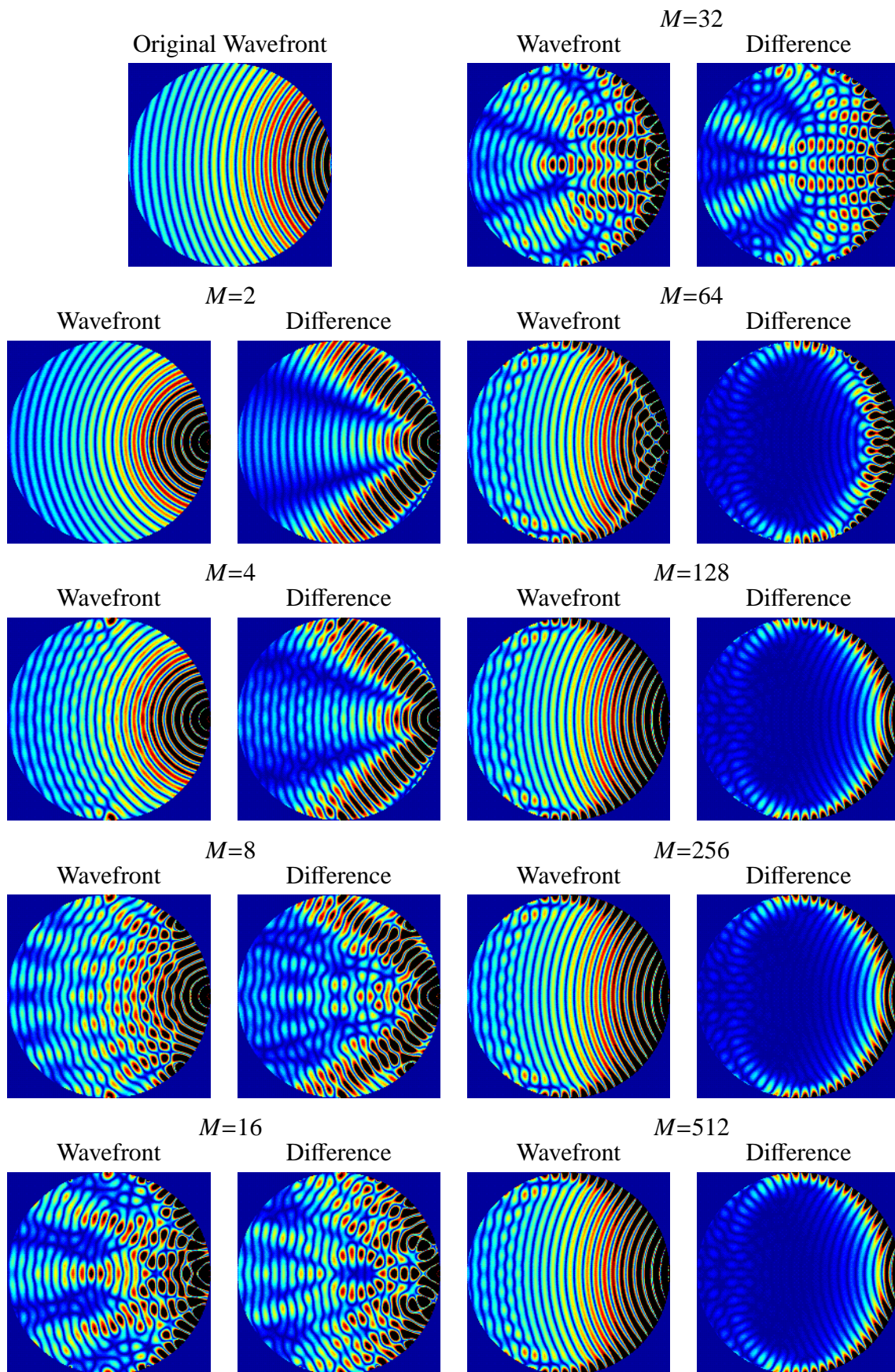


図 A.19 合成した波面の一例 (単一指向性,  $f=1000$  Hz,  $d=3$  m)



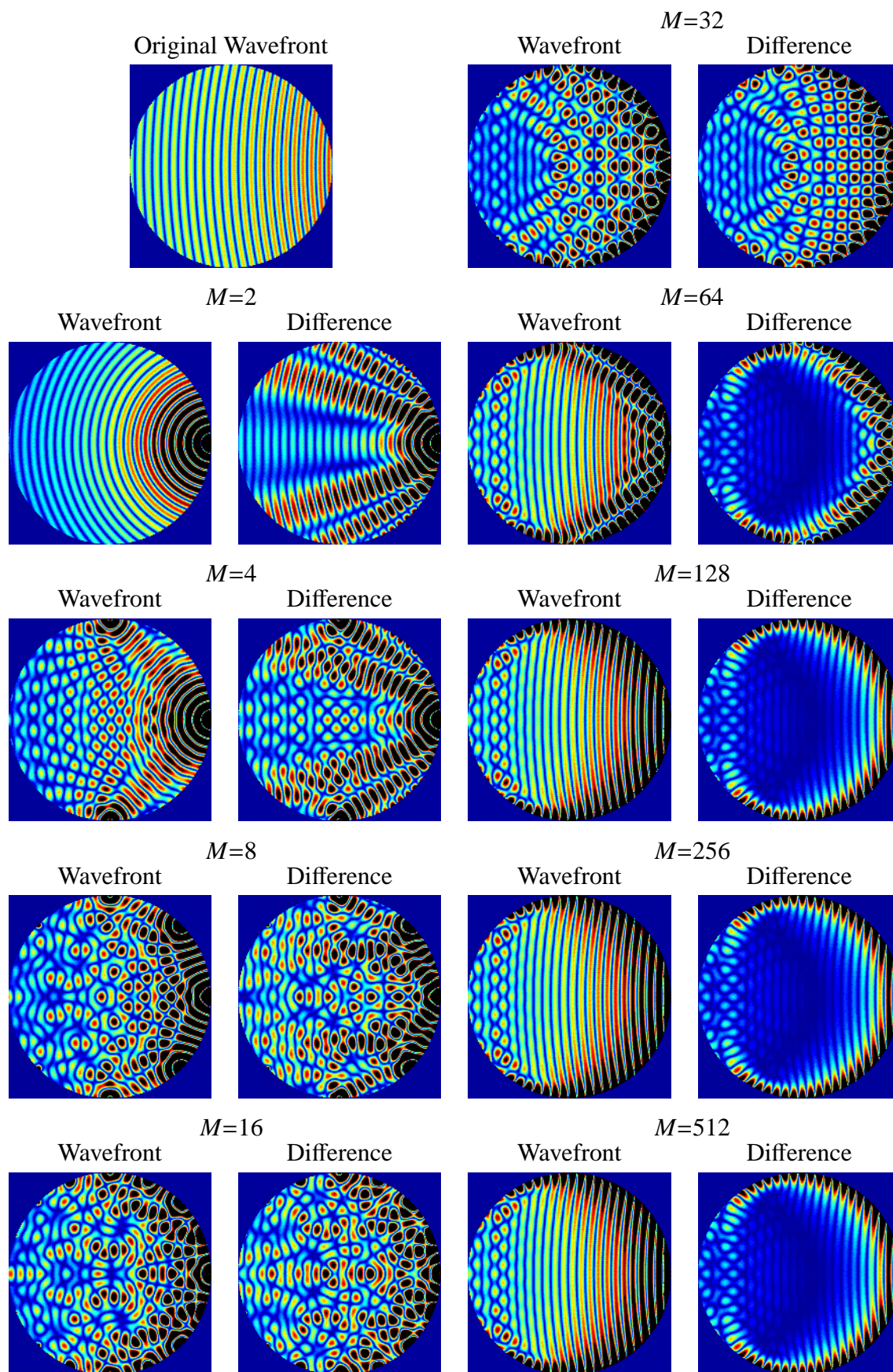


図 A.20 合成した波面の一例 (単一指向性,  $f=1000$  Hz,  $d=10$  m)



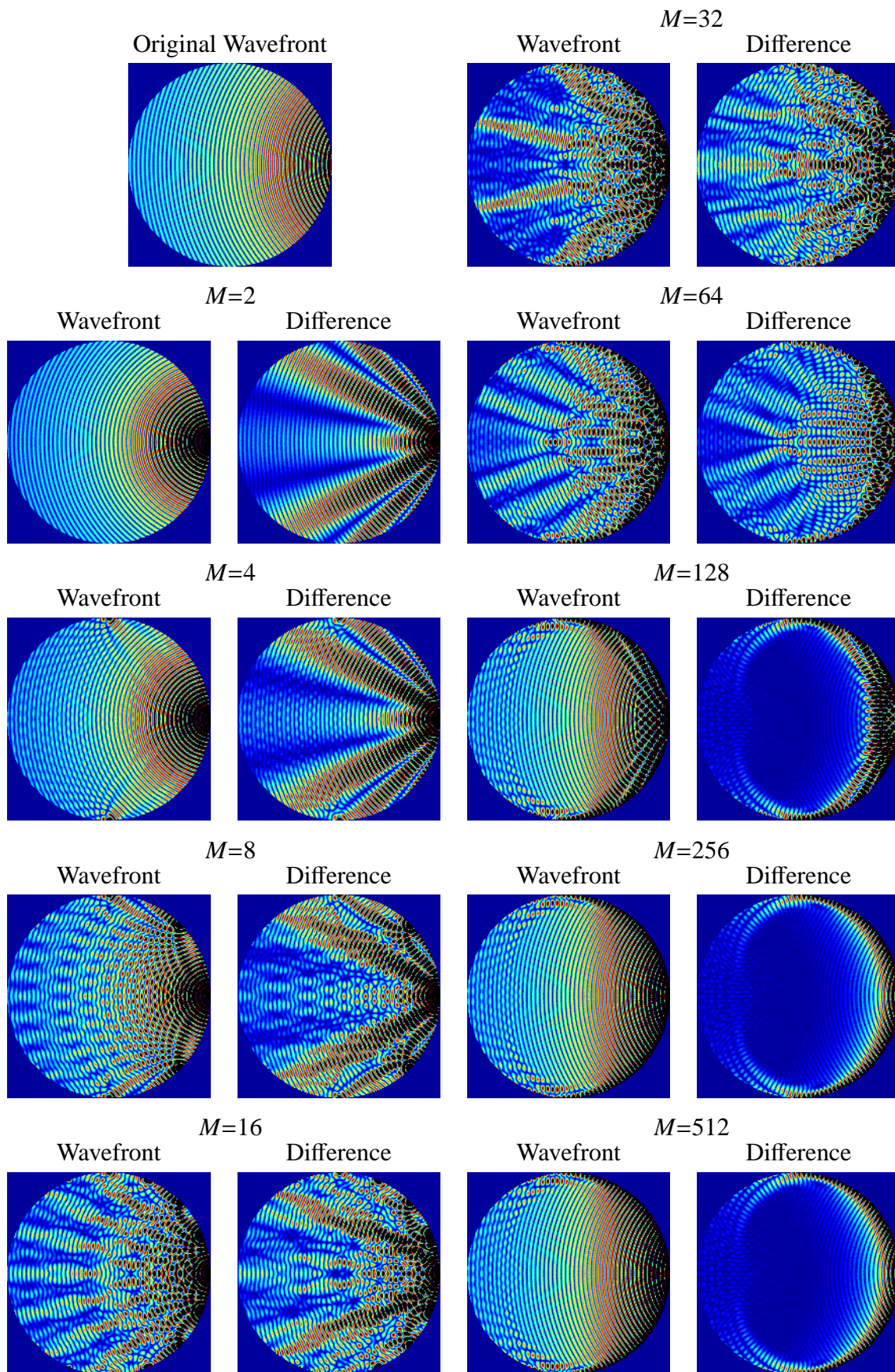


図 A.21 合成した波面の一例 (単一指向性,  $f=2000$  Hz,  $d=3$  m)



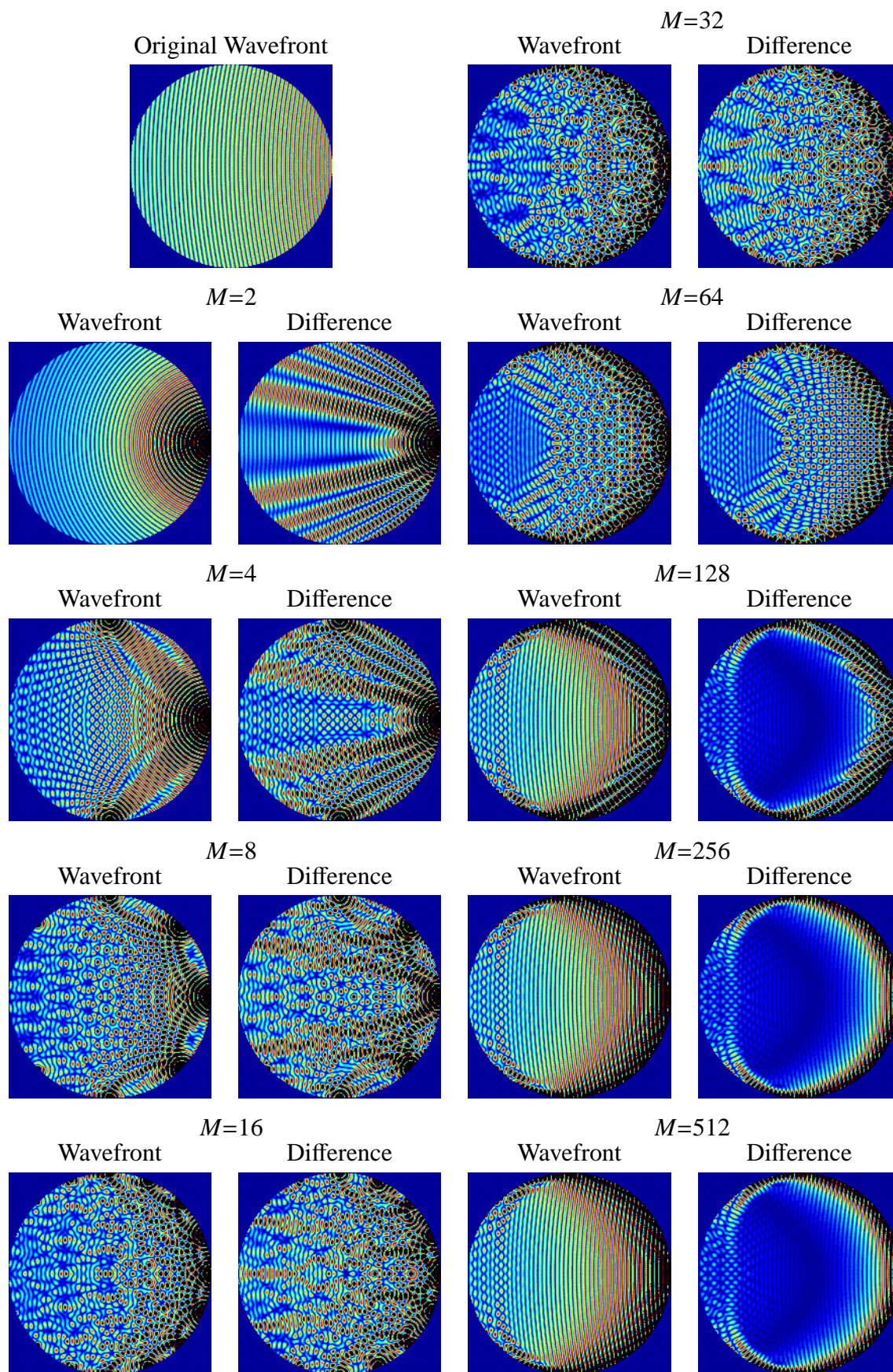


図 A.22 合成した波面の一例 (単一指向性,  $f=2000$  Hz,  $d=10$  m)



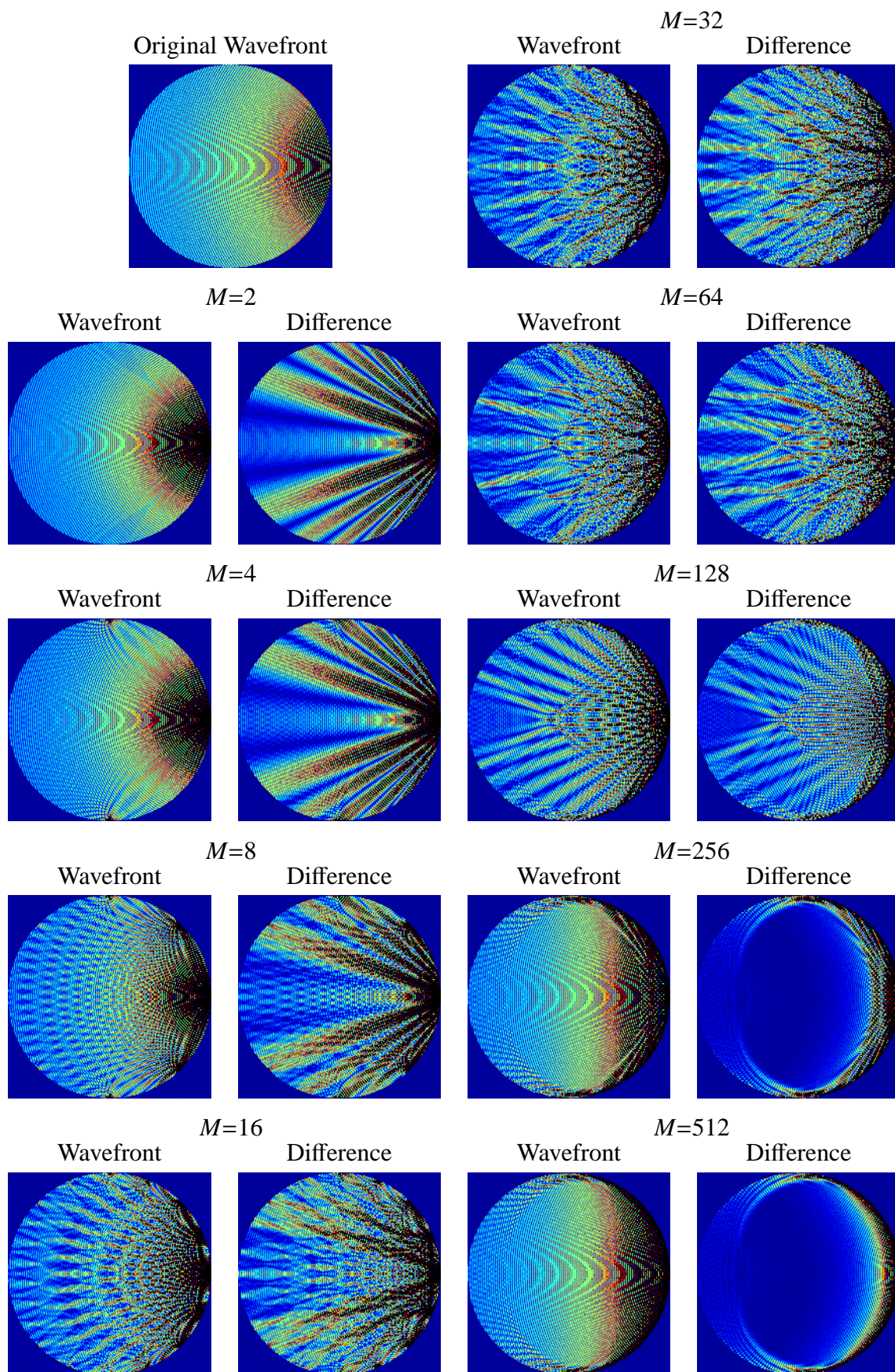


図 A.23 合成した波面の一例 (単一指向性,  $f=4000$  Hz,  $d=3$  m)



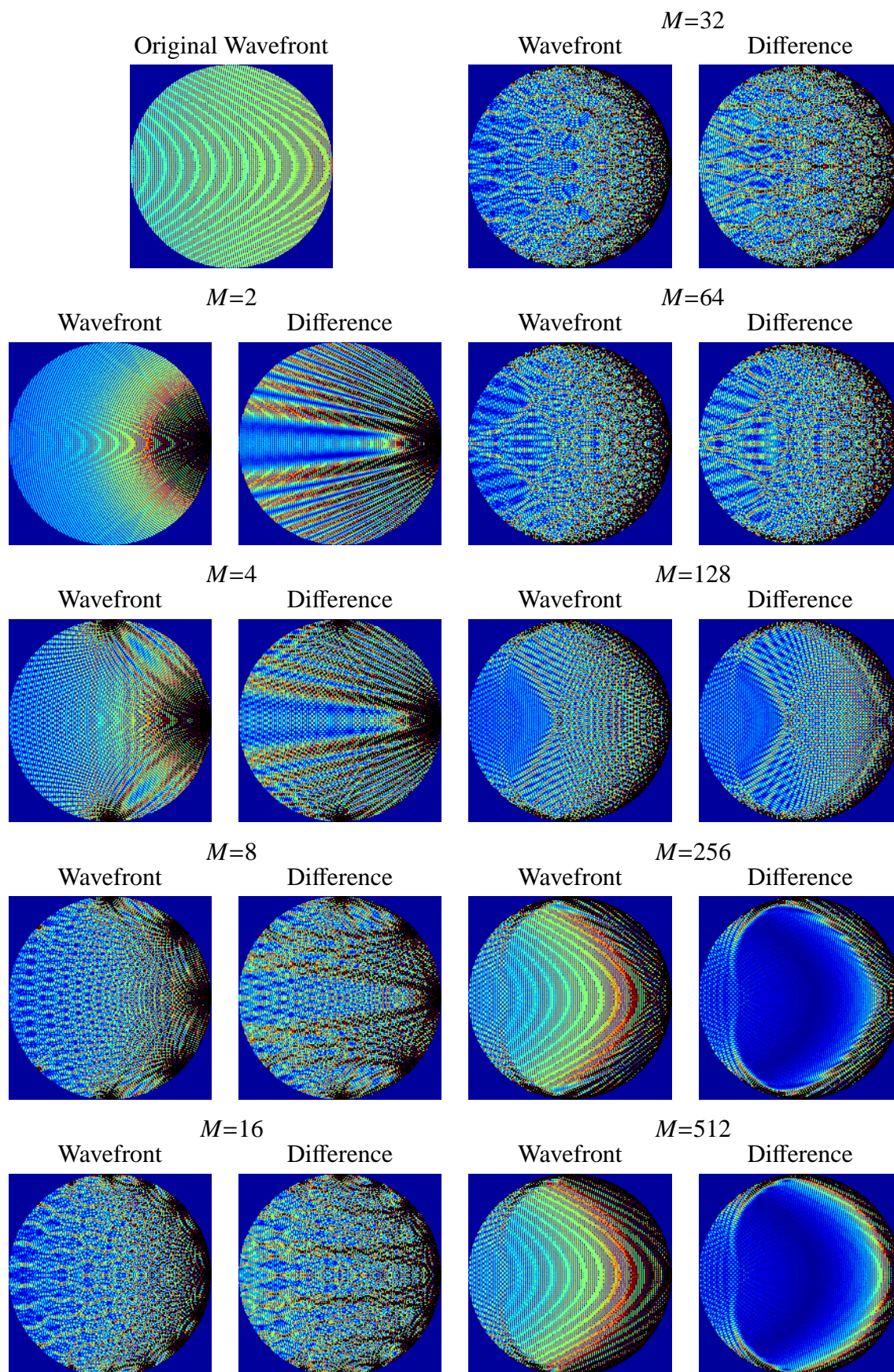


図 A.24 合成した波面の一例 (単一指向性,  $f=4000$  Hz,  $d=10$  m)



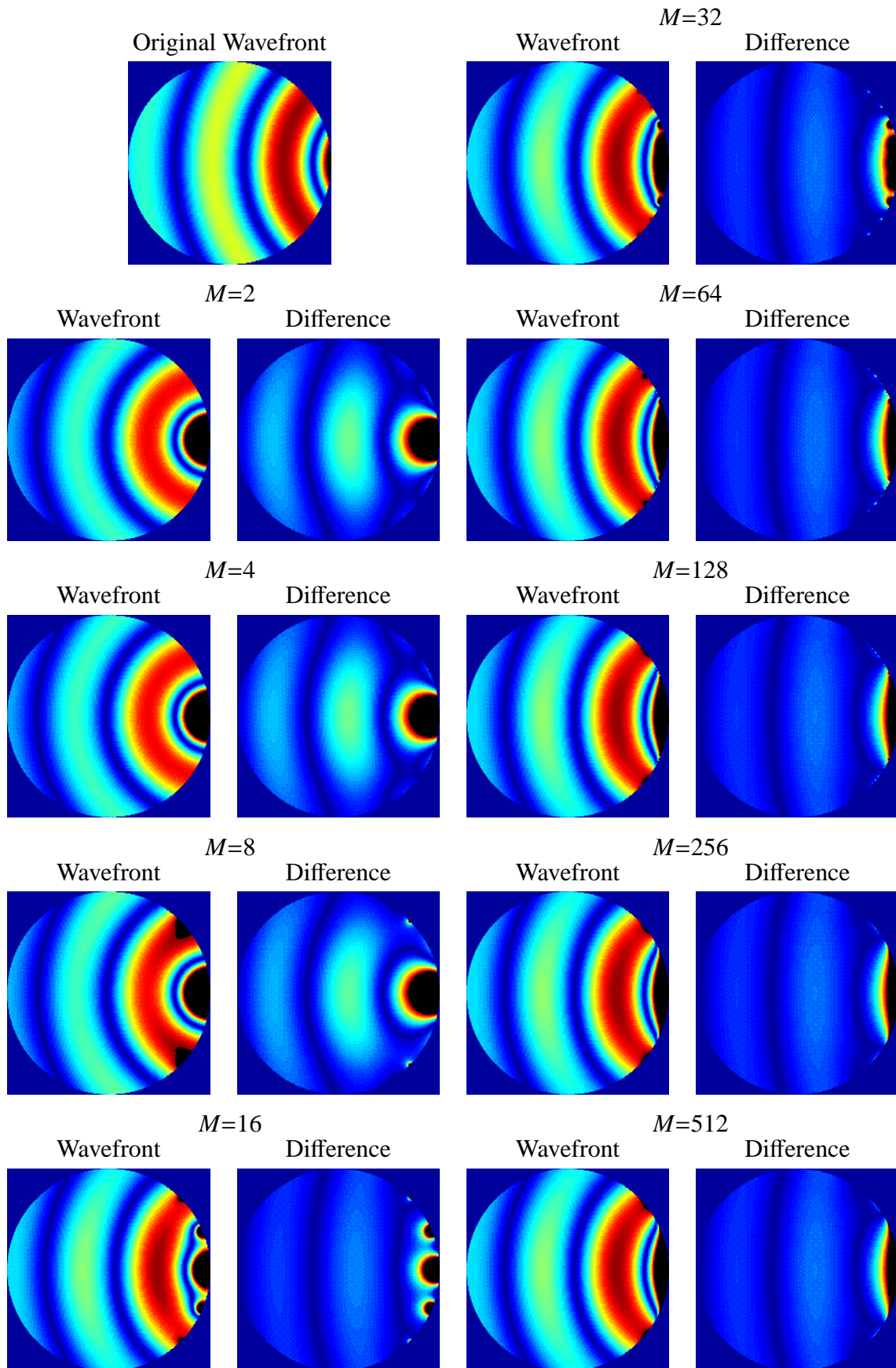


図 A.25 合成した波面の一例 (超指向性,  $f=125$  Hz,  $d=3$  m)

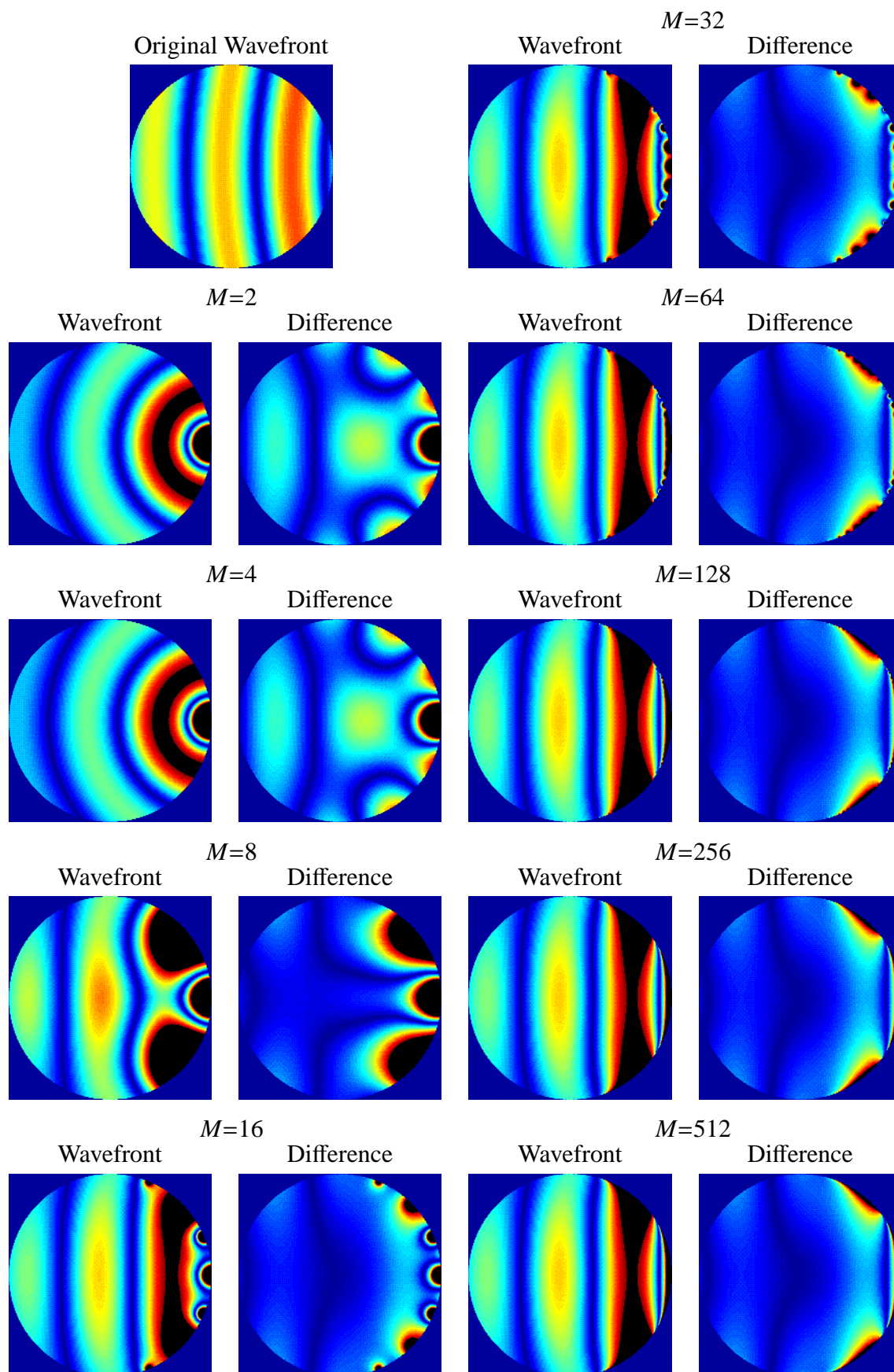


図 A.26 合成した波面の一例 (超指向性,  $f=125$  Hz,  $d=10$  m)



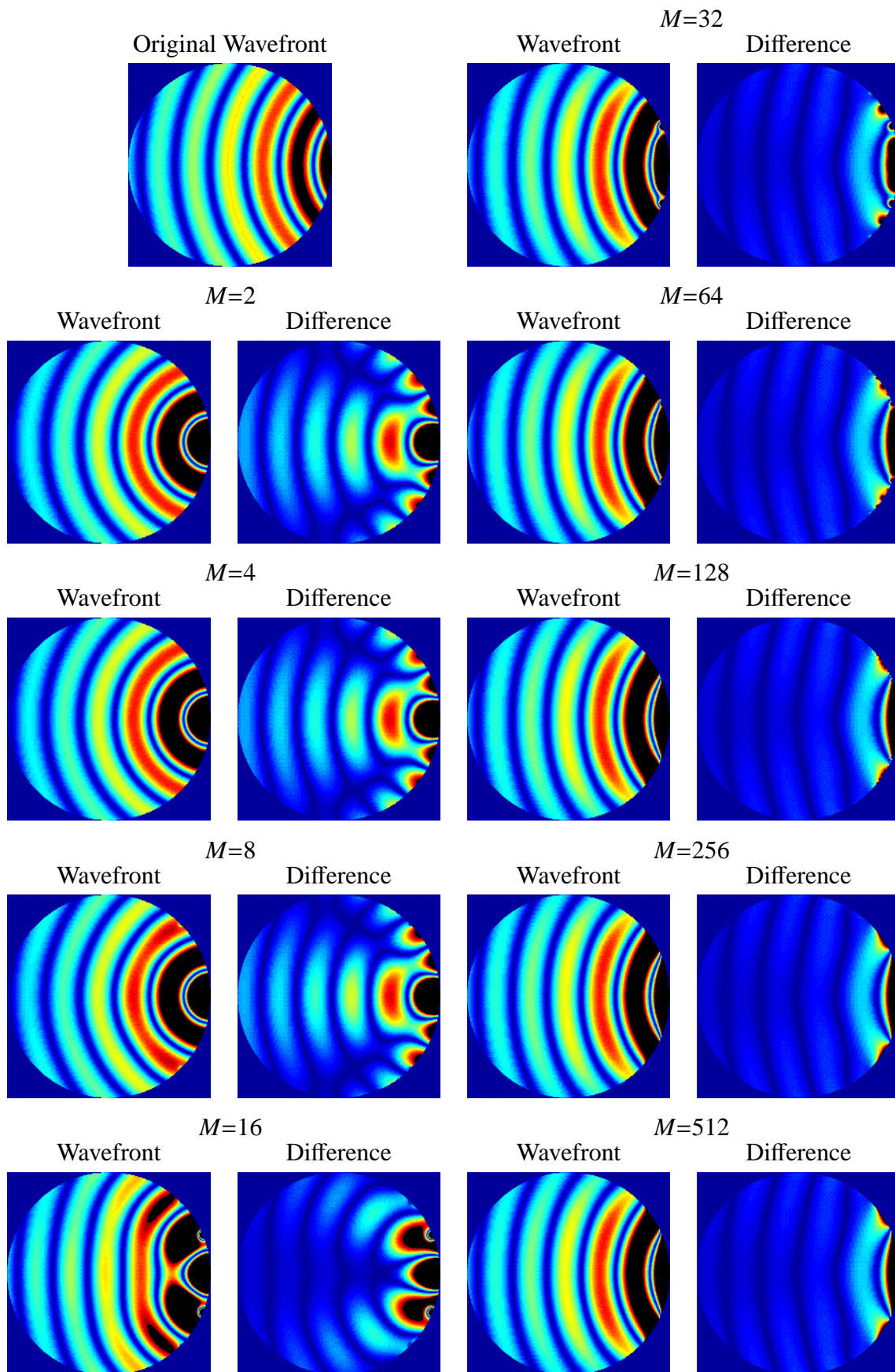


図 A.27 合成した波面の一例 (超指向性,  $f=250$  Hz,  $d=3$  m)



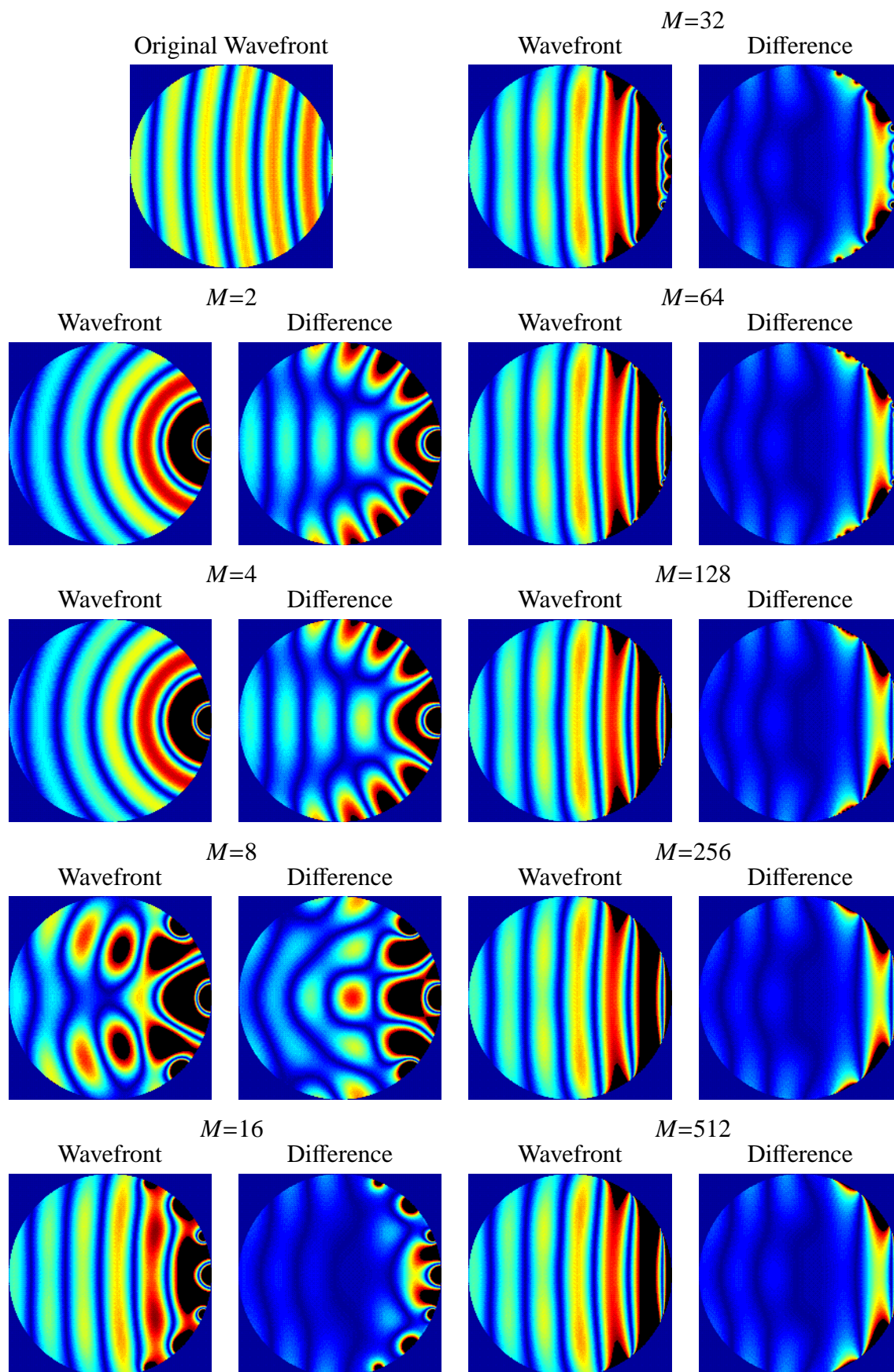


図 A.28 合成した波面の一例 (超指向性,  $f=250$  Hz,  $d=10$  m)

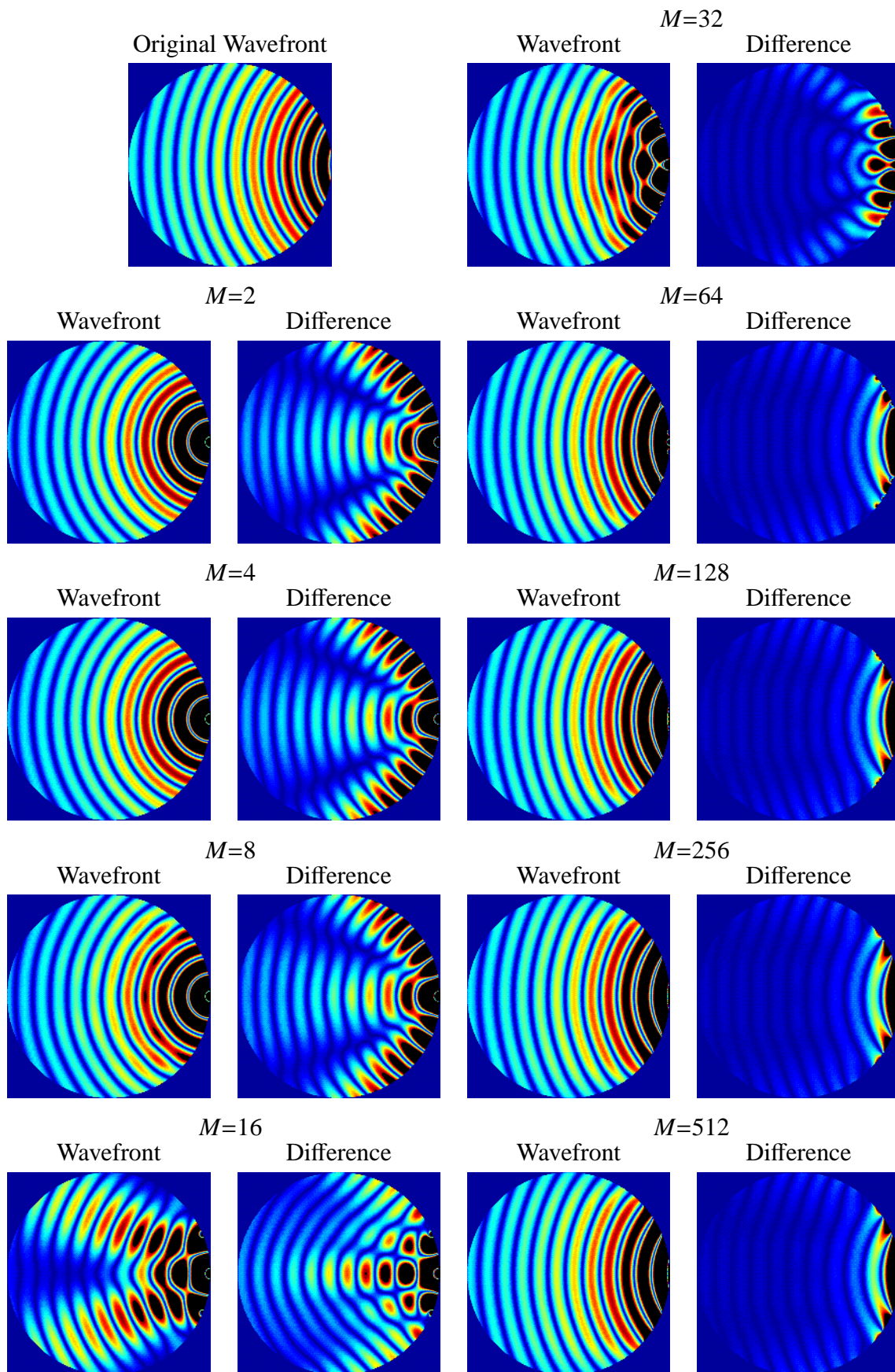


図 A.29 合成した波面の一例 (超指向性,  $f=500$  Hz,  $d=3$  m)



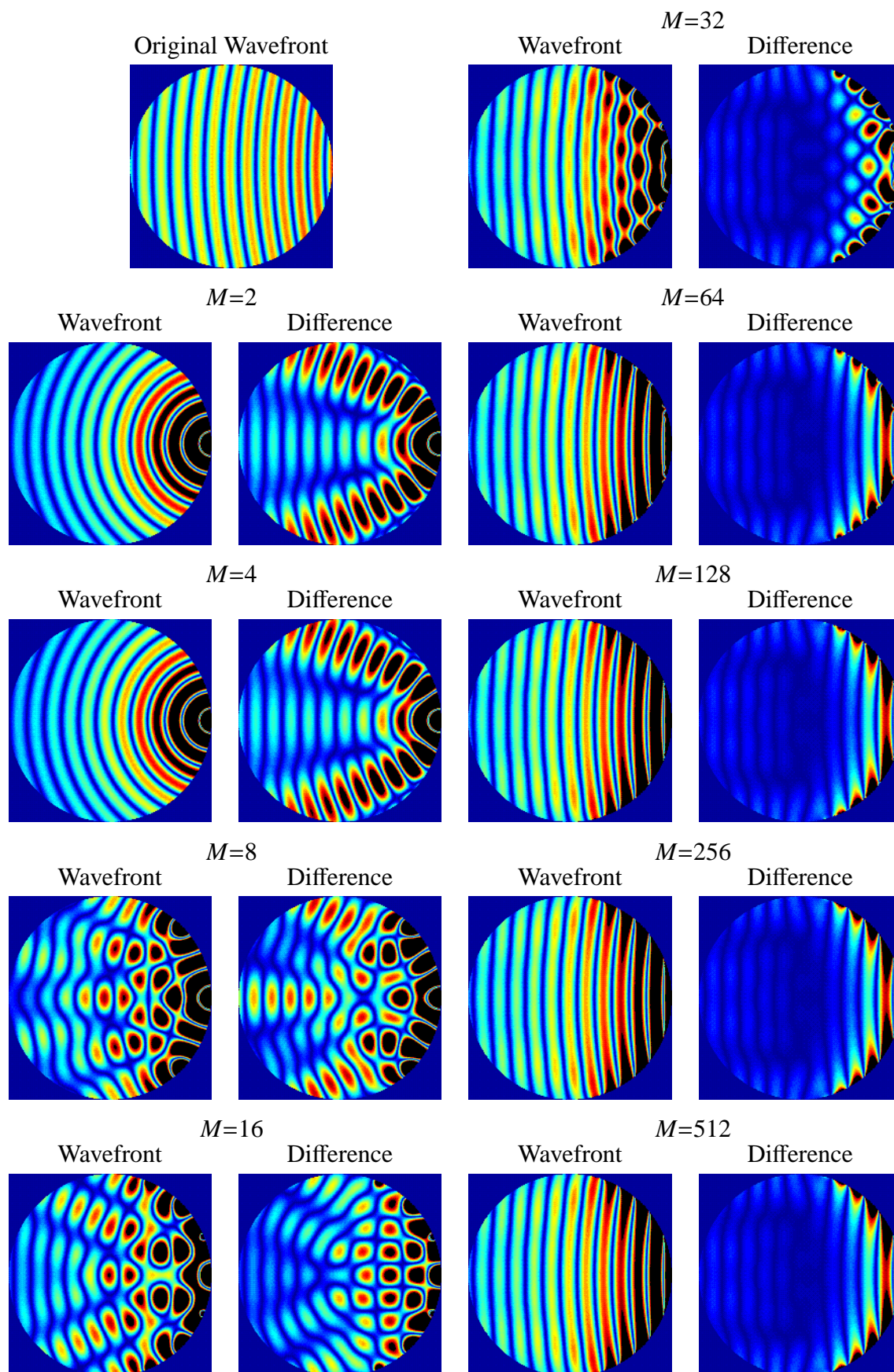


図 A.30 合成した波面の一例 (超指向性,  $f=500$  Hz,  $d=10$  m)

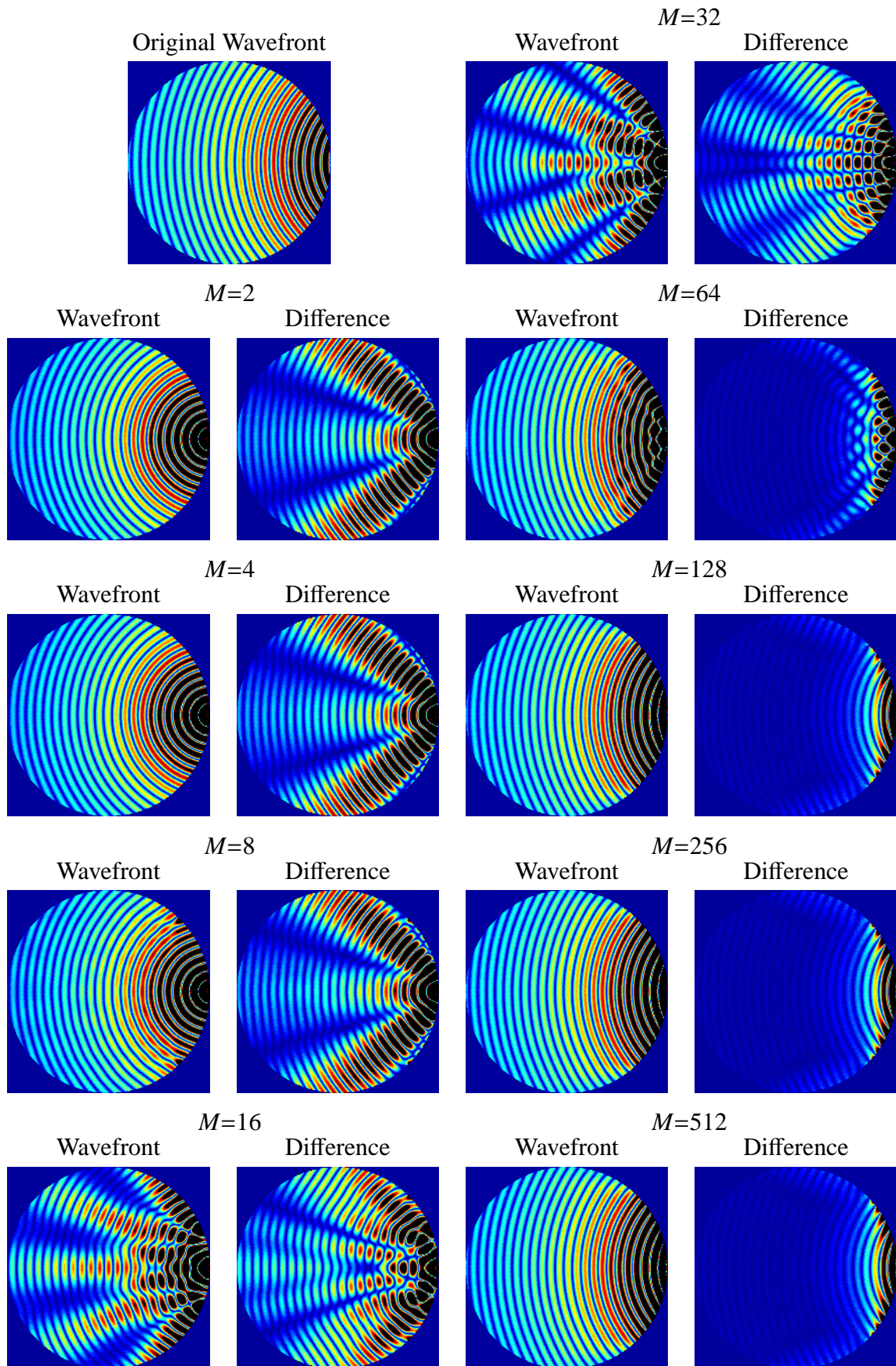


図 A.31 合成した波面の一例 (超指向性,  $f=1000$  Hz,  $d=3$  m)



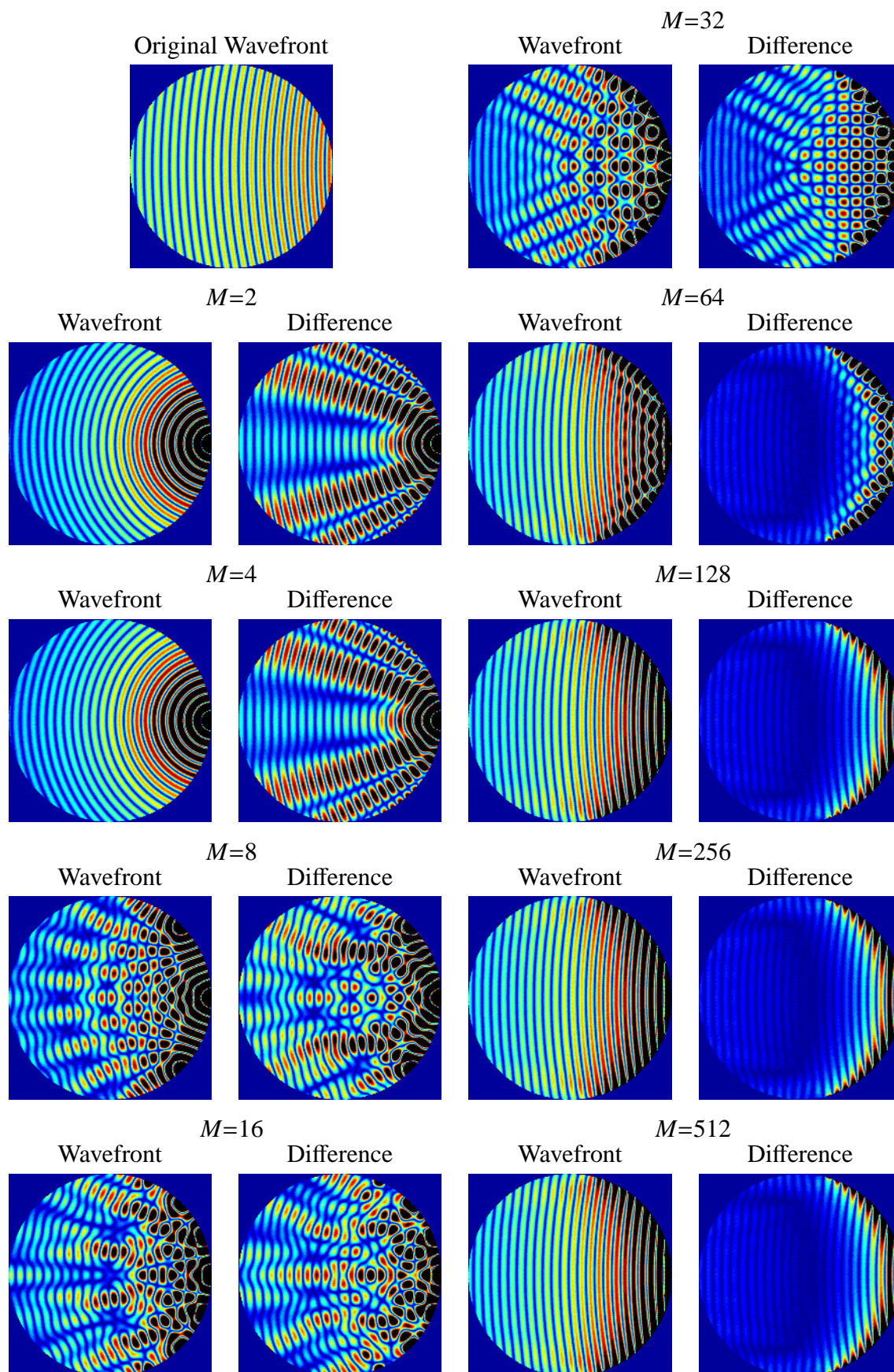


図 A.32 合成した波面の一例 (超指向性,  $f=1000$  Hz,  $d=10$  m)



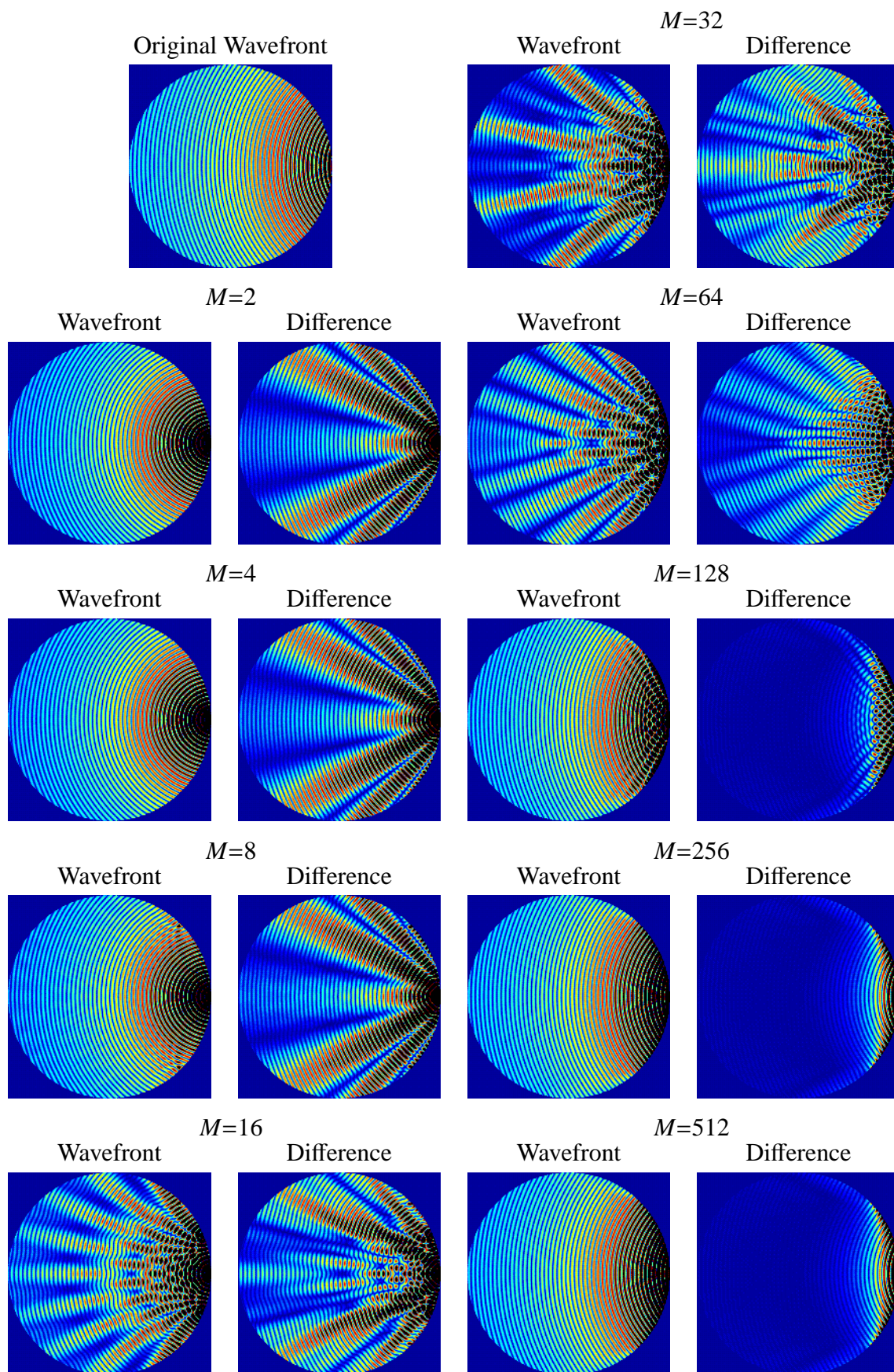


図 A.33 合成した波面の一例 (超指向性,  $f=2000$  Hz,  $d=3$  m)



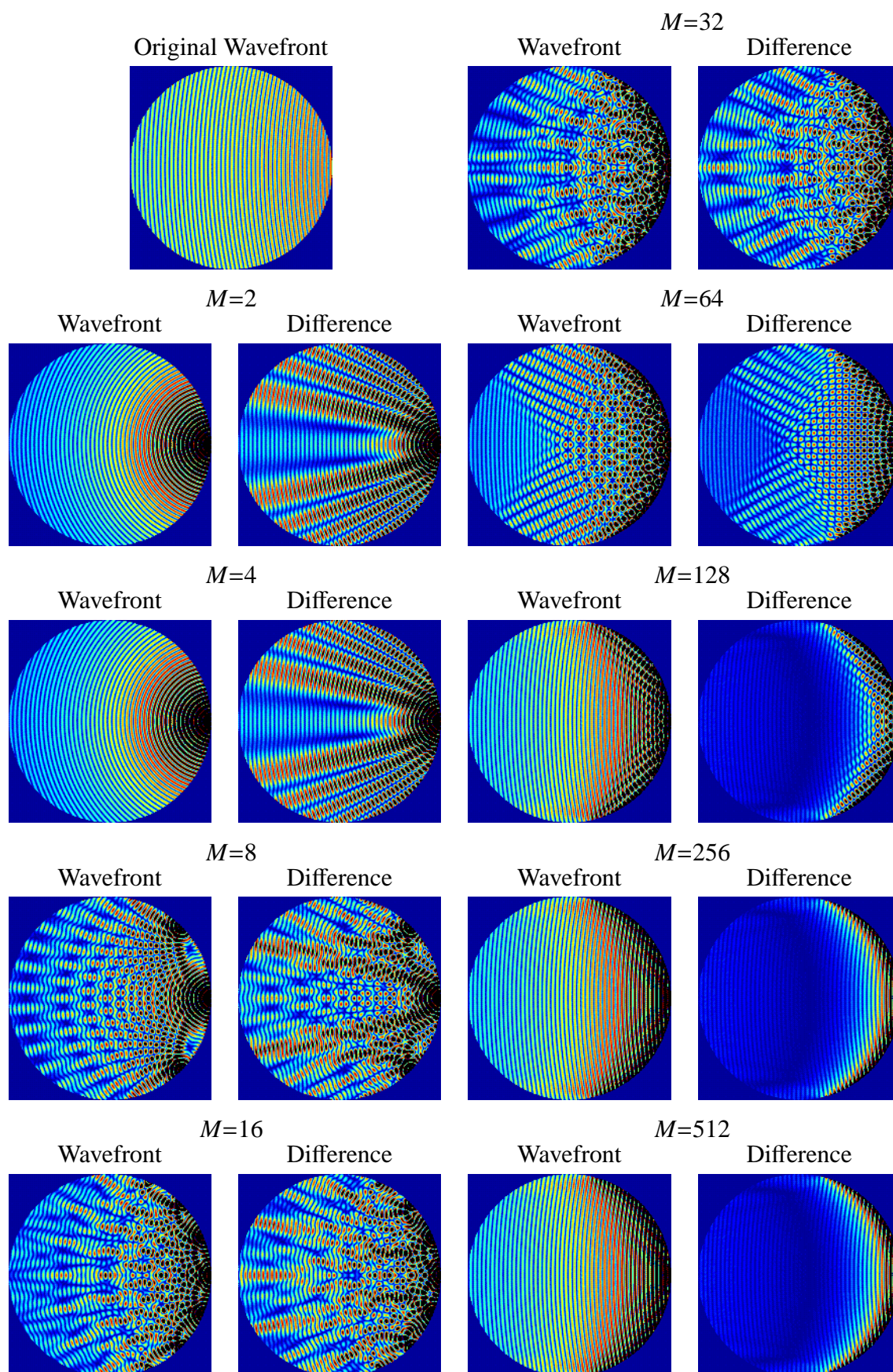


図 A.34 合成した波面の一例 (超指向性,  $f=2000$  Hz,  $d=10$  m)



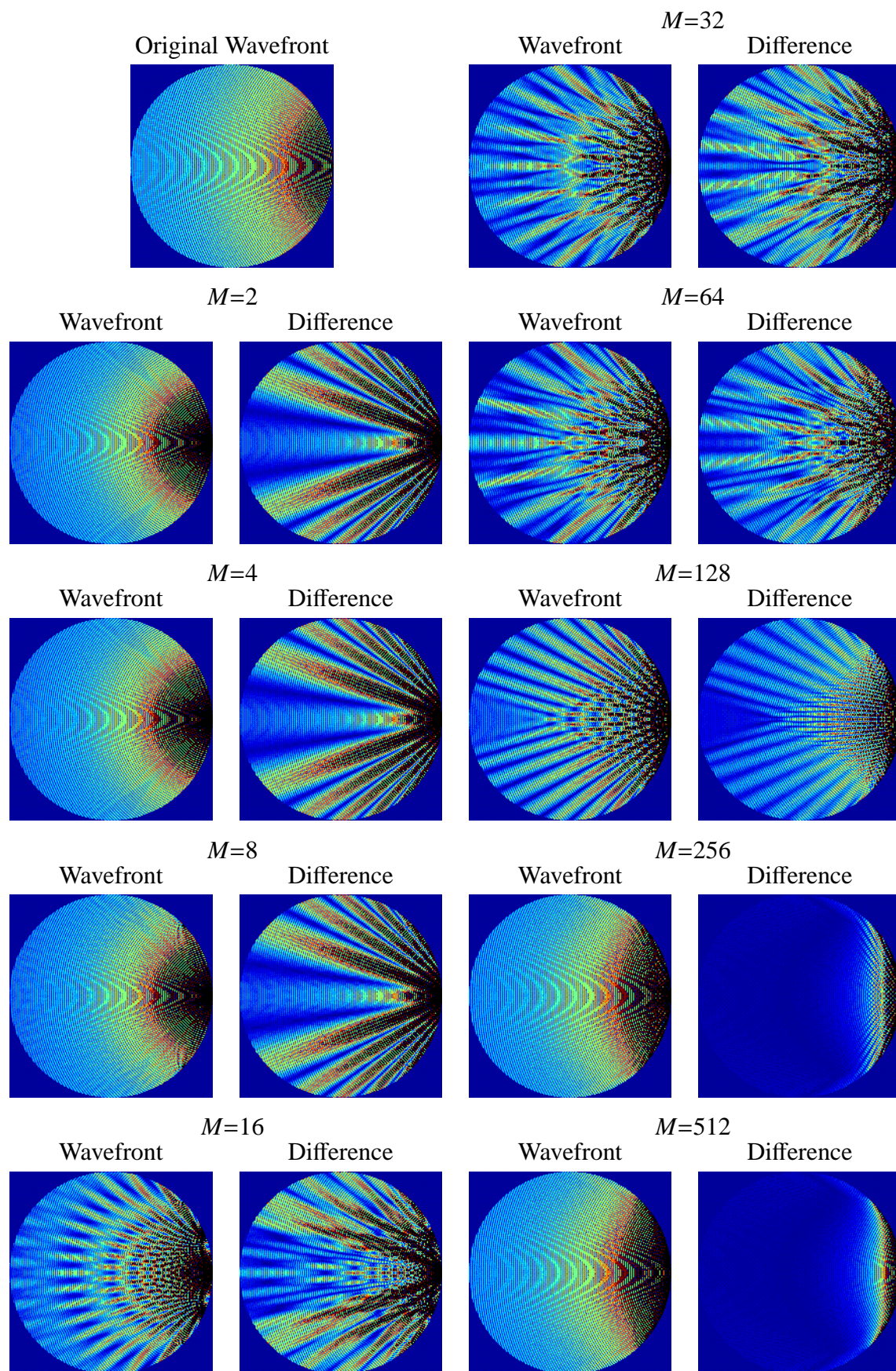


図 A.35 合成した波面の一例 (超指向性,  $f=4000$  Hz,  $d=3$  m)



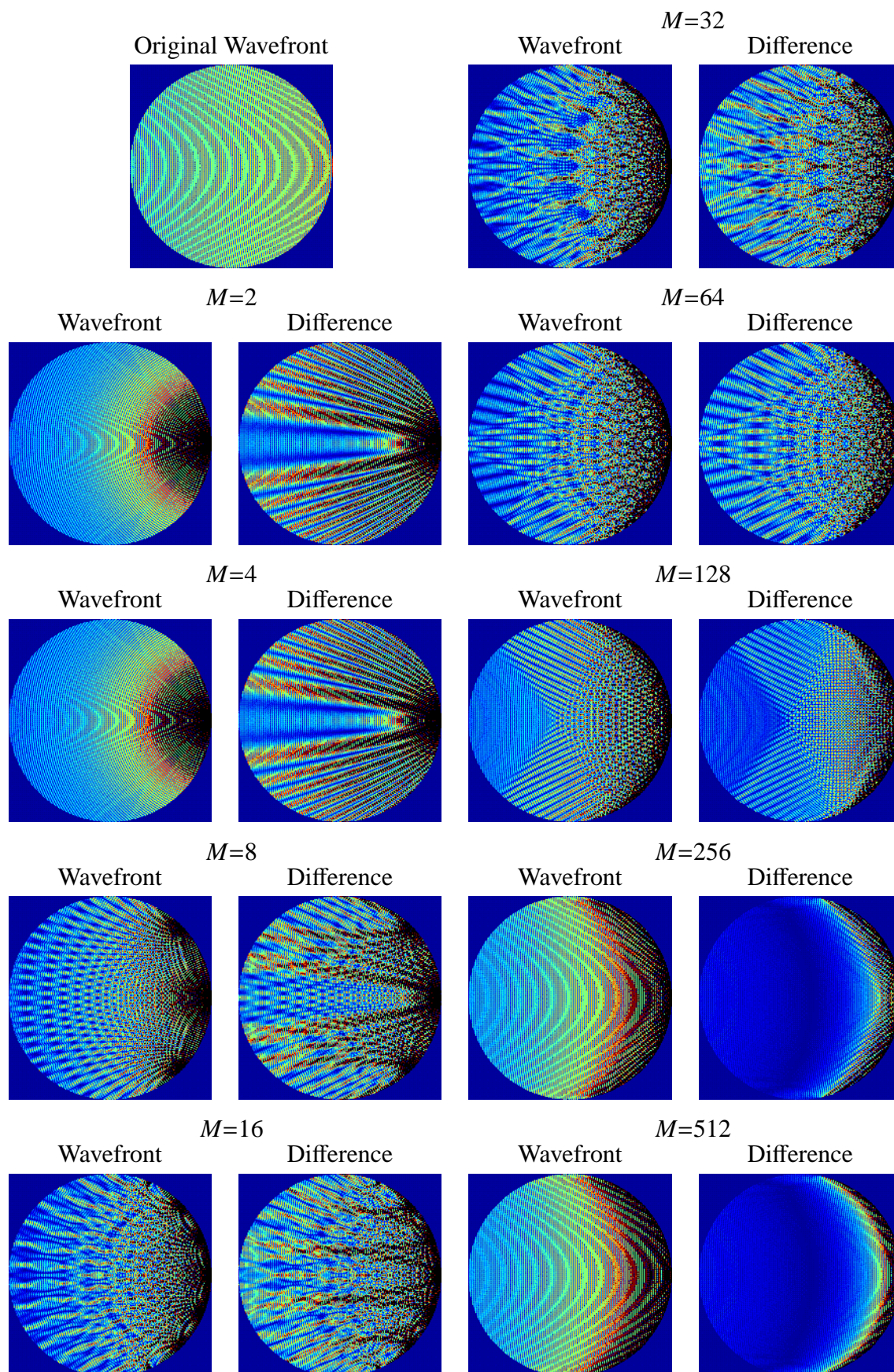


図 A.36 合成した波面の一例 (超指向性,  $f=4000$  Hz,  $d=10$  m)



# 研究業績

## 主な研究業績

### 論文

1. 木村敏幸, 筧一彦, 武田一哉, 板倉文忠, “音場再現におけるマルチチャンネル信号の空間符号化,” 日本バーチャルリアリティ学会論文誌, Vol.8, No.4, pp.433–442, December 2003.
2. 木村敏幸, 筧一彦, 武田一哉, 板倉文忠, “移動音源がある音場の空間符号化 - 音源抽出による伝送量の削減 -,” 日本バーチャルリアリティ学会論文誌, Vol.10, No.1, pp.101–109, March 2005.
3. 木村敏幸, 筧一彦, 武田一哉, 板倉文忠, “波面合成におけるチャンネル数の主観的影響 - 音源が正面付近にある場合 -,” 日本バーチャルリアリティ学会論文誌, Vol.10, No.2, pp. 257–266, June 2005.

### 査読付き国際会議

1. T. Kimura, K. Kakehi, K. Takeda, and F. Itakura, “Spatial Compression of Multi-channel Audio Signals Using Inverse Filters,” Proceedings of International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2002), No. AE-SF01.03, IV, p. 4175, Orlando, FL, May 2002.
2. T. Kimura, K. Kakehi, K. Takeda, and F. Itakura, “Subjective Assessments for the Effect of the Number of Channel Signals on the Sound Field Reproduction Used in Wavefield Synthesis,” Proceedings of the 18th International Congress on Acoustics (ICA 2004), No. Th.P1.18, IV, pp. 3159–3162, Kyoto, Japan, April 2004.
3. T. Kimura, K. Kakehi, K. Takeda, and F. Itakura, “Spatial Coding Based on the Extraction of Moving Sound Sources in Wavefield Synthesis,” Proceedings of International

Conference on Acoustics, Speech, and Signal Processing (ICASSP 2005), No. AE-P5.10, III, pp. 293–296, Philadelphia, PA, March 2005.

## 技術報告

1. 木村敏幸, 筧一彦, 武田一哉, 板倉文忠, “指向性マイクロホンを用いた波面合成法における波面合成精度,” 電子情報通信学会技術研究報告, EA2005-47, pp. 13–18, September 2005.

## 口頭発表

1. 木村敏幸, 筧一彦, 武田一哉, 板倉文忠, “主成分分析によるマルチチャンネル音響符号化の検討,” 日本音響学会春季研究発表会, No. 2-7-15, 筑波大学, March 2001.
2. 木村敏幸, 筧一彦, 武田一哉, 板倉文忠, “空間情報を利用したマルチチャンネル音響信号の符号化と音源定位,” 日本音響学会秋季研究発表会, No. 3-Q-28, 大分大学, October 2001.
3. 木村敏幸, 筧一彦, 武田一哉, 板倉文忠, “逆フィルタを用いた空間音響圧縮技術の拡がり感,” 日本音響学会春季研究発表会, No. 2-4-17, 神奈川大学, March 2002.
4. 木村敏幸, 筧一彦, 武田一哉, 板倉文忠, “逆フィルタを用いた音場空間符号化方式の主観評価,” 日本音響学会秋季研究発表会, No. 3-5-16, 秋田大学, September 2002.
5. 木村敏幸, 筧一彦, 武田一哉, 板倉文忠, “移動音源に対するマルチチャンネル音場空間符号化の検討,” 日本音響学会春季研究発表会, No. 1-8-13, 早稲田大学, March 2003.
6. 木村敏幸, 筧一彦, 武田一哉, 板倉文忠, “波面合成による音場再現における制御点数の影響 - 方向感の検討 -,” 日本音響学会秋季研究発表会, No. 2-5-7, 大同工業大学, September 2003.
7. T. Kimura, K. Kakehi, K. Takeda, and F. Itakura, “Subjective Assessment for the Number of Channel Signals to Realize Sound Field Based on Wavefield Synthesis,” The 146th Meeting of the Acoustical Society of America, No. 2aPP9, Austin, TX, November 2003.
8. 木村敏幸, 筧一彦, 武田一哉, 板倉文忠, “波面合成法による音場再現におけるチャンネル数の主観評価 - 空間印象の検討 -,” 日本音響学会春季研究発表会, No. 3-P-10, 神奈川工科大学, March 2004.

9. 木村敏幸, 箕一彦, 武田一哉, 板倉文忠, “位置情報による移動音源再現手法の主観的影響,” 日本音響学会秋季研究発表会, No. 3-P-29, 琉球大学, September 2004.
10. 木村敏幸, 箕一彦, 武田一哉, 板倉文忠, “波面合成法による立体音響においてチャンネル数が音場知覚に及ぼす影響,” 第2回音空間・HRTF徹底討論ワークショップ, No. P-18, 宮城蔵王, July 2005.
11. 木村敏幸, 箕一彦, 武田一哉, 板倉文忠, “指向性マイクロホンを用いた波面合成法の理論的検討,” 日本音響学会秋季研究発表会, No. 3-Q-10, 東北大学, September 2005.

## その他の研究業績

### 論文

1. N. Inoue, T. Kimura, T. Nishino, K. Itou, and K. Takeda, “Evaluation of HRTFs Estimated Using Physical Features,” *Acoustical Science and Technology*, Vol. 26, No. 5, pp. 453–455, September 2005.

### 査読付き国際会議

1. T. Kimura, W. Mizuno, T. Nishino, K. Itou, and K. Takeda, “Sound Field Auralization System in Free Listening Positions Using Wave Field Synthesis and Head Related Transfer Functions,” *Proceedings of Forum Acusticum 2005 (FA 2005)*, No. 617-0, pp. 2511–2516, Budapest, Hungary, August/September 2005.

### 技術報告

1. 水野渉, 木村敏幸, 西野隆典, 武田一哉, “マイクロホンアレーを用いた自由聴点音場再生システムに関する理論的検討,” 電子情報通信学会技術研究報告, EA2004-118, pp. 7–12, January 2005.

### 口頭発表

1. 細江誠一郎, 木村敏幸, 西野隆典, 武田一哉, “スパーク音源を用いた頭部伝達関数の測定,” 平成16年度電気関係学会東海支部連合大会, No. O-209, 名古屋工業大学, September 2004.

2. 井上直哉, 木村敏幸, 西野隆典, 伊藤克巨, 武田一哉, “身体的特徴量から推定した頭部伝達関数の評価,” 日本音響学会秋季研究発表会, No. 3-P-5, 琉球大学, September 2004.
3. 水野渉, 木村敏幸, 西野隆典, 武田一哉, “頭部伝達関数を用いた自由聴点立体音響の構築と評価,” 日本音響学会春季研究発表会, No. 3-Q-29, 東京農工大学, March 2005.