

音場制御のための信号処理方式の研究

電気通信大学 情報システム学研究科
情報ネットワーク学専攻
9451034 西口 敏行

指導教官

阪口 豊 助教授

安田 稔 教授

出澤 正徳 教授

平成8年2月15日 提出

目次

1	はじめに	1
§1.1	研究の背景	1
1.1.1	従来の音場制御とその問題点	1
1.1.2	スピーカウォールによる音場の能動制御	2
§1.2	本研究の目的と手法	5
§1.3	本論文の構成	7
2	ランダム・ブロック・シャフル法による残響音の生成	8
§2.1	ランダム・ブロック・シャフル法	8
§2.2	主観評価実験	10
2.2.1	実験目的と評価項目	10
2.2.2	予備実験	10
2.2.3	実験計画と条件	13
2.2.4	実験結果	15
§2.3	結果の考察	15
3	インパルス応答の時間伸縮による残響音の生成	31
§3.1	インパルス応答の同期型サンプル・レート変換による時間軸一様伸縮	31
§3.2	インパルス応答の非同期型サンプル・レート変換による時間伸縮	32
§3.3	エンベロープ補正	32
§3.4	試聴実験の結果と考察	35
3.4.1	一様伸縮法	35
3.4.2	一定周期で時間伸張・圧縮を繰り返す方法	36
3.4.3	後部残響音初期とそれ以後で伸縮率を変える方法	37
§3.5	試聴実験のまとめ	38
4	後部残響音の時間シフトによる応答の生成	39
§4.1	後部残響音の時間シフト	39
§4.2	試聴実験の結果と考察	39

5 おわりに	41
§5.1 本研究のまとめ	41
§5.2 今後の課題	41
謝辞	43
参考文献	44

第1章

はじめに

§1.1 研究の背景

コンサートホールや教会など，広い空間内の音場の響きや臨場感などを再現・創造するための音場制御の研究は，建築音響，家庭用音響機器，放送など広い分野での応用が期待され，近年，注目されている研究である。

例えば，多目的ホールの場合，オーケストラのコンサートには残響時間の長い豊かな残響が望まれるが，講演会などのスピーチでは明瞭度を保つことが重要なため，比較的短い残響時間が望まれる。こうしたときに，音場を能動的に制御できれば，ホールの使用目的にあった音場を形成することができる。また，一般家庭においては，従来の2チャンネル・ステレオでは限界のある，空間的な広がり感を再生できるシステムとして，音響機器や放送などに応用が考えられる。

まず，本研究の主題に入る前に，従来から提案されている音場制御の方法とその問題点を述べる。

1.1.1 従来の音場制御とその問題点

既に提案されている音場制御の方法として，図1.1のように1つの聴取位置を取り囲むようにスピーカを多数配置し，コンサートホールの1つの客席に届く反射音をこれらのスピーカから方向別に再生する方法が報告されている[1]。しかし，この方法は，少数のスピーカを離散的に配置するため，聴取位置がずれると方向別の反射音のバランスが崩れてしまう。したがって，聴取位置が1点に限定されるという問題がある。

さらに広い聴取範囲で原音場を再現する方法として，図1.2に示す方法が提案されている[2][3]。この方法は，コンサートホール内の音波の波面を，仮想的なリスニングルー

ムの壁面上にアレイ状に配置されたマイクロフォンで収録し、これをマイクロフォンと同じ配置のアレイ上のスピーカから再生して、リスニングルーム内でコンサートホール内と同様な波面を再生するものである。

この方法は、収録に必要なマイクロフォンの数が多いうえ、再生に用いるソースがアレイ状のマイクロフォンで収録されたものに限定されてしまう。また、基本的に、マイクロフォンの数だけ記録・伝送チャンネルが必要であるなど現実性に乏しいと考えられる。

1.1.2 スピーカウォールによる音場の能動制御

1.1.2.1 スピーカウォールの特徴

新しい音場の制御技術の一つとして、スピーカウォール(部屋の壁一面にアレイ状に配列した多数の小型スピーカ, 図 1.3)による音場制御法が研究されている [4][5]。この方法は、部屋の壁一面に設置した多数の小型スピーカを用いることにより、従来の方法では困難な広い聴取範囲での良好な音場の再現・創造を目指すものである。

この方法は、前述の方法と比較して次の特徴を持っている。

まず第1に、反射音がスピーカウォールを構成する多数のスピーカユニットから放射されるため、スピーカウォールに近づいても急激に反射音のみが大きくなることなく、広い聴取範囲で自然な音場が得られる。これは、反射音のエネルギーがスピーカウォールを構成する多数の小型スピーカユニットに分散されるため、少数の大型スピーカを用いる場合よりも、再生音場内の音圧分布が平坦になり、より自然な音場に近づくためと考えられる。

第2に、反射音は図 1.4のように、原音場の伝達特性(インパルス応答)を収録した

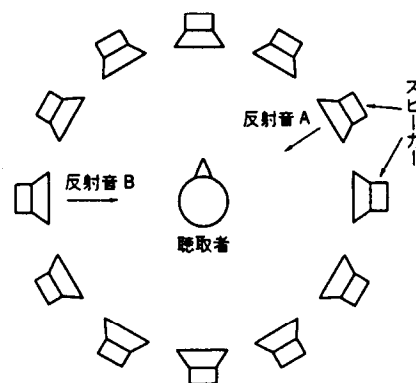


図 1.1: コンサートホールの反射音を方向別に再生する方法の概念図

音源に畳み込むことで得られるため、伝達特性のみを伝送すればよいことになり、記録・伝送に必要なチャンネル数が少なくてすむ。

1.1.2.2 スピーカウォールによる反射音の再生法

図1.5のように、例えば、コンサートホール(原音場)の中に仮想的なリスニングルーム(再生音場)を考える。このとき、原音場内の仮想的な再生音場における反射音の状態を再生音場において再現できれば、再生音場全体において原音場が実現されると考えられる。

一般に残響のある空間のインパルス応答は、直接音が到来した後約0.1秒までに到来する初期反射音と、それ以降に到来する残響音とに分けて考えることができる。

このうち、初期反射音については、その微細な構造(反射音のレベル、遅延時間、音の方向)が聴感上重要であると考えられている。従って、スピーカウォールでは初期反射音の1つ1つを音線追跡法などのシミュレーション手法によって求め、音の方向に対応する方向に設置したスピーカウォールから再生する。さらに、スピーカウォール

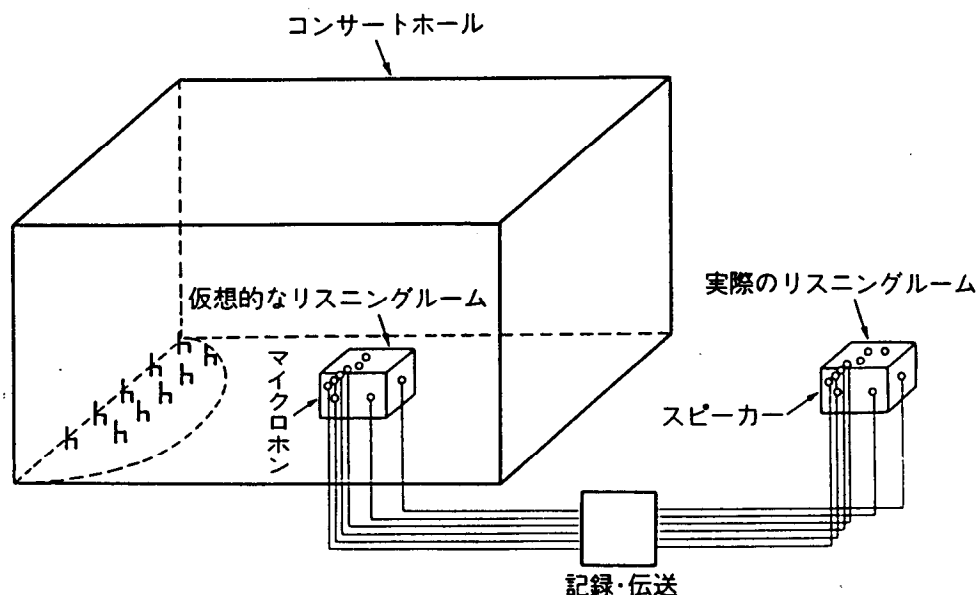


図1.2: 仮想的なリスニングルームを用いる方法

ルを構成する各スピーカユニット相互の位相関係を制御して、初期反射音の波面の制御も行う。

一方、残響音については、反射音の1つ1つが聴感には寄与せず、残響時間周波数特性という統計的な量が聴感上重要であると考えられている。また、コンサートホール等の残響音はあらゆる方向から無相関な音波が到来していると考えられているが（この状態を拡散音場という）、複数の音波の相互相関の低さは聴感上の広がり感にとって重要である。

従ってスピーカウォールでは、残響時間周波数特性に着目し、また原音場の拡散音場性を仮定して制御を行う。このためには、各スピーカから音色（残響音の振幅周波数特性）が等しい残響音を再生し、さらに、残響音が特定の方向性を持たぬよう相互相関を低くなるようにする。

なお、各スピーカから無相関な残響音を再生することは位相干渉防ぐ効果もあり、聴

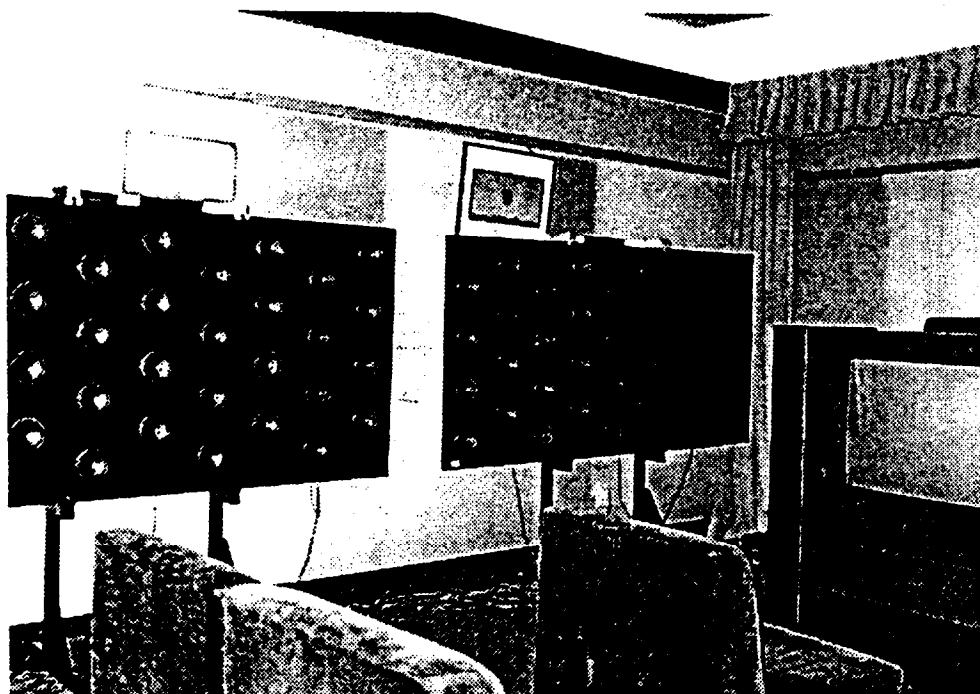


図 1.3: スピーカウォールシステム

取位置による音色の変化を防ぐ役割をも果たす。

§1.2 本研究の目的と手法

本研究では、スピーカウォールによる反射音の再生において、特に聴感上の広がり感や残響音の音色に大きな影響を持つ原音場のインパルス応答の後部残響音の部分の処理を対象とすることにした。

前節で、スピーカウォールによって良好な音場再現を得るには、原音場の残響音の聴感的特徴を保存し、かつ、低相関な残響音をもったインパルス応答が多数必要であることを述べた。しかし、こうした特性を満たすインパルス応答を、原音場での測定によって多数得ることは非現実的で、かつ、技術的にも困難である。

そこで、本研究の目的は、前述の特性を満たす多数のインパルス応答を人工的に生

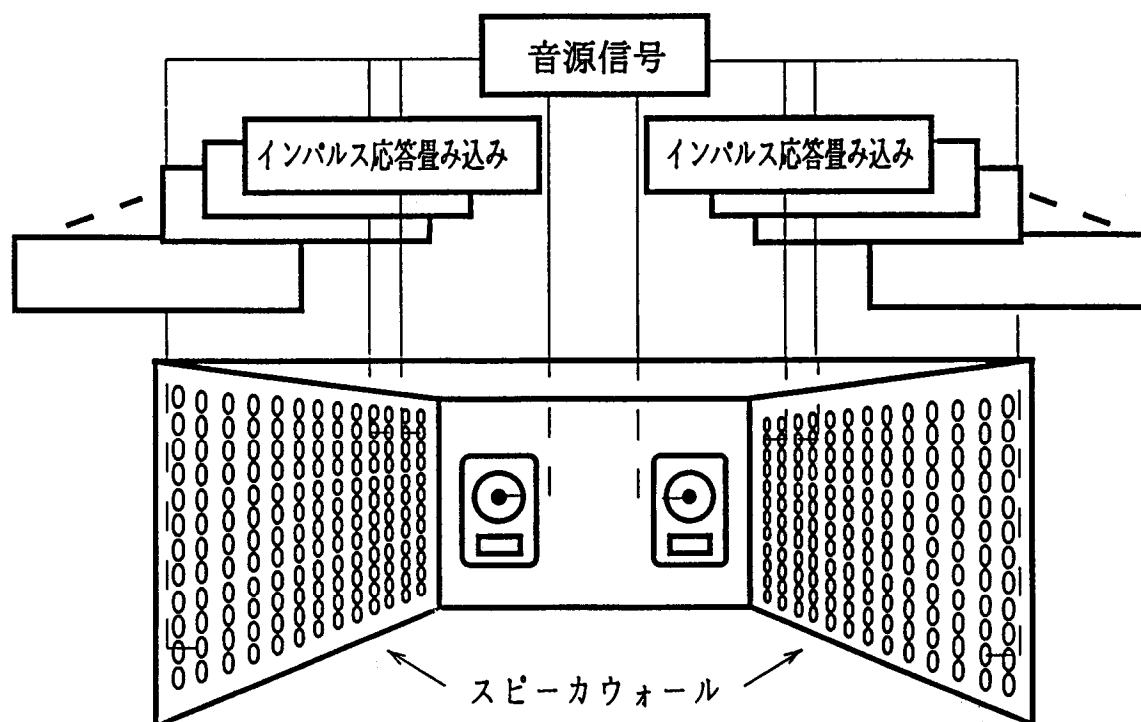


図 1.4: スピーカウォールによる音場の能動制御

成することにある。具体的には、TSP*¹やM系列*²を利用して高いSN比で測定された1つないし少数の原音場のインパルス応答をリファレンスとして、この応答の後部残響音部分を操作することにより、リファレンスの残響音のもつ聴感的特徴を保存した多数の低相関なインパルス応答を生成する信号処理方式を求めることにした。

信号処理の手法として、主にインパルス応答の時間軸を操作し、新たな応答を生成する方法を検討した。また、評価は、応答を畳み込んで残響音を作成し、主観評価実験により行った。

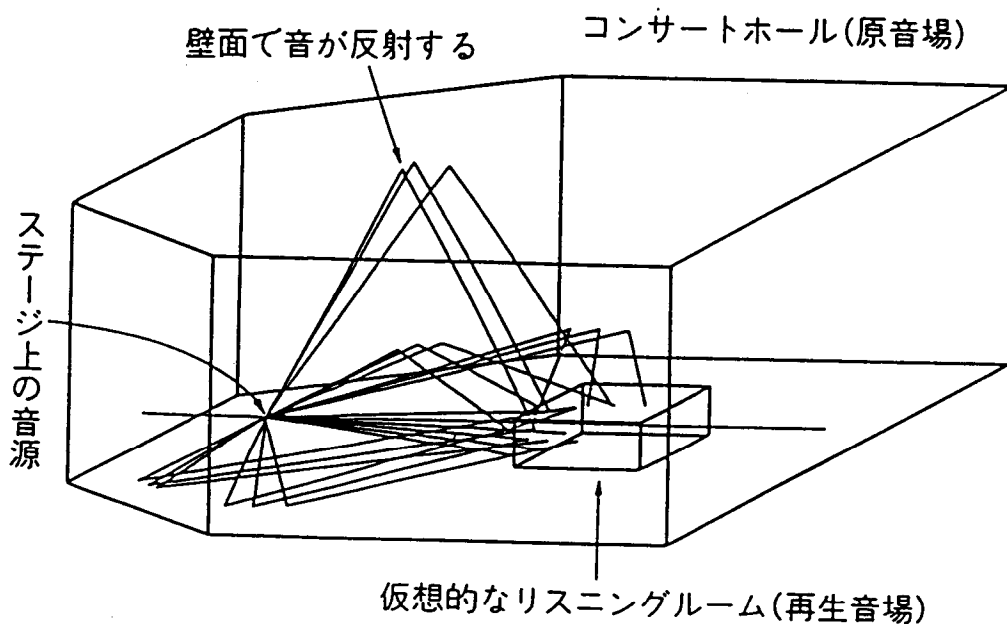


図 1.5: 仮想的なリスニングルームに到達する反射音の伝達経路の例

*¹Time-Stretched Pulse [6].

*²Maximal length sequence [7].

§1.3 本論文の構成

まず，第2章では，生成可能な応答の数が，無制限に存在することを目標として，時間軸をブロックに分割し，このブロックを一様分布を用いて再配置することにより，新たな応答を生成する方法を提案する．また，この方法の主観評価実験についても述べる．

第3章では，生成できる応答の数には制限があるが，残響音の保存性が高く，処理も簡単な方法として，残響音の時間伸縮を行って応答を生成する信号処理方式について述べる．

第4章では，第3章の信号処理を更に簡便化した方法として，後部残響音を時間シフトして応答を生成する信号処理について述べる．

最後に，第5章に，各信号処理方式の特徴と主観評価の結果についてまとめ，今後の課題について述べることにする．

第2章

ランダム・ブロック・シャフル法による 残響音の生成

§2.1 ランダム・ブロック・シャフル法

実際の音場での反射音は、図1.5のように壁や天井での反射をくり返す際に帯域制限を受け、この帯域制限の応答が重なり合っている受音点における応答を形成すると考えられる。ある受音点で観測する方向を変えたときの応答や、受音点の位置を大きくずらした場合の応答の相互相関はかなり低いですが、聴感的な残響音の音色は保存されていることは前に述べた。したがって、こうした応答の変化が解析的に分かれば理想であるが、様々の伝達経路を経て受音点に到達した反射音が、重なり合って構成されている原音場のインパルス応答から、個々反射音の持つ応答を分離して調べることは非常に困難である。そこで、リファレンスのインパルス応答を、ある時間幅をもったブロックに分け、このブロック内で音色を保存させ、相互相関を下げるためにブロックの時間的な並び方を操作することで応答を生成しようと考えた。

具体的には、図2.1のように、リファレンスの応答の後部残響音部分を一定幅をもったブロックに分割し、あるブロックの前後のブロックから一様分布で選んだブロックを、新しい応答のデータとして再配置してゆく。再配置の際には、ブロック間の繋がりが滑らかになるように、ブロック内のデータに窓をかける。ブロックは重なりを許し、重なった部分のデータは加算する。初期反射音部分は、音源の定位保持のためリファレンスの初期反射音をそのまま使用する。本研究では、この方法をランダム・ブロック・シャフル法と呼ぶことにする。パラメータとして、

- (1) 初期反射音の時間幅
- (2) 分布の種類
- (3) 窓の形状
- (4) ブロック間隔
- (5) ブロック幅
- (6) 分布幅

が考えられる。

(1)は一般にホールの容積や形状によらず80-100msといわれる量である。(2)から(6)は、それぞれ残響音の性質に影響を与える量であると考えられるが、こうした複雑な非線形処理の結果は解析的に求め難く、また、聴感との関係を定量的に結びつけることも難しい。そこで、主観評価実験を行って、これらのパラメータが聴感に与える影響を調べることにした。

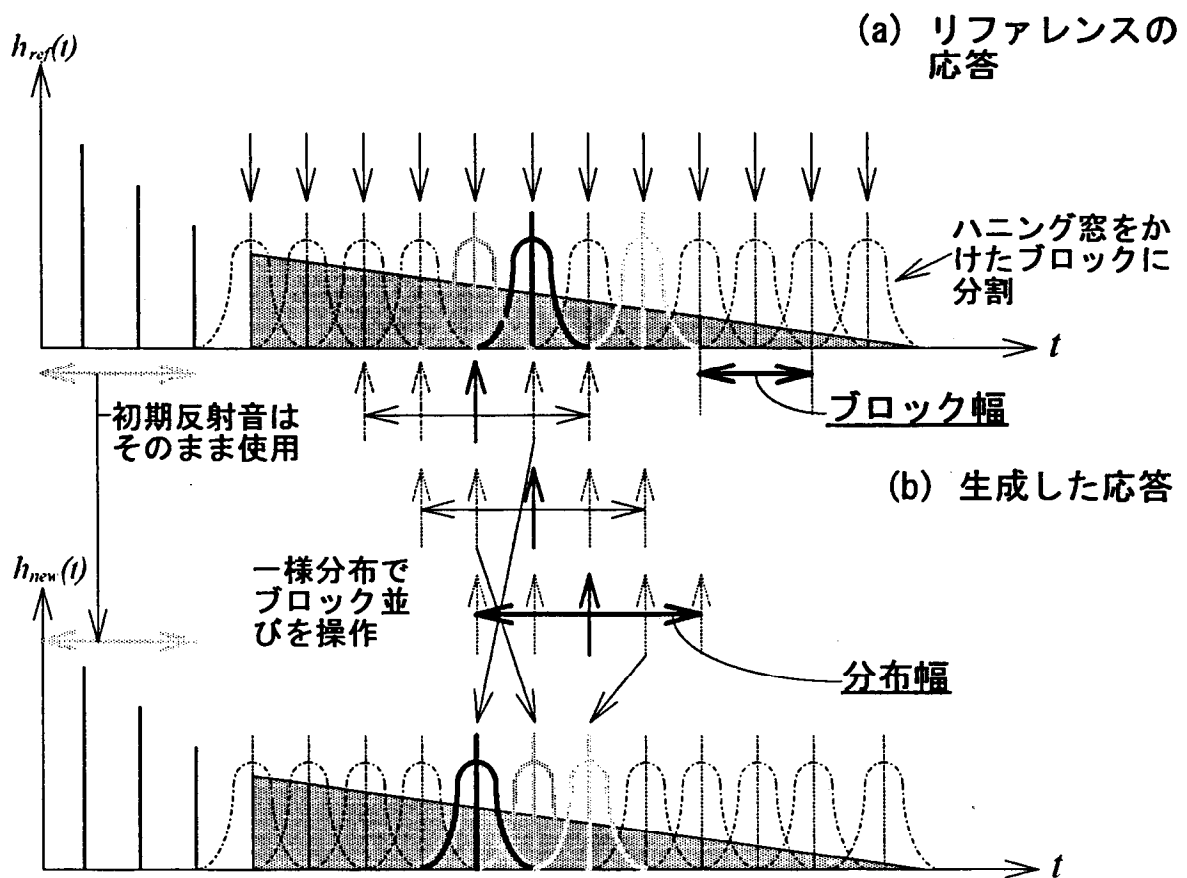


図 2.1: ランダム・ブロック・シャフル法

§2.2 主観評価実験

2.2.1 実験目的と評価項目

主観評価実験の目的は、リファレンスの応答から、ランダム・ブロック・シャフル法を用いて新たな応答を生成する際に、リファレンスの応答による残響音と比較評価し、パラメータの違いが聴感に与える影響を明らかにすることにある。

残響音を評価する際には、

- (a) 残響音の音色
- (b) 残響音の広がり感、定位
- (c) 異なる残響音が混ざり合ったときの音色

という3つの評価項目を考慮することにした。

(a)は、最も基本的な評価項目で、不可欠なものである。評価方法としては、生成の際に用いるリファレンスの応答の片チャンネルの応答と、生成した応答の比較になるので、モノフォニック再生により行う。

(b)は、複数の残響音を再生して音場を構成した際に、聴感上の広がり感を得るために重要であり、ステレオ再生で評価することが望ましい。

(c)は、多数のインパルス応答からなる再生音を多チャンネル再生して音場を形成する際には重要なポイントである。処理の仕方によっては、個々の残響音は良好でも、複数の残響音が混ざり合ったときに非常に不自然な音になってしまう危険がある。

ところで、これら項目別に、様々なパラメータで実験することは事実上困難である。そこで、予備実験を行い、効率的な実験を行えるように計画を立てなければならない。

2.2.2 予備実験

予備実験は、図2.2に示す装置を用いて行った。ワークステーション(WS)で応答の生成処理を行い、これを、たたみ込み器の制御を行うPC/AT互換機を通じてたたみ込み器に送る。さらに、たたみ込み器に、コンパクトディスク(CD)などから無響室収録の評価音源を流すと、リアルタイムでたたみ込みが実行され、残響音が付加される。こうして得た音を場合によってはミキサーでミキシングするなどして、スピーカやヘッドフォンで試聴した。

まず、前節に示した(1)から(6)のパラメータについて検討を行った。

(1) 初期反射音の時間幅については、リファレンスとしている応答波形から 58ms とした。(2) 分布の種類は、低相関な残響音を多数得るのが目的のため、一様分布を採用した。(3) 窓の形状については、ハニング、ブラックマン、正弦波の半波などを試聴した結果、聴感的に大差がないのでハニング窓を採用した。(4) ブロック間隔は、ハニング窓の形状から、データが滑らかに繋がるよう(5) ブロック幅の 1/2 とすることにした。したがって、主観評価実験では、(5) ブロック幅と(6) 分布幅が聴感に与える影響を調べればよい。

次に、評価項目についてだが、予備実験から、(a)については、音色に問題のある応答はミキシングしても音色が良くない傾向があるので、(c)に含めて判断できると考えた。

(b)について、ブロック幅、分布幅の広範囲にわたって、リファレンスの応答がもつ広がり感と比較し、十分な広がり感を持っていることが確認された。また、初期反射音部分はリファレンスのデータをそのまま使用しているため、定位についても、ほぼ問題が無かった。したがって、(b)についての評価実験は割愛できると判断した。

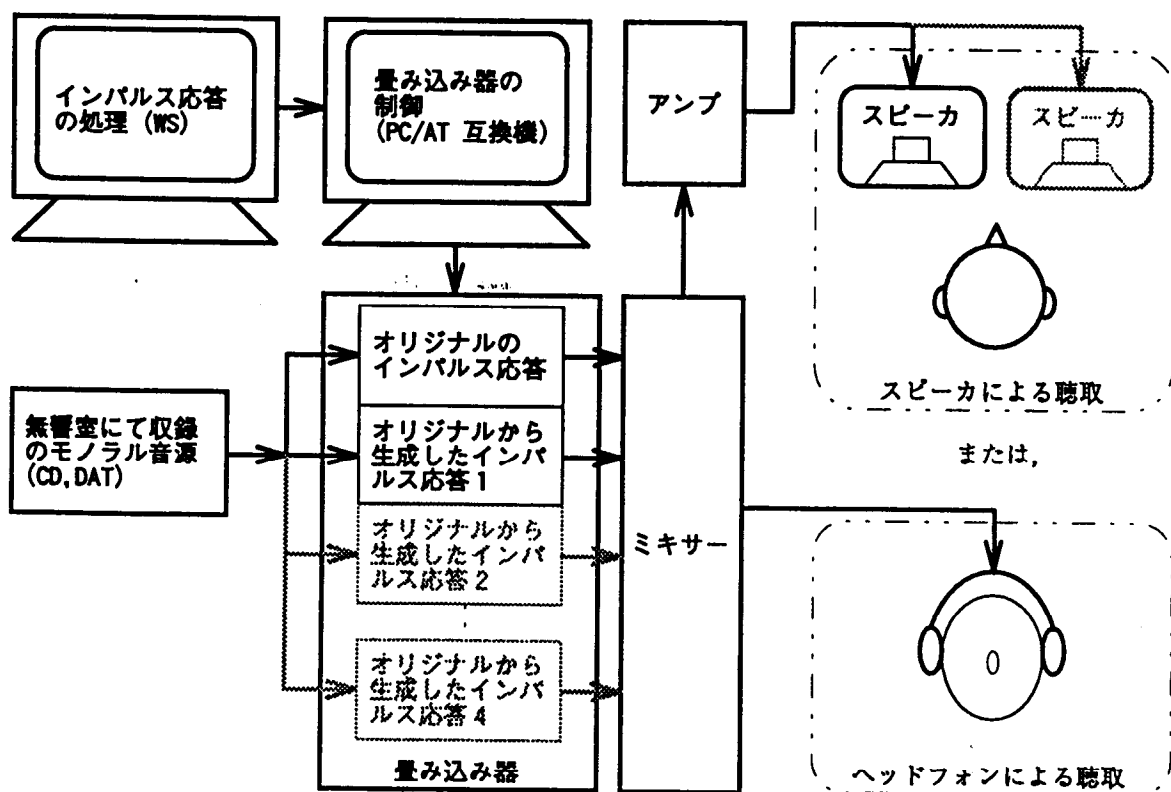


図 2.2: 試聴装置

最後に(c)については、実際に2チャンネル・スピーカ再生を行って空間でミキシングされる状況で試聴した結果、残響音の劣化が感じられたのは、スピーカを非常に近接して設置した場合のみであった。そこで、最も厳しい条件として、電気信号レベルでミキシングして試聴することにした。本方式では、2つの応答をミキシングして劣化が目立たないものは、4つの応答をミキシングしても劣化が目立たなかった。こうしたことから、評価方法として、生成した2つの応答による残響音と、リファレンスの2チャンネル分の測定から得た応答による残響音をそれぞれをミキシングして比較評価すれば充分であると考えられる。

以上のことから、主観評価実験は、分布幅とブロック幅の2つのパラメータについて、異なる残響音をミキシングした場合の音色を評価すれば、妥当かつ効率的な実験を行えるという結論に達した。

この信号処理方式では、生成できる応答の数は理論上ほぼ無限であるが、同じパラメータによってつくられた複数の応答間の聴感的な音質の違いは、パラメータの違いによる音質の違いよりも遥かに小さいことを予備実験で確認した。参考までに、ブロック幅1.33ms、分布幅41.7msで実際に20個の応答を生成し、その残響時間周波数特性の平均と95%信頼区間を調べた結果を図2.3に示す。各帯域において、残響時間のばらつきが小さいことから、生成した応答間の聴感的な違いが小さいことが裏付けられ

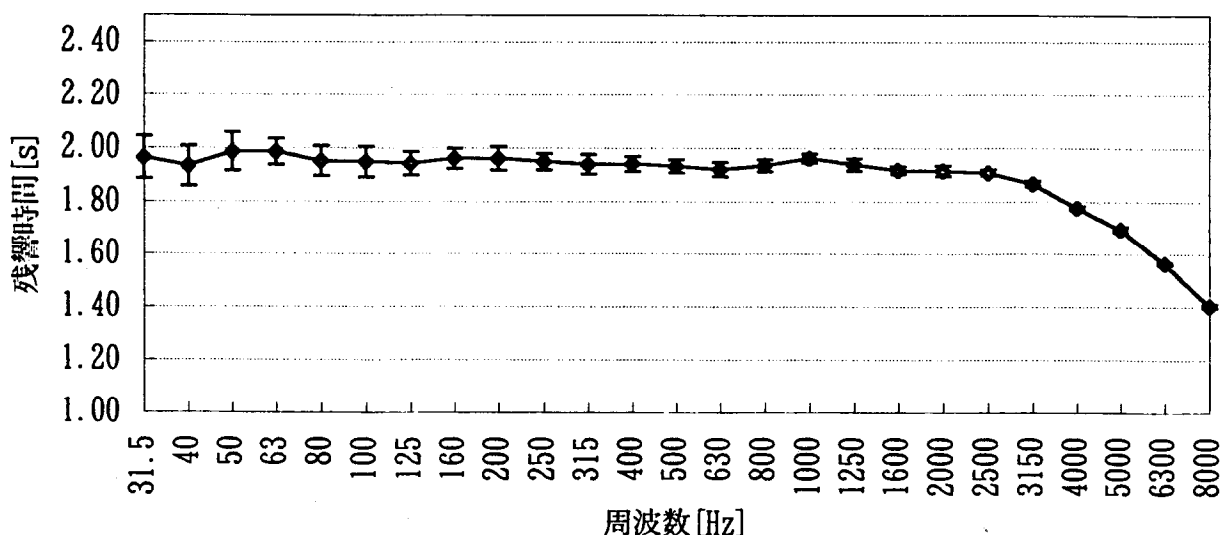


図2.3: ランダム・ブロック・シャフル法で生成した応答の残響時間周波数特性の平均と95%信頼区間

表 2.3: 評定尺度

評点	評価語
5.0	全く差がない The same
4.0	わずかに差がある Slightly different
3.0	差がある Different
2.0	かなり差がある Definitely different
1.0	非常に差がある Very different

ている。

前述の考察を踏まえ、実験計画を立て実験を行った。

2.2.3 実験計画と条件

前述の議論も考慮し、実験条件は次のように定めた。

要因: ブロック幅5種 (0.33, 1.33, 2.67, 5.33, 21.33 ms), 分布幅3種 (10.4, 41.7, 166.7ms).

リファレンスの応答2種 (残響時間2.0秒, 1.0秒), 評定者15名。

更に、追加実験としてブロック幅0.08, 0.17, 0.33 msを加え、ブロック幅は計7種 (0.33msは重複)。追加実験の評定者は5名。

評価法: リファレンスの応答による残響音 (原音) と新たに生成した応答による残響音 (評価音) を比較し、その差をITU-Rの勧告 [8] に準拠した評定尺度法を用い、表 2.3に示した指定の評価語にもとづいて、評点を定めてもらう。

試聴条件: 大型モニタ・スピーカ1台によるモノフォニック試聴。音圧レベルは、騒音計 (A特性,FAST) を使用し、評価音のピークが、聴取位置にて80dB(A)になるように設定した。実験の効率化のため、1回に3名の評定者で試聴する (図 2.4)。試聴室は、評価する残響音の音質に影響を与えないよう吸音処理がなされ、残響時間は200-500Hzで約0.2秒, 1kHz-8kHzで約0.1秒である。

試聴用音源: クリティカル・ソースとして無響室収録のドラムソロに 応答を畳み込んで作成。ただし、残響音の違いを明確にするため、応答の直接音は削除してから畳み込む。

評定者: 聴感実験に習熟した成人男女合わせて15名(追加実験は5名).

評価の繰り返し: あり. 同じ評価音を2回(追加実験は3回)評価する.

対象の提示法: ITU-Rの勧告[8]に準拠して, 1条件について原音(A), 評価音(B)をABABの一組として提示する. 提示順序はランダムとする.

回答の方法: 指定の評価尺度に基づいて5.0から1.0までの評点を小数点第1位まで記入する.

休みの取り方: リファレンスの応答が2種あるので, この応答ごとに2つのセッションに分け, セッションの間に10分の休みを挟む. 1セッションの長さは10~12分程度である.

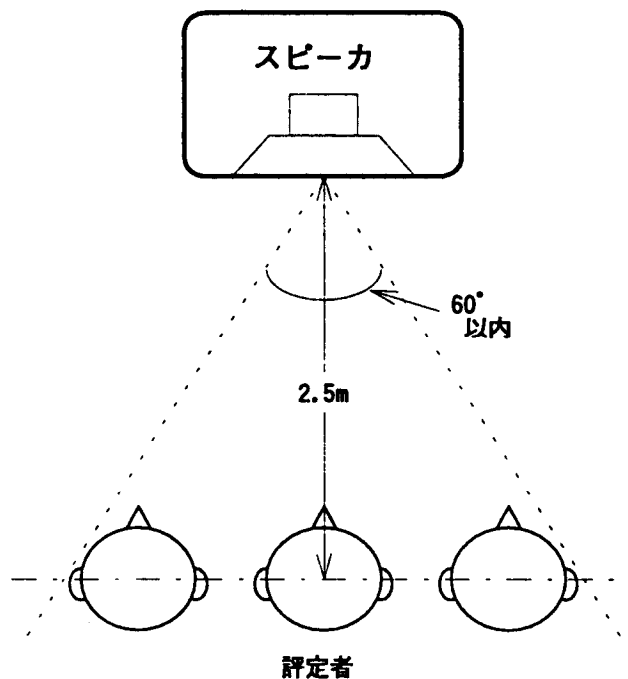


図 2.4: 聴条件

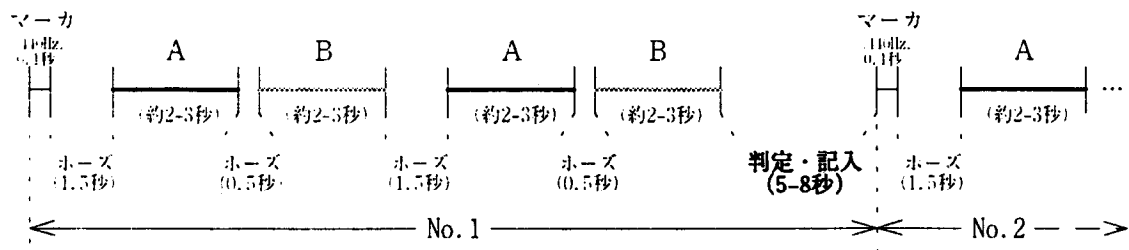


図 2.5: 対象の提示法

リファレンスのインパルス応答は、業務用リバーブレータのインパルス応答を、TSPを用いる方法[6]で測定したものを使用した。

ABABという提示方法では、あらかじめ基準音がAと判っているので、評定者は、評価音Bが必ず劣化をきたしているという先入観をもって試聴することになる。このため、5.0に近い評点は非常につきにくくなり、全体に評点が低い方にシフトしてしまう傾向がある。そこで、評価音Bが基準音と同じ場合を混ぜ、これを教示の際に評定者に伝えることにした。

2.2.4 実験結果

図2.6, 2.7に、実験結果として評点の平均と95%信頼区間を示す。また、今回の実験では、2.2.3に述べたように、評価音の中にリファレンスも含んでいるので、参考までに、リファレンスの評点と95%信頼区間を示した。

分布幅に関しては、残響時間の異なる2つのリファレンスのどちらも、41.7msが良好であり、この時のブロック幅に関しては、リファレンスの残響時間が異なると、評点のピークもやや異なる傾向があるが、0.33msからこの8倍の2.67msまでブロードな傾向にある。どちらのリファレンスにも適していると考えられるのは本実験の結果からは1.33msある。こうした、適当なブロック幅と分布幅を用いれば、リファレンスの評点に迫る結果が得られており、残響音の音質の保存性は、かなり高いといえる。

なお、ブロック幅0.33msのデータは、本実験と追加実験の2つのデータが得られているが、この評点の平均値の違いが最大0.2点であり、本実験の再現性は高いことが確認された。

§2.3 結果の考察

実験結果より、ブロック幅1.33ms、分布幅41.7msの条件のときに、良好な結果が得られた。このときの残響時間周波数特性をリファレンスと比較して図2.8に示した。この図で生成した応答の残響時間周波数特性は、20個の応答の平均と95%信頼区間を示してある。この図から、生成した応答は、2つのリファレンスの特性を良好に保存していることがわかり、生成した個々の応答のばらつきも十分小さいことも確認できる。また、このパラメータにおいて、生成した複数の応答間について、後部残響音部分の相関関数の最大値を求めたところ、約0.02であり、応答の低相関化も良好に実現されていることが確認された。

次に、ブロック幅、分布幅のそれぞれが、最適値からはずれた場合の音質の劣化の原因について考察する。

ブロック幅は、小さ過ぎると個々の反射音の特性を保存しきれず、結果としてリファレンスの応答特性を白色化してしまい、ノイジィな残響音を生みだし評点が低下していると考えられる。ブロック幅が0.33ms以下では、分布幅に依らず評点が低下していくことから、ブロック幅が小さい場合は、この残響音の白色化が劣化の大きな要因を占めていると考えられる。図2.9に残響時間2.0sのリファレンスの周波数スペクトルと残響曲線^{*1}を、図2.10にブロック幅0.08ms、分布幅41.7msで処理した応答の周波数スペクトルと残響曲線を示す。同様に、図2.11に残響時間1.0sのリファレンスの周波数スペクトルと残響曲線を、図2.12にブロック幅0.08ms、分布幅41.7msで処理した応答の周波数スペクトルと残響曲線を示す。加えて、図2.13に、残響時間周波数特性の比較を示す。残響時間周波数特性は、生成した応答20個の各周波数での残響時間の平均であり、95%信頼区間を付して示してある。これらの図から、生成した応答は、残響曲線には大きな変化が見られないが、周波数スペクトル上では高域にノイズが乗っており、残響時間周波数特性でも高域の残響時間が減少していないことが明らかで、リファレンスの応答を白色化している様子がわかる。

一方、ブロック幅が大きい場合は、ブロック内の反射音の特性の保存性は良くなるはずだが、ブロック間の応答のつながりが滑らかでなくなり、また、一定の分布幅の中で選択できるブロックの数が減少するため、重複使用されるブロックが増え、残響音の音色を変化させてしまっていると考えられる。図2.14, 2.15, 2.18にブロック幅21.33ms、分布幅41.7msで処理した応答の周波数スペクトル、残響曲線、残響時間周波数特性をそれぞれ示す。これらの図では、リファレンスと比較して、生成した応答の残響曲線と周波数スペクトルには大きな変化が見られない。生成した応答の劣化に対応する物理量については今後の検討課題としたい。

分布幅は、小さ過ぎると、選択されるブロックが限られてしまい、重複使用されるブロックも増え、応答の聴感的特徴を崩してしまうと考えられる。しかしながら、残響時間が1sの場合は、分布幅が小さい場合の評点の差が明確ではない。この原因についても、今後の検討課題である。

一方、大きすぎると、ブロック間のエネルギーの差が大きくなる場合が生じ、残響エネルギーが滑らかに減衰しなくなるため、聴感的に不自然な残響音になったと考えられる。これは、結果から、減衰の早い残響時間が1.0秒の方が2.0秒の場合より、分布

^{*1}残響曲線の詳細は§3.3で述べる。

幅が大きいときの評点の低下が目立つことから裏付けられる。残響が滑らかに減衰してない例として、**図 2.16, 2.17, 2.19**にブロック幅21.33ms, 分布幅166.7msで生成した応答の残響曲線, 周波数スペクトル, 残響時間周波数特性をそれぞれ示す。これらの図から, 生成した応答は, 周波数スペクトルや残響時間周波数特性では大きな変化が見られないが, 残響曲線より, 残響エネルギーが滑らかに減衰していない様子がわかる。

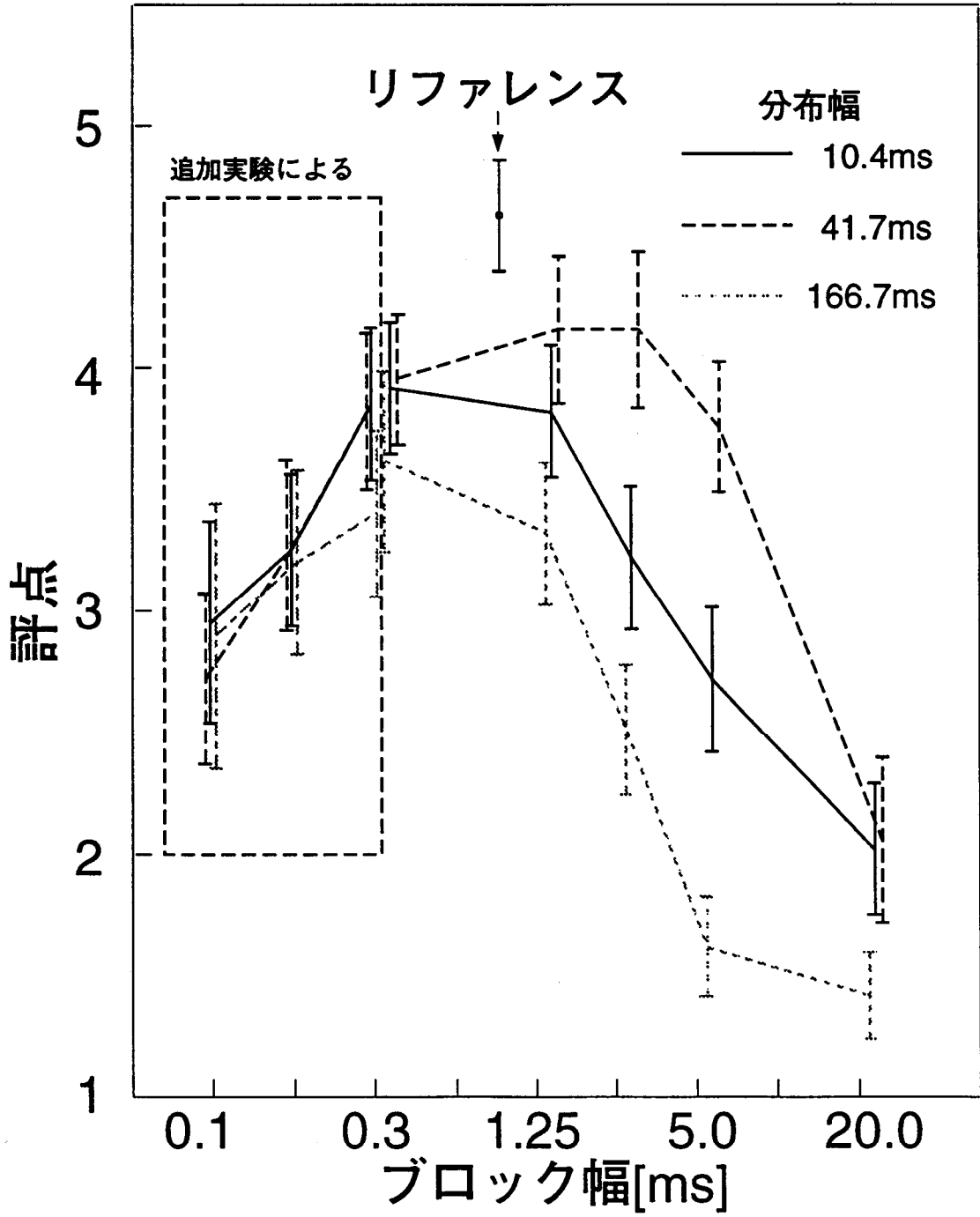


図 2.6: 評価結果(残響時間:2s)

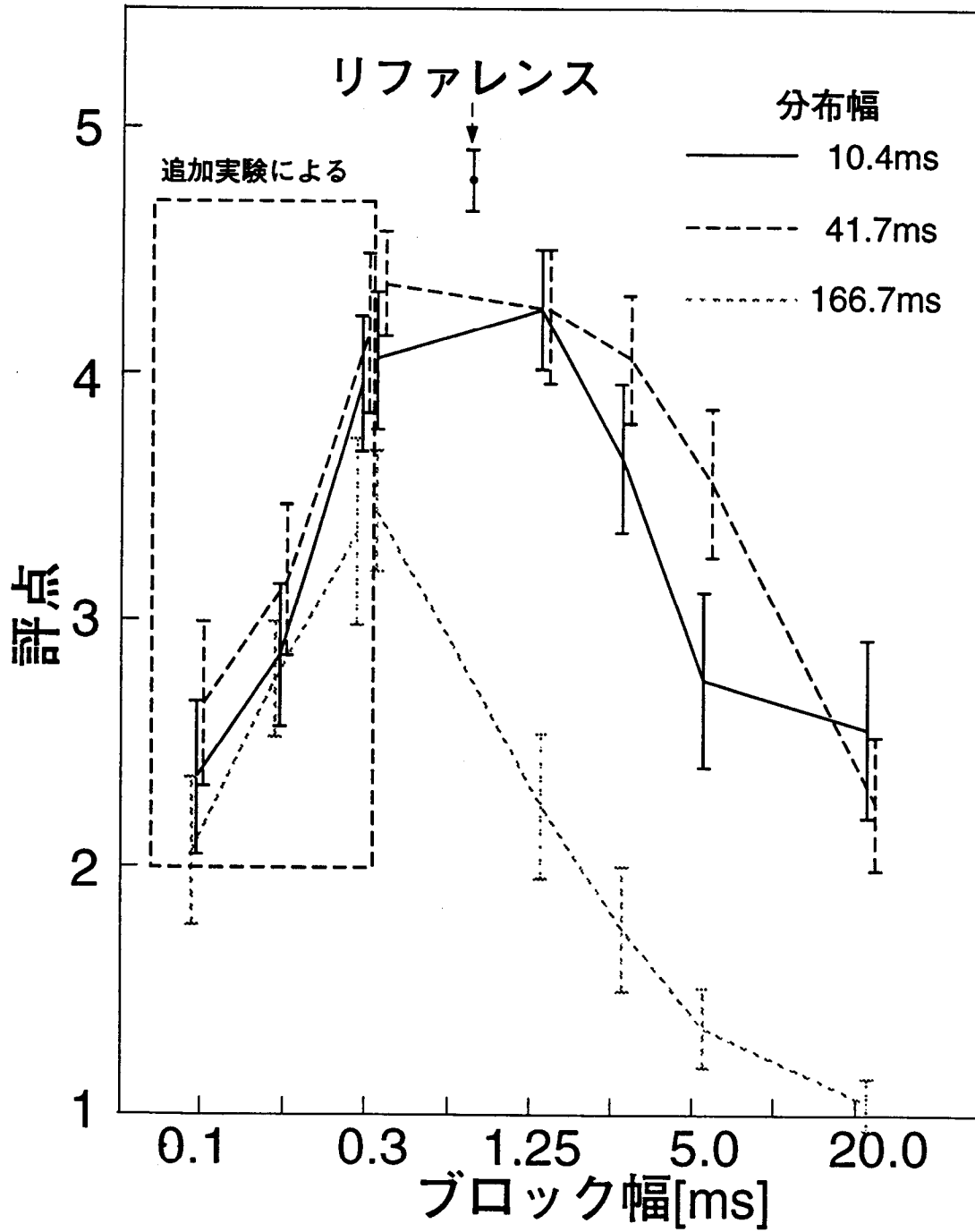


図 2.7: 評価結果(残響時間:1s)

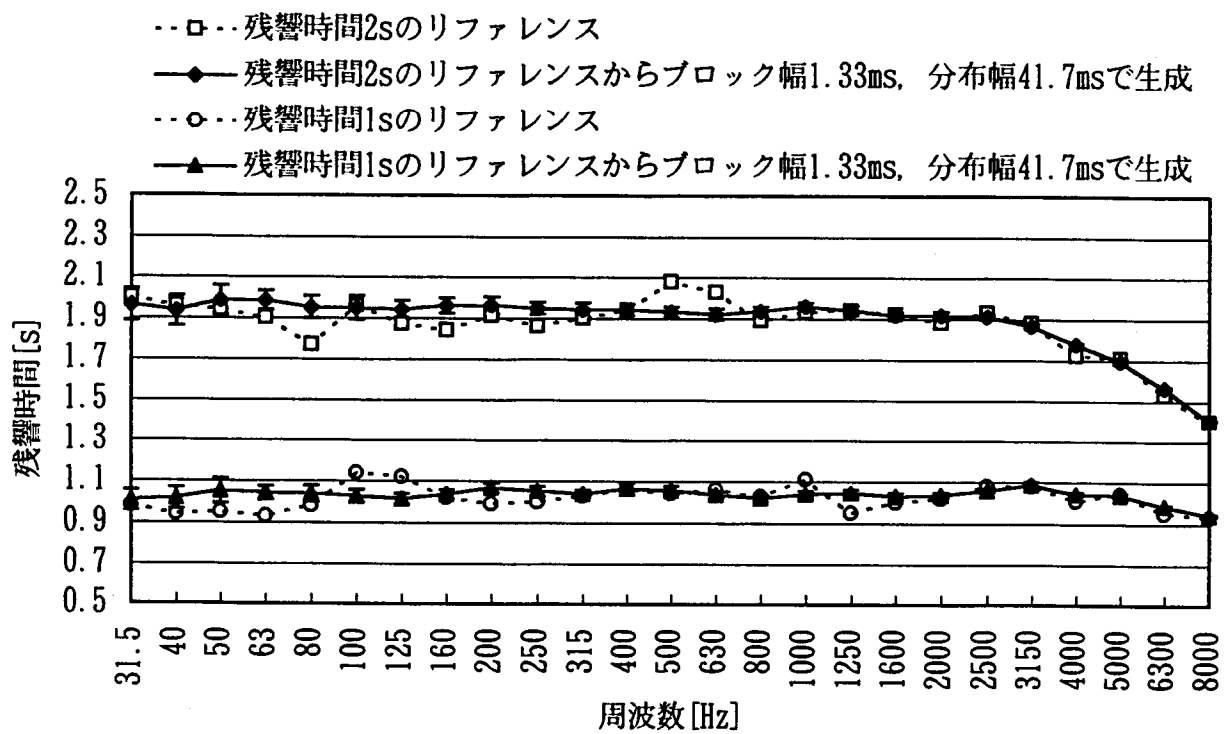
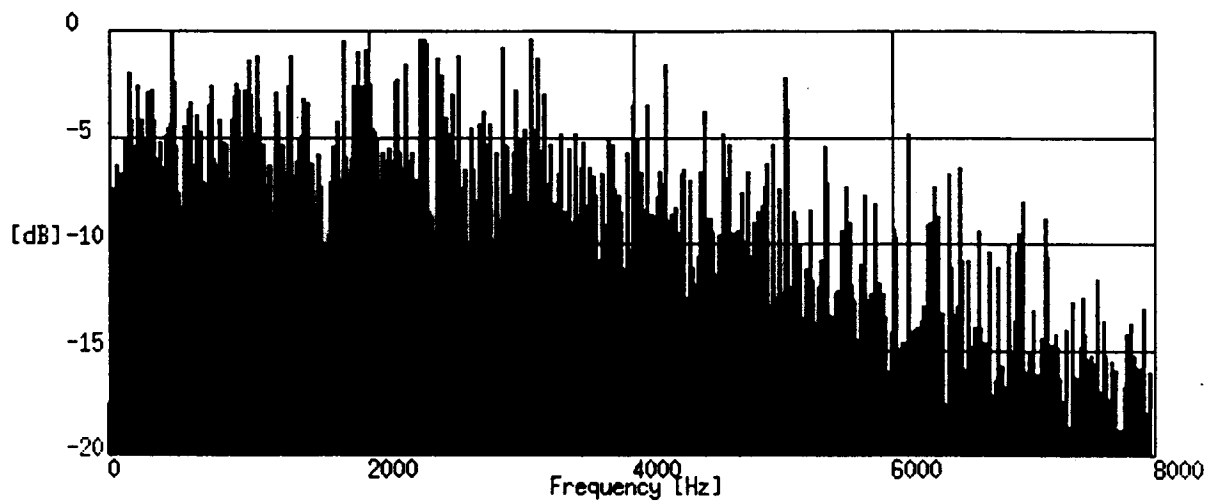
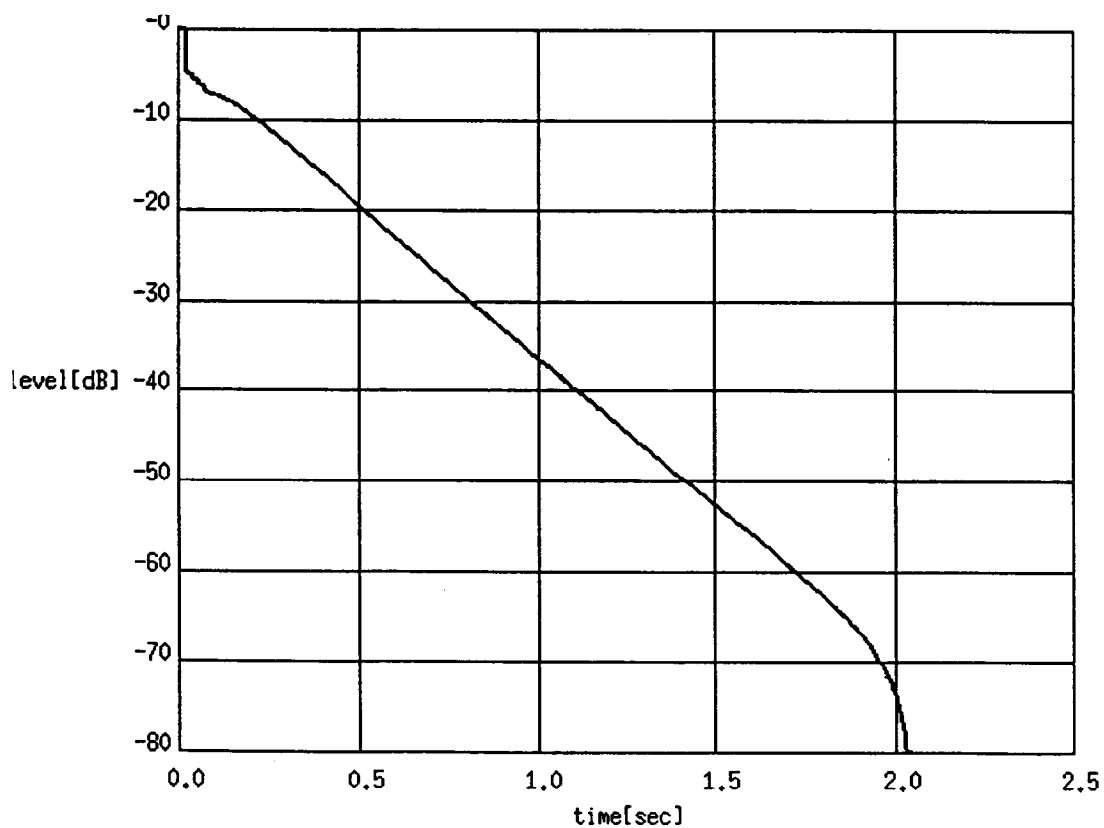


図 2.8: リファレンスと生成した応答の残響時間周波数特性の比較

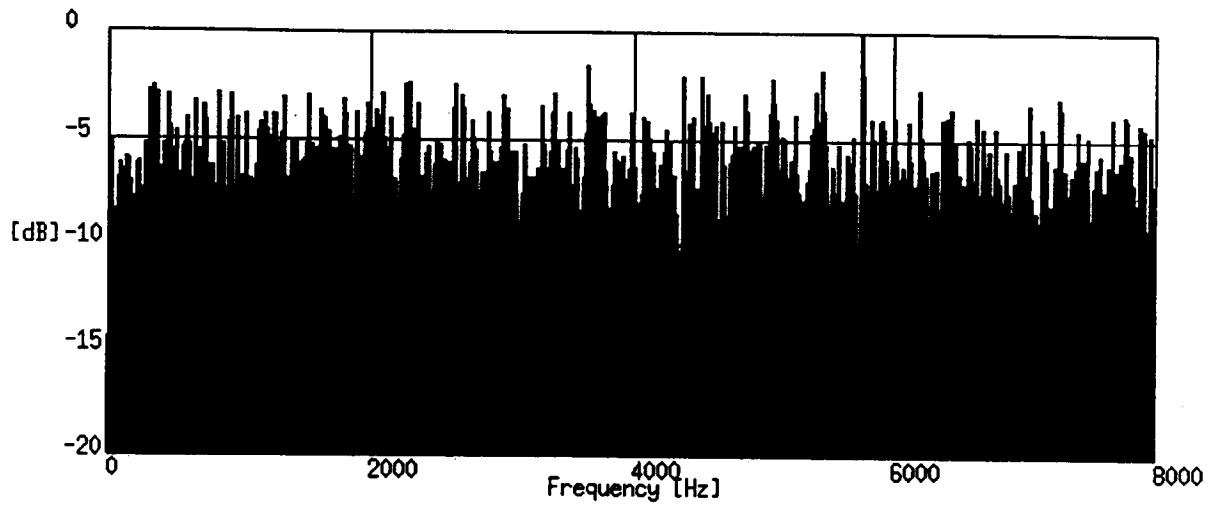


(a) 周波数スペクトル(0.5s以後)

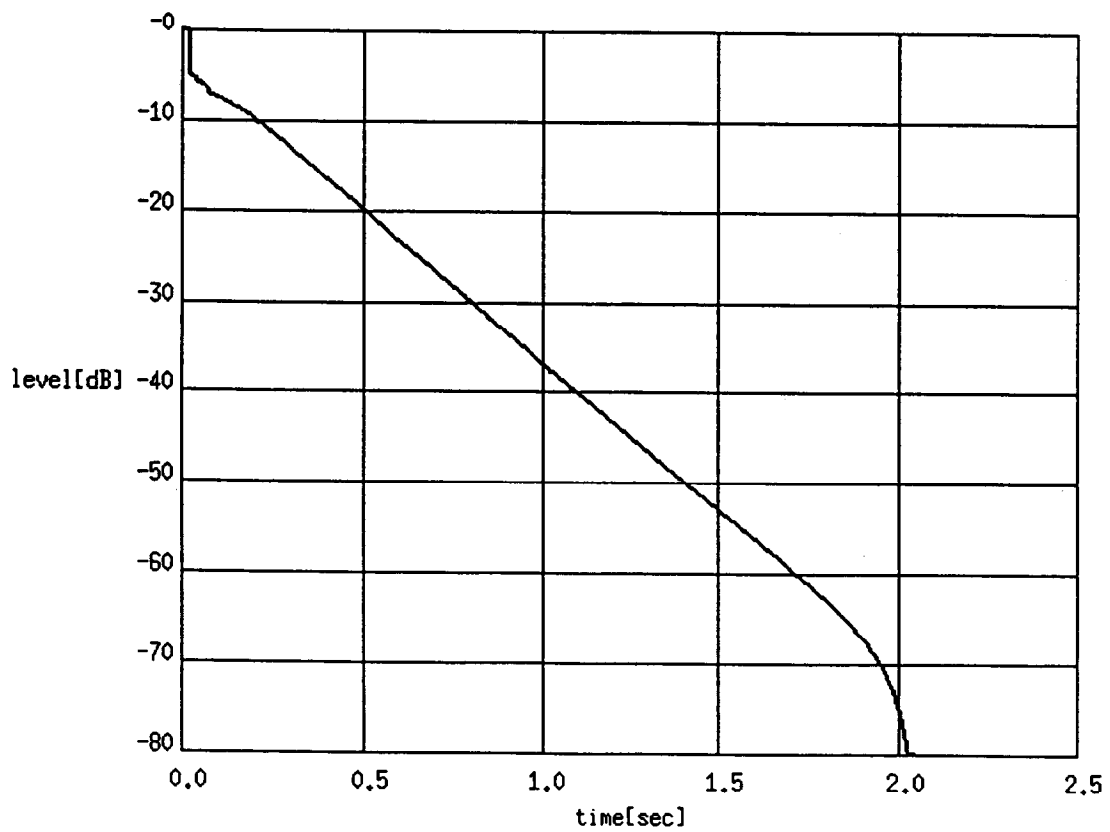


(b) 残響曲線

図 2.9: 残響時間2sのリファレンスの特性

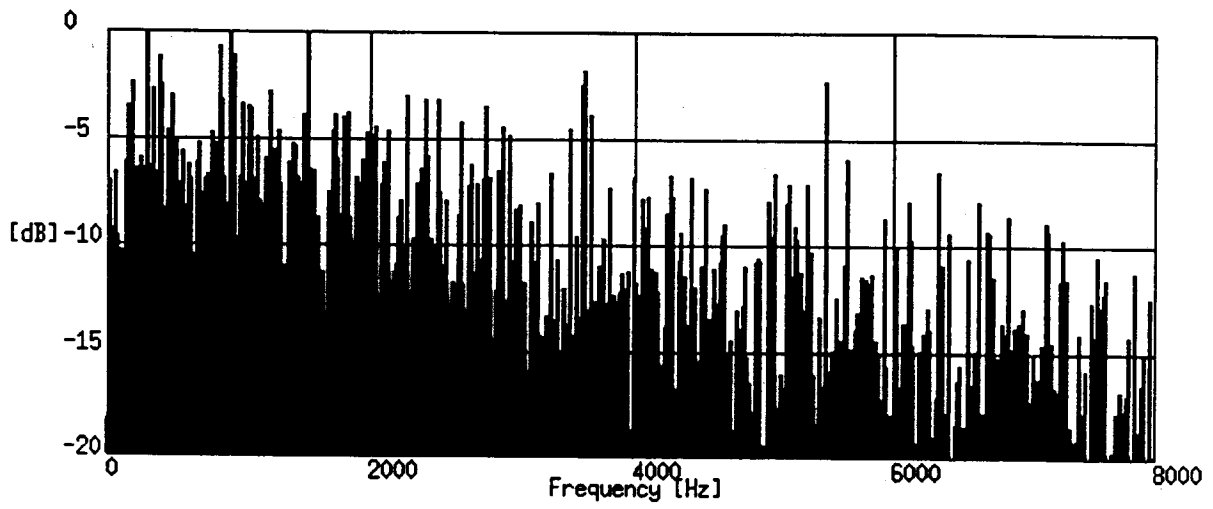


(a) 周波数スペクトル(0.5s以後)

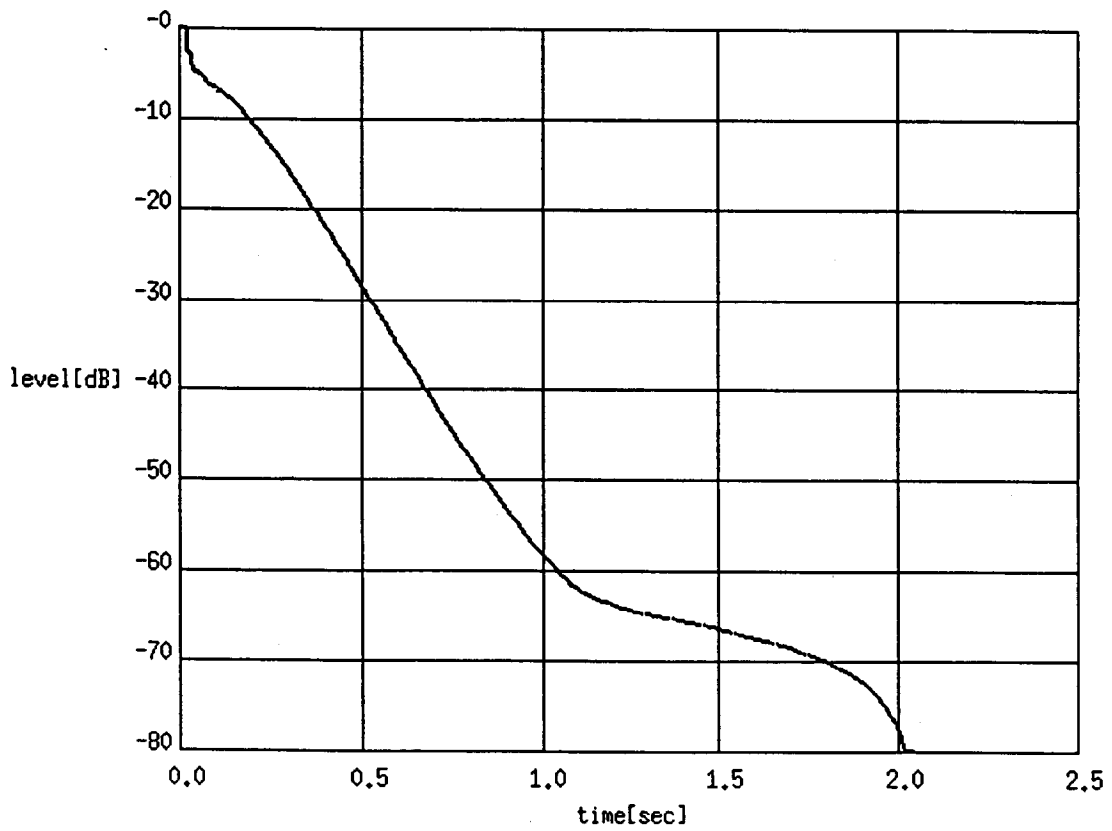


(b) 残響曲線

図 2.10: 残響時間2sのリファレンスからブロック幅0.08ms, 分布幅41.7msで処理した応答の特性

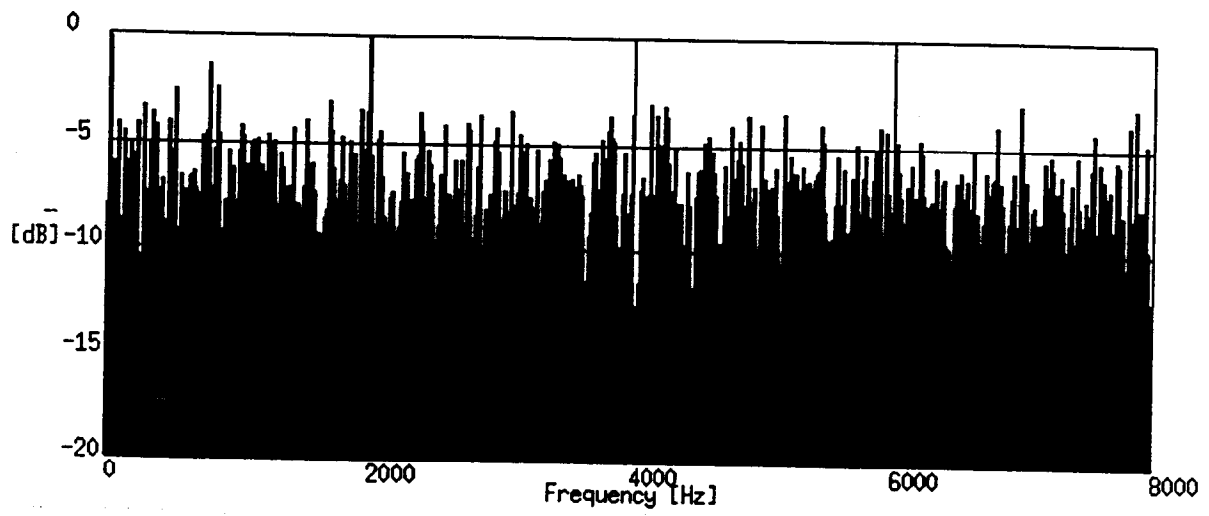


(a) 周波数スペクトル(0.3s以後)

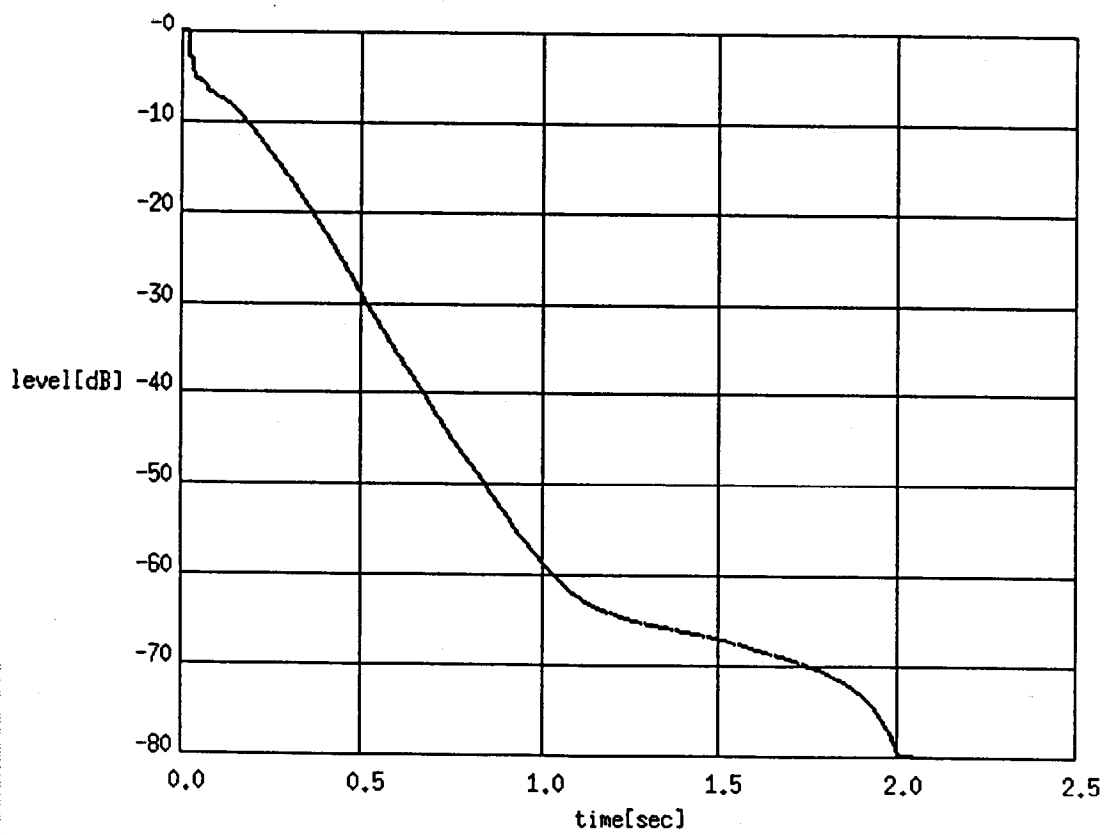


(b) 残響曲線

図 2.11: 残響時間 1s のリファレンスの特性



(a) 周波数スペクトル(0.3s以後)



(b) 残響曲線

図 2.12: 残響時間1sのリファレンスからブロック幅0.08ms, 分布幅41.7msで処理した応答の特性

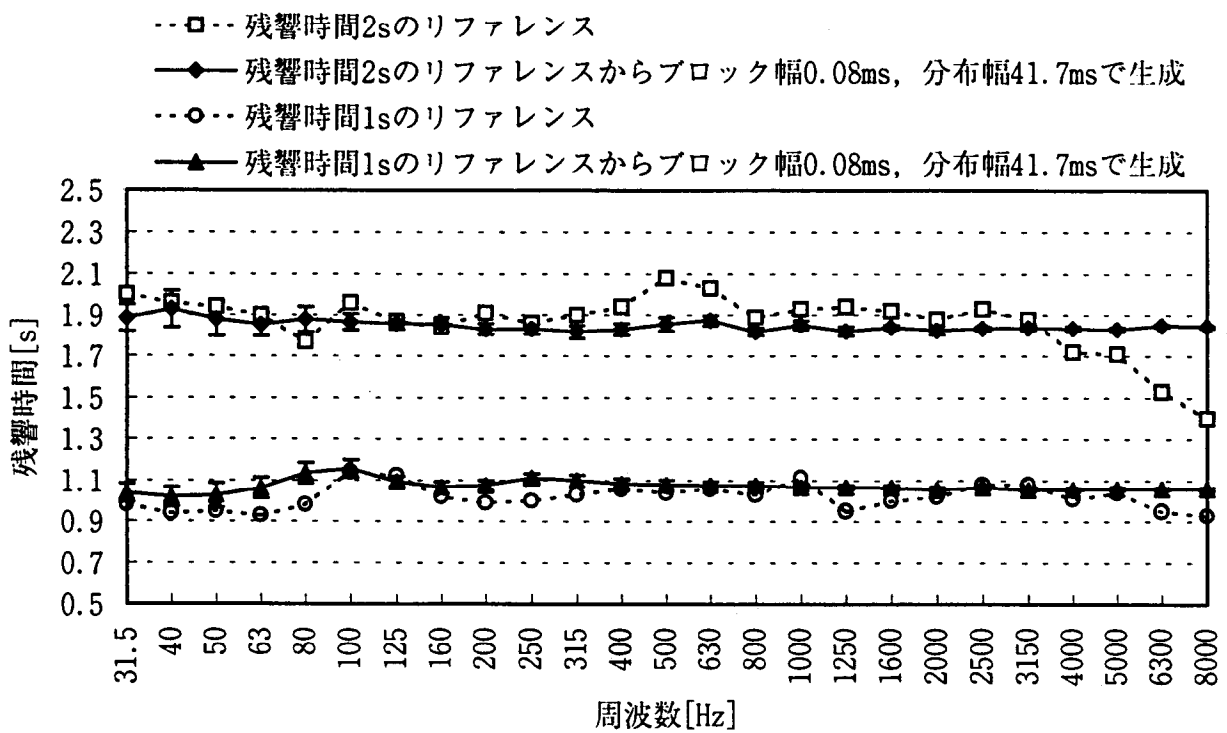
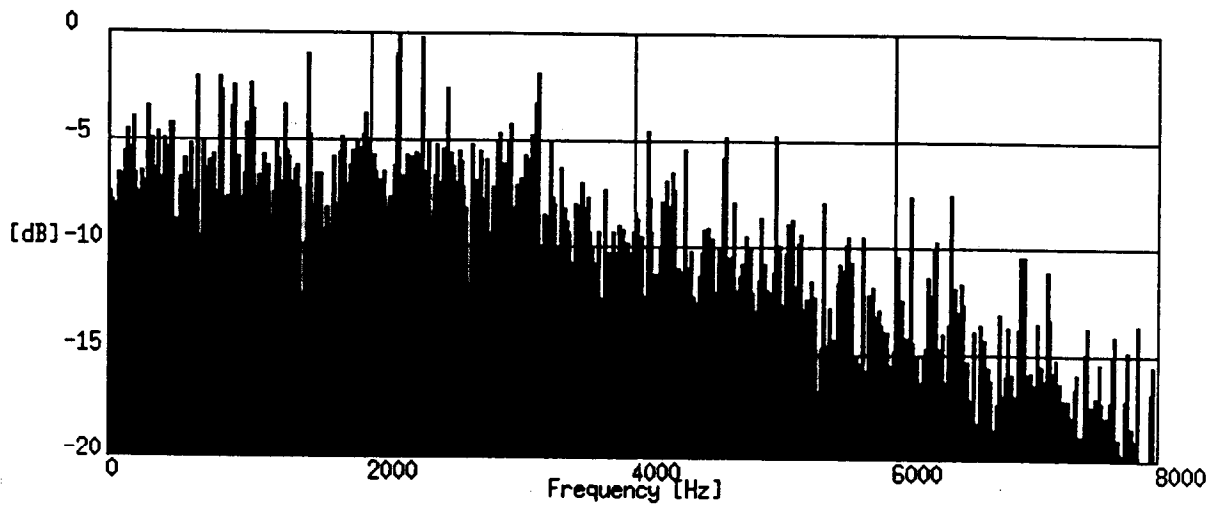
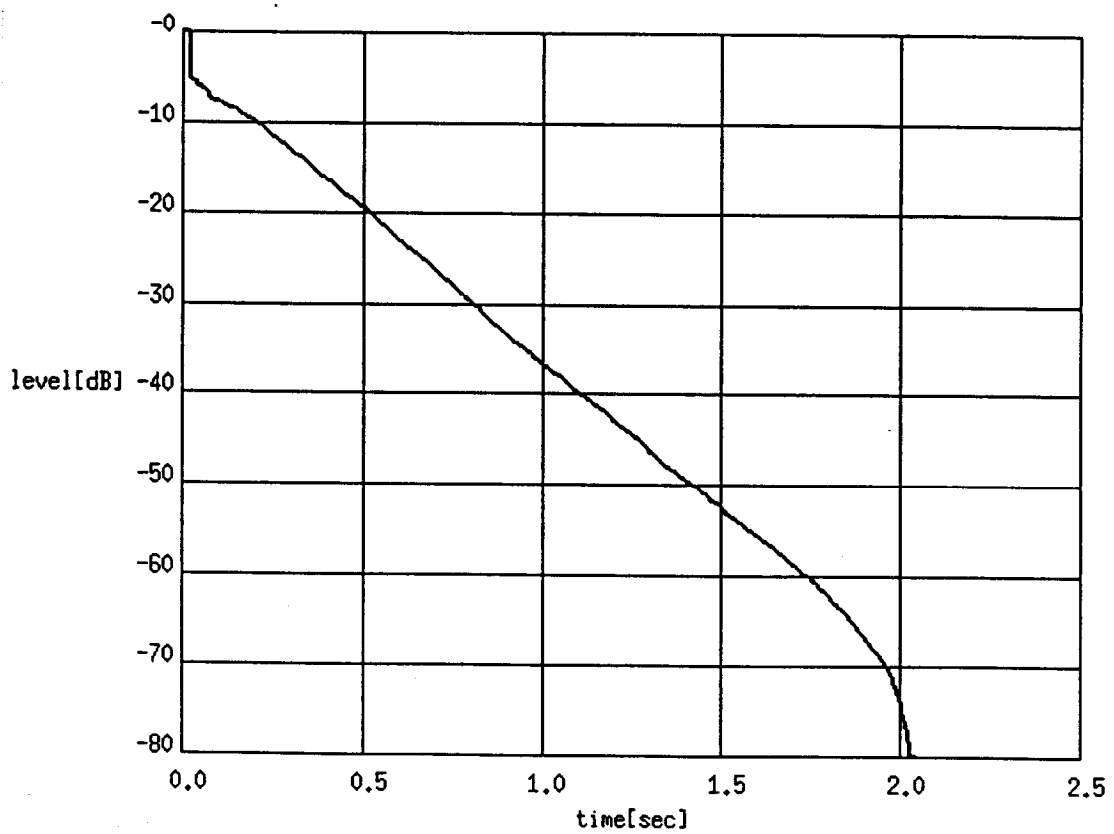


図 2.13: リファレンスとブロック幅 0.08ms, 分布幅 41.7ms で処理した応答の残響時間周波数特性の比較

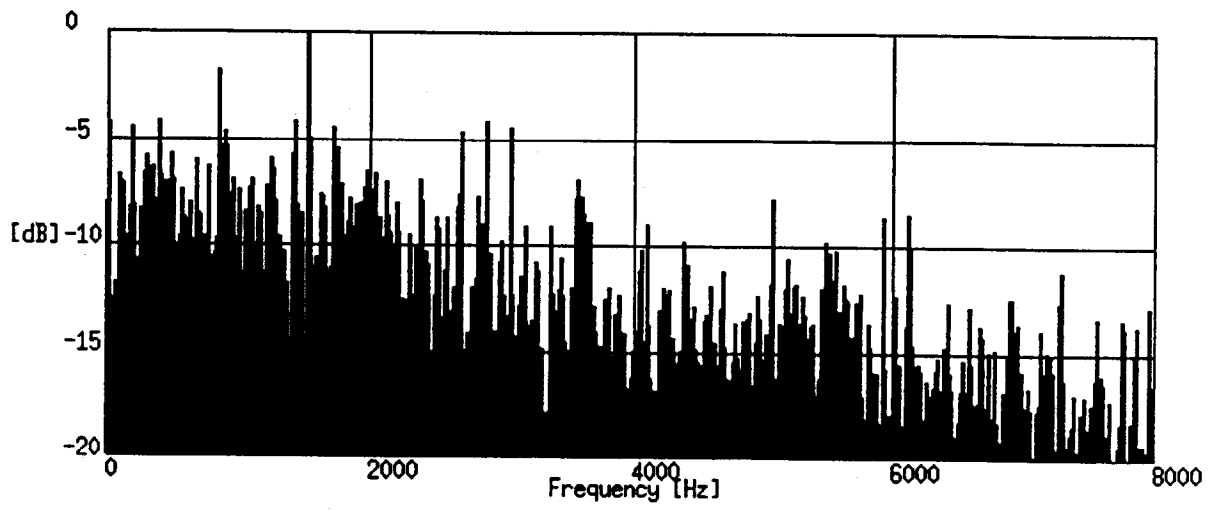


(a) 周波数スペクトル(0.5s 以後)

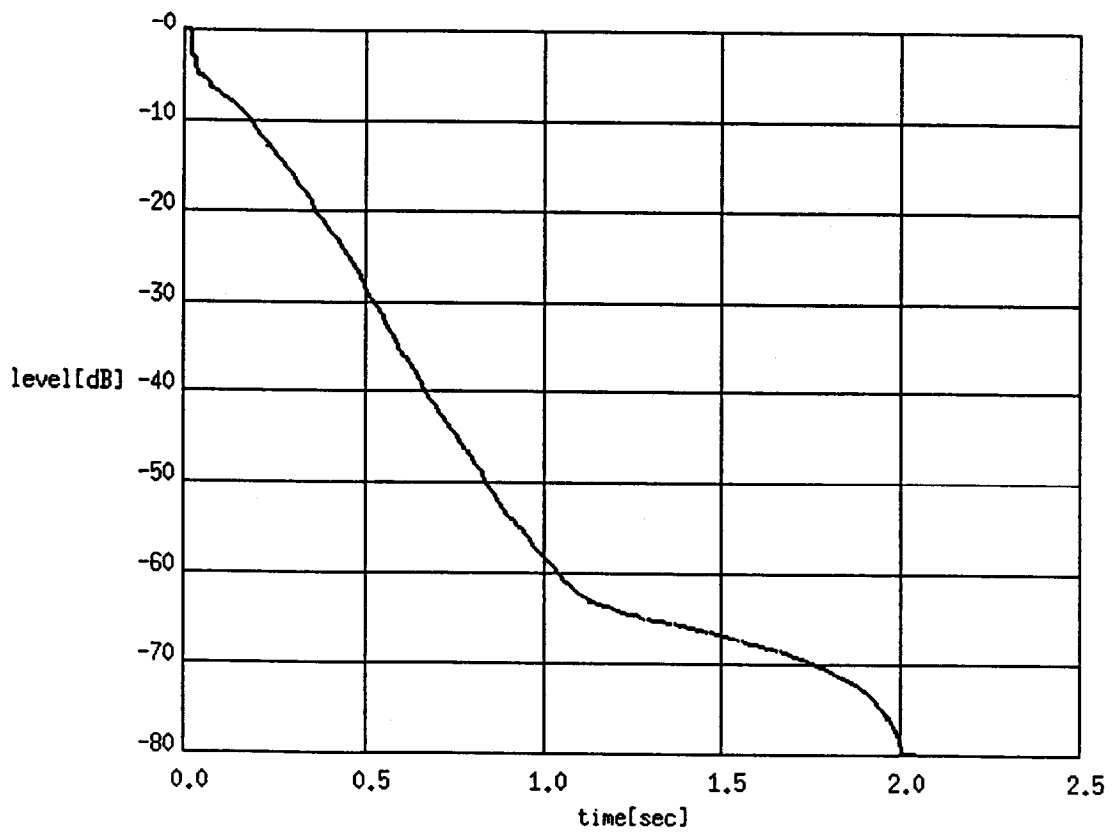


(b) 残響曲線

図 2.14: 残響時間2sのリファレンスからブロック幅21.33ms, 分布幅41.7msで処理した応答の特性

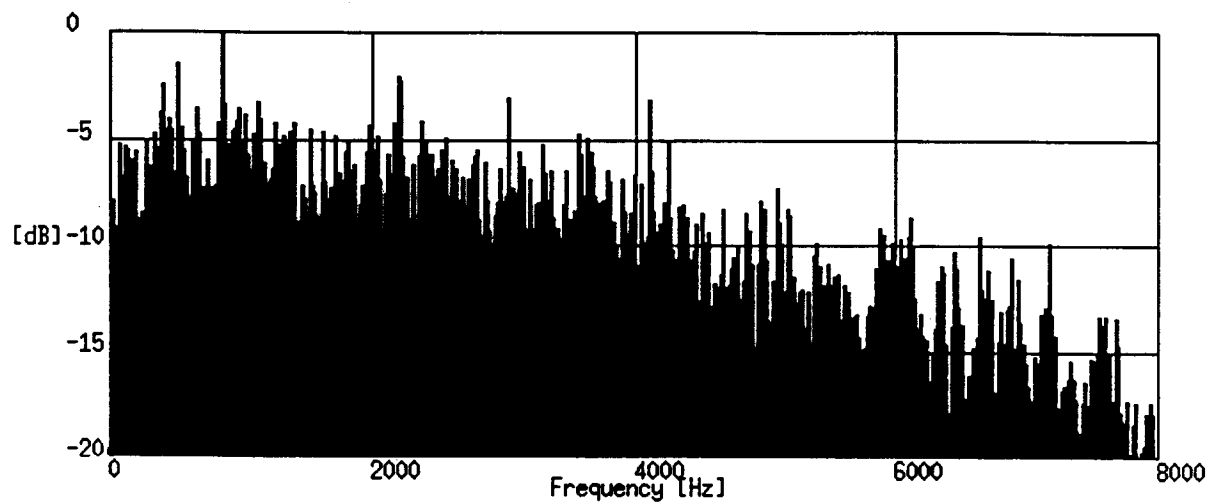


(a) 周波数スペクトル(0.3s以後)

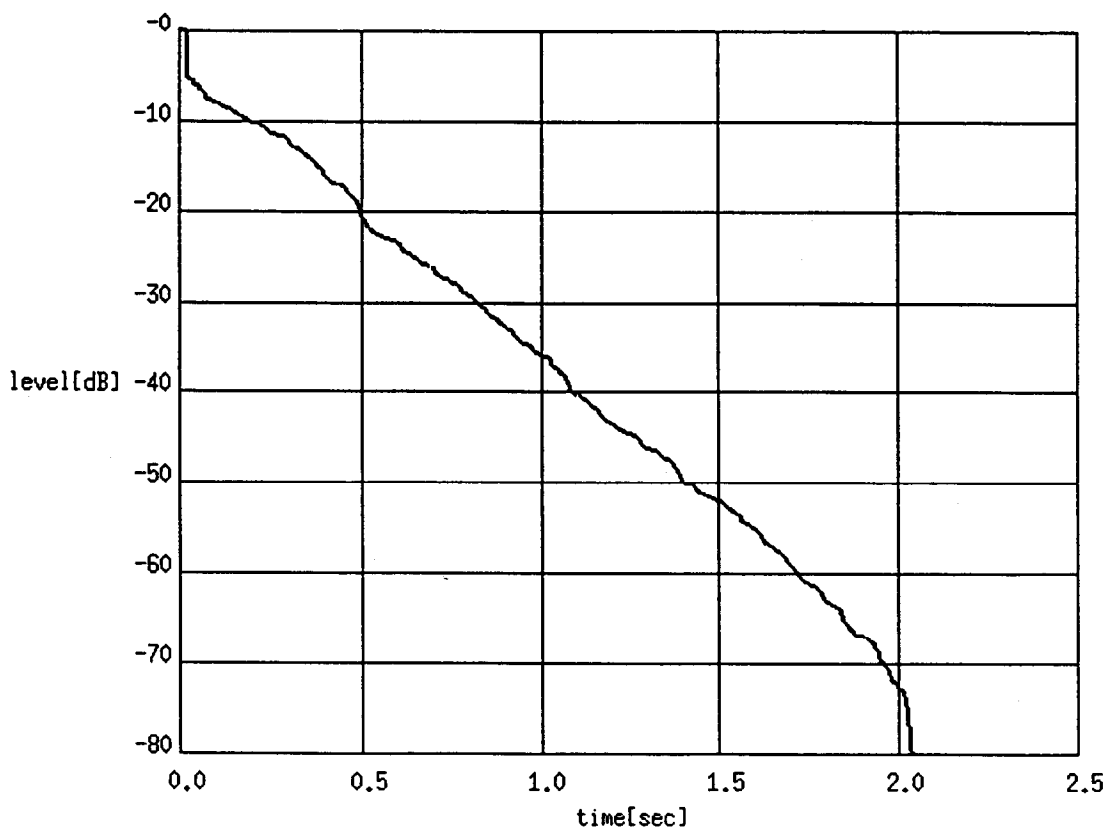


(b) 残響曲線

図 2.15: 残響時間1sのリファレンスからブロック幅21.33ms, 分布幅41.7msで処理した応答の特性

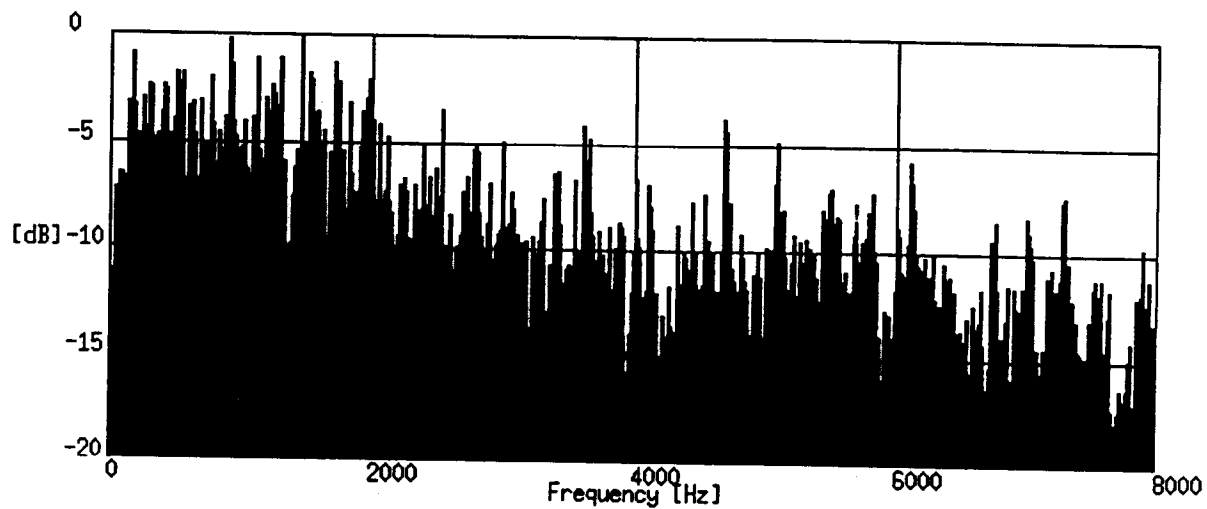


(a) 周波数スペクトル(0.5s以後)

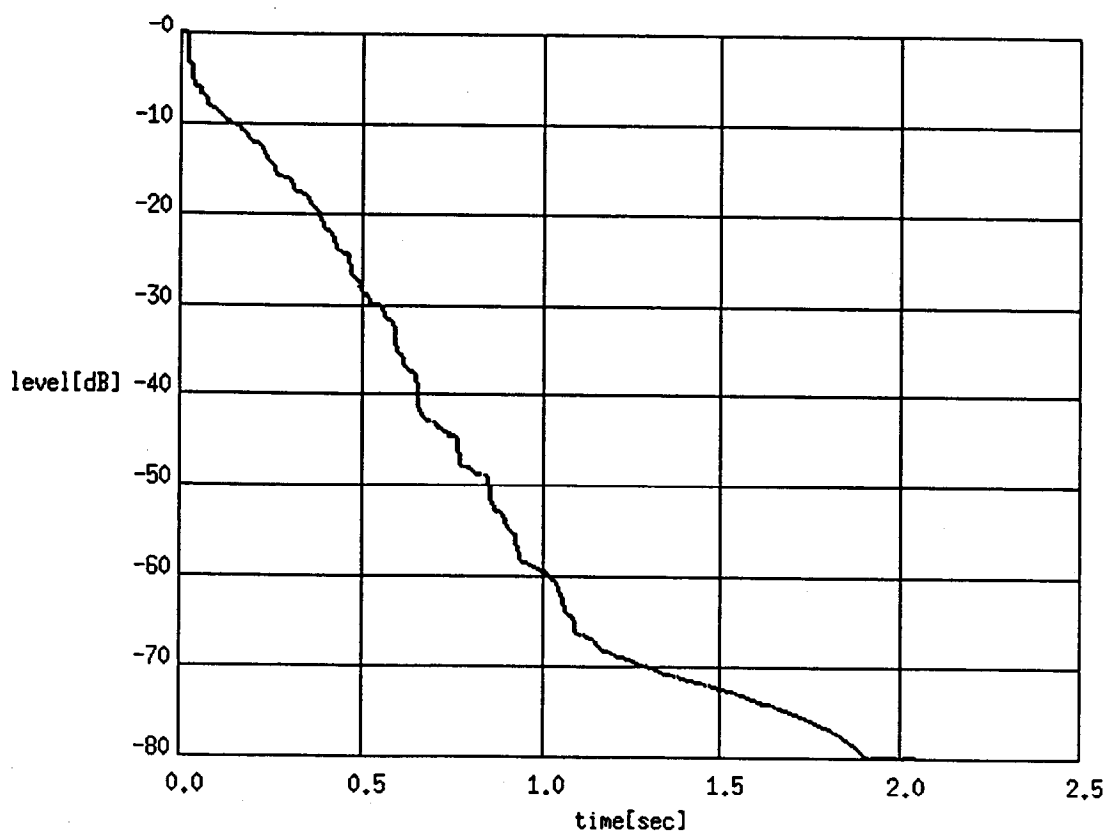


(b) 残響曲線

図 2.16: 残響時間2sのリファレンスからブロック幅21.33ms, 分布幅166.7msで処理した応答の特性



(a) 周波数スペクトル(0.3s以後)



(b) 残響曲線

図 2.17: 残響時間1sのリファレンスからブロック幅21.33ms, 分布幅166.7msで処理した応答の特性

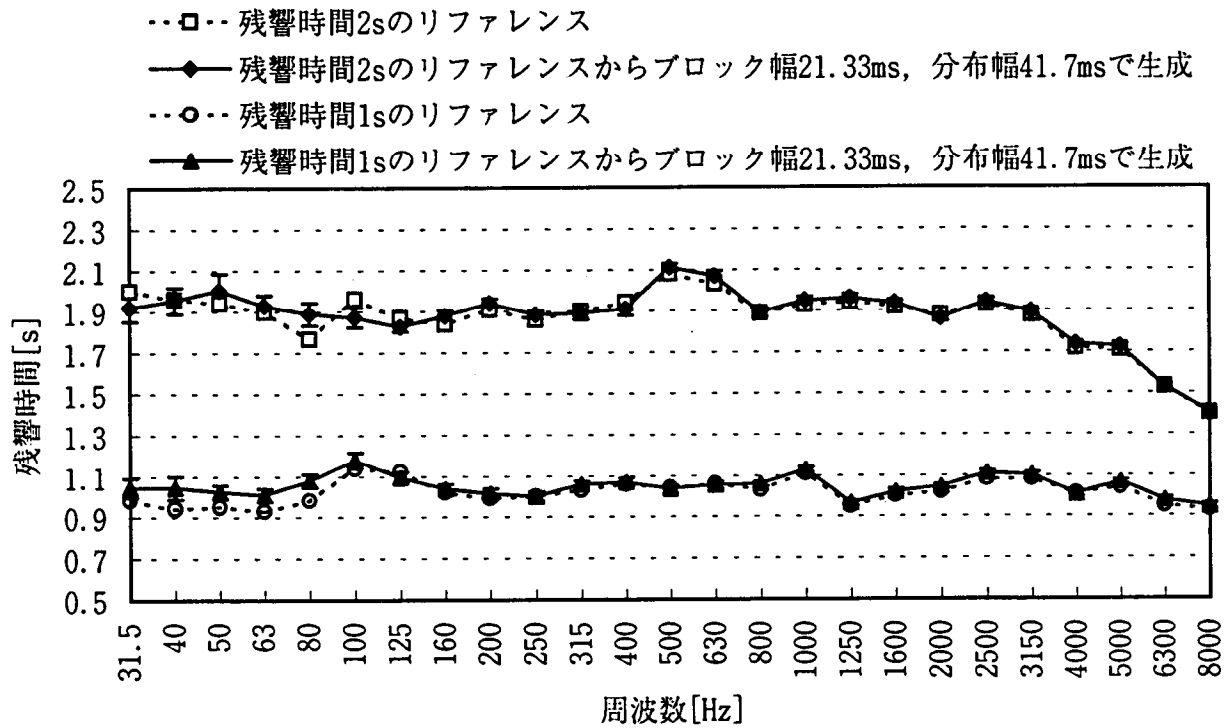


図 2.18: リファレンスとブロック幅 21.33ms, 分布幅 41.7ms で処理した応答の残響時間周波数特性の比較

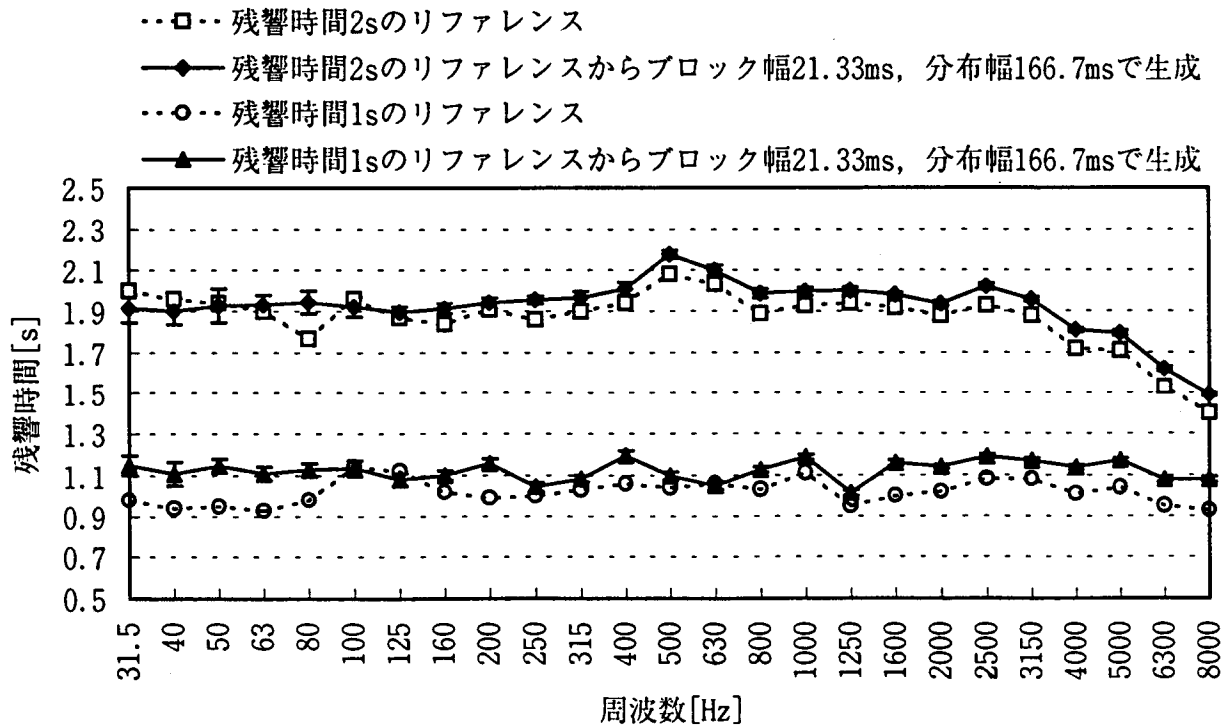


図 2.19: リファレンスとブロック幅 21.33ms, 分布幅 166.7ms で処理した応答の残響時間周波数特性の比較

第3章

インパルス応答の時間伸縮による残響音の生成

§3.1 インパルス応答の同期型サンプル・レート変換による時間軸一様伸縮

離散時間信号として得られているインパルス応答のサンプル・データの時間軸を伸縮するために、サンプリング周波数の変換を行うことにした。最も単純な変換方法として、図3.1に示すような同期型サンプル・レート変換のアルゴリズムを採用した [9][10]。元のデータのサンプリング周波数 f_s とする。まず、サンプル・データに $L - 1$ 個の0データを等間隔で挿入し、カット・オフ周波数 f_c のFIR型のローパス・フィルタ(LPF)で補間することにより、 L 倍のオーバー・サンプリングを行う。このデータを今度は M サンプルごとに出力することによりサンプリング周波数を $(L/M)f_s$ に変換できる。

ただし、 L 、 M は正整数で、 f_c は標本化定理から f_s の $1/2$ 以下でなければならず、本研究では f_s が 48kHz のため、 f_c を 20kHz とした。計算機資源や計算時間の制約から、本アルゴリズムでは、 1% 刻みの時間伸縮処理を施し試聴した。試聴結果は §3.4 で述べる。

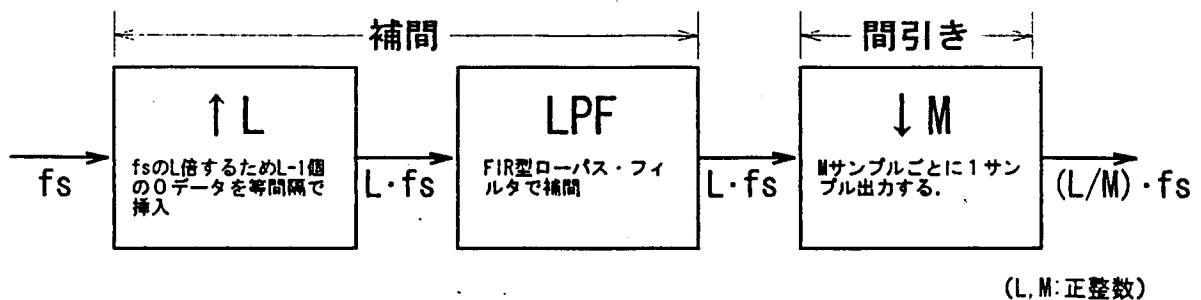


図3.1: 同期型サンプル・レート変換

§3.2 インパルス応答の非同期型サンプル・レート変換による時間伸縮

前節の方法では、伸縮率を小刻みに変えることが難しい。例えば、0.1%の伸縮を行うためには1000倍程度のオーバー・サンプリングと補完処理が必要になる。そこで、非同期型のサンプル・レート変換を利用することにした[10][11]。

この方法は、サンプル・レート変換用のLSIなどに採用されているもので、図3.2で示すように、入力的时间離散信号を、ある一定の高い補完レートで補間し、さらに、サンプルの間隔は一次補間することによって、連続時間信号とみなし、所望の時間間隔でリサンプルすることにより、サンプリング周波数の変換を行う。リサンプルのタイミングによっては誤差が存在するが、補間した信号の隣接サンプル間の差が1LSB以下となるような十分に高い補間レートをを用いることによって、必要な精度を保つことが出来る。本研究では、256倍の補間レートをを用いた。

このアルゴリズムを用いて、0.2%および0.5%の時間伸長や、伸縮率が時間変化する処理を施し、試聴実験を行った。詳細は§3.4で述べる。

§3.3 エンベロープ補正

時間伸縮によって残響時間も変化してしまうため、何らかの補正が必要である。残響時間は、インパルス応答のエネルギーの減衰の様子を表わす残響曲線から決定されるので、残響曲線がリファレンスと重なるように補正をすればよいことになる。これを本論文では、残響音のエンベロープ補正と呼ぶことにする。

残響曲線を求めるにあたり、インパルス応答から直接求められ、かつ、計算の容易な

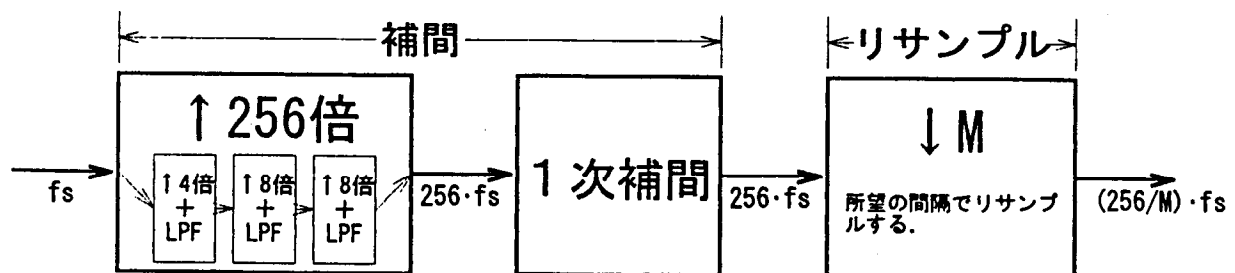


図3.2: 非同期型サンプル・レート変換 (M:実数)

Schroederの方法[12]を採用した。ここでは、残響曲線 $S(t)$ [dB]を次のように定義する。

$$S(t) = 10 \log_{10} \left\{ \int_t^{t+T} h^2(\tau) d\tau \right\}. \quad (3.1)$$

ただし、 T はインパルス応答 $h(t)$ が収束するに十分な時間とする。

リファレンスのインパルス応答と、時間伸縮処理したインパルス応答の残響曲線が同様の減衰特性を示すように、伸縮処理したインパルス応答のエンベロープを補正する。残響音のエネルギーは、指数関数的に減衰する性質があるため、残響曲線 $S(t)$ は、ほぼ直線的に減衰するように描くことができる。実際には、初期反射音の減衰と後部残響音の減衰の仕方は異なり、また、残響エネルギーは十分に減衰しても、測定時に重畳する暗騒音などのノイズの影響を受け、残響曲線はある一定値以下には減衰しない。そこで、直接音が到来する時刻からエネルギーが -5 dBになるまでの初期反射音部分と、それ以降からエネルギーが -30 dBになる後部残響音部分に分け、最小2乗法により2折れ線近似して補正関数を求め、伸縮したインパルス応答に乗じることにした。このような分け方は、建築音響の分野で一般的に採用されているものである。

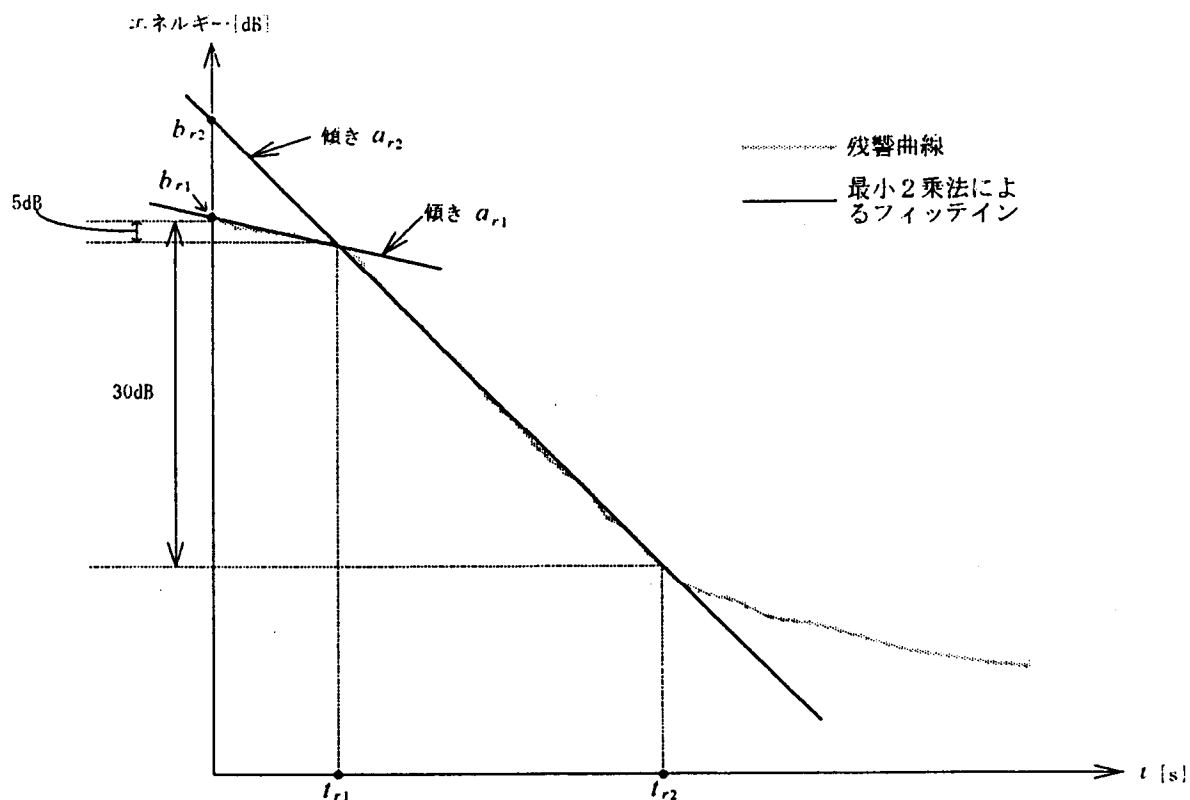
リファレンスのインパルス応答を $h_r(t)$ 、エンベロープ補正する前のインパルス応答を $h_i(t)$ 、補正後のインパルス応答を $h(t)$ とする。図3.3のように $h_r(t)$ 、 $h_i(t)$ それぞれの残響曲線を計算し、先に述べたように最小2乗法によって、2折れ線近似する。図3.3に定めた近似直線の傾き、切片を用いて、 $h(t)$ を次のように求めることができる。

$$h(t) = \begin{cases} \exp\{c(b_{r1} - b_{i1})\} \cdot \exp\{c(a_{r1} - a_{i1})t\} \cdot h_i(t) & (0 \leq t < t_{r1}) \\ \exp\{c(b_{r2} - b_{i2})\} \cdot \exp\{c(a_{r2} - a_{i2})t\} \cdot h_i(t) & (t \geq t_{r1}) \end{cases} \quad (3.2)$$

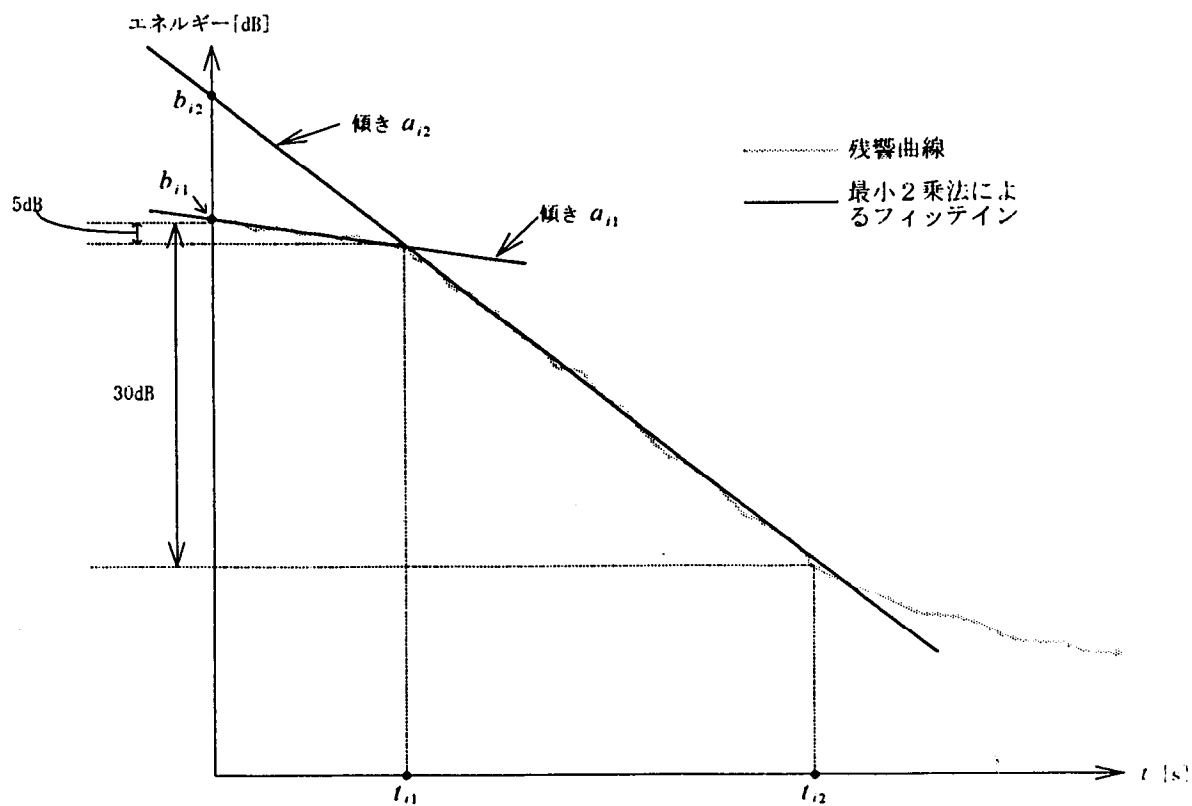
ただし、 c は定数で、残響曲線を式(3.1)のように定めた場合は、

$$c = \frac{\log_e 10}{20} \quad (3.3)$$

である。



(a) リファレンスとするインパルス応答の残響曲線



(b) エンベロープ補正するインパルス応答の残響曲線

図 3.3: エンベロープ補正のための残響曲線の2折れ線近似

§3.4 試聴実験の結果と考察

前述の方法を用いてインパルス応答を時間伸縮し、さらに、エンベロープ補正を施して新たな応答を生成し、この応答による残響音を試聴した。試聴は、図2.2に示した装置を用い、主にヘッドフォンを使用した。リファレンスの応答の残響時間は2.0秒である。評価は、2.2.1で述べた評価項目にしたがって行った。

時間伸縮の方法としては、一様伸縮法、一定周期で時間伸張・圧縮を繰り返す方法、後部残響音初期とそれ以後で伸縮率を変える方法の3種を試みた。

以下、それぞれの方法について、試聴結果と考察を述べる。

3.4.1 一様伸縮法

時間伸縮率は一定で、0.2%、0.5%、および、1%から1%刻みで10%まで変化させて試聴した。

まず、モノ聴取における残響音の音色は、伸縮率が数%までならリファレンスの音色を非常に良好に保存するが、10%では、リファレンスとの差が感じられるようになった。これは、ひとつには、時間を伸長すれば残響音を形成する反射音の密度が低下し、圧縮すれば上昇するするためと考えられる。また、伸縮に伴い、応答の周波数スペクトルも変化する。インパルス応答を $h(t)$ 、そのフーリエ変換が $H(\omega)$ とし、この変換対を

$$h(t) \leftrightarrow H(\omega) \quad (3.4)$$

と表わすことにすると、 $h(t)$ を伸縮する操作は、実数 a を用いて $h(at)$ と表わすことができ、この時、変換対は

$$h(at) \leftrightarrow \frac{1}{|a|} H\left(\frac{\omega}{a}\right) \quad (3.5)$$

となる。したがって、伸縮率が大きいほど、スペクトルの変化も大きく、聴感に影響を与える考えられる。

また、エンベロープ補正に施すことによりスペクトルの変化が起こるが、本実験のように10%程度の伸縮では、式(3.2)から分かるように、極めて緩やかに変化する指数関数窓をかけているのすぎないので、エンベロープ補正による音色変化は小さいと考えられる。

ステレオ聴取による広がり感は、0.2%の伸縮率でも小さいながら存在し、1%以上では十分な広がり感を得られ、10%まで広がり感の大きさは大差がなかった。定位につ

いては、生成した応答間の伸縮率の差が1%以下では不安定で、特に0.5%では激しい定位の振れが感じられた。この定位の振れは、音像の融合[13]によって生じると考えられる。音像の融合とは、左右の耳に100-600 μ sの時間差をもった音が入る場合、両耳の間に音像が生じる現象である。定位の振れが生じるのは、伸縮率が小さい場合、2つの応答の時間差が音像の融合が生じる時間差の範囲に入っている期間が長いとめと考えられる。

伸縮率の異なる残響をミキシングしたときの音色については、伸縮率の差が4%以上ならば問題ないが、4%より小さい場合は、バンド・ノイズのような不自然な残響音が聞こえた。伸縮率の異なる応答をミキシングした場合、ある程度短い時間間隔でみれば、ほぼ同一の信号を時間シフトしてミキシングしていることになる。この場合、時間シフトの幅を τ とすれば、ミキシングする前の応答に対して、 $1/\tau$ Hz毎に最大・最小となる周波数特性をもったフィルタをかけていることになる。このようなフィルタは、その規則的な周波数特性が楕円の形状に似ているため、楕形フィルタと呼ばれる。伸縮率の差が小さい応答をミキシングした際には、この楕形フィルタの楕の幅がゆっくり変化してゆくため、聴感的に不自然な残響になると考えられる。

3.4.2 一定周期で時間伸張・圧縮を繰り返す方法

この方法は、図3.4に示すように、一定の周期 T で伸張と圧縮を繰り返して処理するものがある。実験では、 T は100msとし、伸縮率 a は0.2%、および、1.2%の2種で試聴した。

0.2%の時間伸張・圧縮を繰り返す場合では、個々の音色と定位に関しては、ほぼ良好であるが、広がり感は極めて小さく、特に低域では、広がり感が感じられず、ミキシング時の音色も、唸るような不自然な残響音となってしまった。

1.2%の時間伸張・圧縮を繰り返す場合は、音色は良好で、広がり感も得られるようになるが、低域の広がり感は、リファレンスより小さい。定位に関しては、片側に寄って聞こえ、ミキシング時の音色は、やはり、不自然であった。

この方法は、伸縮率を変えても、全体的に良い結果が得られない。時間伸張・圧縮を繰り返すと、リファレンスの応答と生成した応答の時間差が小さい部分が続出し、広がり感が得られないと同時に、前述の楕形フィルタの効果が強く出てミキシング時の音色を劣化させていると考えられる。

3.4.3 後部残響音初期とそれ以後で伸縮率を変える方法

図3.5に示すように、応答の後部残響音の初期 t_1 は比較的大きな伸縮率 a で時間伸縮し、それ以後は小さな伸縮率 $0.1a$ で処理を行う方法である。前述の音像融合による定位の振れを防ぐため、初期反射音部分の伸縮は行わない。

実験では、 t_1 を100ms、300msの2種とし、初期の伸縮率 a を4-64%まで、4%刻みで変えて試聴した。

個々の音色がほぼ自然なのは、 $t_1 = 100\text{ms}$ の場合は $a = 64\%$ まで、 $t_1 = 300\text{ms}$ の場

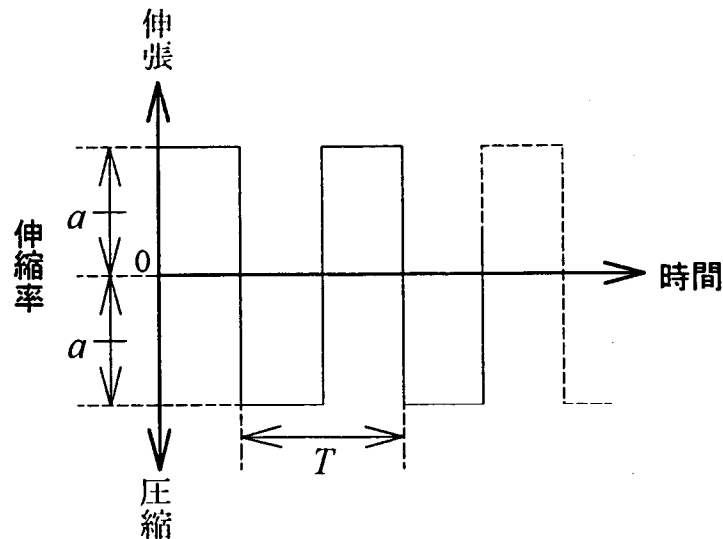


図3.4: 一定周期で時間伸張・圧縮を繰り返す方法

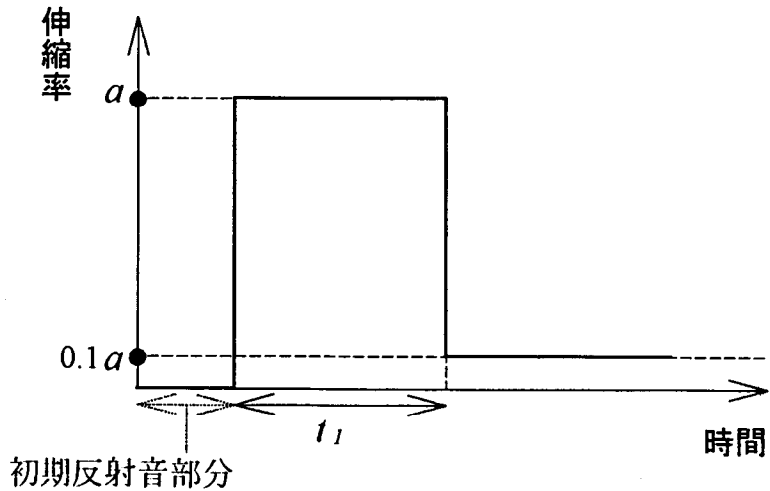


図3.5: 後部残響音初期とそれ以後で伸縮率を変える方法

合は $a = 32\%$ までである。これは、これ以上伸長した場合、残響音の密度が疎になるのが聴感的に気になり、逆に圧縮した場合は、応答のデータ長が短くなり、残響時間も短くなってしまうためである。ミキシング時の音色が良好なのは、 $t_1 = 300\text{ms}$ で伸縮率 a の差が8%以上でなければならなかった。広がり感は、 $t_1 = 100\text{ms}$ の場合8-12%以上、 $t_1 = 300\text{ms}$ の場合4%以上ならば良好に得られた。定位については、いずれのパラメータでもほぼ良好な結果が得られた。

この方法は、後部残響音初期の伸縮率を大きく取ることにより、前述の楕形フィルタの効果と考えられるミキシング時の音色の劣化を回避しつつ、応答全体の時間的長さの変化を押さえて、生成できる応答の数を増やすことをねらったが、試聴結果から、複数の応答を生成する際の伸縮率の差を大きくとらなければならず、生成可能な応答の数を増やすことは出来なかった。

§3.5 試聴実験のまとめ

様々な伸縮方法で試聴実験を行った結果、2.2.1の評価項目のすべてに良好な結果を得られるのは、一様伸縮法、後部残響音初期とそれ以後で伸縮率を変える方法であった。しかし、前者の方法では、伸縮率は10%程度までで、かつ、生成した個々の応答の伸縮率の差が4%以上でなければミキシング時の音色が劣化する。同様に、後者の方では、後部残響音の初期300msの伸縮率が32%程度までが限度で、かつ、応答間の伸縮率の差が8%以上必要であるという制限がある。このため、多数の応答を生成することは望めないが、特に一様伸縮法は処理が簡単で、音色の保存性もよい。したがって、必要な応答の数が数チャンネル分でよい場合には、有効な信号処理方法といえるだろう。

第4章

後部残響音の時間シフトによる応答の生成

前章の同様伸縮法は，新たに生成できる応答の数に制限があるものの，伸縮率を適切に選べば音質的には良好であった．この方法は，处理的には応答の後部残響音を徐々に時間シフトしていたと考えられる．そこで，この方法を簡略化し，時間シフトによって応答を生成する方法を試みた．

§4.1 後部残響音の時間シフト

処理方法を図4.1に示す．図4.1(b)のようにリファレンスの応答 $h_{org}(t)$ を T_s だけ時間シフトし，かつ，後部残響音の減衰特性が $h_{org}(t)$ と揃うようにレベル補正した応答 $h_{sft}(t)$ をつくる．さらに，§3.4で述べた音像の融合による初期音の定位のふらつきを避けるため，初期反射音の部分は $h_{org}(t)$ をそのまま用い，(c)のように後部残響音に入ったところで $h_1(t)$ にクロスフェードした応答 $h_{new}(t)$ を生成する．初期反射音の時間幅は58ms， T_c はクロスフェード時間で50msとした．レベル補正には§3.3のエンベロープ補正を用いた．

§4.2 試聴実験の結果と考察

試聴は，図2.2に示した装置を用いた．リファレンスの応答の残響時間は2.0秒である．

試聴の結果，図4.1の T_s で示したシフト時間が約30ms以上ならば2.2.1の各評価項目をほぼ満足する．しかし，シフト時間を大きくすると使用できる応答のデータが短

くなり残響時間も短くなってしまふ。エンベロープ補正にも限界があり、シフト時間の上限は応答の長さやダイナミック・レンジにもよるが数百msであると考えられる。したがって、この方法単独では、生成できる応答の数は多くはないが、時間シフトとクロスフェードという簡単な処理であるので、一様伸縮法との組み合わせで用いることにより、生成可能な応答の数を増やすことができると考えられる。

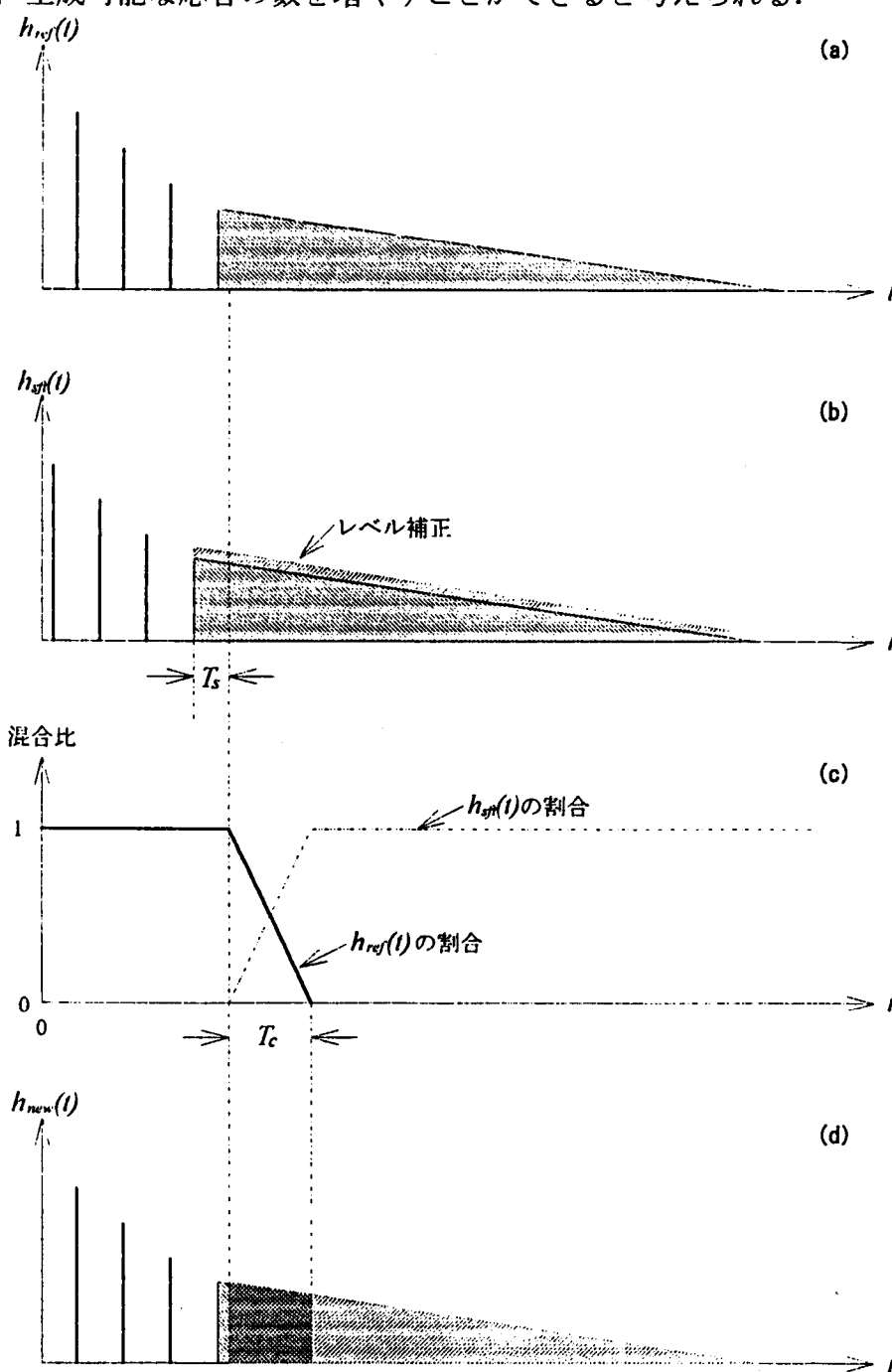


図4.1: 時間シフト法. (a) リファレンスの応答, (b) 後部残響音の時間シフトとレベル補正, (c) クロスフェード, (d) 生成した応答.

第5章

おわりに

§5.1 本研究のまとめ

スピーカウォールによる残響音の再生に必要となる、共通の音色をもった多数のインパルス応答を、原音場のインパルス応答から生成する信号処理方式を考案した。

原音場の残響音がもつ聴感的な特徴を保存し、かつ、低相関な応答を生成するために、本研究では、応答の時間軸を操作する手法を採用した。生成できる応答の数が、ほぼ無制限にある方式としてランダム・ブロック・シャッフル法を考案し、主観評価実験の結果、各種パラメータを適切に選べば音色の保存性も良好であることが確認された。また、生成する応答の数が少なくてよい場合には、一様伸縮、時間シフト、エンベロープ補正を組み合わす方式が、音色の保存性も良好で、かつ、処理も簡単であることが判った。

本研究の信号処理方式は、スピーカウォールだけでなく、放送やレコーディングのマルチチャンネル化に伴い、ニーズが高まってきているマルチチャンネルのリバーブレータの実現にも応用できると考えられる。

§5.2 今後の課題

今まで述べた処理方式には、リファレンスとする応答に重畳したノイズを抑圧する効果はない。ところが、コンサートホールなどで実測された応答の場合、音場内の暗騒音に加え、マイクロファンや増幅器などの測定装置が発生する雑音などが重畳していることが多く、こうした応答から直接新たな応答を生成しても、良好な残響音は期待できない。したがって、今後、実際の音場のインパルス応答をより高いSN比で測

定する技術や、応答からノイズを除去する処理などの研究が必要だと考えられる。
また、本研究では、反射音の後部残響音部分のみを処理の対象としたが、初期反射音の処理も行えば、再生音場内での音の定位や方向性など、より高度な音場制御も可能になると考えられる。

スピーカウォールによる音場制御を実現するには、先に述べたような問題やスピーカ配置の検討など解決すべき課題が残されているが、こうした電気音響的な音場制御は、現在の建築音響的な手法では実現の難しい、スイッチひとつで音場の特性を変えることができるといった魅力的な可能性を秘めているため、さらなる研究が期待される。

謝辞

本研究は，NHK放送技術研究所との共同研究として行われました。共同研究の機会を与えてをくださいました電気通信大学情報システム学研究科 出澤 正徳 教授，安田 稔 教授，NHK放送技術研究所 西澤 台次 所長，中林 克己 次長，音響聴覚研究部 宮坂 栄一 部長，ヒューマンサイエンス研究部 三宅 誠 部長 に深く感謝いたします。そして，本研究の直接のご指導，ご助言を賜りました電気通信大学情報システム学研究科 阪口 豊 助教授，NHK放送技術研究所 音響聴覚研究部 小宮山 摂 主任研究員，小野 一穂 研究員に深く感謝いたします。また，様々なご助言をいただいた電気通信大学情報システム学研究科 栗田 泰一郎 助教授，主観評価実験に際し，評定者としてご協力いただいた音響聴覚研究部の研究員の皆様にお礼申し上げます。

参考文献

- [1] 是永 雄二, “音場シミュレータ”, センサ技術 **9**, pp.17-23(1989).
- [2] H.F.Olson, “Home Entertainment : Audio 1988”, J.Audio Eng. Soc. Am., **17**, pp.390-404(1969).
- [3] M.Camras, “Approach to Recreating a Sound Field”, J.Acoust. Soc. Am., **43**, pp.1425-1431(1968).
- [4] 小野 一穂, 井上 友幸, 西 隆司, 中林 克己, 三上 淳一, “アレイスピーカによる音場の能動制御”, 日本音響学会秋期研究発表会講演論文集, pp.769-770(1991).
- [5] 小野 一穂, 小宮山 撰, “アレイ状に配置したスピーカによる音像・音場の制御法に関する一考察”, 信学会春期全国大会講論集, pp.528-531(1991).
- [6] Nobuharu Aoshima, “Computer-Generated Pulse Signal Applied of Sound Measurement”, J. Acoust. Soc. Am., **69**(5), pp.1484-1488(1981).
- [7] Jeffrey Borish, James B. Angell, “An Efficient Algorithm for Measuring the Impulse Response Using Pseudorandom Noise”, J. Audio Eng. Soc. Am., **31**(7), pp.478-487(1983).
- [8] ITU-R Rec.562-2, “Subjective Assessment of Sound Quality” (1986).
- [9] 城戸 健一, デジタル信号処理入門, 丸善(1985).
- [10] NHK放送技術研究所 編, マルチメディア時代のデジタル放送技術辞典, 丸善(1994).
- [11] 佐野 伸一, “SM5844AF サンプルレート・コンバータの制作”, ラジオ技術 **10**, pp.60-72, ラジオ技術社(1994).

- [12] M.R.Schroeder, "New Method of Measuring Reverberation Time", J. Acoust. Soc. Am., **37**, pp.409-412(1965).
- [13] イェンス ブラウエルト, 森本 政之, 後藤 敏幸, 空間音響, 鹿島出版会(1986).