JAIST Repository

https://dspace.jaist.ac.jp/

Title	雑音残響環境における室内音響特性のブラインド推定 法の研究
Author(s)	宮崎,晃和
Citation	
Issue Date	2015-03
Туре	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/12648
Rights	
Description	Supervisor: 鵜木 祐史, 情報科学研究科, 修士



Japan Advanced Institute of Science and Technology

修士論文

雑音残響環境における室内音響特性の ブラインド推定法の研究

北陸先端科学技術大学院大学 情報科学研究科情報科学専攻

宮崎 晃和

2015年3月

修士論文

雑音残響環境における室内音響特性の ブラインド推定法の研究

指導教員 鵜木祐史 准教授

審査委員主査	鵜木祐史 准教授
審査委員	赤木正人 教授
審査委員	党建武 教授

北陸先端科学技術大学院大学 情報科学研究科情報科学専攻

1210052 宮崎 晃和

提出年月: 2015年2月

Copyright © 2015 by Miyazaki Akikazu

概 要

音声コミュニケーションを目的として設計された室は、高い音声伝送品質を持つ必要があ る.これを知るためには、主に音声了解度といった主観評価実験を行わなければならな い.この評価実験は、非常に大がかりなものであり、労力や時間がかかる.そこで、必要 最小限の労力や時間で,かつ正確に室内音声伝送品質を評価するために客観評価尺度が ある.主要な客観評価尺度として,残響時間,D値,音声伝送指標(STI),RASTIなど があり、これらの室内音響特性の算出には室内インパルス応答(RIR)の測定が必要であ る.しかし、この測定は高い音圧レベルの信号を利用して行われるため、聴力保護の観点 から、人を排除して行われなければならない. そのため、人が常に居て人を排除できない ような環境(駅構内や空港といった公共施設など)で RIR を測定し、これらの室内音響 特性を知ることは困難である。そこで本研究ではRIRの測定を行わずに、身の回りにあ る音から室内音響特性のブラインド推定を行う手法を提案することを目的とする.これま でに、Unokiらは、変調伝達関数の概念に基づいた室内音響特性のブラインド推定法の検 討を行ってきた(以後,従来法と呼ぶ).その成果として,(1)実測のRIRがSchroeder のRIRモデルで近似できない場合でも一般化RIRモデルを用いることで正しくSTIを推 定できること、(2) 観測した残響音声から正しく STI を推定できること、(3) 人が居る環 境でもSTIを推定できることを確認した.しかし、従来法の問題点として、変調周波数 領域で推定を行った結果から、時間領域の評価尺度である残響時間、D値の推定を行って いること、音声信号を利用した室内音響特性の推定において、時間的なパワーエンベロー プの山を少なくとも二つ以上取り出す必要があるため、ある程度の時間長の信号が必要で あること、室に暗騒音や背景雑音がないものとして残響による影響のみの音環境を想定し ていたため、背景雑音のある環境で室内音響特性の推定を行った場合に、雑音の影響を受 け室内音響特性の推定結果に誤差が生じる問題があった.本研究では、上述した従来法の 問題点を解決するために、時間領域の評価指標において最適な一般化 RIR モデルのパラ メータの検討,時間的なパワーエンベロープの山を一つだけ利用する方法の検討,実環境 で用いる場合を考え雑音の影響を考慮した室内音響特性の推定法への拡張を行い、雑音残 響環境における室内音響特性のブラインド推定法を提案した.提案法の推定評価を雑音残 響 AM 信号と雑音残響音声信号を用いて行った.その結果,STI,RASTI については雑 音残響環境においても高精度に推定可能であることが明らかになった.また,残響に関す る時間領域の評価指標である残響時間T_RとD値は、雑音残響AM信号において、雑音の 影響を受けずに提案法で推定可能であることを示した.

目 次

第1章	序論	1
1.1	はじめに	1
1.2	室内音響特性の主観評価尺度	2
	1.2.1 明瞭度	2
	1.2.2 了解度	2
	1.2.3 聴き取りにくさ	2
1.3	室内音響特性の客観評価尺度	3
	1.3.1 明瞭度指数(Articulation Index, AI)	3
	1.3.2 残響時間 (Reverberation Time, RT)	3
	1.3.3 D値(Deutlichkeit)	3
	1.3.4 音声伝送指標(STI)	3
	1.3.5 RASTI	4
1.4	関連研究....................................	4
1.5	問題点....................................	4
1.6	本研究の目的	5
1.7	本論文の構成	5
第2章	室内音響特性の算出方法	6
2.1	残響時間の算出方法	6
2.2	D 値の算出方法	6
2.3	STI の算出方法	6
	2.3.1 変調伝達関数(MTF)の概念	6
	2.3.2 MTF に基づく雑音・残響信号のモデル化	7
	2.3.3 残響環境における MTF	8
	2.3.4 雑音環境における MTF	10
	2.3.5 雑音残響環境における MTF	11
	2.3.6 STIの算出	11
2.4	RASTIの算出方法	14
生っ辛	<i>祥 </i>	17
売り早 91	ルイム 安内インパルフトダ(DID)エデルの一郎化	17
ა.1 ვე	エバイン/ パレハル合 (IUII) ビノノビジー 限1 ·····家内 - 一 郷 特性の ブライン ド 推 完 注 の 柵 更	17

	3.3 3 4	残響音 問題占	· 声信号による MTF 推定	19 19
<i>h</i> .h-	0. 1			10
第4	1草	提案法		23
	4.1	MTF 7		23
		4.1.1		23
		4.1.2	MTFの最適な近似	24
		4.1.3	MTF 推定の条件設定	24
	4.2	首声信		29
		4.2.1	MTF 推定に利用する時間的なエンベローブの検討	29
		4.2.2	人工的な信号による変調スペクトルの調査	30
		4.2.3	音声信号による MTF 推定	30
	4.3	雑音の	影響への対応	30
		4.3.1	雑音が MTF 推定に与える影響	30
		4.3.2	雑音の影響を考慮した室内音響特性の推定法	31
		4.3.3	雑音の影響を考慮した MTF 推定の確認	31
第5	5章	室内音	響特性の推定評価	37
	5.1	AM 信	号を用いた評価	37
		5.1.1	評価条件	37
		5.1.2	残響時間推定結果	37
		5.1.3	D 值推定結果	37
		5.1.4	STI 推定結果	38
		5.1.5	RASTI 推定結果	38
	5.2	音声信	・ 号を用いた評価(残響信号にデータベースの RIR を用いた場合)	43
		5.2.1	評価条件	43
		5.2.2	残響時間推定結果	43
		5.2.3	D 値推定結果	43
		5.2.4	STI 推定結果	43
		5.2.5	RASTI 推定結果	43
	5.3	音声信	「号を用いた評価(残響信号に大学内で測定した RIR を用いた場合)	49
		5.3.1	評価条件..................................	49
		5.3.2	残響時間推定結果	49
		5.3.3	D 値推定結果	49
		5.3.4	STI 推定結果	49
		5.3.5	RASTI 推定結果	49
	5.4	考察	·····	50
	5.5	まとめ		52

第6章	結論	54
6.1	本研究で明らかになったこと	54
6.2	残された課題	55
付録		61
付録A	評価に利用した実測室内インパルス応答(RIR)の詳細	61
付録B	4.1 節で行った RIR モデルの近似結果の詳細	66

図目次

2.1	残響時間の定義	7
2.2	D 値の定義	8
2.3	音声明瞭度予測理論の概要図	9
2.4	残響環境の変調伝達関数 $m_R(f_m)$ の特性	10
2.5	雑音環境の変調伝達関数 $m_N(f_m)$ の特性	11
2.6	雑音残響環境の変調伝達関数 f_m の特性	12
2.7	STI 算出方法のブロックダイアグラム	13
2.8	RASTI 算出方法のブロックダイアグラム	16
3.1	RIR モデルの近似例	18
3.2	従来法のブロックダイアグラム	19
3.3	MTF 推定のコンセプト : (a) 入力信号の変調スペクトル, (b) 出力信号の	
	変調スペクトル.実線は残響の $\operatorname{MTF}m_R(f_m)$ を示す..........	20
3.4	パワーエンベロープと変調スペクトルの関係 : (a) 山のピークの間隔 0.1 s,	
	(b)山のピークの間隔 0.5 s, (c)エンベロープ間の間隔 1 s	21
3.5	帯域分割された音声信号のパワーエンベロープと変調スペクトルの例	22
4.1	時間領域,線形スケールフィッティングでの残響時間算出結果	25
4.2	時間領域,対数スケールフィッティングでの残響時間算出結果	25
4.3	時間領域,線形スケールフィッティングでの D 値算出結果	26
4.4	時間領域,対数スケールフィッティングでの D 値算出結果	26
1 5		$_{20}$
4.0	変調周波数領域、線形スケールでの残響時間算出結果	20 27
$4.5 \\ 4.6$	変調周波数領域,線形スケールでの残響時間算出結果 変調周波数領域,対数スケールでの残響時間算出結果	20 27 27
4.54.64.7	変調周波数領域,線形スケールでの残響時間算出結果 変調周波数領域,対数スケールでの残響時間算出結果 変調周波数領域,線形スケールでのD値算出結果	20 27 27 28
4.54.64.74.8	変調周波数領域,線形スケールでの残響時間算出結果 変調周波数領域,対数スケールでの残響時間算出結果 変調周波数領域,線形スケールでのD値算出結果 変調周波数領域,対数スケールでのD値算出結果	20 27 27 28 28
 4.5 4.6 4.7 4.8 4.9 	変調周波数領域,線形スケールでの残響時間算出結果	20 27 27 28 28 28 29
 4.5 4.6 4.7 4.8 4.9 4.10 	変調周波数領域,線形スケールでの残響時間算出結果 変調周波数領域,対数スケールでの残響時間算出結果 変調周波数領域,線形スケールでのD値算出結果 変調周波数領域,対数スケールでのD値算出結果 従来法で注意する必要がある入力信号のMTF パワーエンベロープがインパルス信号1本と2本の場合のMTF	 20 27 27 28 28 29 32
 4.5 4.6 4.7 4.8 4.9 4.10 4.11 	変調周波数領域,線形スケールでの残響時間算出結果 変調周波数領域,対数スケールでの残響時間算出結果 変調周波数領域,線形スケールでのD値算出結果 変調周波数領域,対数スケールでのD値算出結果 従来法で注意する必要がある入力信号のMTF パワーエンベロープがインパルス信号1本と2本の場合のMTF パワーエンベロープが残響を付加した信号1本と2本の場合のMTF	20 27 27 28 28 29 32 33
4.5 4.6 4.7 4.8 4.9 4.10 4.11 4.12	変調周波数領域,線形スケールでの残響時間算出結果 変調周波数領域,対数スケールでの残響時間算出結果 変調周波数領域,線形スケールでのD値算出結果 変調周波数領域,対数スケールでのD値算出結果 従来法で注意する必要がある入力信号のMTF パワーエンベロープがインパルス信号1本と2本の場合のMTF パワーエンベロープが残響を付加した信号1本と2本の場合のMTF パワーエンベロープから山を一つ取り出した場合と二つ取り出した場合の	20 27 27 28 28 29 32 33
4.5 4.6 4.7 4.8 4.9 4.10 4.11 4.12	変調周波数領域,線形スケールでの残響時間算出結果 変調周波数領域,対数スケールでの残響時間算出結果 変調周波数領域,線形スケールでのD値算出結果 変調周波数領域,対数スケールでのD値算出結果 従来法で注意する必要がある入力信号のMTF パワーエンベロープがインパルス信号1本と2本の場合のMTF パワーエンベロープが残響を付加した信号1本と2本の場合のMTF パワーエンベロープから山を一つ取り出した場合と二つ取り出した場合の MTFと残響環境のMTF	20 27 27 28 28 29 32 33 34
4.5 4.6 4.7 4.8 4.9 4.10 4.11 4.12 4.13	 変調周波数領域,線形スケールでの残響時間算出結果 変調周波数領域,対数スケールでの及響時間算出結果 変調周波数領域,線形スケールでのD値算出結果 変調周波数領域,対数スケールでのD値算出結果 従来法で注意する必要がある入力信号のMTF パワーエンベロープがインパルス信号1本と2本の場合のMTF パワーエンベロープが残響を付加した信号1本と2本の場合のMTF パワーエンベロープから山を一つ取り出した場合と二つ取り出した場合のMTF 従来法にSNR = 20 dBの雑音残響信号を入力したときのMTF 	 20 27 27 28 28 29 32 33 34 35

4.15 提案法に SNR = 20 dB の雑音残響信号を入力したときの MTF 36

5.1	雑音残響 AM 信号からの残響時間 T _R 推定の結果 (SNR = 20 dB)	39
5.2	雑音残響 AM 信号からの残響時間 T_R 推定の結果 (SNR = 5 dB)	39
5.3	雑音残響 AM 信号からの D 値推定の結果 (SNR = 20 dB)	40
5.4	雑音残響 AM 信号からの D 値推定の結果 (SNR = 5 dB)	40
5.5	雑音残響 AM 信号からの推定の結果 (SNR = 20 dB)	41
5.6	雑音残響 AM 信号からの STI 推定の結果 (SNR = 5 dB)	41
5.7	雑音残響 AM 信号からの RASTI 推定の結果 (SNR = 20 dB)	42
5.8	雑音残響 AM 信号からの RASTI 推定の結果 (SNR = 5 dB)	42
5.9	雑音残響音声信号(SMILE2004)からの残響時間 <i>T_R</i> 推定の結果 (SNR =	
	20 dB)	44
5.10	雑音残響音声信号(SMILE2004)からの残響時間 <i>T_R</i> 推定の結果 (SNR =	
	5 dB)	44
5.11	雑音残響音声信号(SMILE2004)からの残響時間T _R 推定の結果 (提案法	
	$\mathcal{O}\mathcal{P}$, SNR = 20 dB)	45
5.12	雑音残響音声信号(SMILE2004)からの残響時間T _R 推定の結果 (提案法	
	$\mathcal{O}\mathcal{P}$, SNR = 5 dB)	45
5.13	雑音残響音声信号(SMILE2004)からの D 値推定の結果 (SNR = 20 dB)	46
5.14	雑音残響音声信号(SMILE2004)からの D 値推定の結果 (SNR = 5 dB)	46
5.15	雑音残響音声信号(SMILE2004)からの STI 推定の結果 (SNR = 20 dB)	47
5.16	雑音残響音声信号(SMILE2004)からの STI 推定の結果 (SNR = 5 dB)	47
5.17	雑音残響音声信号(SMILE2004)からの RASTI 推定の結果 (SNR = 20	
	dB)	48
5.18	雑音残響音声信号(SMILE2004)からの RASTI 推定の結果 (SNR = 5 dB)	48
5.19	雑音残響音声信号(大学内の室)からの残響時間 T_R 推定の結果	50
5.20	雑音残響音声信号(大学内の室)からのD値推定の結果	51
5.21	雑音残響音声信号(大学内の室)からの推定の結果	52
5.22	雑音残響音声信号(大学内の室)からの推定の結果	53
Λ 1	十学内で測定した実測 RIR (No. 44~40)の字環境	69
A.1	大子門で側定した実側 $\operatorname{RIR}(No. 44°49)$ の主衆現 \ldots \ldots \ldots + 学内で測定した実測 $\operatorname{RIR}(No. 50~53)$ の字環境	62 63
A.2	八千円 () () () () () () () () () (05
B.1	時間領域,線形スケールで近似した結果(No. 1)	67
B.2	時間領域,線形スケールで近似した結果(No. 2)	67
B.3	時間領域,線形スケールで近似した結果(No. 3)	67
B.4	時間領域,線形スケールで近似した結果(No. 4)	67
B.5	時間領域,線形スケールで近似した結果(No. 5)	68
B.6	時間領域,線形スケールで近似した結果(No. 6)	68
B.7	時間領域,線形スケールで近似した結果(No. 7)	68

時間領域,	線形スケールで近似した結果	(No.	8)		68
時間領域,	線形スケールで近似した結果	(No.	9)		69
時間領域,	線形スケールで近似した結果	(No.	10)		69
時間領域,	線形スケールで近似した結果	(No.	11)		69
時間領域,	線形スケールで近似した結果	(No.	12)		69
時間領域,	線形スケールで近似した結果	(No.	13)		70
時間領域,	線形スケールで近似した結果	(No.	14)		70
時間領域,	線形スケールで近似した結果	(No.	15)		70
時間領域,	線形スケールで近似した結果	(No.	16)		70
時間領域,	線形スケールで近似した結果	(No.	17)		71
時間領域,	線形スケールで近似した結果	(No.	18)		71
時間領域,	線形スケールで近似した結果	(No.	19)		71
時間領域,	線形スケールで近似した結果	(No.	20)		71
時間領域,	線形スケールで近似した結果	(No.	21)		72
時間領域,	線形スケールで近似した結果	(No.	22)		72
時間領域,	線形スケールで近似した結果	(No.	23)		72
時間領域,	線形スケールで近似した結果	(No.	24)		72
時間領域,	線形スケールで近似した結果	(No.	25)		73
時間領域,	線形スケールで近似した結果	(No.	26)		73
時間領域,	線形スケールで近似した結果	(No.	27)		73
時間領域,	線形スケールで近似した結果	(No.	28)		73
時間領域,	線形スケールで近似した結果	(No.	29)		74
時間領域,	線形スケールで近似した結果	(No.	30)		74
時間領域,	線形スケールで近似した結果	(No.	31)		74
時間領域,	線形スケールで近似した結果	(No.	32)		74
時間領域,	線形スケールで近似した結果	(No.	33)		75
時間領域,	線形スケールで近似した結果	(No.	34)		75
時間領域,	線形スケールで近似した結果	(No.	35)		75
時間領域,	線形スケールで近似した結果	(No.	36)		75
時間領域,	線形スケールで近似した結果	(No.	37)		76
時間領域,	線形スケールで近似した結果	(No.	38)		76
時間領域,	線形スケールで近似した結果	(No.	39)		76
時間領域,	線形スケールで近似した結果	(No.	40)		76
時間領域,	線形スケールで近似した結果	(No.	41)		77
時間領域,	線形スケールで近似した結果	(No.	42)		77
時間領域,	線形スケールで近似した結果	(No.	43)		77
時間領域,	対数スケールで近似した結果	(No.	1)		78
時間領域,	対数スケールで近似した結果	(No.	2)		78
	時時時時時時時時時時時時時時時時時時時時時時時時時時時時時時時時時時時時時時時	時間領域、線形スケールで近似した結結果 時間領域、線形スケールで近似した結結果 時間間領域、線形スケールで近似した結結果	時間領域,線形スケールで近似した結果 (No. 時間領域,線形スケールで近似した結果 (No.	時間領域,線形スケールで近似した結果 (No. 8) 時間領域,線形スケールで近似した結果 (No. 10) 時間領域,線形スケールで近似した結果 (No. 11) 時間領域,線形スケールで近似した結果 (No. 13) 時間領域,線形スケールで近似した結果 (No. 13) 時間領域,線形スケールで近似した結果 (No. 14) 時間領域,線形スケールで近似した結果 (No. 15) 時間領域,線形スケールで近似した結果 (No. 16) 時間領域,線形スケールで近似した結果 (No. 17) 時間領域,線形スケールで近似した結果 (No. 17) 時間領域,線形スケールで近似した結果 (No. 17) 時間領域,線形スケールで近似した結果 (No. 19) 時間領域,線形スケールで近似した結果 (No. 20) 時間領域,線形スケールで近似した結果 (No. 21) 時間領域,線形スケールで近似した結果 (No. 21) 時間領域,線形スケールで近似した結果 (No. 22) 時間領域,線形スケールで近似した結果 (No. 23) 時間領域,線形スケールで近似した結果 (No. 24) 時間領域,線形スケールで近似した結果 (No. 25) 時間領域,線形スケールで近似した結果 (No. 26) 時間領域,線形スケールで近似した結果 (No. 27) 時間領域,線形スケールで近似した結果 (No. 27) 時間領域,線形スケールで近似した結果 (No. 30) 時間領域,線形スケールで近似した結果 (No. 33) 時間領域,線形スケールで近似した結果 (No. 34) 時間領域,線形スケールで近似した結果 (No. 35) 時間領域,線形スケールで近似した結果 (No. 37) 時間領域,線形スケールで近似した結果 (No. 33) 時間領域,線形スケールで近似した結果 (No. 34) 時間領域,線形スケールで近似した結果 (No. 34) 時間領域,線形スケールで近似した結果 (No. 34) 時間領域,線形スケールで近似した結果 (No. 34) 時間領域,線形スケールで近似した結果 (No. 34) 時間領域,線形スケールで近似した結果 (No. 34)	時間領域,線形スケールで近似した結果 (No. 8)

時間領域,	対数スケールで近似した結果	(No.	3) .		78
時間領域,	対数スケールで近似した結果	(No.	4) .		78
時間領域,	対数スケールで近似した結果	(No.	5) .		79
時間領域,	対数スケールで近似した結果	(No.	6) .		79
時間領域,	対数スケールで近似した結果	(No.	7) .		79
時間領域,	対数スケールで近似した結果	(No.	8) .		79
時間領域,	対数スケールで近似した結果	(No.	9) .		80
時間領域,	対数スケールで近似した結果	(No.	10)		80
時間領域,	対数スケールで近似した結果	(No.	11)		80
時間領域,	対数スケールで近似した結果	(No.	12)		80
時間領域,	対数スケールで近似した結果	(No.	13)		81
時間領域,	対数スケールで近似した結果	(No.	14)		81
時間領域,	対数スケールで近似した結果	(No.	15)		81
時間領域,	対数スケールで近似した結果	(No.	16)		81
時間領域,	対数スケールで近似した結果	(No.	17)		82
時間領域,	対数スケールで近似した結果	(No.	18)		82
時間領域,	対数スケールで近似した結果	(No.	19)		82
時間領域,	対数スケールで近似した結果	(No.	20)		82
時間領域,	対数スケールで近似した結果	(No.	21)		83
時間領域,	対数スケールで近似した結果	(No.	22)		83
時間領域,	対数スケールで近似した結果	(No.	23)		83
時間領域,	対数スケールで近似した結果	(No.	24)		83
時間領域,	対数スケールで近似した結果	(No.	25)		84
時間領域,	対数スケールで近似した結果	(No.	26)		84
時間領域,	対数スケールで近似した結果	(No.	27)		84
時間領域,	対数スケールで近似した結果	(No.	28)		84
時間領域,	対数スケールで近似した結果	(No.	29)		85
時間領域,	対数スケールで近似した結果	(No.	30)		85
時間領域,	対数スケールで近似した結果	(No.	31)		85
時間領域,	対数スケールで近似した結果	(No.	32)		85
時間領域,	対数スケールで近似した結果	(No.	33)		86
時間領域,	対数スケールで近似した結果	(No.	34)		86
時間領域,	対数スケールで近似した結果	(No.	35)		86
時間領域,	対数スケールで近似した結果	(No.	36)		86
時間領域,	対数スケールで近似した結果	(No.	37)		87
時間領域,	対数スケールで近似した結果	(No.	38)		87
時間領域,	対数スケールで近似した結果	(No.	39)		87
時間領域,	対数スケールで近似した結果	(No.	40)		87
	時時時時時時時時時時時時時時時時時時時時時時時時時時時時時時時時時時時時時時時	時間領領域、対数スケールで近似した結結果果時間領領域、対数スケールで近似した結結果果時間領領域、対数スケールで近似した結結果の一次で近似した結結果時間領領域、対数スケールで近似した結結果の一次で近似したた結果果時間間領領域、対数スケールで近似した結結果と時間間領領域、対数スケールで近似したた結結果の一次で近似したた結結果の一次で近似したた結結果の一次で近似したた結結果の一次で近似したた結結果の一次で近似したた結結果の一次で近似したた結結果の一次で近似したた結結果の一次で近似したた結結果の一次で近似したた結結果の一次で近似したた結結果の一次対数スケールで近似したた結結果の一次対数スケールで近似したた結結果の一次対数スケールで近似した結結果との一次対数スケールで近似した結結果との一次対数スケールで近似した結結果との一次対数スケールで近似した結結果との一次対数スケールで近似した結結果との一次対数スケールで近似した結結果との一次対数スケールで近似した結結果との一が対数スケールで近似した結結果との一次対数スケールで近似した結結果との一次対数スケールで近似した結結果との一次対数スケールで近似した結結果との一次対数スケールで近似した結結果との一次対数スケールで近似した結結果との一次対数スケールで近似した結結果との一次対数スケールで近似した結結果との一次が近似した結果との一人で近似した結結果との一人で近似した結素との一人で近似した結素との一人ので近似した結素との一人で近似した結素との一人ので近似したに結果との一人ので近似したに結果との一人ので近似した結果との一人ので近似したに結果との一人ので近似したに結果との一人ので近似したに結果との一人ので近似した結果との一人ので近似したに結果との一人ので近似したに結果との一人ので近似した結果との一人ので近似したに結果との一人ので近似したに結果との一人の一人の一人の一人の一人の一人の一人に結果との一人の一人の一人の一人にに結果との一人の一人に近似したに結果との一人の一人の一人の一人の一人にに結果との一人の一人の一人に近似したに結果との一人の一人の一人の一人の一人にに結果との一人の一人の一人に近似したに結果との一人の一人の一人の一人にに結果との一人の一人の一人の一人にににに、	時間領域、対数スケールで近似した結果 (No. 時間領域、対数スケールで近似した結果 (No.	時間領域,対数スケールで近似した結果 (No. 3) . 時間領域,対数スケールで近似した結果 (No. 5) . 時間領域,対数スケールで近似した結果 (No. 6) . 時間領域,対数スケールで近似した結果 (No. 7) . 時間領域,対数スケールで近似した結果 (No. 9) . 時間領域,対数スケールで近似した結果 (No. 9) . 時間領域,対数スケールで近似した結果 (No. 10) 時間領域,対数スケールで近似した結果 (No. 11) 時間領域,対数スケールで近似した結果 (No. 12) 時間領域,対数スケールで近似した結果 (No. 13) 時間領域,対数スケールで近似した結果 (No. 13) 時間領域,対数スケールで近似した結果 (No. 13) 時間領域,対数スケールで近似した結果 (No. 14) 時間領域,対数スケールで近似した結果 (No. 15) 時間領域,対数スケールで近似した結果 (No. 16) 時間領域,対数スケールで近似した結果 (No. 17) 時間領域,対数スケールで近似した結果 (No. 17) 時間領域,対数スケールで近似した結果 (No. 20) 時間領域,対数スケールで近似した結果 (No. 21) 時間領域,対数スケールで近似した結果 (No. 22) 時間領域,対数スケールで近似した結果 (No. 23) 時間領域,対数スケールで近似した結果 (No. 23) 時間領域,対数スケールで近似した結果 (No. 23) 時間領域,対数スケールで近似した結果 (No. 23) 時間領域,対数スケールで近似した結果 (No. 23) 時間領域,対数スケールで近似した結果 (No. 23) 時間領域,対数スケールで近似した結果 (No. 30) 時間領域,対数スケールで近似した結果 (No. 31) 時間領域,対数スケールで近似した結果 (No. 33) 時間領域,対数スケールで近似した結果 (No. 33)	時間領域、対数スケールで近似した結果(No. 3)

B.84 時間領域,対数ス	ケールで近似した結果(No. 4	41)	 88
B.85 時間領域,対数ス	ケールで近似した結果(No. 4	42)	 88
B.86 時間領域,対数ス	ケールで近似した結果(No. 4	43)	 88
B.87 変調周波数領域,	線形スケールで近似した結果	(No. 1)	 89
B.88 変調周波数領域,	線形スケールで近似した結果	(No. 2)	 89
B.89 変調周波数領域,	線形スケールで近似した結果	(No. 3)	 89
B.90 変調周波数領域,	線形スケールで近似した結果	(No. 4)	 89
B.91 変調周波数領域,	線形スケールで近似した結果	(No. 5)	 90
B.92 変調周波数領域,	線形スケールで近似した結果	(No. 6)	 90
B.93 変調周波数領域,	線形スケールで近似した結果	(No. 7)	 90
B.94 変調周波数領域,	線形スケールで近似した結果	(No. 8)	 90
B.95 変調周波数領域,	線形スケールで近似した結果	(No. 9)	 91
B.96 変調周波数領域,	線形スケールで近似した結果	(No. 10)	 91
B.97 変調周波数領域,	線形スケールで近似した結果	(No. 11)	 91
B.98 変調周波数領域,	線形スケールで近似した結果	(No. 12)	 91
B.99 変調周波数領域,	線形スケールで近似した結果	(No. 13)	 92
B.100変調周波数領域,	線形スケールで近似した結果	(No. 14)	 92
B.101変調周波数領域,	線形スケールで近似した結果	(No. 15)	 92
B.102変調周波数領域,	線形スケールで近似した結果	(No. 16)	 92
B.103変調周波数領域,	線形スケールで近似した結果	(No. 17)	 93
B.104変調周波数領域,	線形スケールで近似した結果	(No. 18)	 93
B.105変調周波数領域,	線形スケールで近似した結果	(No. 19)	 93
B.106変調周波数領域,	線形スケールで近似した結果	(No. 20)	 93
B.107変調周波数領域,	線形スケールで近似した結果	(No. 21)	 94
B.108変調周波数領域,	線形スケールで近似した結果	(No. 22)	 94
B.109変調周波数領域,	線形スケールで近似した結果	(No. 23)	 94
B.110変調周波数領域,	線形スケールで近似した結果	(No. 24)	 94
B.111変調周波数領域,	線形スケールで近似した結果	(No. 25)	 95
B.112変調周波数領域,	線形スケールで近似した結果	(No. 26)	 95
B.113変調周波数領域,	線形スケールで近似した結果	(No. 27)	 95
B.114変調周波数領域,	線形スケールで近似した結果	(No. 28)	 95
B.115変調周波数領域,	線形スケールで近似した結果	(No. 29)	 96
B.116変調周波数領域,	線形スケールで近似した結果	(No. 30)	 96
B.117変調周波数領域,	線形スケールで近似した結果	(No. 31)	 96
B.118変調周波数領域,	線形スケールで近似した結果	(No. 32)	 96
B.119変調周波数領域,	線形スケールで近似した結果	(No. 33)	 97
B.120変調周波数領域,	線形スケールで近似した結果	(No. 34)	 97
B.121変調周波数領域,	線形スケールで近似した結果	(No. 35)	 97

B.122変調周波数領域,	線形スケールで近似した結果	(No. 36)	 97
B.123変調周波数領域,	線形スケールで近似した結果	(No. 37)	 98
B.124変調周波数領域,	線形スケールで近似した結果	(No. 38)	 98
B.125変調周波数領域,	線形スケールで近似した結果	(No. 39)	 98
B.126変調周波数領域,	線形スケールで近似した結果	(No. 40)	 98
B.127変調周波数領域,	線形スケールで近似した結果	(No. 41)	 99
B.128変調周波数領域,	線形スケールで近似した結果	(No. 42)	 99
B.129変調周波数領域,	線形スケールで近似した結果	(No. 43)	 99
B.130変調周波数領域,	対数スケールで近似した結果	(No. 1)	 100
B.131変調周波数領域,	対数スケールで近似した結果	(No. 2)	 100
B.132変調周波数領域,	対数スケールで近似した結果	(No. 3)	 100
B.133変調周波数領域,	対数スケールで近似した結果	(No. 4)	 100
B.134変調周波数領域,	対数スケールで近似した結果	(No. 5)	 101
B.135変調周波数領域,	対数スケールで近似した結果	(No. 6)	 101
B.136変調周波数領域,	対数スケールで近似した結果	(No. 7)	 101
B.137変調周波数領域,	対数スケールで近似した結果	(No. 8)	 101
B.138変調周波数領域,	対数スケールで近似した結果	(No. 9)	 102
B.139変調周波数領域,	対数スケールで近似した結果	(No. 10)	 102
B.140変調周波数領域,	対数スケールで近似した結果	(No. 11)	 102
B.141変調周波数領域,	対数スケールで近似した結果	(No. 12)	 102
B.142変調周波数領域,	対数スケールで近似した結果	(No. 13)	 103
B.143変調周波数領域,	対数スケールで近似した結果	(No. 14)	 103
B.14-変調周波数領域,	対数スケールで近似した結果	(No. 15)	 103
B.145変調周波数領域,	対数スケールで近似した結果	(No. 16)	 103
B.146変調周波数領域,	対数スケールで近似した結果	(No. 17)	 104
B.147変調周波数領域,	対数スケールで近似した結果	(No. 18)	 104
B.148変調周波数領域,	対数スケールで近似した結果	(No. 19)	 104
B.149変調周波数領域,	対数スケールで近似した結果	(No. 20)	 104
B.150変調周波数領域,	対数スケールで近似した結果	(No. 21)	 105
B.151変調周波数領域,	対数スケールで近似した結果	(No. 22)	 105
B.152変調周波数領域,	対数スケールで近似した結果	(No. 23)	 105
B.153変調周波数領域,	対数スケールで近似した結果	(No. 24)	 105
B.154変調周波数領域,	対数スケールで近似した結果	(No. 25)	 106
B.155変調周波数領域,	対数スケールで近似した結果	(No. 26)	 106
B.156変調周波数領域,	対数スケールで近似した結果	(No. 27)	 106
B.157変調周波数領域,	対数スケールで近似した結果	(No. 28)	 106
B.158変調周波数領域,	対数スケールで近似した結果	(No. 29)	 107
B.159変調周波数領域,	対数スケールで近似した結果	(No. 30)	 107

B.160変調周波数領域,	対数スケールで近似した結果	(No. 31)	107
B.161変調周波数領域,	対数スケールで近似した結果	(No. 32)	107
B.162変調周波数領域,	対数スケールで近似した結果	(No. 33)	108
B.163変調周波数領域,	対数スケールで近似した結果	(No. 34)	108
B.164変調周波数領域,	対数スケールで近似した結果	(No. 35)	108
B.165変調周波数領域,	対数スケールで近似した結果	(No. 36)	108
B.166変調周波数領域,	対数スケールで近似した結果	(No. 37)	109
B.167変調周波数領域,	対数スケールで近似した結果	(No. 38)	109
B.168変調周波数領域,	対数スケールで近似した結果	(No. 39)	109
B.169変調周波数領域,	対数スケールで近似した結果	(No. 40)	109
B.170変調周波数領域,	対数スケールで近似した結果	(No. 41)	110
B.171変調周波数領域,	対数スケールで近似した結果	(No. 42)	110
B.172変調周波数領域,	対数スケールで近似した結果	(No. 43)	110

表目次

1.1	室の音声伝送品質と STI の関係	4
$2.1 \\ 2.2$	STIのMTFデータマトリクス RASTIのMTFデータマトリクス	$\frac{15}{16}$
A.1 A.2	SMILE2004 データベースの各 RIR 条件	$\begin{array}{c} 64 \\ 65 \end{array}$

第1章 序論

1.1 はじめに

室の設計を行う際には、利用目的に合わせた室内音響特性を持つように設計する必要が ある.たとえば、コンサートホールなどでは楽器音が客席で美しく聞こえるような響きを 重視して設計しなければならない.また教室や講演室のような室では講師や講演者の音声 を受聴者に聞こえやすくするために響きすぎないような設計をしなければらない.そのた め、設計段階から室の用途に合わせて、室の形や壁や天井に使われる材質の吸音率を考え コンピュータや模型によるシミュレーションを行いながら室内音響の設計を行っている. そして、実際に完成した室が設計通りの室内音響特性を持っているのか測定を行い、もし 室内音響特性が設計値以下である場合には吸音材や反射板の追加などで対策を行う必要 がある.

室内音響特性には、主観評価尺度と客観評価尺度を利用したものがある. 実際にコン サートホールや教室,講演室で音を聞くのは人であるため,単純な物理現象ではなく,人 がどのような聴覚的印象を持つのかを評価する必要がある. 主観評価尺度では人の聴覚 的印象を聴取実験により評価を行い,客観評価尺度では測定した物理指標から人の聴覚 的印象と相関のある指標を求め評価するものである. 主観評価尺度には、明瞭度や了解 度,聴き取りにくさといった評価尺度がある.これら主観評価尺度の測定には被験者によ る聴取実験によって人が直接感じる品質を評価を行う必要がある.人が直接的に部屋の 評価を行うため非常に重要な評価尺度である.しかし、被験者の聴力検査の実施や個人 差による評価のばらつきを解消するために多数の被験者を用意しなければならない.ま た、信号によって人の感じ方が異なるために多数の信号を用いて評価をしなければなら ないといった問題があるため、労力や時間がかかる.客観評価尺度には、残響時間、D値 (Deutlichkeit) [8], 音声伝送指標 [13] といった評価尺度がある.客観評価尺度は,音の物 理的な特徴から主観評価で得られる結果と同等の結果を求めるものである.これにより、 多くの被験者を利用した聴取実験を行うことなく室内音響の測定解析用機材を用いること で主観評価よりもコストや時間を大幅に削減し簡便に得ることができる。また、同じ機材 を用いれば同じ結果が得られるため、異なる場所や日時間での結果の比較を行う際にも客 観評価尺度は有用である、そのため、主観評価尺度と客観評価尺度は評価目的を考慮し、 適切に使い分けることが重要である.

1.2 室内音響特性の主観評価尺度

1.2.1 明瞭度

明瞭度は、音声の分節単位の一つである音節を被験者に提示する聴取実験を行い、どれ だけ正確に聴き取れたかを算出する [1, 2]. 音節には、直音、濁音、半濁音、拗音を合計 した 100 音節を利用する. この 100 音節をランダムに並べ替えたものを被験者に提示して 聴き取った結果の正答率によって音声明瞭度が求められる. 明瞭度試験には単音節を提示 する単音節明瞭度試験や音節を無意味に組み合わせて連音節を提示する連音節明瞭度試 験がある.

1.2.2 了解度

了解度は、単語または文章を被験者に提示する聴取実験を行い、どれだけ正確に聴き取 れたかの正答率を算出する[2].明瞭度との違いは、明瞭度では単音節もしくは無意味な 連音節であるのに対して了解度は有意味な単語または文章を用いる点である.有意味な音 声の場合は、一部聴き取れない音節があった場合でも前後の音節から類推することができ るため明瞭度より実際のコミュニケーションに即した評価尺度となる.

1.2.3 聴き取りにくさ

聴き取りにくさ [3, 4, 5] は、森本らが提案した音声伝送品質の評価尺度である.明瞭度 や了解度は試験で提示する音声信号によって正答率が変わるため、再現性を得るためには 試験用音声の基準を定め必要がある.そこで、実用的な音声伝送品質の評価として、より 実際の状況に近いよる試験用音声として、「なじみ」の程度を表す親密度の高い音声が考 えられる.しかし、親密度の高い音声の場合は、音声伝送品質が多少劣化していても親密 度が高いために類推しやすく了解度の正答率が 100%になってしまい、音声伝送品質が劣 化しているかどうかの評価ができない.そこで、森本らは明瞭度や了解度では差が出ない が聴覚上は差がある音声伝送品質の評価として聴き取りにくさを提案した.聴き取りにく さは高親密度の単語リストを試験音として提示し、「聴き取りにくい」、「やや聴き取りに くい」、「かなり聴き取りにくい」、「非常に聴き取りにくい」という4段階の評価尺度で評 価を行う.この聴き取りにくさの評価により、了解度では測定できない音声伝送品質の測 定が可能になった.

1.3 室内音響特性の客観評価尺度

1.3.1 明瞭度指数(Articulation Index, AI)

明瞭度指数は, French & Steinberg [6] によって提案された電話受聴における伝送路の 評価として信号対雑音比(SNR)や伝送周波数特性などの物理指標から明瞭度を求める 方法であり, Kryter [7] によって対象をスピーカー受聴とした明瞭度指数や簡易計算法が 提案された.

1.3.2 残響時間(Reverberation Time, RT)

音は、室において壁や床、天井に何度も反射して残響音を作る.このときの残響音の 度合いを数値として表したものが残響時間であり、音源から出力を止めてから音のエネル ギーが 60 dB 減衰すまでの時間として定義され T₆₀ とも表現される [11].残響時間は、壁 や床、天井の材質による吸音率や、室の表面積、体積によって変化する.そのため、室を 設計する際には、利用目的を考慮して材質や室の形状を考える必要がある.残響時間の測 定方法としては室内インパルス応答(Room Impulse Response, RIR)の減衰特性を調べ る方法がある.

1.3.3 D值(Deutlichkeit)

D値[8]は、Thieleが提案した言葉の明瞭性に関する客観評価尺度である.これは、RIR の直接音到来時から 50 ms までのエネルギーと信号の全エネルギーとの比で定義されて いる.従って、信号の全エネルギーが 50 ms 以内にある場合に最も明瞭性が高いというこ とになる.これは、直接音から 50 ms 以内の初期反射音には、直接音を聞こえやすくする 効果があるためである.

1.3.4 音声伝送指標(STI)

音声伝送指標(Speech Transmission Index, STI)は、室内音響における音声伝送品質 を評価するために利用される客観的評価尺度である[12].表1.1に示すように、室内音響 における音声伝送品質と対応しており、「音声明瞭度」の予測に利用されている[2].最近 では、STIが「音声明瞭度」よりも「聴き取りにくさ」と高い相関があることが知られて おり、室の音声伝送の評価において、その利用の重要性が増している[3,4,5].STIの計 算法は、IEC 60268-16 [13]によって標準化されている.これは、Houtgast & Steeneken [14,15]によって提案された変調伝達関数(Modulation Transfer Function, MTF)の概念 に基づいている.STIは、7つのオクターブ帯域でRIRのMTFを求め、それらの荷重和 から求められる[13].そのため、RIR(あるいはMTF)の測定が必要となる.

表 1.1: 室の音声伝送品質と STI の関係.

Quality	Bad	Poor	Fair	Good	Excellent
STI	0.0	0.3	0.45	0.6	0.75
	~ 0.3	~ 0.45	~ 0.6	~ 0.74	~ 1.0

1.3.5 RASTI

RASTI (Room Acoustic Speech Transmission Index) は,STIの簡易測定法であり,STI 同様に IEC 60268-16 [13] によって標準化されている.STI は七つのオクターブ帯域を利用するが,RASTI のそのうち 500 Hz と 2 kHz の二つの帯域,また,STI は各オクターブ帯域で 14 個の変調周波数から MTF を利用するため,計 98 個の MTF を求めるが RASTI は 5 Hz で 4 個,2 kHz では 5 個の計 9 個の MTF を利用するため STI よりも高速に求められる.

1.4 関連研究

室内音響特性のブラインド推定法について関連している研究として, 残響時間の推定法と STIの推定法がある. 残響時間の推定法として, 最尤法(Maximum Likelihood Estimation, MLE)を利用した手法[17]があるが,これは残響時間の推定に長時間の観測信号を必要と する問題がある.また, 混合ガウス分布(Gaussian Mixture Model, GMM)を利用した方 法[18, 19]や,スペクトル減衰分布を利用した方法[20]もあるが,これらは残響時間の推 定に事前にトレーニングを行ったデータの利用を必要とする問題がある.また,統計的機 械学習による推定手法として,人工ニューラルネットワーク(Artificial Neural Network, ANN)を用いて残響時間やSTIなどの室内音響特性を推定する手法[21, 22, 23]がある. これらの推定法は事前に大規模な強化学習を行う必要がある.

事前にデータを利用しない室内音響特性のブラインド推定法として、Unoki らによって MTF の概念に基づいた残響音声回復法 [24] や残響時間推定法 [25] が提案され、これらの研究結果に基づいて STI ブラインド推定法(以後、従来法と呼ぶ)が検討されてきた [26, 27, 28]. これらの研究結果から(1)実測の RIR が Schroeder の RIR モデル [29] で近似できない場合でも STI を正しく推定できること,(2) 観測した残響音声から STI を正しく推定できることを確認した.

1.5 問題点

主観評価尺度は、聴取実験を行う際に多数の評価者が必要で、実験を行うには労力と時間がかかる.また、客観評価尺度では主に RIR の測定を前提としている.しかし、測定は高い音圧レベルの信号を利用して行われるため、聴力保護の観点から人を排除して行わ

れなければならない.そのため、人が常に居て排除できないような環境(駅構内や空港といった公共施設など)で RIR を測定し、これらの室内音響特性を知ることは困難である. また、室内音響特性は人の密度や配置によって変化するため、時々刻々と変化する環境に 追従して室内音響特性を求める必要がある.このため、上述したように室内音響特性を推 定する研究が行われている.これらの関連研究において、残響環境において事前データを 用いることなく室内音響特性のブラインド推定法は提案されているが、実環境には背景雑 音も存在しているため、雑音残響環境における室内音響特性のブラインド推定を行う必要 がある.

1.6本研究の目的

RIRの測定を行い室内音響特性の算出を行う場合には、人を排除して行う必要がある. そのため、人を排除した状態で測定した室内音響特性から室の評価を行わなければならない.しかし、実際の室の運用は人がいる状態であるため、人がいる状態において設計通りの室内音響特性が確保されているか評価する必要がある.また、人の往来により時々刻々と変化する室内音響特性を求める必要がある.そこで本研究では RIR の測定を行わずに、身の回りにある音から室内音響特性のブラインド推定を行う手法を提案することを目的とする.

1.7 本論文の構成

本論文は、6章で構成される

第1章 本研究の背景,代表的な評価尺度の概要,研究の問題点と目的,また関連研究 について述べる.

第2章 客観評価尺度の算出法について述べる.

第3章 本研究で従来法とする手法の概要と問題点について述べる.

第4章 従来法における問題点を解決するための手法の提案を行う.

第5章 提案法の評価として入力信号に AM 信号と音声信号を用いた場合で行い考察 する.

第6章 本研究で明らかになったこと、残された課題について述べる.

第2章 室内音響特性の算出方法

2.1 残響時間の算出方法

残響時間は時間領域で評価する音響特性である. 図 2.1 に示すように,定常状態から発音を止めて 60 dB 減衰するまでの時間が残響時間で T_{60} と呼ばれる.また,残響時間を求める方法の一つとして,測定された RIR の減衰曲線から得る手法がある. T_{60} はダイナミックレンジの確保が必要であり,案騒音の影響により 60 dB までの減衰が測定不可能の場合には,30 dB 減衰した時間を 2 倍にして残響時間を算出する T_{30} などがある.

2.2 D 値の算出方法

D値は,図2.2に示すように直接音到来から50msまでのエネルギーと音全体のエネル ギーの比を表しており、下記に示す式(2.1)で表される.

$$D = \frac{\int_0^{50ms} h^2(t)dt}{\int_0^\infty h^2(t)dt}$$
(2.1)

2.3 STIの算出方法

2.3.1 変調伝達関数(MTF)の概念

変調伝達関数(Modulation Transfer Function, MTF)の概念は Houtgast & Steeneken [15]によって提案されたに基づいている.この概念は、図 2.3 に示すように、室内の伝達 特性を入力と出力の時間的な包絡線情報の関係で説明するものである.この概念は室内 音響を拡散音場と仮定し、室内の伝達特性を残響による影響(重畳性)と雑音による影響 (加法性)に分類した.そして、次式に示すように入力信号 x(t) のパワーエンベロープを $\overline{I_i^2}$,出力信号 y(t) のパワーエンベロープを $\overline{I_o^2}$ とした.

$$\mathbf{Input}: \ \overline{I_i^2} \left\{ 1 + \cos(2\pi f_m t) \right\}$$
(2.2)

$$\mathbf{Output}: \ \overline{I_o^2} \left\{ 1 + m(f_m) \cos(2\pi f_m(t-\theta)) \right\}$$
(2.3)



図 2.1: 残響時間の定義

ただし, $\overline{I_i^2} \ge \overline{I_o^2}$ は入出力信号の強度, f_m は変調周波数, θ は位相である. このとき, 入力信号のパワーエンベロープ $\overline{I_i^2}$ は変調度 $m(f_m) = 1$ (100% 振幅変調) であり, $\overline{I_o^2}$ の変調度 $m(f_m)$ が入出力間の MTF を表している.

2.3.2 MTFに基づく雑音・残響信号のモデル化

MTF における雑音・残響信号のモデルは下記のようになる.まず,入力信号 $\mathbf{x}(t)$,出力 信号 $\mathbf{y}(t)$,残響インパルス応答 $\mathbf{h}(t)$,雑音信号 $\mathbf{n}(t)$ をそれぞれ次のようにモデル化する.

$$\mathbf{x}(t) = e_x(t)\mathbf{c}_x(t) \tag{2.4}$$

$$\mathbf{y}(t) = \mathbf{x}(t) * \mathbf{h}(t) + \mathbf{n}(t) = e_y(t)\mathbf{c}_y(t)$$
(2.5)

$$\mathbf{h}(t) = e_h(t)\mathbf{c}_h(t) = a\exp(-6.9t/T_R)\mathbf{c}_h(t)$$
(2.6)

$$\mathbf{n}(t) = e_n(t)\mathbf{c}_n(t) \tag{2.7}$$

ただし, $e_x(t)$, $e_y(t)$, $e_h(t)$, $e_n(t)$ は, それぞれ, $\mathbf{x}(t)$, $\mathbf{y}(t)$, $\mathbf{h}(t)$, $\mathbf{n}(t)$ のエンベロープ であり, $\mathbf{c}_x(t)$, $\mathbf{c}_y(t)$, $\mathbf{c}_h(t)$, $\mathbf{c}_n(t)$ はそれぞれ, $\mathbf{x}(t)$, $\mathbf{y}(t)$, $\mathbf{h}(t)$, $\mathbf{n}(t)$ のキャリアである.



図 2.2: D 値の定義

これらのキャリアは白色ガウス性のランダム変数である.

次に,系の線形性とキャリアの相互独立性を仮定する.この場合,入出力信号のパワー エンベロープは次式のようなモデルで表すことができる [30].

$$e_y^2(t) = e_x^2(t) * e_h^2(t) + e_n^2(t)$$
(2.8)

2.3.3 残響環境における MTF

残響環境における入出力信号のパワーエンベロープは、次式のようになる.

$$e_x^2(t) = \overline{e_x^2} \{1 + \cos(2\pi f_m t)\}$$
 (2.9)

$$e_y^2(t) = e_x^2(t) * e_h^2(t)$$
 (2.10)

$$= \frac{e_x^2}{\alpha} \left\{ 1 + m_R(f_m) \cos(2\pi f_m t) \right\}$$
(2.11)



図 2.3: 音声明瞭度予測理論の概要図

ただし、 $\alpha = \int_0^\infty \mathbf{h}^2(t) dt$ である. $m_R(f_m)$ は次式で定義される.

$$m_R(f_m) = \frac{\int_0^\infty \mathbf{h}^2(t) \exp(-j2\pi f_m t) dt}{\int_0^\infty \mathbf{h}^2(t) dt}$$
(2.12)

式(2.12)の残響インパルス応答hを次式のように定義する.これは,RIRを統計的に近似した Schroeder の RIR モデルである.

$$\mathbf{h}(t) = a \exp\left(-\frac{6.9t}{T_R}\right) \mathbf{c}_h(t) \tag{2.13}$$

ここで, aは RIR のゲイン項であり, T_R は残響時間である.式 (2.12)に式 (2.13)の RIR モデルを代入することで, 残響環境における MTF は次式で定義される.

$$m_R(f_m) = \left[1 + \left(2\pi f_m \frac{T_R}{13.8}\right)^2\right]^{-\frac{1}{2}}$$
(2.14)

図 2.4 に,式 (2.14)の MTF の特性を示す.図 2.4 から,残響環境の MTF は $f_m \ge T_R$ を パラメータとした,低域通過フィルタの特性を持っていることがわかる.



図 2.4: 残響環境の変調伝達関数 m_R(f_m)の特性

2.3.4 雑音環境における MTF

雑音環境における入出力信号のパワーエンベロープは、次式のようになる.

$$e_x^2(t) = \overline{e_x^2} \{1 + \cos(2\pi f_m t)\}$$
 (2.15)

$$e_y^2(t) = e_x^2(t) + e_n^2(t)$$
 (2.16)

$$= (\overline{e_x^2} + \overline{e_n^2}) \{ 1 + m_N(f_m) \cos(2\pi f_m t) \}$$
(2.17)

 e_n^2 が定常であると雑音環境における MTF は次式で定義される.

$$m_N(f_m) = \frac{\overline{e_x^2}}{\overline{e_x^2} + \overline{e_n^2}} = \frac{1}{1 + 10^{-\frac{\text{SNR}}{10}}}$$
(2.18)

図 2.5 に、式(2.18)の MTF の特性を示す. 図 2.5 から、雑音環境の MTF は、 f_m に依存 せずに一定の減衰特性を持っていることがわかる.



図 2.5: 雑音環境の変調伝達関数 m_N(f_m)の特性

2.3.5 雑音残響環境における MTF

雑音残響環境における MTF は式(2.14)と式(2.18)から次式で定義される.

$$m(f_m) = m_R(f_m) \cdot m_N(f_m) = \left[1 + \left(2\pi f_m \frac{T_R}{13.8}\right)^2\right]^{-\frac{1}{2}} \left(\frac{1}{1 + 10^{-\frac{\text{SNR}}{10}}}\right)$$
(2.19)

図 2.6 に、式(2.19)の MTF の特性を示す.図 2.6 から、雑音残響環境の MTF は、雑音・ 残響の影響を同時に受け、 $f_m \ge T_R \varepsilon$ パラメータとした、低域通過フィルタの特性と f_m に依存しない一定の減衰特性を持っていることがわかる.

2.3.6 STIの算出

室内音響における音声伝送の物理的評価指標(音声伝送指標,STI)の計算法は、図2.7 のブロックダイアグラムで表される.これは、IEC60268-16[13]に基づくものであり、室



図 2.6: 雑音残響環境の変調伝達関数 fm の特性

は、統計的室内インパルス応答(Schroeder の RIR モデル [29])

$$\mathbf{h}(t) = e_h(t)\mathbf{c}_h(t) = a \exp(-6.9t/T_R)\mathbf{c}_h(t)$$
(2.20)

で表されるものとした.ただし、aはRIRのゲイン項であり、 $\mathbf{c}_h(t)$ はランダム変数となる白色雑音キャリアである.MTFは

$$m(f_m) = \frac{\int_0^\infty \mathbf{h}^2(t) \exp(-j2\pi f_m t) dt}{\int_0^\infty \mathbf{h}^2(t) dt}$$
(2.21)

で定義されるため、残響環境(式(2.20))のMTFは

$$m_R(f_m, T_R) = \left[1 + \left(2\pi f_m \frac{T_R}{13.8}\right)^2\right]^{\left(-\frac{1}{2}\right)}$$
(2.22)

となる. ただし, aは1に正規化されている. ここで, T_R は残響時間である. $m(f_m)$ は 変調周波数 f_m と残響時間 T_R をパラメータとした, 低域通過フィルタの特性をもつ.



Octave-band Eq. (2.23) Eq. (2.24) Eqs. (2.25)&(2.26) filterbank

図 2.7: STI 算出方法のブロックダイアグラム

STIの計算法は、次の5つのステップに要約される.

(i) 7 つのオクターブ帯域での MTF の計算:

7つのオクターブ帯域 (キャリア周波数が125 Hz~8 kHz, チャンネル番号が $k = 1, 2, 3, \dots, 7$, 変調周波数が $f_i = 0.63 \sim 12.5$ Hz, 変調周波数の添字が $i = 1, 2, 3, \dots, 14$) で次式のよう に変調伝達関数 $m_k(f_i)$ を求める(式 (2.22)と同じ).ここで, f_i は変調周波数 f_m を離散 化したものである.求める 98 個の $m_k(f_i)$ を表 2.1 に示す.

$$m_k(f_i) = 1/\sqrt{1 + (2\pi f_i T_R/13.8)^2}$$
(2.23)

(ii) MTF から信号対雑音比(SNR)の計算:

次式のように, $m_k(f_i)$ から SNR N(k,i)を求める.

$$N(k,i) = 10 \log_{10} m_k(f_i) / (1 - m_k(f_i))$$
(2.24)

(iii) 伝送指標(TI)の計算:

次式のように、SNR N(k,i)を正規化することで、伝送指標 T(k,i)を求める.

$$T(k,i) = \begin{cases} 1, & (15 < N(k,i)) \\ \frac{N(k,i)+15}{30}, & (-15 \le N(k,i) \le 15) \\ 0, & (N(k,i) < -15) \end{cases}$$
(2.25)

(iv) 変調伝達指標 (MTI) の計算:

次式のように、伝送指標 T(k,i) を平均化することで、変調伝達指標 M(k) を求める.

$$M(k) = \frac{1}{14} \sum_{i=1}^{14} T(k,i)$$
(2.26)

(v) STIの計算:

次式のように、音声伝送指標 STI を求める.

$$STI = \sum_{k=1}^{7} W(k)M(k)$$
 (2.27)

ただし、荷重係数 W(k) は、W(1) = 0.129、W(2) = 0.143、W(3) = W(4) = 0.114、W(5) = 0.186、W(6) = 0.171、W(7) = 0.143 である.

2.4 RASTIの算出方法

RASTIの算出方法は、STIの算出方法と同様に IEC60268-16[13] で標準化されている (図 2.8). STIの算出方法と RASTIの算出方法の違いは、STIでは表 2.1 に示す 98 個の $m_k(f_i)$ を利用するが、RASTIでは表 2.2 に示す 9 個の $m_k(f_i)$ を利用する.

RASTIの計算法は、次の5つのステップに要約される.

(i) 二つのオクターブ帯域での MTF の計算:

二つのオクターブ帯域(キャリア周波数が 500 Hz のときチャンネル番号 k = 1, 変調周 波数が $f_i = 1.0, 2.0, 4.0$, Hz, 変調周波数の添字が i = 1, 2, 3, 4, キャリア周波数が 2 kHz のときチャンネル番号 k = 2, 変調周波数が $f_j = 0.7, 1.4, 2.8, 5.6, 11, 2$ Hz, 変調周波数の 添字が j = 1, 2, 3, 4, 5)で次式のように変調伝達関数 $m_k(f_{i \text{ or } j})$ を求める(式(2.22)と同 じ). なお, $f_{i \text{ or } j}$ は変調周波数 f_m を離散化したものである. 求める 4 個の $m_1(f_i)$ と 5 個 の $m_2(f_j)$ を表 2.2 に示す.

$$m_k(f_{i \text{ or } j}) = 1/\sqrt{1 + (2\pi(f_{i \text{ or } j})T_R/13.8)^2}$$
(2.28)

(ii) MTF から信号対雑音比(SNR)の計算:

次式のように, $m_k(f_{i \text{ or } j})$ から SNR N(k, i or j)を求める.

$$N(k, iorj) = 10 \log_{10} m_k(f_{i \text{ or } j}) / (1 - m_k(f_{i \text{ or } j}))$$
(2.29)

Oct. band [Hz]	125	250	500	1000	2000	4000	8000
$f_1 = 0.63$	$m_1(f_1)$	$m_2(f_1)$	$m_3(f_1)$	$m_4(f_1)$	$m_5(f_1)$	$m_6(f_1)$	$m_7(f_1)$
$f_2 = 0.80$	$m_1(f_2)$	$m_2(f_2)$	$m_3(f_2)$	$m_4(f_2)$	$m_5(f_2)$	$m_6(f_2)$	$m_7(f_2)$
$f_3 = 1.00$	$m_1(f_3)$	$m_2(f_3)$	$m_3(f_3)$	$m_4(f_3)$	$m_5(f_3)$	$m_6(f_3)$	$m_7(f_3)$
$f_4 = 1.25$	$m_1(f_4)$	$m_2(f_4)$	$m_3(f_4)$	$m_4(f_4)$	$m_5(f_4)$	$m_6(f_4)$	$m_7(f_4)$
$f_5 = 1.60$	$m_1(f_5)$	$m_2(f_5)$	$m_3(f_5)$	$m_4(f_5)$	$m_5(f_5)$	$m_6(f_5)$	$m_7(f_5)$
$f_6 = 2.00$	$m_1(f_6)$	$m_2(f_6)$	$m_3(f_6)$	$m_4(f_6)$	$m_5(f_6)$	$m_6(f_6)$	$m_7(f_6)$
$f_7 = 2.50$	$m_1(f_7)$	$m_2(f_7)$	$m_3(f_7)$	$m_4(f_7)$	$m_5(f_7)$	$m_6(f_7)$	$m_7(f_7)$
$f_8 = 3.15$	$m_1(f_8)$	$m_2(f_8)$	$m_3(f_8)$	$m_4(f_8)$	$m_5(f_8)$	$m_6(f_8)$	$m_7(f_8)$
$f_9 = 4.00$	$m_1(f_9)$	$m_2(f_9)$	$m_3(f_9)$	$m_4(f_9)$	$m_5(f_9)$	$m_6(f_9)$	$m_7(f_9)$
$f_{10} = 5.00$	$m_1(f_{10})$	$m_2(f_{10})$	$m_3(f_{10})$	$m_4(f_{10})$	$m_5(f_{10})$	$m_6(f_{10})$	$m_7(f_{10})$
$f_{11} = 6.30$	$m_1(f_{11})$	$m_2(f_{11})$	$m_3(f_{11})$	$m_4(f_{11})$	$m_5(f_{11})$	$m_6(f_{11})$	$m_7(f_{11})$
$f_{12} = 8.00$	$m_1(f_{12})$	$m_2(f_{12})$	$m_3(f_{12})$	$m_4(f_{12})$	$m_5(f_{12})$	$m_6(f_{12})$	$m_7(f_{12})$
$f_{13} = 10.00$	$m_1(f_{13})$	$m_2(f_{13})$	$m_3(f_{13})$	$m_4(f_{13})$	$m_5(f_{13})$	$m_6(f_{13})$	$m_7(f_{13})$
$f_{14} = 12.50$	$m_1(f_{14})$	$m_2(f_{14})$	$m_3(f_{14})$	$m_4(f_{14})$	$m_5(f_{14})$	$m_6(f_{14})$	$m_7(f_{14})$

表 2.1: STIの MTF データマトリクス

(iii) 伝送指標(TI)の計算:

次式のように, SNR N(k, i or j)を正規化することで, 伝送指標 T(k, i or j)を求める.

$$T(k, i \text{ or } j) = \begin{cases} 1, & (15 < N(k, i \text{ or } j)) \\ \frac{N(k, i \text{ or } j) + 15}{30}, & (-15 \le N(k, i \text{ or } j) \le 15) \\ 0, & (N(k, i \text{ or } j) < -15) \end{cases}$$
(2.30)

(iv) 変調伝達指標(MTI)の計算:

次式のように、伝送指標 T(k, i or j) を平均化することで、変調伝達指標 M(k) を求める.

$$M(k) = \begin{cases} \frac{1}{4} \sum_{i=1}^{4} T(k, i) \\ \frac{1}{5} \sum_{j=1}^{5} T(k, j) \end{cases}$$
(2.31)

次式のように、音声伝送指標 STI を求める.

$$RASTI = \sum_{k=1}^{2} W(k)M(k)$$
(2.32)

ただし、荷重係数W(k)は、W(1) = 0.45、W(2) = 0.55であり、最終的には計9個のMTFは同じ重みで扱われる.

Oct. band [Hz]	500	2000
$f_m = 0.7$	—	$m_2(f_1)$
$f_m = 1.0$	$m_1(f_1)$	—
$f_m = 1.4$	—	$m_2(f_2)$
$f_m = 2.0$	$m_1(f_2)$	—
$f_m = 2.8$	—	$m_2(f_3)$
$f_m = 5.0$	$m_1(f_3)$	—
$f_m = 5.6$	—	$m_2(f_4)$
$f_m = 8.0$	$m_1(f_4)$	_
$f_m = 11.2$	_	$m_2(f_5)$

表 2.2: RASTIのMTFデータマトリクス



図 2.8: RASTI 算出方法のブロックダイアグラム

第3章 従来法

3.1 室内インパルス応答(RIR)モデルの一般化

従来法では、無雑音・拡散音場を仮定し、次式に示す一般化 RIR モデル(Schroeder の RIR モデルを次式のように拡張版)ならびに MTF を用いて、室内音響特性を模擬した.

$$\mathbf{h}(t) = t^{b} \exp\left(-\frac{6.9t}{T_{R}}\right) \mathbf{c}(t)$$
(3.1)

ここで、bは0以上の実数をとる次数である.また、bが0のとき、一般化RIRモデルと SchroederのRIRモデルは等価な式になるため、SchroederのRIRモデルよりも自由度の 高いモデルになっている.このときのMTFは、式(3.1)を式(2.12)に代入することで 求められ、次式で表される.

$$m(f_m) = \left[1 + \left(2\pi f_m \frac{T_R}{13.8}\right)^2\right]^{-(2b+1)/2}$$
(3.2)

ー般化 RIR モデルの MTF と Schroeder の RIR モデルの MTF の違いは,指数部の-(2b+1)/2である. 一般化 RIR モデルを用いることにより,図 3.1 に示すように RIR の立ち上がりとその傾きを表現できるため,Schroeder の RIR モデルよりも実測した RIR のエンベロープの形状を正確に近似できるようになり,MTF も正確に近似できる.

3.2 室内音響特性のブラインド推定法の概要

従来法のブロックダイアグラムを図 3.2 に示す.従来法における室内音響特性の推定方法は,MTFの推定を行い推定された RIR から室内音響特性の推定を行っている.まず,MTFの推定を行うために,入力信号の変調度が1であると仮定し,次の特性を利用する.

- 1. 変調周波数0 Hz における MTF の変調度は入出力間で変わらないこと.
- 2. 入力信号において,主な変調周波数での変調スペクトルレベルは0 Hz のものと等 しいこと.
- 3. 出力信号の変調スペクトルが、 $MTFm_R(f_m)$ に沿って残響時間の増加とともに減衰 すること.



図 3.1: RIR モデルの近似例

これらの特性を表したものを図 3.3 に示す.そして、これらの特性を利用し、出力信号の 主要な二つの変調周波数スペクトルレベルの減衰を入力信号のスペクトルレベルまで回 復する逆特性 $m_R^{-1}(f_m)$ を求めることで、残響時間 T_R ならびに次数 bを推定する.具体的 には、次式の自乗平均平方根(Root Mean Square, RMS)誤差が最小になるときの最適 な T_R と b を求める.

$$\{\hat{T}_R, \hat{b}\} = \arg\min_{T_R, b} \text{RMS}(T_R, b)$$
(3.3)

$$RMS(T_R, b) = \sqrt{\frac{1}{L} \sum_{\ell=1}^{L} [|E_y(f_{m\ell})| - m(f_{m\ell}, T_R, b)]^2}$$
(3.4)

ここで, $E_y(f_{m\ell})$ は出力信号の主要な変調周波数 $f_{m\ell}$ での変調スペクトル, $m(f_{m\ell}, T_R, b)$ は主要な変調周波数 $f_{m\ell}$, 残響時間 T_R , 次数 b をパラメータとした, 式 (3.2)より算出される MTF である. 推定パラメータ数が 2 であるため, L = 2 として推定した.



図 3.2: 従来法のブロックダイアグラム

3.3 残響音声信号による MTF 推定

ー般化モデルでは変数が T_R とbの二つあるため、従来法でMTF推定を行うには、入力 信号の変調スペクトルにおいて二つ以上の変調周波数が変調度が1でなければならない。 そこで、従来法では信号の時間的なパワーエンベロープとMTFの調査を行い、パワーエ ンベロープ上で周期的な少なくとも二つ以上の山を持つ場合にMTF上で変調度1の変調 周波数が複数観測されることを発見した(図 3.4).従来法ではこの結果を基に残響音声 信号のパワーエンベロープから図 3.4 のような二つ以上の山を得る方法を調査し音声信号 の高周波帯域に周期的な山が現れることを発見し(図 3.5)、次に示すような操作が提案 された。

- 1. オクターブフィルタバンクを用いて、残響音声信号を周波数帯域分割する.
- 2. 帯域分割された残響音声信号の時間的なパワーエンベロープ内から二つの山を取り 出す.
- 3. パワーエンベロープ内から二つの山を取り出す条件は,各山のピーク間の距離が最 も長いものとする.
- この操作を行った音声信号から MTF 推定を行い,室内音響特性の推定に利用する.



図 3.3: MTF 推定のコンセプト: (a) 入力信号の変調スペクトル, (b) 出力信号の変調スペクトル. 実線は残響の MTF $m_R(f_m)$ を示す.

3.4 問題点

従来法は,残響環境における室内音響特性の推定において,AM信号と音声信号を用いた評価実験で,STIを高い精度で推定することを確認した.しかし,実環境における室内音響特性の推定を行うには次に示す問題点が考えられる.

- 1. 変調周波数領域で推定を行った残響時間*T_R*と次数*b*で,時間領域の評価尺度の推定 を行っている.
- 2. 音声信号を利用した MTF 推定において,時間的なパワーエンベロープの山を少な くとも二つ以上取り出す必要がある.
- 3. 無雑音の室を仮定し、残響による影響のみを考慮している.

従来法では、MTFの形状を合わせることを考えて一般化 RIR モデルを用いた MTF 推定 を行っている.これにより、Schroeder の RIR モデルよりもより正確に MTF 推定を行え るようになったため、MTF から算出される STI の推定精度を向上させることが可能になっ た.しかし、変調周波数領域で推定した *T_R* と *b* では、時間領域における評価指標である 残響時間や D 値の推定に適した RIR の推定が行われているとは限らないため、時間領域 における評価指標への対応を考慮する必要がある.従来法で音声信号を利用した室内音響 特性の推定を行う場合においては、帯域分割された音声信号から時間的なパワーエンベ ロープ上で少なくとも二つ以上の山を取り出してくる必要があるが、二つ以上の山を取り 出すためにはある程度の時間長の信号が必要である.従来法では、室に暗騒音や背景雑音 がないものとして、純粋に残響による影響を考慮して室内音響特性の推定を行う枠組みを 検討している.そのため,背景雑音がある場合においては,推定精度が低下してしまうため,実環境で用いる場合を考えると背景雑音の影響を考慮する必要がある.


図 3.4: パワーエンベロープと変調スペクトルの関係: (a) 山のピークの間隔 0.1 s, (b) 山 のピークの間隔 0.5 s, (c) エンベロープ間の間隔 1 s.



図 3.5: 帯域分割された音声信号のパワーエンベロープと変調スペクトルの例

第4章 提案法

4.1 MTF 推定の再検討

従来法において変調周波数領域で推定を行う際に、時間領域の評価尺度への影響を考慮 していないという問題を検討するために、一般化 RIR モデルの次数 b の値を制限するこ とで改善されるか検討を行った.従来法で用いている一般化 RIR モデルでは、Schroeder の RIR モデルを拡張し、立ち上がりがある RIR に対応できる自由度の高いモデルである. この自由度の高い一般化 RIR モデルを用いることで、従来法は変調周波数領域上で MTF の形状を Schroeder の RIR モデルを用いた場合よりも正確に推定し、高精度な STI の推 定を実現した.しかし、変調周波数領域で T_R と b の値を推定し MTF の形状を高精度に推 定できていても、時間領域である RIR の形状が正確に推定できているとは限らない.そ のため、残響時間や D 値といった時間領域における評価指標の推定を行う上では、時間 領域での RIR の形状も考慮することが重要である.そこで、提案法では、従来法におい て b は 0 以上の実数としていた一般化 RIRI モデルの b の値を 0 ≤ b < 1 に制限した一般化 RIR モデルの再検討を行った.これは、一般化 RIR モデルの t^b の項は RIR の立ち上がり を表現したものであり、0 < b < 1 では立ち上がった後の変化が緩やかであり、1 < b の場 合よりも立ち下りの表現へ与える影響が少なくなり時間領域の室内音響特性の推定に効 果があると考えられるためである.

4.1.1 RIR の最適な近似

ー般化 RIR モデルの b 値を 0 ≤ b < 1 に制限した場合の効果を確認するために、時間 領域において Schroeder の RIR モデル、従来法(0 ≤ b)の一般化 RIR モデル、提案法 (0 ≤ b < 1)の一般化 RIR モデルを用いて実測の RIR に対して RMS 誤差が最小になるよ うにフィッティングを行った。使用した実測 RIR は SMILE2004[34]に収録されいてる 43 個の実測 RIR である。図 4.1 と図 4.3 に線形スケールで実測 RIR とフィッティングを行い 求めた残響時間と D 値を、図 4.2 と図 4.4 に対数スケールで実測 RIR とフィッティングを 行い求めた残響時間と D 値を示す。これらの結果から、線形スケールでフィッティングを 行い求めた残響時間 T_R の RMS 誤差が対数スケールでフィッティングを行った場 合よりも大きくなっていた。これは、線形スケールの場合はパワーエンベロープのピーク 部分の形状の誤差が少なくなるようにフィッティングするが、対数スケールの場合は RIR の減衰特性の誤差が少なくなるようにフィッティングするためで残響時間 T_R の精度が高 くなることがわかる.また、フィッティング精度は提案法が従来法よりも同等かそれ以上 の精度であった.

4.1.2 MTF の最適な近似

ー般化 RIR モデルのb値を0 ≤ b < 1に制限した場合の効果を確認するために、変調周 波数領域において Schroeder の RIR モデル、従来法(0 ≤ b) の一般化 RIR モデル、提案 法(0 ≤ b < 1)の一般化 RIR モデルを用いて実測の RIR に対して RMS 誤差が最小にな るようにフィッティングを行った。使用した実測 RIR は SMILE2004[34] に収録されいて る 43 個の実測 RIR である。図 4.1 と図 4.3 に線形スケールで実測 RIR とフィッティング を行い求めた残響時間と D 値を、図 4.2 と図 4.4 に対数スケールで実測 RIR とフィッティ ングを行い求めた残響時間と D 値を示す。これらの結果から、変調周波数領域では対数 スケールよりも線形スケールでフィッティングを行った場合の方が残響時間 T_R の精度の 方が高いことがわかった。また、フィッティング精度は提案法が従来法よりも同等かそれ 以上の精度であった。

4.1.3 MTF 推定の条件設定

時間領域でフィッティングを行った場合、変調周波数領域でフィッティングを行った場合の両方で一般化RIRモデルのbの値を $0 \leq b < 1$ に制限した提案法の方がbの値を $0 \leq b < 1$ に制限した提案法の方がbの値を $0 \leq b < 1$ に制限した従来法よりも残響時間 T_R やD値の精度が同等か高い値となった.このため、提案法では一般化RIRモデルのbの値を $0 \leq b < 1$ に制限してMTF推定を行い、室内音響特性の推定を行うこととした.



図 4.1:時間領域,線形スケールフィッティングでの残響時間算出結果



図 4.2: 時間領域,対数スケールフィッティングでの残響時間算出結果



図 4.3:時間領域,線形スケールフィッティングでのD 値算出結果



図 4.4: 時間領域,対数スケールフィッティングでのD 値算出結果



図 4.5: 変調周波数領域,線形スケールでの残響時間算出結果



図 4.6: 変調周波数領域,対数スケールでの残響時間算出結果



0.4 0.6 Calculated D

図 4.8: 変調周波数領域,対数スケールでのD 値算出結果



図 4.9: 従来法で注意する必要がある入力信号の MTF

4.2 音声信号を利用した MTF 推定

4.2.1 MTF 推定に利用する時間的なエンベロープの検討

従来法では、図3.4に示すように時間的なパワーエンベロープの山を二つ取り出すこと により、主要な変調周波数において変調度1のピークを得ていた.これは、周期 T のイ ンパルス列のフーリエ変換は周期1/T のインパルス列になるという性質を利用したもの である.このとき、取り出してきたパワーエンベロープの中に、ピーク値が小さくても取 り出す二つの山以外の成分が含まれていると、その成分も含めた周期で、変調周波数上に ピークが現れてしまう(図4.9).また、山を二つ取り出さないといけないため、入力信号 が条件を満たすのに時間がかかる場合があるという問題がある.そのため、提案法では、 時間的なエンベロープから山を一つ取り出す方法の検討を行った.

4.2.2 人工的な信号による変調スペクトルの調査

人工的な信号から、変調スペクトルの形状を調査し時間的なパワーエンベロープから 山を一つ取り出すことで MTF 推定が行えるか検討した. インパルスとインパルス列のパ ワーエンベロープと MTF の比較結果を図 4.10 に示す. インパルス列の MTF は上述のよ うにインパルス列となり、インパルスは全周波数成分を含んでいるため全ての変調度が 1 となり、パワーエンベロープの山がインパルス列の場合の変調スペクトルのピークに沿っ て現れる. 次に、MTF が減衰する場合について検討するためにパワーエンベロープに人 工的な残響信号利用した. 残響信号には、残響時間 $T_R = 1$ s の Schroeder の RIR モデル (式 (2.13))を用いた. パワーエンベロープに山を 1本用いた場合と 2本用いた場合の MTF を求めた結果を図 4.11 に示す. 結果から、時間的なパワーエンベロープとして山を 二つ取り出した場合の変調スペクトルは山を二つ取り出してきた場合の変調スペクトル のピークに沿った形となっており、1本のみ取り出すことで、2本取り出した場合と同じ 変調度が得られることがわかった.

4.2.3 音声信号による MTF 推定

実際の残響音声信号を用いた場合において時間的なパワーエンベロープから山を一つ取 り出した場合と二つ取り出した場合の MTF を調査した結果を図 4.12 に示す.結果から, 人工的な信号の場合と同様に,時間的なパワーエンベロープから山を一つ取り出した場 合の変調スペクトルは山を二つ取り出した場合の変調スペクトルのピークに沿っており, 音声信号を用いた MTF 推定を行う上で取り出す時間的なパワーエンベロープの山は一つ で良いことがわかった.

4.3 雑音の影響への対応

従来法では,無雑音を仮定しているため,入力信号に背景雑音が存在した場合に室内音響特性の推定精度が低下するという問題があった.そのため,提案法では,背景雑音の影響を考慮した室内音響特性の推定法に拡張する.

4.3.1 雑音が MTF 推定に与える影響

従来法の MTF 推定において, 残響音声信号に背景雑音が含まれる場合の影響について 調査した.入力信号に SNR = 20 dB の雑音残響音声信号を入力した場合の結果を図 4.13 に示す.結果から,従来法では雑音の影響によりパワーエンベロープが持ち上がっており, その影響で変調周波数の直流分が大きくなり MTF の主要な周波数成分が相対的に小さく なっていることがわかった.

4.3.2 雑音の影響を考慮した室内音響特性の推定法

従来法を雑音残響環境に対応した室内音響特性推定法に拡張するために,Unokiらが提案した MTF の概念に基づいたパワーエンベロープ回復処理 [30] を利用し,雑音残響環境のパワーエンベロープから雑音の影響の回復処理を行った.図 4.14 に提案法のブロック ダイアグラムを示す.まず,観測信号 y(t)のパワーエンベロープ $e_y^2(t)$ を次式のように求める.

 $\hat{e}_{y}^{2}(t) = \mathbf{LPF}\left[\left|y(t) + j \cdot \mathbf{Hilbert}(y(t))\right|^{2}\right]$ (4.1)

ただし, **Hilbert()**は Hilbert 変換, **LPF**[·] は低域通過フィルタであり, **LPF**[·] のカット オフ周波数は 20 Hz とする [24].

次に、森田らによって提案された雑音残響にロバストな音声区間検出法(VAD) [31] を 利用して、音声区間(SSs)/非音声区間(NSs)を推定する.これらの区間を知ること で、音声区間では音声+雑音のパワー(S+N)を、雑音区間では雑音パワー(N)を推定 できるため、SNR 推定では、観測信号 y(t)に含まれる SNR を求める.

次に, 推定された音声区間/非音声区間を利用して, 雑音のパワーエンベロープ $e_n^2(t)$ の時間平均 $\overline{e_n^2}$ を求め, これを観測されたパワーエンベロープ $e_y^2(t)$ から減算することで, 残響信号のパワーエンベロープを求める. この信号は従来法の MTF 推定で利用される入 力信号そのものになるため, 提案法では, 従来法 (図 4.14)の MTF 推定ならびに RIR 推 定をそのまま利用し, 残響時間 T_R , D 値の推定を行う.

次に、SNR 推定で得られた SNR を利用して、式(2.18)から雑音環境の MTF $m_N(f_m)$ を求める. 原則的に、提案法では、SIT 計算も従来法と同様に利用することができるが、 $m_N(f_m)$ を7オクターブ帯域分割後の $m_{kR}(f_m), k = 1, 2, \cdots, 7$ に乗算して式(2.19)を求める. 最後に、2章で説明した計算法と全く同じ計算をして、STI・RASTIを推定する.

4.3.3 **雑音の影響を考慮した MTF 推定の確認**

雑音の影響を考慮した提案法の処理を行い,雑音残響環境のパワーエンベロープから雑音の影響が回復されているか調査を行った.入力信号に図 4.13 と同じ雑音残響音声信号を用いた場合の結果を図 4.15 に示す.結果から,入力信号のパワーエンベロープが雑音の影響から回復されたことで実測 RIR の MTF に沿ったピークが観測さており,従来法の 残響の MTF 推定を利用可能であることがわかった.



図 4.10: パワーエンベロープがインパルス信号1本と2本の場合のMTF



図 4.11: パワーエンベロープが残響を付加した信号1本と2本の場合のMTF



図 4.12: パワーエンベロープから山を一つ取り出した場合と二つ取り出した場合の MTF と残響環境の MTF



図 4.13: 従来法に SNR = 20 dB の雑音残響信号を入力したときの MTF



図 4.14: 推定法のブロックダイアグラム



図 4.15: 提案法に SNR = 20 dB の雑音残響信号を入力したときの MTF

第5章 室内音響特性の推定評価

従来法における問題点が提案法において解決されているか推定評価実験を行うことで 確認した.評価には入力信号にAM信号と音声信号を用いた雑音残響信号を利用した.

5.1 AM 信号を用いた評価

5.1.1 評価条件

最初の評価として、AM 音(白色雑音の搬送波に振幅変調をかけたもの)に RIR を畳 み込むことで得られた残響信号に、背景雑音として白色雑音を付加した雑音残響信号を 利用した.このAM 音は、パワーエンベロープで周期的な情報をもつように設計された ものである.パワーエンベロープ上の周期は 0.2 s になるように設定したため、基本変調 周波数は5 Hz となる.畳み込んだ残響には SMILE2004 データセット [34] にある 43 個の RIRA.1 を利用した.背景雑音は SNR を 20 dB と 5 dB とした 2 種類の白色雑音を利用し た.比較のために従来法も利用して推定結果の比較を行った.図の横軸は実測の RIR か ら計算された測定値を、縦軸はそれに対応した推定値を示す."△"と"o"の記号は、従来 法と提案法を利用して得られた推定値を示す.図中の数字は、43 個の RIR の番号を示す. 図中の破線は、最適な推定結果を示す.

5.1.2 残響時間推定結果

図 5.1 と図 5.2 に、雑音残響 AM 信号からブラインド推定した残響時間 T_R の結果を示す. SNR = 20 dB のとき(図 5.1)、従来法と提案法ともに同程度の推定結果であることがわかる.一方で、SNR = 5 dB のとき(図 5.2)、従来法では背景雑音の影響により、実際の残響時間 T_R よりも長く過小評価してしてしまうのに対して、提案法では背景雑音の影響を受けずに残響時間 T_R を推定できることがわかる.

5.1.3 D 值推定結果

図 5.3 と図 5.4 に, 雑音残響 AM 信号からブラインド推定した D 値の結果を示す. SNR = 20 dB のとき(図 5.3), 従来法と提案法ともに同程度の推定結果で測定値よりも過大推

定していることがわかる.一方で、SNR = 5 dBのとき(図 5.4)、従来法ではSNR = 20 dBのときよりも過小推定されて測定値に近づいていることがわかる.

5.1.4 STI 推定結果

図5.5と図5.6に、雑音残響AM信号からブラインド推定したSTIの結果を示す.SNR = 20 dBのとき(図5.5)、従来法と提案法の結果とも良好にSTIを推定できることがわかる.一方で、SNR = 5 dBのとき(図5.6)、従来法は背景雑音の影響により過大推定してしまうのに対し、提案法は背景雑音の影響を受けず、良好にSTIを推定できることがわかる.

5.1.5 RASTI 推定結果

図 5.7 と図 5.8 に、雑音残響 AM 信号からブラインド推定した STI の結果を示す. SNR = 20 dB のとき (図 5.7)、従来法と提案法の結果とも良好に STI を推定できることがわかる. 一方で、SNR = 5 dB のとき (図 5.8)、従来法は背景雑音の影響により過大推定してしまうのに対し、提案法は背景雑音の影響を受けず、良好に RASTI を推定できることがわかる.



図 5.1: 雑音残響 AM 信号からの残響時間 T_R 推定の結果 (SNR = 20 dB)



図 5.2: 雑音残響 AM 信号からの残響時間 T_R 推定の結果 (SNR = 5 dB)



図 5.3: 雑音残響 AM 信号からの D 値推定の結果 (SNR = 20 dB)



図 5.4: 雑音残響 AM 信号からの D 値推定の結果 (SNR = 5 dB)



図 5.6: 雑音残響 AM 信号からの STI 推定の結果 (SNR = 5 dB)



図 5.7: 雑音残響 AM 信号からの RASTI 推定の結果 (SNR = 20 dB)



図 5.8: 雑音残響 AM 信号からの RASTI 推定の結果 (SNR = 5 dB)

5.2 音声信号を用いた評価(残響信号にデータベースのRIRを用いた場合)

5.2.1 評価条件

雑音残響音声信号を利用した評価シミュレーションを行った.ここでは,ATR音声デー タベース [33] から8話者(男性話者4名,女性話者4名)が発した8個の日本語文章を 音声信号として利用した.残響音声に関しては,これらの音声信号にSMILE2004データ セット [34] にある43 個の RIR(表A.1)を利用した.背景雑音には上述のAM 信号によ る評価と同様に,SNRを20 dB と5 dB とした2種類の白色雑音を利用した.

5.2.2 残響時間推定結果

図 5.9 と図 5.10 に、雑音残響音声信号からブラインド推定した残響時間 T_R の結果を示 す. SNR = 20 dB のとき、SNR = 5 dB のときのいずれも従来法は大幅に過小推定して いるため、提案法のみ表示したものを図 5.11 と図 5.12 に示す. 従来法では、残響時間に 大幅に推定誤差が生じていたが、提案法では、SNR = 20 dB のとき、SNR = 5 dB のい ずれでも大きな誤差が無く残響時間の推定が行われていることがわかる.

5.2.3 D 值推定結果

図 5.13 と図 5.14 に、雑音残響音声信号からブラインド推定した D 値の結果を示す. SNR = 20 dB のとき(図 5.13)、従来法では大幅に過小推定しているのに対して、提 案法では破線に沿って推定結果が現れている傾向があるのがわかる.一方で、SNR = 5 dB のとき(図 5.14)、従来法は SNR = 20 dB の場合と同様に大幅に過小推定しており、 提案法に関しても推定値の値が D = 0.4 付近に固まっている傾向がわかる.

5.2.4 STI 推定結果

図 5.15 と図 5.16 に、雑音残響音声信号からブラインド推定した STI の結果を示す. SNR = 20 dB のとき(図 5.15), SNR = 5 dB のとき(図 5.16)のいずれも従来法で は大幅に過小推定しているのに対して提案法では良好に STI を推定できることがわかる.

5.2.5 RASTI 推定結果

図 5.17 と図 5.16 に、雑音残響音声信号からブラインド推定した RASTI の結果を示す. SNR = 20 dB のとき (図 5.17), SNR = 5 dB のとき (図 5.18) のいずれも従来法では大幅に過小推定しているのに対して提案法では良好に RASTI を推定できることがわかる.



図 5.9: 雑音残響音声信号 (SMILE2004) からの残響時間 T_R 推定の結果 (SNR = 20 dB)



図 5.10: 雑音残響音声信号 (SMILE2004) からの残響時間 T_R 推定の結果 (SNR = 5 dB)



図 5.11: 雑音残響音声信号(SMILE2004)からの残響時間 T_R 推定の結果 (提案法のみ, SNR = 20 dB)



図 5.12: 雑音残響音声信号(SMILE2004) 46らの残響時間 T_R 推定の結果 (提案法のみ, SNR = 5 dB)



図 5.13: 雑音残響音声信号 (SMILE2004) からのD 値推定の結果 (SNR = 20 dB)



図 5.14: 雑音残響音声信号 (SMILE2004) からの D 値推定の結果 (SNR = 5 dB)



図 5.15: 雑音残響音声信号 (SMILE2004) からの STI 推定の結果 (SNR = 20 dB)



図 5.16: 雑音残響音声信号 (SMILE2004) からの STI 推定の結果 (SNR = 5 dB)



図 5.17: 雑音残響音声信号 (SMILE2004) からの RASTI 推定の結果 (SNR = 20 dB)



図 5.18: 雑音残響音声信号(SMILE2004)からの RASTI 推定の結果 (SNR = 5 dB)

5.3 音声信号を用いた評価(残響信号に大学内で測定した RIR を用いた場合)

5.3.1 評価条件

RIR 測定装置 [35] (B&K Omni-power Omnidirectional Sound Source: Type 4292-L, B&K Power Amplifier: Type 2734, B&K Hand-held analyzer: Type 2250, and B&K DIRAC Room acoustics software: Type 7841, ver. 5.0) を利用して,大学内の室で RIR の測定 [36] を行った 10 個の RIR (表 A.2) を利用した.測定した RIR を残響信号として, ATR 音声データベース [33] から 8 話者 (男性話者 4 名,女性話者 4 名)が発した 8 個の日 本語文章の音声信号として畳み込んだものを雑音残響音声として利用した.

5.3.2 残響時間推定結果

図 5.19 に、雑音残響音声信号からブラインド推定した残響時間 T_Rの結果を示す.従来 法では、全体的に過小推定が若干起きているが、提案法では逆に若干であるが過大推定が 起きていることがわかる.しかし、従来法と提案法のいずれも推定値に大きな誤差は無い ことがわかる.

5.3.3 D 值推定結果

図 5.20 に、雑音残響音声信号からブラインド推定した D 値の結果を示す. 従来法では、 全体的に過小推定が若干起きているが、提案法では逆に若干であるが過大推定が起きてい ることがわかる. また、提案法の RMS 誤差は 0.163 であるのに対して従来法の RMS 誤 差は 0.305 と提案法よりも約 2 倍ほど高い値となっている.

5.3.4 STI 推定結果

図 5.21 に, 雑音残響音声信号からブラインド推定した STI の結果を示す. 従来法では, 全体的に過小推定が若干起きているが,提案法では良好に推定できることがわかる.

5.3.5 RASTI 推定結果

図 5.22 に, 雑音残響音声信号からブラインド推定した RASTI の結果を示す. 従来法では, 全体的に過小推定が若干起きているが, 提案法では良好に推定できることがわかる.



図 5.19: 雑音残響音声信号(大学内の室)からの残響時間 T_R 推定の結果

5.4 考察

AM 信号を用いた残響時間 T_R の推定評価において, SNR = 5 dB の場合(図 5.2)に従来法が提案法よりも RMS が大きくなっていた.これは,MTF 推定において実際には雑音残響環境の MTF であるが,従来法は残響環境の MTF だとして推定を行っているため,雑音の影響を残響の影響として見えてしまい実際の残響時間よりも長い残響時間を推定してしまっていると考えられる.AM 信号を用いた D 値の推定評価において,SNR = 5 dB の場合(図 5.4)において,提案法の RMS 誤差が 0.1777 なのに対して従来法の RMS 誤差が 0.1121 と提案法よりも小さくなっていた.しかし,D 値は SNR に関係なく算出される室内音響特性である.そのため,SNR = 20 dB の場合(図 5.3)と SNR = 5 dB の場合(図 5.4) で従来法と提案法同士を見てみると,提案法では SNR が変化してもD 値が大きく変化してないことがわかる.これにより,提案法は SNR が変化してもD 値が安定して推定可能であるのに対して,従来法ではD 値の推定が雑音の影響を受けてしまっていることがわかる.また,雑音と残響の影響を同時に評価する STI, RASTI については,SNR = 20 dB の場合(図 5.5 と図 5.7)は従来法でも提案法と同程度に推定できていたが,



図 5.20: 雑音残響音声信号(大学内の室)からの D 値推定の結果

SNR = 5 dB の場合(図 5.8 と図 5.8)では提案法は良好に推定できているが従来法では RMS 誤差が大きくなった.これは、STI, RASTIの推定に必要な残響時間 T_R の推定結果 が SNR = 5 dB の場合に悪くなっていること、図 2.5 で示したように雑音環境の MTF の 影響が大きくなることが原因であると考えられる.

音声信号(残響信号にデータベースの RIR を用いた場合)を用いた従来法による室内 音響特性推定は、図 4.13 で示したように SNR = 20 dB の時点で、入力信号の変調スペク トルが MTF の概形から大きく外れているため、室内音響特性の推定値が実測から大きく 外れてしまっている.提案法では、残響時間 T_R の推定結果において、SNR = 20 dB の場 合(図 5.11)は過大推定している傾向はあるが比較的破線に沿って推定できていることが わかる.しかし、SNR = 5 dB の場合(図 5.12)は大きな誤差は無いが過大推定、過小推 定の両方が起きている.これにより、D 値の推定結果も SNR = 20 dB の場合(図 5.13) では過大推定している傾向があるが、破線に沿った結果が見られる.しかし、SNR = 5 dB の場合(図 5.14)では推定値が D = 0.4付近に集まっている.また STI と RASTIの 推定結果(図 5.15~図 5.18)については提案法は良好に推定できていることがわかる.音 声信号(残響信号に大学内で測定した RIR を用いた場合)も音声信号(残響信号にデー



図 5.21: 雑音残響音声信号(大学内の室)からの推定の結果

タベースの RIR を用いた場合)と同じ傾向が従来法,提案法ともに見られたが,大学内 で測定した RIR を用いた場合は SNR が高いため,従来法においても大きな誤差は見られ なかった.

これらの結果から、変調周波数領域で評価を行う STI, RASTI は雑音残響環境で高い推定精度で推定することがわかったが、時間領域の評価尺度である残響時間 T_R, D 値の推定結果は提案法においても雑音に頑健ではあったが、推定誤差が生まれていた.

5.5 まとめ

本章では、提案法が従来法において問題となっていた点が解決できているか評価を行った。その結果、従来法では室内音響特性の推定誤差が大幅に大きくなる雑音残響環境において、提案法では、雑音残響 AM 信号、雑音残響音声信号の両方において STI、RASTI を高精度に推定可能であることを確認した。また、時間領域の評価指標である残響時間 T_R , D 値は雑音残響環境の AM 信号においては雑音の影響を受けることなく推定可能であったが、推定値と実測値との間に誤差が生じていた。



図 5.22: 雑音残響音声信号(大学内の室)からの推定の結果

第6章 結論

6.1 本研究で明らかになったこと

本研究では、実環境で RIR の測定を行わずに、身の回りにある音から室内音響特性の ブラインド推定を行う手法を提案することを目的として、雑音残響環境下における室内音 響特性のブラインド推定法の提案を行った.従来法の三つの問題を解決するために以下に 示す提案を行った.

- 一般化 RIR モデルのパラメータを調査し、次数 b を制限することを提案.
- 時間的なパワーエンベロープ上で一つの山で MTF 推定を行うことを提案.
- 従来法を拡張した雑音の影響を考慮した室内音響特性のブラインド推定法を提案.

一つ目の提案では、実測された RIR に対して一般化 RIR モデルのフィッティングを、時間領域と変調周波数領域において線形スケールと対数スケールのそれぞれで行い、一般化 RIR モデルのパラメータである残響時間 T_R と次数bの値を調査した.その結果、時間領域の客観評価尺度である残響時間 T_R と次数bの近似精度は、従来法の $0 \leq b$ の場合よりも次数bを制限した提案法の $0 \leq b < 1$ の場合の方が高いこと、近似を行う領域やスケールによって実測 RIR の形状に対して優先して近似される部分が異なること明らかにした.

二つ目の提案では、従来法では時間領域上で周期的なインパルス列のフーリエ変換は周 波数領域上でも周期的なインパルス列という特性を利用して、音声信号の時間的なパワー エンベロープ上で周期的な少なくとも二つ以上の山を探してきて MTF 推定を行っていた のに対して、時間的なパワーエンベロープ上で一つの山を探してきて MTF 推定を行う方 法を提案した.これにより、従来法では RIR の MTF に沿ってピークが現れるパルス状の 変調周波数しか得られなかったの対して、提案法において時間的なパワーエンベロープ上 の一つの山を用いることで、音声信号から RIR の MTF に沿った変調周波数が得られるこ とを明らかにした.

三つめの提案では、MTFの概念に基づいたパワーエンベロープ回復処理と雑音残響に ロバストな音声区間検出法を利用し、無雑音を仮定していた従来法を雑音の影響を考慮し た室内音響特性のブラインド推定法に拡張した.これらの提案法と従来法を雑音残響 AM 音と雑音残響音声信号を用いて推定評価を行った.その結果、残響に関する時間領域の評 価指標である残響時間 T_R と D 値は、雑音残響 AM 信号において、雑音の影響を受けずに 提案法で推定可能であることを示した.また、STI、RASTI については雑音残響環境にお いて高精度に推定可能であることが明らかになった.

6.2 残された課題

本研究では、時間領域の評価指標である残響時間 T_R , D 値の推定についても考慮する ために、一般化 RIR モデルの次数 b の値についてフィッティングにより検討し、制限を 行った.しかし、実際の室内音響特性の推定は従来法と同様に変調周波数領域で行ってい るため、次数 b の値の制限に効果は限定的であった.そのため、時間領域の評価指標を推 定するために、時間領域での室内音響特性の推定法の検討や、変調周波数領域でも時間領 域の形状をより考慮可能な推定法を検討を行う必要がある.また、提案法では、音声信号 から時間的なパワーエンベロープ上の一つの山を取り出して用いることで MTF 推定が可 能であることを示したが、音声信号から時間的なパワーエンベロープ上の山を取り出す前 処理として従来法と同様に中心周波数の高いの一つのオクターブバンドを利用している. しかし、オクターブバンドごとの残響時間は同一ではないこと、STI の算出においても七 つのオクターブバンドを利用することを考えると、各オクターブバンドにおいて残響時間 の推定を行う必要があると考えられる.

謝辞

本研究を進めるにあたり,多大なる御指導,御助言を賜りました主指導教員である鵜木 祐史准教授に深く感謝致します.また,研究室会議などの場において数多くの御助言を頂 いた赤木正人教授,宮内良太助教,森川大輔助教をはじめとした鵜木研究室,赤木研究室 のメンバーの皆様に感謝致します.
参考文献

- [1] 飯田茂隆, "明瞭度試験法について,"日本音響学会誌, 43(7), 532-536, 1987.
- [2] 戸井田 義徳, "空間内における音声情報伝達," (小特集-音声の明瞭度と認識率-) 日本音響学会誌, 51(4), 312-316, 1995.
- [3] Sato, H., Morimoto, M., and Sato, H., "Evaluation of speech transmission performance using listening difficulty ratings," J. Acoust. Soc. Jpn., 63(5), 275–280, 2007.
- [4] 佐藤 逸人, 森本 正之, 佐藤 洋, "聴き取りにくさによる音声伝達性能の評価,"日本音 響学会誌, 63(5), 275-280, 2007.
- [5] Sato, H., Morimoto, M., Sato, H., Wada, M., "Relationship between listening difficulty and acoustical objective measures in reverberation fields," J. Acoust. Soc. Am., 123(4), 2087–2093, 2008.
- [6] French, N. R. and Steinberg, J. C., "Factors governing the intelligibility of speech sounds," J. Acoust. Soc. Am. 19, 90–119, 1947.
- [7] Kryter, K. D., "Methods for the calculation and use of the articulation index," J. Acoust. Soc. Am. 34, 1689–1697, 1962.
- [8] Thiele, R., "Richtungsverteilung und Zeitfolge der Schallruckwurfe in Raumen," Acustica., 3, 291–302, 1973.
- [9] 佐藤 洋, "音声伝達研究の新しい動き", 日本音響学会誌. 60(3), 141-146, 2004.
- [10] 小椋 靖夫, 浜田 晴夫, 三浦 種敏, "音場における音声伝送品質のための MTF と STI について," 日本音響学会誌, 40(3), 181-191, 1984.
- [11] ISO 3382, Acoustics-Measurement of the Reverberation Time of Rooms with Reference to Other Acoustical Parameters, 2nd ed. Géneve, 1997.
- [12] Kuttruff, H., Room Acoustics, 3rd ed. (Elsevier Science Publishers Ltd., Lindin), 1991.
- [13] IEC 60268-16:2003. Sound system equipment Part 16: Objective rating of speech intelligibility by speech transmission index.

- [14] Houtgast, T. and Steeneken, H. J. M., "The Modulation Transfer Function in Room Acoustics as a Predictor of Speech Intelligibility," Acustica., 28, 66–73, 1973.
- [15] Houtgast, T., Steeneken, H. J. M., and Plomp, R., "Predicting speech intelligibility in rooms from the Modulation Transfer Function. I. General Room Acoustics," *Acustica*, 46, 60–72, 1980.
- [16] Schroeder, M. R., "Modulation transfer functions: definition and measurement," Acustica, 49, 179–182, 1981.
- [17] Ratnam, R., Jones, D. L. and Brine, W. D. O'., "Fast Algorithms for Blind Estimation of Reverberation Time," *IEEE SIGNAL PROCESSING LETTERS*, vol. 11,no. 6, pp. 537-540, 2007.
- [18] Falk, T. H., Yuan, H., and Chan, W., "Spectro-Temporal Processing for Blind Estimation of Reverberation Time and Single-Ended Quality Measurement of Reverberant Speech," *Proc. INTERSPEECH2007*, 514–517, 2007.
- [19] Talmon, R. and Habets, E. P. A., "Blind reverberation time estimation by intrinsic modeling of reverberant speech," *Proc. ICASSP2013*, 156–160, 2013.
- [20] Eaton, J., Gaubitch, N. D, and Naylor, P.A, "Noise-robust reverberation time estimation using spectral decay distributions with reduced computational cost," *Proc. ICASSP2013*, 161–165, 2013.
- [21] Li, F. F. and Cox, T. J., "Speech transmission index from running speech: A neural network approach," J. Acoust. Soc. Am. 113, 1999-2008, 2003.
- [22] Kendrick, P., Cox, T. J., Zhang, Y., Chambers, J. A., and Li, F. F., "Room acoustic Parameter extraction from music signals," *Proc. ICASSP2006* V, 801–804, 2008.
- [23] Kendrick, P., Cox, T. J., Li, F. F., Zhang, Y. and Chambers, J. A., "Monaural room acoustic parameters from music and speech," J. Acoust. Soc. Am. 124(1), 278–287, 2008.
- [24] Unoki, M., Sakata, K., Furukawa, M., and Akagi, M., "A speech dereverberation method based on the MTF concept in power envelope restoration," Acoust. Sci. & Tech., 25(4), 243–254, 2004.
- [25] Unoki, M. and Hiramatsu, S., "MTF-based method of blind estimation of reverberation time in room acoustics," Proc. EUSIPCO2008, Lausanne, Switzerland, CDROM, 2008.

- [26] Unoki, M. Ikeda, T., and Akagi, M., "Blind Estimation Method of Speech Transmission Index in Room Acoustics," Prof. Forum Acousticum 2011, CDROM, 2011.
- [27] Unoki, M., Ikeda, T., Sasaki, K., Miyauchi, R., Akagi, M., and Kim, N. S., "Blind method of estimating Speech Transmission Index in room acoustics based on concept of modulation transfer function," *Proc. ChinaSIP2013*, 2013.
- [28] Unoki, M., Sasaki, K., Miyauchi, R., Akagi, M., and Kim, N. S., "Blind Method of Estimating Speech Transmission Index From Reverberant Speech Signals," *Proc. EUSIPCO2013*, 2013.
- [29] Schroeder, M. R., "New method of measuring reverberation time," J. Acoust. Soc. Am, 37, 409, 1965.
- [30] Unoki, M., Yamasaki, Y., and Akagi, M., "MTF-based power envelope restoration in noisy reverberant environments," *Proc. EUSIPCO2009*, 228–232, 2009.
- [31] 森田 翔太, 鵜木 祐史, 赤木 正人, "雑音残響にロバストな音声区間検出法の検討,"日本音響学会 2013 年秋季研究発表会, 講演論文集, 1-P-22b, 2013.
- [32] Arai, T., Pavel, M., Hermansky, H., and Avendano, C., "Syllable intelligibility for temporally filtered LPC cepstral trajectories," J. Acoust. Soc. Am., 105(5), 2738-2791, 1999.
- [33] Takeda, T., Sagisaka, Y., Katagiri, K., Abe, M., and Kuwabara, H., "Speech Database User's Manual," ATR Technical Report, TR-I-0028, 1988.
- [34] 日本建築学会(編),建築と環境のサウンドライブラリ SMILE2004,技報堂出版, 2004.
- [35] Room acoustics measurements DIRAC. http://www.bksv.com/Products/analysis-software/acoustics/ building-acoustics/
- [36] 金田豊, "インパルス応答測定の際の留意点,"日本音響学会誌, 55(5), 364-369, 1999.



- Akikazu Myazaki, Shota Morita, Masashi Unoki, "Study on Blind Method of Estimating Speech Transmission Index from Noisy Reverberant Amplitude-Modulated-Signals," *Journal of Signal Processing*, 18(4), 201–204, 2014.
- Akikazu Myazaki, Shota Morita, Masashi Unoki, "Study on Blind Method of Estimating Speech Transmission Index from Noisy Reverberant Amplitude-Modulated-Signals," 2014 RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing, 2014.
- 3. 宮崎 晃和, 森田 翔太, 鵜木 祐史, "実環境における音声伝送指標のブラインド推定 法の検討,"2014年日本音響学会春季研究発表会, 1-P4-18, 2014.
- 4. 宮崎 晃和, 森田 翔太, 鵜木 祐史, "背景雑音を考慮した音声伝送指標のブラインド 推定法の検討,"信学技報, 113(349), 1-6, 2013.
- 5. 宮崎 晃和, 森田 翔太, 鵜木 祐史, "背景雑音の影響を考慮した STI ブラインド推定 法の検討,"平成 25 年度電気関係学会北陸支部連合大会, G-4, 2013.

付録A 評価に利用した実測室内インパルス応答(RIR)の詳細

評価に利用した SMILE2004 データベース [34] の実測 RIR (No. 1~43) の詳細を表 A.1 に、大学内で測定した実測 RIR (No. 44~53) の詳細を表 A.2 に示す. また、大学内で測 定した実測 RIR の室環境について図 A.1 (No. 44~49)、図 A.2 (No. 50~53) に示す.



図 A.1: 大学内で測定した実測 RIR (No. 44~49)の室環境



図 A.2: 大学内で測定した実測 RIR (No. 50~53)の室環境

No.	RIR No.	室の名称・条件	T_{60} (s)
1	301	MRH 1 (with RB)	1.09
2	302	MRH 1 (without RB)	0.80
3	303	MRH 2 (with RB)	1.44
4	304	MRH 2 (without RB)	1.04
5	305	MRH 3 (with RB)	1.93
6	306	MRH 3 (without RB)	1.35
7	307	MRH 4 (with RB)	1.42
8	308	MRH 4 (without RB)	1.54
9	319	MRH 5 $(14,000 \ m^3)$	1.47
10	320	MRH 6 (19,000 m^3)	2.16
11	309	CCH 1 $(5, 600 \text{ m}^3)$	2.35
12	310	CCH 1 (d = 6 m)	2.34
13	311	CCH 1 $(d = 11 \text{ m})$	2.35
14	312	CCH 1 $(d = 15 \text{ m})$	2.39
15	313	CCH 1 $(d = 19 \text{ m})$	2.38
16	314	CCH 2 $(6, 100 \text{ m}^3)$	1.14
17	315	CCH 3 $(20,000 \text{ m}^3)$	1.96
18	316	CCH 4 (with AC)	1.92
19	317	CCH 4 (with AC)	2.55
20	323	CCH 5 $(17,000 \text{ m}^3)$	2.32
21	324	CCH 6 (1F front)	1.77
22	325	CCH 6 (2F side)	1.74
23	326	CCH 6 (3F)	1.69
24	201	Lecture room	1.36
25	318	Theater hall $(3,900 \text{ m}^3)$	0.85
26	401	Meeting room (130 m^3)	0.62
27	402	Lecture room (400 m^3)	1.12
28	403	Lecture room $(2, 400 \text{ m}^3)$	1.09
29	404	$GSH (11,000 \text{ m}^3)$	1.54
30	405	Church 1 $(1, 200 \text{ m}^3)$	0.71
31	406	Church 2 $(3, 200 \text{ m}^3)$	1.30
32	407	Event hall 1 $(28,000 \text{ m}^3)$	3.03
33	408	Event hall 2 $(41,000 \text{ m}^3)$	3.62
34	409	Gym 1 $(12,000 \text{ m}^3)$	2.82
35	410	$Gym 2 (29,000 m^3)$	1.70
36	411	Living room (110 m^3)	0.36
37	412	Movie theater (560 m^3)	0.38
38	413	Atrium $(4,000 \text{ m}^3)$	1.57
39	414	Tunnel $(5,900 \text{ m}^3)$	2.72
40	415	Concourse in train station	1.95
41	416	GSH 2 (1F front)	1.53
42	417	GSH 2 (1F center)	1.49
43	418	GSH 2 (1F balcony)	1.40

表 A.1: SMILE2004 データベースの各 RIR 条件

No.	室の名称・条件	T_{60} (s)
44	I2 講義室(中央にマイク、中央に人を集中、0人、153.6 m ³)	0.70
45	I2 講義室(中央にマイク,中央に人を集中,18 人,153.6 m ³)	0.55
46	I2 講義室(部屋の隅にマイク,部屋の隅に人を集中,0 人,153.6 m ³)	0.69
47	I2 講義室(部屋の隅にマイク, 部屋の隅に人を集中, 18 人, 153.6 m ³)	0.61
48	I2 講義室(中央にマイク,部屋に人を分散,0 人,153.6 m ³)	0.71
49	I2 講義室(中央にマイク,部屋に人を分散,18 人,153.6 m ³)	0.58
50	I34 講義室(スピーカーから 4m 地点にマイク,0 人,307.2 m ³)	0.63
51	I34 講義室(スピーカーから 4m 地点にマイク,18 人,307.2 m ³)	0.60
52	I34 講義室(スピーカーから 8m 地点にマイク,0 人,307.2 m ³)	0.62
53	I34 講義室 スピーカーから 8m 地点にマイク,18 人,307.2 m ³)	0.63

表 A.2: JAIST 内で測定した室の各 RIR 条件

付録B 4.1節で行ったRIRモデルの近(以結果の詳細)

4.1 節において Schroeder の RIR モデル,従来法 $(0 \le b)$ の一般化 RIR モデル,提案法 $(0 \le b < 1)$ の一般化 RIR モデルの形状を SMILE2004 データベース [34] の実測 RIR (No. 1~43) に大して RMS 誤差が最小になるように近似した結果の詳細を以下に示す.時間領域において線形スケールで実測 RIR と近似を行い求めた結果を図 B.1~図 B.43 に, 変調周波数領域において線形スケールで実測 RIR と近似を行い求めた結果を図 B.86 に,変調周波数領域において線形スケールで実測 RIR と近似を行い求めた結果を図 B.87 ~図 B.129 に,変調周波数領域において対数スケールで実測 RIR と近似を行い求めた結果を図 B.87 目に変調周波数領域において対数スケールで実測 RIR と近似を行い求めた結果を図 B.87 定日 B.130~図 B.172 に示す.それぞれの図は1段目に時間領域の線形スケール,2段目に時間領域の対数スケール,3段目に変調周波数領域の線形スケール,4段目に変調周波数領域の対数スケールを表している.



図 B.1: 時間領域,線形スケールで近似した 図 B.2: 時間領域,線形スケールで近似した 結果 (No. 1) 結果 (No. 2)



図 B.3: 時間領域,線形スケールで近似した 図 B.4: 時間領域,線形スケールで近似した 結果 (No. 3) 結果 (No. 4)



図 B.5: 時間領域,線形スケールで近似した 図 B.6: 時間領域,線形スケールで近似した 結果 (No. 5) 結果 (No. 6)



図 B.7: 時間領域,線形スケールで近似した 図 B.8: 時間領域,線形スケールで近似した 結果 (No. 7) 結果 (No. 8)



図 B.9: 時間領域,線形スケールで近似した 図 B.10: 時間領域,線形スケールで近似した 結果 (No. 9) 結果 (No. 10)



図 B.11: 時間領域,線形スケールで近似した 図 B.12: 時間領域,線形スケールで近似した 結果 (No. 11) 結果 (No. 12)



図 B.13: 時間領域,線形スケールで近似した 図 B.14: 時間領域,線形スケールで近似した 結果 (No. 13) 結果 (No. 14)



図 B.15: 時間領域,線形スケールで近似した 図 B.16: 時間領域,線形スケールで近似した 結果 (No. 15) 結果 (No. 16)





MI [dB]

-20

-40

0

-2

-40

-60

MI [dB]





図 B.23: 時間領域,線形スケールで近似した 図 B.24: 時間領域,線形スケールで近似した 結果 (No. 23) 結果 (No. 24)



図 B.25: 時間領域,線形スケールで近似した 図 B.26: 時間領域,線形スケールで近似した 結果 (No. 25) 結果 (No. 26)



図 B.27: 時間領域,線形スケールで近似した 図 B.28: 時間領域,線形スケールで近似した 結果 (No. 27) 結果 (No. 28)



図 B.29: 時間領域,線形スケールで近似した 図 B.30: 時間領域,線形スケールで近似した 結果 (No. 29) 結果 (No. 30)



図 B.31: 時間領域,線形スケールで近似した 図 B.32: 時間領域,線形スケールで近似した 結果 (No. 31) 結果 (No. 32)



結果(No. 33)

結果 (No. 34)



図 B.35: 時間領域,線形スケールで近似した 図 B.36: 時間領域,線形スケールで近似した 結果 (No. 35) 結果 (No. 36)



図 B.37: 時間領域,線形スケールで近似した 図 B.38: 時間領域,線形スケールで近似した 結果 (No. 37) 結果 (No. 38)



図 B.39: 時間領域,線形スケールで近似した 図 B.40: 時間領域,線形スケールで近似した 結果 (No. 39) 結果 (No. 40)



図 B.41: 時間領域,線形スケールで近似した 図 B.42: 時間領域,線形スケールで近似した 結果 (No. 41) 結果 (No. 42)



図 B.43: 時間領域,線形スケールで近似した 結果(No. 43)



図 B.44: 時間領域,対数スケールで近似した 図 B.45: 時間領域,対数スケールで近似した 結果 (No. 1) 結果 (No. 2)



図 B.46: 時間領域,対数スケールで近似した 図 B.47: 時間領域,対数スケールで近似した 結果 (No. 3) 結果 (No. 4)



図 B.48: 時間領域,対数スケールで近似した 図 B.49: 時間領域,対数スケールで近似した 結果(No. 5) 結果(No. 6)



図 B.50: 時間領域,対数スケールで近似した 図 B.51: 時間領域,対数スケールで近似した 結果 (No. 7) 結果 (No. 8)



図 B.52: 時間領域,対数スケールで近似した 図 B.53: 時間領域,対数スケールで近似した 結果 (No. 9) 結果 (No. 10)



図 B.54: 時間領域,対数スケールで近似した 図 B.55: 時間領域,対数スケールで近似した 結果 (No. 11) 結果 (No. 12)



図 B.56: 時間領域,対数スケールで近似した 図 B.57: 時間領域,対数スケールで近似した 結果(No. 13) 結果(No. 14)



図 B.58: 時間領域,対数スケールで近似した 図 B.59: 時間領域,対数スケールで近似した 結果 (No. 15) 結果 (No. 16)







図 B.62: 時間領域,対数スケールで近似した 図 B.63: 時間領域,対数スケールで近似した 結果 (No. 19) 結果 (No. 20)



図 B.64: 時間領域,対数スケールで近似した 図 B.65: 時間領域,対数スケールで近似した 結果 (No. 21) 結果 (No. 22)



図 B.66: 時間領域,対数スケールで近似した 図 B.67: 時間領域,対数スケールで近似した 結果 (No. 23) 結果 (No. 24)



図 B.68: 時間領域,対数スケールで近似した 図 B.69: 時間領域,対数スケールで近似した 結果 (No. 25) 結果 (No. 26)



図 B.70: 時間領域,対数スケールで近似した 図 B.71: 時間領域,対数スケールで近似した 結果 (No. 27) 結果 (No. 28)



図 B.72: 時間領域,対数スケールで近似した 図 B.73: 時間領域,対数スケールで近似した 結果 (No. 29) 結果 (No. 30)



図 B.74: 時間領域,対数スケールで近似した 図 B.75: 時間領域,対数スケールで近似した 結果 (No. 31) 結果 (No. 32)



結果 (No. 33)

結果 (No. 34)



図 B.78: 時間領域,対数スケールで近似した 図 B.79: 時間領域,対数スケールで近似した 結果 (No. 35) 結果 (No. 36)



図 B.80: 時間領域,対数スケールで近似した 図 B.81: 時間領域,対数スケールで近似した 結果(No. 37) 結果(No. 38)



図 B.82: 時間領域,対数スケールで近似した 図 B.83: 時間領域,対数スケールで近似した 結果 (No. 39) 結果 (No. 40)



図 B.84: 時間領域,対数スケールで近似した 図 B.85: 時間領域,対数スケールで近似した 結果 (No. 41) 結果 (No. 42)



図 B.86: 時間領域,対数スケールで近似した 結果(No. 43)





-80 L

0.1

0.2 0.3

0.4 0.5

Time [s]

0.6 0.7

0.8

-80 L

0.2

0.4

0.6

Time [s]

0.8

1

図 B.89: 変調周波数領域,線形スケールで近 図 B.90: 変調周波数領域,線形スケールで近 似した結果 (No. 3) 似した結果 (No. 4)





図 B.93: 変調周波数領域,線形スケールで近 図 B.94: 変調周波数領域,線形スケールで近 似した結果 (No. 7) 似した結果 (No. 8)





20

₩ 0.5

0 L 0

0

-20

-40

MI [dB]

₹ 0.5

0 L 0

0

MI [dB] -2

-40

-60

10 15 Modulation frequency [Hz]

Time [s]

10 15 Modulation frequency [Hz]

20



図 B.101: 変調周波数領域,線形スケールで 図 B.102: 変調周波数領域,線形スケールで 近似した結果 (No. 15) 近似した結果 (No. 16)


図 B.105: 変調周波数領域,線形スケールで 図 B.106: 変調周波数領域,線形スケールで 近似した結果 (No. 19) 近似した結果 (No. 20)



図 B.109: 変調周波数領域,線形スケールで 図 B.110: 変調周波数領域,線形スケールで 近似した結果 (No. 23) 近似した結果 (No. 24)



図 B.113: 変調周波数領域,線形スケールで 図 B.114: 変調周波数領域,線形スケールで 近似した結果 (No. 27) 近似した結果 (No. 28)



図 B.117: 変調周波数領域,線形スケールで 図 B.118: 変調周波数領域,線形スケールで 近似した結果(No. 31) 近似した結果(No. 32)



図 B.121: 変調周波数領域,線形スケールで 図 B.122: 変調周波数領域,線形スケールで 近似した結果 (No. 35) 近似した結果 (No. 36)



図 B.125: 変調周波数領域,線形スケールで 図 B.126: 変調周波数領域,線形スケールで 近似した結果 (No. 39) 近似した結果 (No. 40)



近似した結果(No. 41)

近似した結果(No. 42)



No. 43

図 B.129: 変調周波数領域,線形スケールで 近似した結果 (No. 43)



図 B.132: 変調周波数領域,対数スケールで図 B.133: 変調周波数領域,対数スケールで 近似した結果 (No. 3) 近似した結果 (No. 4)



図 B.136: 変調周波数領域,対数スケールで図 B.137: 変調周波数領域,対数スケールで 近似した結果 (No. 7) 近似した結果 (No. 8)



図 B.140: 変調周波数領域,対数スケールで図 B.141: 変調周波数領域,対数スケールで 近似した結果(No. 11) 近似した結果(No. 12)



図 B.144: 変調周波数領域,対数スケールで図 B.145: 変調周波数領域,対数スケールで 近似した結果(No. 15) 近似した結果(No. 16)



図 B.148: 変調周波数領域,対数スケールで図 B.149: 変調周波数領域,対数スケールで 近似した結果(No. 19) 近似した結果(No. 20)



図 B.152: 変調周波数領域,対数スケールで図 B.153: 変調周波数領域,対数スケールで 近似した結果(No. 23) 近似した結果(No. 24)



図 B.156: 変調周波数領域,対数スケールで図 B.157: 変調周波数領域,対数スケールで 近似した結果(No. 27) 近似した結果(No. 28)



図 B.160: 変調周波数領域,対数スケールで図 B.161: 変調周波数領域,対数スケールで 近似した結果(No. 31) 近似した結果(No. 32)



図 B.164: 変調周波数領域,対数スケールで図 B.165: 変調周波数領域,対数スケールで 近似した結果(No. 35) 近似した結果(No. 36)



図 B.168: 変調周波数領域,対数スケールで図 B.169: 変調周波数領域,対数スケールで 近似した結果(No. 39) 近似した結果(No. 40)



近似した結果(No. 41)

近似した結果(No. 42)



No. 43

図 B.172: 変調周波数領域,対数スケールで 近似した結果 (No. 43)