

Title	室の音響測定を必要としない音声伝達指標と室内音響特性の推定法の研究
Author(s)	佐々木, 恭平
Citation	
Issue Date	2013-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/11332
Rights	
Description	Supervisor: 鶴木 祐史, 情報科学研究科, 修士

修 士 論 文

室の音響測定を必要としない音声伝達指標と
室内音響特性の推定法の研究

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

佐々木 恭平

2013年3月

修士論文

室の音響測定を必要としない音声伝達指標と
室内音響特性の推定法の研究

指導教員 鷓木 祐史 准教授

審査委員主査 鷓木 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

1110028 佐々木 恭平

提出年月: 2013年2月

概要

様々な室の音環境を改善するためには、目的に即した室の設計を必要とする。室の設計に際して、例えばスピーチホールや教室では音声伝送品質の評価が行われる。その評価のためには「音声明瞭度」あるいは「聴き取りにくさ」に関する大規模な聴取実験が実施される。しかし、それらの実施には時間や労力を要するため、聴取実験を行うことなく算出が可能な物理指標（客観評価尺度）から主観評価を得ることが望ましい。その物理指標の一つに音声伝達指標（STI）がある。これは聴き取りにくさ（主観評価尺度）と高い相関があることが示されており、主観評価に近い結果を得ることができる。

STIは変調伝達関数（MTF）から算出される。その算出法は、Houtgast & Steenekenによって提案されたものであり、IECで規格化されている。MTF/STIの算出には、室内インパルス応答（RIR）を実測して算出するか、振幅変調音を利用して実測する必要がある。しかし、聴力保護の観点から人を排除してこれらを測定しなければならない。そのため、駅構内や空港内といった人を排除できない音環境では、これらの実測が不可能であるためMTF/STIを算出できない。この問題を解決するために、UnokiらはMTFの概念に基づきRIRの測定を必要としない、STIと室内音響特性（ T_{60} とD値）をブラインド推定する手法を提案した。これは観測信号の変調スペクトルからMTFを推定し、推定されたMTFからSTIを算出する手法である。彼らはAM信号を利用した推定シミュレーションで提案法が高い推定精度をもつことを示したが、実環境での利用を考えると、次の問題点が挙げられる。(i) 彼らはSchroederの統計的RIRモデルに基づき推定法を提案したが、実環境ではこの仮定から外れる場合があるため、推定精度が低下すること。(ii) 身の回りの音を利用した場合、同程度の推定精度を得られるか不明であること。(iii) 人がいる音環境でもSTIと室内音響特性は推定可能であるが、その実証がされていないこと。

本研究は、上述の問題点を解決するために、(1) 指数関数的減衰をするSchroederの統計的RIRモデルを、Lu & Unokiにより提案されたガンマ分布関数を概形としたRIRモデルに拡張し、そのMTFの推定法を確立すること、(2) 身の回りの音（音声）を利用したSTIと室内音響特性の推定を検証し、必要に応じて改良すること、(3) IEC/ISOで規格されたSTIと室内音響特性を、その実測値と同程度の精度で推定できることをフィールドテストにより実証し（推定精度の担保）、人を排除できない音環境でのSTIと室内音響特性も推定可能であることを実証することに取り組んだ。そして、室の音響測定を必要としない、STIと室内音響特性のブラインド推定法を提案した。

本研究で提案した手法によって、AM信号、人工の残響音声信号、実測された残響音声信号を利用したSTIと室内音響特性の推定を行い、推定結果を利用して評価を行った。評価の結果、高精度なSTIと室内音響特性の推定、音声を利用したSTIと室内音響特性の推定を実現でき、実測値と推定値の精度についても、大きな差がないことを示した。また、人が存在する室で、提案法は観測した音声信号から、STIと室内音響特性を推定できることを明らかにした。これらのことから3つの問題について解決できたことを示した。以上

によって、室の音響測定を必要としない音声伝達指標と室内音響特性の推定が実現でき、これによって、実環境での STI と室内音響特性の推定が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	室の音声伝送品質を評価する評価尺度	2
1.2.1	音声明瞭・単語了解度	2
1.2.2	聴き取りにくさ	2
1.2.3	明瞭度指数	3
1.2.4	Deutlichkeit (D 値) D_{50}	3
1.3	音声伝達指標	3
1.4	音声伝達指標を得るための手法	4
1.4.1	室内インパルス応答を利用した手法	4
1.4.2	AM 音を利用した手法	4
1.5	測定法での問題点	4
1.6	研究の目的	5
第2章	従来法	6
2.1	音声伝達指標の算出方法	6
2.1.1	変調伝達関数の概念	6
2.1.2	残響環境における変調伝達関数	6
2.1.3	音声伝達指標の算出	8
2.2	音声伝達指標の推定法に関する研究	9
2.3	変調伝達関数の概念に基づいた音声伝達指標の推定法	10
2.3.1	変調伝達関数の推定アルゴリズム	10
2.3.2	音声伝達指標のブラインド推定法	12
2.4	従来法での問題点	12
第3章	提案法	17
3.1	室内インパルス応答モデルの一般化	17
3.2	一般化 RIR を利用した変調伝達関数の推定アルゴリズム	17
3.3	音声伝達指標のブラインド推定法	19
3.4	室内音響特性の推定アルゴリズム	20
3.5	音声信号を利用した音声伝達指標の推定	21
3.5.1	音声信号の変調スペクトル	21

3.5.2	音声信号の音声伝達指標の推定への適用法	21
第4章	評価実験	24
4.1	実験の目的	24
4.2	評価1：従来法との比較評価	24
4.2.1	実験条件	24
4.2.2	実験結果	24
4.3	評価2：音声信号を利用した音声伝達指標の推定の評価	25
4.3.1	実験条件	25
4.3.2	実験結果	26
4.4	評価3：実測した残響音声信号を利用した評価	36
4.4.1	実験条件	36
4.4.2	実験結果	36
第5章	結論	42
5.1	本研究で明らかになったこと	42
5.2	残された課題	43

目次

2.1	変調伝達関数と音声伝達指標	7
2.2	残響環境における MTF の特性	8
2.3	STI 算出のブロックダイアグラム	10
2.4	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)	12
2.5	STI 推定のブロックダイアグラム	13
2.6	従来法による残響時間 T_R の推定結果	14
2.7	従来法による STI の推定結果	14
2.8	従来法による D 値の推定結果	15
2.9	実測 RIR とそれを推定した Schroeder の RIR のパワーエンベロープ	16
3.1	2つの RIR モデルによる実測 RIR へ近似例 ((a) 正規化されたパワーエンベロープ, (b) dB 変換されたパワーエンベロープ)	18
3.2	2つの RIR モデルによる実測 RIR の MTF への近似例	19
3.3	一般化 RIR を利用した 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)	20
3.4	提案法の処理手順	20
3.5	パワーエンベロープと変調スペクトルの形状の関係	22
3.6	オクターブフィルタバンクで帯域分割した音声信号の変調スペクトル	22
3.7	STI 推定に利用する音声区間	23
4.1	提案法による残響時間 T_R の推定結果	26
4.2	提案法による次数 n の推定結果	27
4.3	提案法による STI の推定結果	28
4.4	提案法による D 値の推定結果	28
4.5	従来法による音声を利用した残響時間 T_R の推定結果	30
4.6	提案法による音声を利用した残響時間 T_R の推定結果	30
4.7	提案法による音声を利用した次数 n の推定結果	31
4.8	従来法による音声を利用した STI の推定結果	34
4.9	提案法による音声を利用した STI の推定結果	34
4.10	従来法による音声を利用した D 値の推定結果	35
4.11	提案法による音声を利用した D 値の推定結果	35

4.12	評価 3 に用いた実験装置	37
4.13	残響時間 T_R の測定結果と推定結果	38
4.14	次数 n の測定結果と推定結果	38
4.15	STI の測定結果と推定結果	39
4.16	D 値の測定結果と推定結果	39
4.17	残響時間 T_R の推定結果 (縦軸: 人が存在する環境, 横軸: 人が存在しない環境)	40
4.18	次数 n の推定結果 (縦軸: 人が存在する環境, 横軸: 人が存在しない環境)	40
4.19	STI の推定結果 (縦軸: 人が存在する環境, 横軸: 人が存在しない環境)	41
4.20	D 値の推定結果 (縦軸: 人が存在する環境, 横軸: 人が存在しない環境)	41

表 目 次

1.1	STI と音声伝送品質の評価基準	4
2.1	MTF のデータマトリクス	11
4.1	SMILE2004 データベースの各 RIR 条件	32
4.2	評価 2 に用いた音声信号	33
4.3	評価 3 に用いた室	36

第1章 序論

1.1 はじめに

音声コミュニケーションにおいて重要なことは、音声に含まれる言語情報を相手に確実に伝えることである。これは室内において同様のことがいえる。例えば、教室やスピーチホールでは、教師、講演者による教示や講演の内容が聴講者に伝わらなければならない。また、駅のコンコースや空港内では、運行状況のアナウンスや災害時の避難放送などを旅客に確実に伝えることが重要である。

室内において、音信号は室の壁や床、天井や設置物などによって反射する。その反射音が残響となって響きが生まれる。コンサートホールなどでは、この響きを利用して楽器音などを大きく、広く聞かせている。そのため、反射音が拡散するように、狙った位置に集まるように室形を構成している。しかし、教室やスピーチホールでは、残響によって音声が歪んでしまい、音声聴き取りにくくなり、情報を伝えることが困難になる。そのため、反射音の発生しにくい室形を構成する、室の壁に吸音材を設けるなど、室の音響設計によって、音声の歪みを回避する必要がある。音響設計を行う際、音声伝送品質の評価が行われる。音声伝送品質は、音響設計の方針を決定する重要な評価値である。音響設計は、音声伝送品質が一定水準を得るよう設計される。また、音声伝送品質は、決められた評価尺度を用いて評価される。

音声伝送品質の評価尺度には、大きく分けて主観評価尺度と客観評価尺度（物理指標）がある。主観評価尺度は、音声明瞭度、単語理解度、聴き取りにくさ等がある。これらは人の主観による評価であるため、重要な評価尺度である。しかし、主観評価尺度を得るには大規模な聴取実験を行う必要がある。そのため、実験のために多大な時間と労力がかかる。一方、客観評価尺度は、明瞭度指数（AI: Articulation Index）、Deutlichkeit（D値）、音声伝達指標（STI: Speech Transmission Index）などがあり、これらは聴取実験を行うことなく、音響計測によって得ることができる。

客観評価尺度の一つとして音声伝達指標（STI: Speech Transmission Index）がある。STIは主観評価尺度と相関があり、特に聴き取りにくさと高い相関がある。これより、STIは聴取実験を行うことなく、主観評価尺度に近い評価を得ることができる。以上からSTIは、室の音声伝送品質を最も有効に評価できる指標として用いられている。

1.2 室の音声伝送品質を評価する評価尺度

音声伝送品質の評価尺度には，主観評価尺度と客観評価尺度がある．主観評価尺度には，音声明瞭度，単語了解度，聴き取りにくさ等がある [1]．また，客観評価尺度には，明瞭度指数（AI: Articulation Index），Deutlichkeit（D 値），音声伝達指標（STI: Speech Transmission Index）などがある．

1.2.1 音声明瞭・単語了解度

音声明瞭度 [2, 3]，単語了解度 [4, 5] は，音声伝送品質を評価する主観評価尺度である．単音や単語，音節がどの程度相手に伝わるかを示す．これらは，聴取実験により評価値を得ることができる．音声明瞭度は，百音表（五十音，濁音，半濁音，拗音を含む語）から無作為に被験者に提示し，書き取り試験を行う．書き取り試験の成績によって，評価値が決まる．単語了解度は，最初あるいは最後の一音素のみを変化させた，意味のある複数の単語を被験者に提示し，選択や口頭による回答を行う．試験の正答率によって，評価値が決まる．音声の明瞭性や了解性は背景雑音の量や，周波数スペクトルの変化に影響する．主に前者では信号対雑音比での雑音のレベルが増すほど，後者では信号内の周波数成分が取り除かれるほど，明瞭性や了解性は低下する．しかし，提示する音声が続音であった場合，明瞭性，了解性の低下が抑えられる．音声が歪むことによって，言語情報がいくらか破壊されるが，破壊されずに残った情報を手がかりに，人は破壊された情報を補うことができるためである．

1.2.2 聴き取りにくさ

聴き取りにくさ [6, 7] は，森本らによって提案された主観評価尺度である．音声明瞭・単語了解度は，提示する音声信号を変えると正答率が変わる，親密度（「なじみ」の程度を表す尺度）の高い単語ほど正答率が高くなるなど，再現性に問題がある．特に親密度の高い単語を用いた場合，伝達経路による劣化がある，「聴き取りにくい」と感じる音声信号でも，100%の正答率を出してしまう．このことから，音声伝送品質が最高に達していないにも関わらず，音声明瞭・単語了解度の正答率が，100%に達してしまい，それ以上の評価ができなくなってしまう．

音声明瞭・単語了解度の評価は相対的のみの評価であり，どの程度聴き取りやすいかという絶対評価はできない．この問題を解決するために，森本らによって聴き取りにくさという評価尺度が提案された．聴き取りにくさは，聴き取りにくいと判断された割合で音声伝送品質を評価する．人は音声を，聴き取りやすいと感じるより，聴き取りにくいと感じることが，日常において多く，主観評価尺度として相応しいものだとしている．実際，実験から室内の音声伝送品質の主観評価尺度として，相対的および絶対的な両方の観点から評価が可能であることが示されており，主観評価尺度としての優位性を示している．親密

度の高い単語を被験者に提示し、「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」の4つの選択肢により回答する。聴き取りにくさは、その中の「やや聴き取りにくい」から「非常に聴き取りにくい」までの割合を算出することで評価値を得る。

1.2.3 明瞭度指数

明瞭度指数 [8] は French と Steinberg によって提案された，明瞭度に関する尺度である。彼らは，スピーチレベル，ノイズレベル，伝送周波数特性などの物理量から，明瞭度を計算によって求める方法を提案した。後に Kryter により検討が加えられ，適用範囲の拡大，簡易計算法の開発によって，より実用的な形となった。1969 年には ANSI: American National Standards Institute で規格化されている。明瞭度指数は，音声伝達に重要な周波数帯域をいくつかに分け，それぞれの周波数帯域で信号対雑音比を 0 ~ 1 にスケールし，加重平均することで得られる。

1.2.4 Deutlichkeit (D 値) D_{50}

D 値 [9] は，V.R.Thiele によって提案された客観評価尺度であり，直接音到来から t ms までのエネルギーと，音信号全体のエネルギーとの比で定義される。これは，式 (1.1) で表される。

$$D = \frac{\int_0^{50\text{ms}} [h(t)]^2 dt}{\int_0^{\infty} [h(t)]^2 dt} \quad (1.1)$$

音声伝送品質の評価には，直接音到来後 50 ms までのエネルギーとの比が用いられる (D_{50})。50 ms までに含まれる反射音には，音声の明瞭性を高める働きがあるためである。

1.3 音声伝達指標

STI (STI: Speech Transmission Index) [11] は室の音声伝送品質を評価する客観評価尺度である。STI は単語了解度等の主観評価尺度との相関があり，特に聴き取りにくさと高い相関がある [10]。STI は 0 ~ 1 の値を取り，その値によって表 1.1 のように評価される。STI の算出法は Houtgast & Steeneken [12, 13, 14] によって提案され，その手法は IEC 60268-16 [15] によって規格化されている。STI の算出には同じく Houtgast & Steeneken によって提案された変調伝達関数 (MTF: Modulation Transfer Function) が用いられ，MTF の概念に基づいて行われる。MTF の概念において，残響環境下での観測信号の MTF は，残響によって減衰する。この特性を評価尺度として利用するため，オクターブバンドフィルタバンクにて帯域分割された，7つの MTF の荷重和をとることで，STI が算出される。

表 1.1: STI と音声伝送品質の評価基準

STI	0 ~ 0.45	0.45 ~ 0.6	0.6 ~ 0.75	0.75 ~ 1.0
評価基準	不可	可	良	優

1.4 音声伝達指標を得るための手法

STI は中心周波数 0.125 ~ 8 kHz の 7 つのオクターブバンドの MTF から算出される。したがって、STI を得るにはこれら MTF を得る必要がある。MTF を得る方法には大きく分けて 2 種類あり、RIR を計測してそれから MTF を算出する方法と、IEC60268-16 にて規格されている AM 信号を利用して MTF を測定する方法がある。

1.4.1 室内インパルス応答を利用した手法

MTF は RIR のパワーエンベロープのスペクトルで表現される伝達関数である。RIR を利用した STI の測定は、RIR から MTF を算出することによって STI を得る。まず RIR を測定し、中心周波数 0.125 ~ 8 kHz の 7 つのオクターブバンドオクターブフィルタバンクに通して、RIR を 7 つのオクターブ帯域に分割する。そして、帯域分割された各 RIR のパワーエンベロープに対してフーリエ変換を施すことによって MTF を算出する。STI は算出された各 MTF の荷重和をとることで得られる。

1.4.2 AM 信号を利用した手法

この手法は、AM 信号を利用することにより、MTF を直接測定する手法である。室内にて、MTF のデータマトリクス 98 個に対応する AM 信号を発生させる。その入出力間のパワー比により、各帯域および各変調周波数における変調度を測定する。そして、帯域毎の MTF を、各変調周波数での変調度の平均をとることで求め、それらの荷重和をとり、STI を算出する。

1.5 測定法での問題点

RIR を利用した測定法では、十分な信号対雑音比 (SNR) を確保しなくてはならない。よって、雑音による測定誤差を低減するため、高音圧のインパルス信号を発生させる必要がある。また、AM 信号を利用した測定法では、AM 信号の測定を繰り返し行う必要がある。これらは、人のいる環境で行う場合、聴覚保護の観点および労力の程度から、測定が困難である。よって STI は現状、人が排除された環境でのものであり、教室やスピーチホールなど多数の人が存在することを目的とした室や、駅のコンコースや空港内など人の

排除できない音環境では，正確な STI を得ることができない．また，現状の測定法では測定装置が大掛かりである．これより，物の位置や気温，湿度の変化についても，測定時間と測定にかかる労力から対応しがたい．以上から，現状の STI の値は実際の音環境でのものを得ているとは言いがたい．

1.6 研究の目的

音響測定を利用して STI を得る方法では，人がいるまたは人を排除できない音環境や，刻一刻と伝達特性が変化する音環境に対して正確な STI を得ることができない．音響設計時に得た STI と実際の環境での STI に大きな差があると，設計通りの音環境が得られない可能性がある．これは音響設計が無駄になることを意味するため，早急に STI を得る方法の改善が必要である．人がいるまたは人を排除できない音環境や，刻一刻と伝達特性が変化する音環境に対して，正確な STI や室内音響特性を得るには，音響測定を行わないで STI と室内音響特性を得る手法を提案する他はないと考えられる．そこで，本研究では，室の音響測定を必要としない，STI と室内音響特性のブラインド推定法を提案することを目的とする．これにより，MTF/STI 算出のために人を排除する必要がないため，多様な音環境での STI と室内音響特性を得ることができる．また，インパルス信号などを発生させる必要がなくなるため，測定装置の小型化が期待でき，短時間で少労力の推定を行うことができる．

第2章 従来法

2.1 音声伝達指標の算出方法

STIの算出方法はHoutgast & Steeneken[12, 13, 14]によって提案された。この手法はIEC 60268-16[15]によって規格化されている。STIの算出にはMTFが用いられ、MTFの概念に基づいて行われる(図2.1)。

2.1.1 変調伝達関数の概念

MTFの概念は、音声明瞭度を予測する処理体系として、Houtgast & Steenekenによって提案されたものである。彼らは、想定する室内音響を拡散音場と仮定し、室内の変調伝達関数(変調度)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。これは、次式のように入出力信号の時間的なパワーエンベロープで定義される。

$$\bar{I}_i^2 \{1 + \cos(2\pi f_m t)\} \quad (2.1)$$

$$\bar{I}_o^2 \{1 + m(f_m) \cos(2\pi f_m(t - \tau))\} \quad (2.2)$$

ここで、 \bar{I}_i^2 と \bar{I}_o^2 は入出力の信号強度、 f_m は変調周波数、 τ は位相情報、 $m(f_m)$ は変調度を表す。このときの $m(f_m)$ はMTFとして参照される。

2.1.2 残響環境における変調伝達関数

入力信号、出力信号(残響信号)、インパルス応答を、それぞれ、 $x(t)$ 、 $y(t)$ 、 $h(t)$ とする。出力信号は入力信号とインパルス応答の畳み込みで表され、 $y(t) = x(t) * h(t)$ となる。各信号は、時間包絡線(エンベロープ) $e(t)$ とキャリア $c(t)$ の積で構成されるものとする。線形システムと信号のキャリアの相互独立性を仮定すると、出力信号のパワーエンベロープ $e_y^2(t)$ は、次式のように入力信号と残響インパルス応答のそれぞれのパワーエンベロープの畳み込みとして表すことができる。これらの関係から、入力信号と出力信号のパワーエンベロープを表現しなおすと式(2.4)、式(2.5)となる。

$$e_y^2(t) = e_x^2(t) * e_h^2(t) \quad (2.3)$$

$$e_x^2(t) = \bar{e}_x^2(1 + \cos(2\pi f_m t)) \quad (2.4)$$

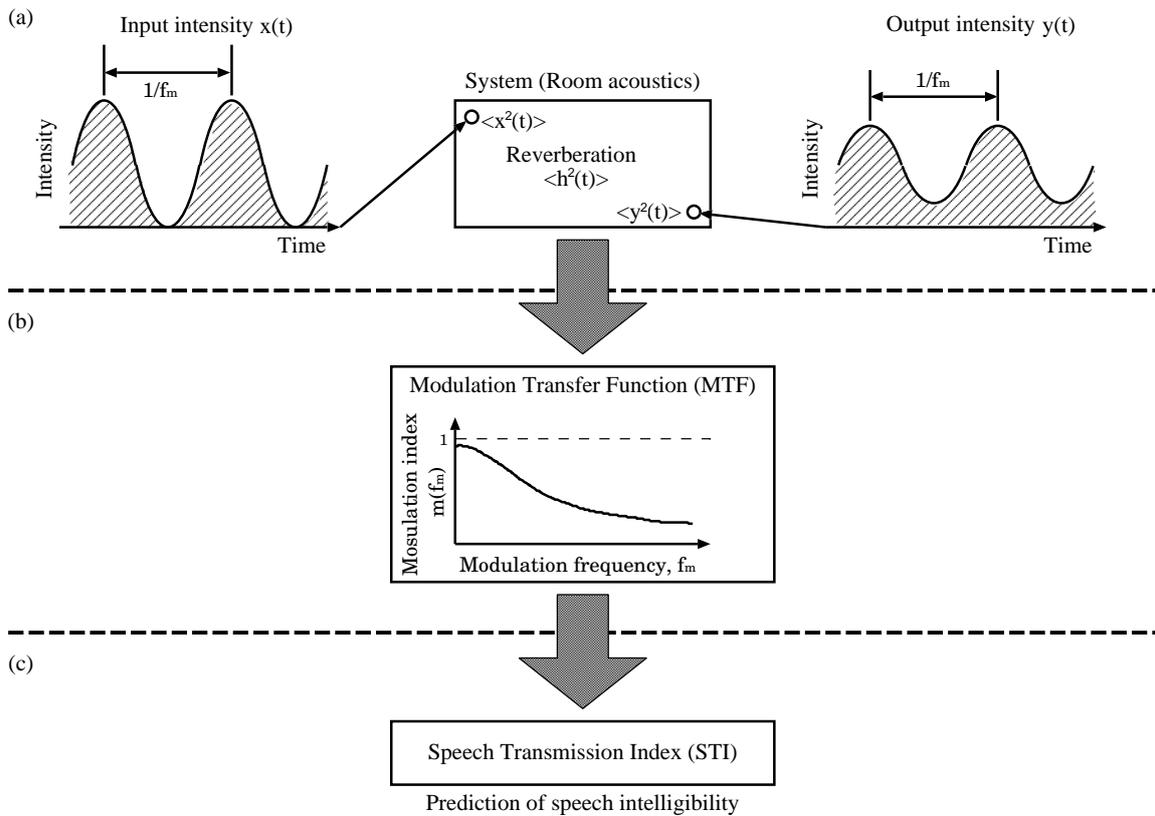


図 2.1: 変調伝達関数と音声伝達指標

$$e_y^2(t) = \frac{\overline{e_x^2}}{\alpha} \{1 + m(f_m) \cos(2\pi f_m t)\} \quad (2.5)$$

ただし, $\alpha = \int_0^\infty h^2(t) dt$ である. また, $m(f_m)$ は MTF であり, 次式で定義される.

$$m(f_m) = \frac{\int_0^\infty h^2(t) \exp(-j2\pi f_m t) dt}{\int_0^\infty h^2(t) dt} \quad (2.6)$$

ここで, 残響インパルス応答 $h(t)$ を, 次式で定義される Schroeder の RIR モデル [25]

$$h(t) = e_h(t)c_h(t) = a \exp(-6.9t/T_R) c_h(t) \quad (2.7)$$

として利用すると, MTF は次式のように表される.

$$m(f_m) = \left[1 + \left(2\pi f_m \frac{T_R}{13.8} \right)^2 \right]^{(-1/2)} \quad (2.8)$$

図 2.2 は, $m(f_m)$ の特性を示す. この図から, MTF は変調周波数 f_m に対して, 残響時間 T_R とともに変化する一種の低域通過フィルタの特性を持つことがわかる.

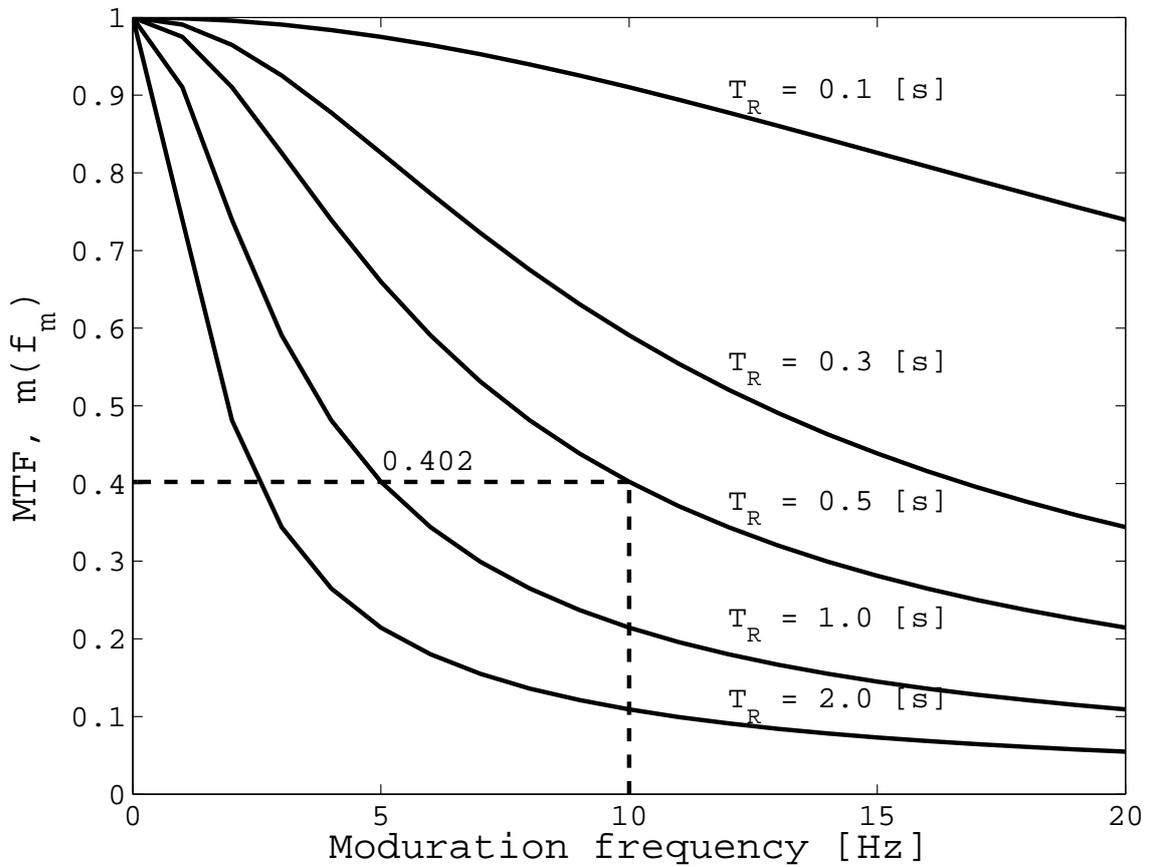


図 2.2: 残響環境における MTF の特性

2.1.3 音声伝達指標の算出

STI の算出方法は IEC 60268-16 にて規格化されており，図 2.3 に示す手順で算出される．具体的には以下の 5 つのステップで得られる．

(1) オクターブ帯域ごとの MTF の測定:

中心周波数 0.125 ~ 8 kHz の 7 個のオクターブバンド ($k = 1, 2, \dots, 7$) の各バンドで，0.63 ~ 12.5 Hz までの 1/3 オクターブ間隔の 14 個の変調周波数 F_i ($i = 1, 2, \dots, 14$) の，計 98 個の MTF $m_k(F_i)$ を測定する．これら MTF を表 2.1 に示す．

(2) MTF からの等価 SNR の算出:

測定された MTF $m_k(F_i)$ から次式より信号対雑音比 (SNR) $N(k, i)$ を求める．

$$N(k, i) = 10 \log_{10} \frac{m_k(F_i)}{1 - m_k(F_i)} \quad (2.9)$$

(3) 伝送指標の算出:

SNR $N(k, i)$ を次式で正規化し, 伝送指標 $I(k, i)$ を求める .

$$I(k, i) = \begin{cases} 1 & (15 < N(k, i)) \\ (N(k, i) + 15)/30 & (-15 \leq N(k, i) \leq 15) \\ 0 & (N(k, i) < -15) \end{cases} \quad (2.10)$$

(4) 変調伝達指標の算出:

各オクターブバンドの伝送指標 $I(k, i)$ の平均値を次式により求め, 変調伝達指標 $M(k)$ を得る .

$$M(k) = \frac{1}{14} \sum_{i=1}^{14} I(k, i) \quad (2.11)$$

(5) STI の算出:

変調伝達指標 $M(k)$ それぞれにに荷重をかけたものを平均して次式により STI を求める .

$$\text{STI} = \sum_{k=1}^7 W(k)M(k) \quad (2.12)$$

なお, 荷重 $W(k)$ は次式のとおりである .

$$\begin{cases} W(1) = 0.129 \\ W(2) = 0.143 \\ W(3) = W(4) = 0.114 \\ W(5) = 0.186 \\ W(6) = 0.171 \\ W(7) = 0.143 \end{cases} \quad (2.13)$$

2.2 音声伝達指標の推定法に関する研究

STI およびそれに関連するブラインド推定法について, 報告されているものを挙げる . STI に関連する値である残響時間のブラインド推定法に関する研究として, 最尤法を利用した方法 [16] や混合ガウス分布モデル (GMM: Gaussian Mixture Model) [17] を利用した方法がある . しかし, 前者は相当な数の観測信号が必要で, 残響時間を求めるに際して長時間を要するという問題がある . 後者についても, 事前に用意されている情報を参照していることから, ブラインド推定法とは言い難い . また, 観測信号のエンベロープ [18] や変調スペクトル [19] の減衰を利用して, 残響時間を推定する方法も提案されている . これらはブラインドで残響時間を推定できることをシミュレーションにより示した . これらの手法は, STI 推定への応用が望めるが, 前者については現在報告されていない .

STI のブラインド推定法に関する研究としては, ハイブリッドニューラルネットワークを利用した統計的な機械学習によって, STI や T_{60} などの客観評価尺度を推定する手法

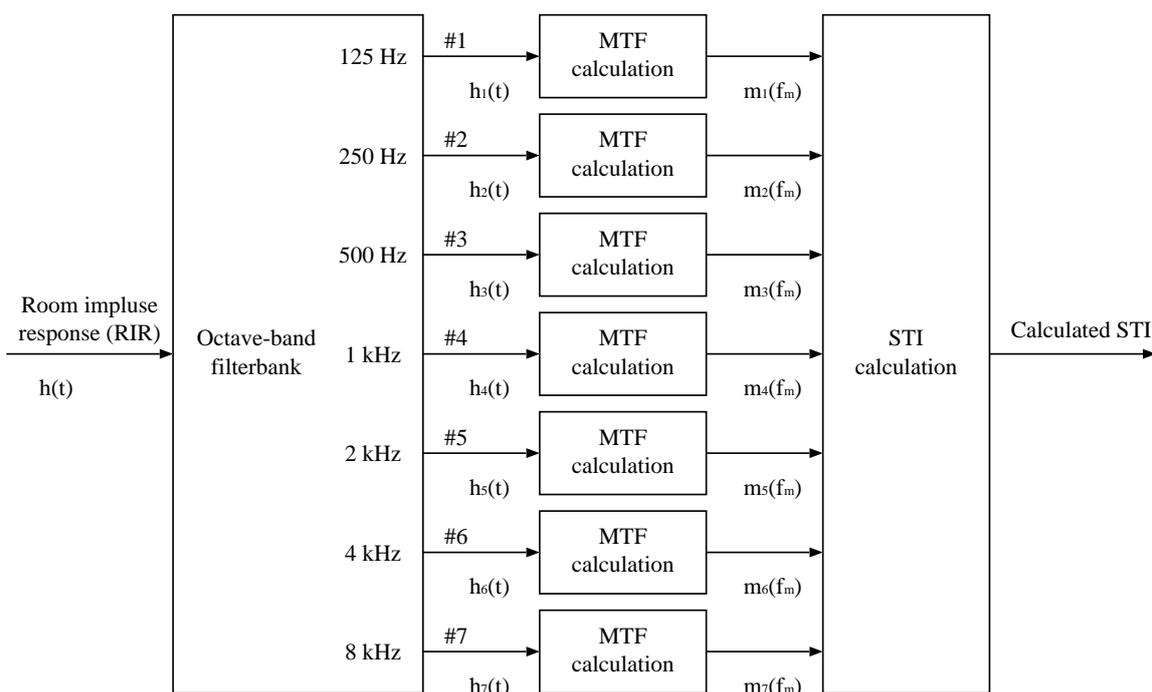


図 2.3: STI 算出のブロックダイアグラム

[20, 21] が報告されている．しかし，事前に音環境に関する大規模な学習データを必要とする．そのため，この方法では未知の音環境，特に人が居る環境での推定に適しているとはいえない．また，観測した楽器音や音声から残響時間や初期反射時間を推定することで，STI を推定する方法 [22] もある．しかし，入力信号の情報をを用いることから，ブラインド推定であるとは言い難い．

STI のブラインド推定法として，最も有力な研究としては，Unoki らによって提案された，変調伝達関数の概念に基づいた STI のブラインド推定法 [23, 24] (以下，従来法と呼ぶ) である．この方法は，先に挙げた変調スペクトルの減少を利用して残響時間を推定する手法 [19] を応用し，MTF を推定する．そして，推定された MTF から IEC 60268-16 に規格されている STI の算出法を用いて，STI を推定する手法である．この手法は，事前の学習や情報の用意を必要とせず，観測信号のみを利用して STI や T_{60} ， D_{50} などの客観評価尺度を得ることができるブラインド推定法である．また，AM 信号を利用したシミュレーションにおいて，この手法は高い推定精度を示した．

2.3 変調伝達関数の概念に基づいた音声伝達指標の推定法

2.3.1 変調伝達関数の推定アルゴリズム

まず，MTF は次の 4 つの特性を利用して推定される．

表 2.1: MTF のデータマトリクス

Oct. band [Hz]	125	250	500	1000	2000	4000	8000
$F_1 = 0.63$	$m_1(F_1)$	$m_7(F_1)$
$F_2 = 0.80$	$m_1(F_2)$	$m_7(F_2)$
$F_3 = 1.00$	$m_1(F_3)$	$m_7(F_3)$
$F_4 = 1.25$	$m_1(F_4)$	$m_7(F_4)$
$F_5 = 1.60$	$m_1(F_5)$	$m_7(F_5)$
$F_6 = 2.00$	$m_1(F_6)$	$m_7(F_6)$
$F_7 = 2.50$	$m_1(F_7)$	$m_7(F_7)$
$F_8 = 3.15$	$m_1(F_8)$	$m_7(F_8)$
$F_9 = 4.00$	$m_1(F_9)$	$m_7(F_9)$
$F_{10} = 5.00$	$m_1(F_{10})$	$m_7(F_{10})$
$F_{11} = 6.30$	$m_1(F_{11})$	$m_7(F_{11})$
$F_{12} = 8.00$	$m_1(F_{12})$	$m_7(F_{12})$
$F_{13} = 10.00$	$m_1(F_{13})$	$m_7(F_{13})$
$F_{14} = 12.50$	$m_1(F_{14})$	$m_7(F_{14})$

1. 変調周波数 0 Hz における変調スペクトルレベルは，入出力間で変わらないこと．
2. 主要な変調周波数における変調スペクトルレベルは，入力信号の変調スペクトルにおいて，0 Hz のものと等しいこと．
3. 出力信号の変調スペクトルにおいて，主要な変調周波数における変調スペクトルレベルの減衰は $m(f_m)$ に対応していること．
4. (3) で観測された減衰を，元のスペクトルレベルまで回復するために必要な逆特性 $m^{-1}(f_m)$ の T_R が，MTF の残響時間そのものであること．

これらの特性から，入出力信号の MTF は図 2.3.1(a) と (b) のような概形となる．以上から，観測した信号の主要な変調周波数の変調スペクトルに対して，式 (2.8) の残響時間 T_R を変化させ， $m(f_m)$ をその変調スペクトルに近似する．そして，近似された値を最小自乗の意味で最適化することで，最適な残響時間 \hat{T}_R を推定する． \hat{T}_R は次式によって推定される．

$$\hat{T}_R = \arg \min_{T_R} (|\log |E_y(f_d)| - \log |E_y(0)| - \log \hat{m}(f_d, T_R)|), \quad (2.14)$$

ここで， $\log |E_y(f_d)| - \log |E_y(0)|$ は変調周波数 f_d での減衰された変調スペクトル， $\hat{m}(f_d, T_R)$ は， T_R の関数とする f_d での MTF である．

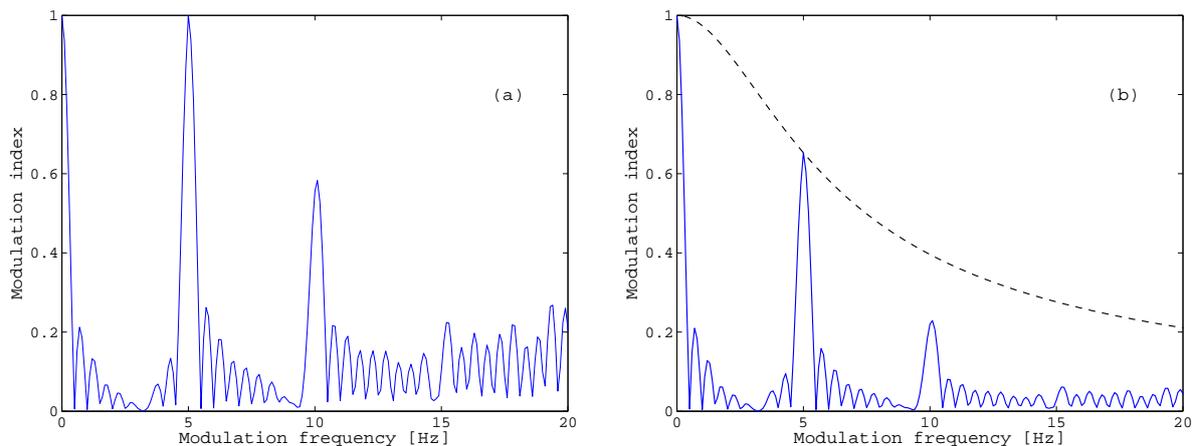


図 2.4: MTF 推定のコンセプト ((a) 入力信号の変調スペクトル, (b) 出力信号の変調スペクトル)

2.3.2 音声伝達指標のブラインド推定法

図 2.5 は STI 推定法の推定手順を示す。まず、式 (2.14) によって \hat{T}_R を推定する。次に、推定された \hat{T}_R を Schroeder の RIR モデル (式 (2.7)) に代入し、RIR $\hat{h}(t)$ を推定する。そして、推定された RIR を、中心周波数が 125 Hz から 8 kHz の範囲とするオクターブフィルタバンクを利用して 7 つの帯域に分割し、各帯域内の MTF を求める。最後に、推定された MTF から IEC60268-16 の計算方法を利用して STI を算出する。

図 2.6~図 2.8 に AM 信号を利用した推定結果を示す。残響時間の推定値が過小推定になる傾向が見られるが、STI と D 値については、推定値に大きなばらつきはなく、実測値に近い値を推定できていることがわかる。これより、観測信号が AM 信号のような、主要な変調周波数成分を検出しやすい信号を用いた場合において、従来法が有効であることを明らかにした。

2.4 従来法での問題点

従来法は、AM 信号を利用した推定実験で、高い推定精度をもつことを示したが、実環境での利用を考えると、次の 3 つの問題点が挙げられる。

1. Schroeder の統計的 RIR モデルに基づき推定法を提案したが、実環境ではこの仮定から外れる場合があるため、推定精度が低下すること。
2. 身の回りの音を利用した場合、同程度の推定精度を得られるか不明であること。
3. 人がいる音環境でも STI と室内音響特性は推定可能であるが、その実証がされていないこと。

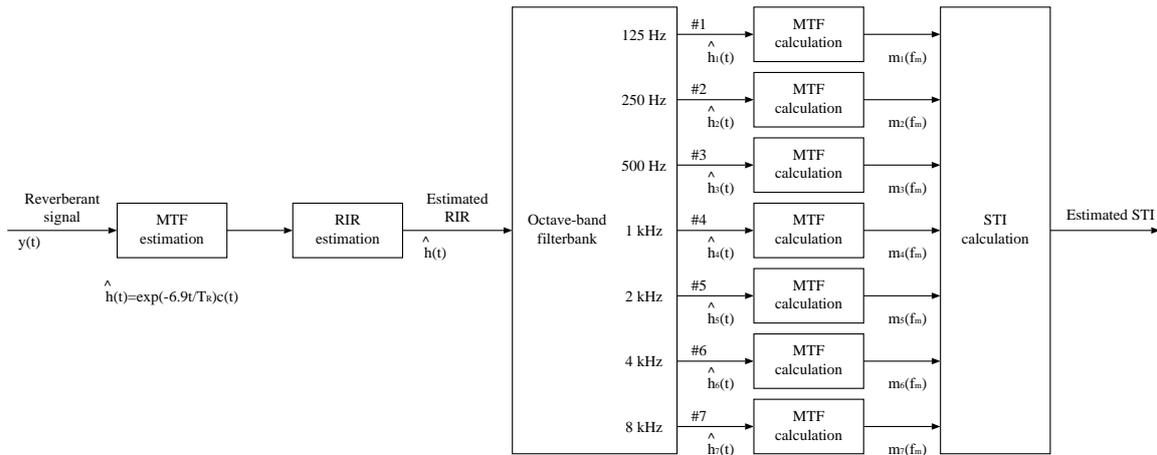


図 2.5: STI 推定のブロックダイアグラム

1. について，従来法は Schroeder の統計的近似インパルス応答モデルを用いている．このモデルの特性は時刻 $t = 0$ のときにエンベロープが最大値をとり，その後は指数関数的に減衰するというものである．しかし，実測の RIR の形状はその限りではない．エンベロープの最大値のある時刻 t が 0 であることはほとんどなく，Schroeder の RIR モデルにはない立ち上がりが発生している．また，立ち上がりの傾きも Schroeder の RIR モデルから外れた実測の RIR が存在する．これらから，Schroeder の RIR モデルを利用した推定では，図 2.9 に示すように，実測の RIR の形状に近似すると大きなズレが生じる（誤差が生じる）．RIR の近似でズレが生じると，実測の RIR から得られる MTF とは，減衰の傾きが異なる MTF が推定される．STI 推定の際は，その MTF から算出を行うため，STI の推定精度が低下してしまう．この問題から，推定精度の解決には，より実測 RIR との近似性が高い RIR モデルを利用することが必要である．

2. について，従来法は観測信号から STI を推定する手法であるが，評価は AM 信号を利用した推定結果で行っており，観測信号を利用した STI 推定を行ってはいない．観測信号を利用する場合，その信号が 2.2.1 に示した MTF の特性全てを満たすかどうか，明かされていない．従来法はそれについて触れられておらず，観測信号を利用して STI 推定を行った場合，AM 信号を利用したときの推定結果を同程度の推定精度が得られるか不明である．よって，観測信号を利用した STI 推定について，検証を行い，必要であれば推定手法を提案する必要がある．

3. について，従来法の目的が人がいる音環境でも STI の推定が行えることであるにも関わらず，人がいる音環境での STI 推定が行われていない．これより，人がいる音環境での STI 推定の実証がされていないことになる．この問題については，何らかの方法で，人がいる音環境での STI 推定が可能であることを実証する必要がある．

以上の問題を解決するため，提案法では，(1) Schroeder の統計的近似インパルス応答に自由度を 1 つ増した，一般化 RIR モデル [26] を用いた MTF/STI 推定法の提案，(2) 身の周りの音を利用した STI と室内音響特性の推定を検証し，必要に応じて改良，(3) IEC/ISO で

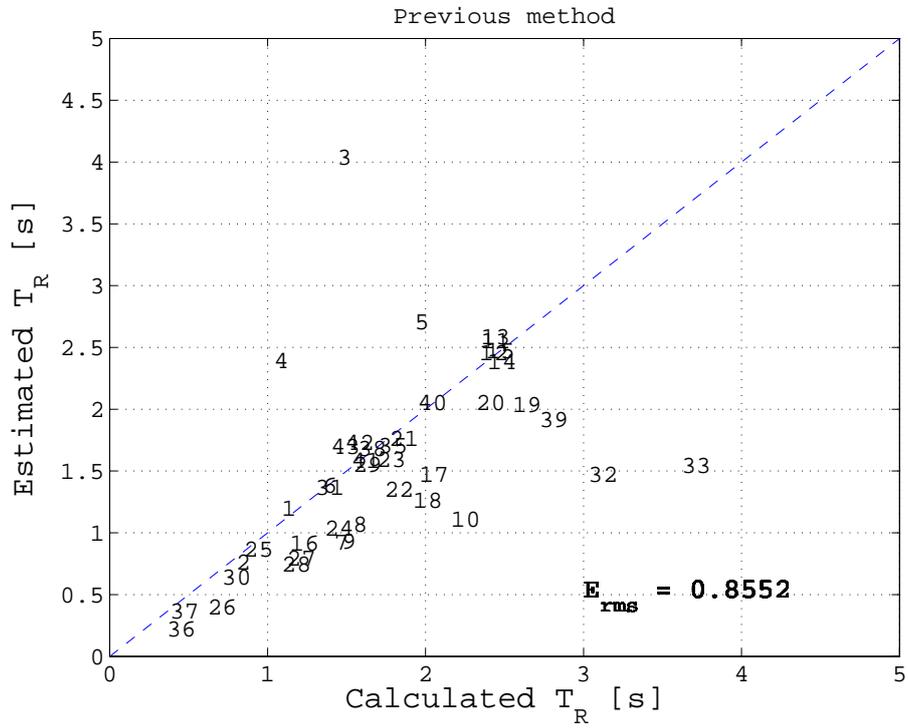


図 2.6: 従来法による残響時間 T_R の推定結果

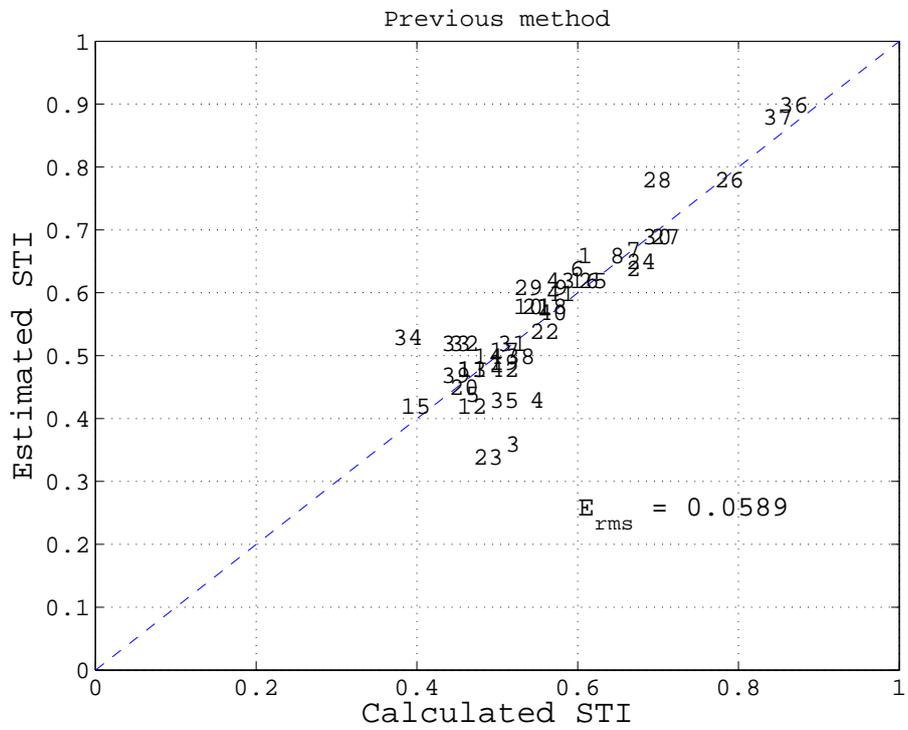


図 2.7: 従来法による STI の推定結果

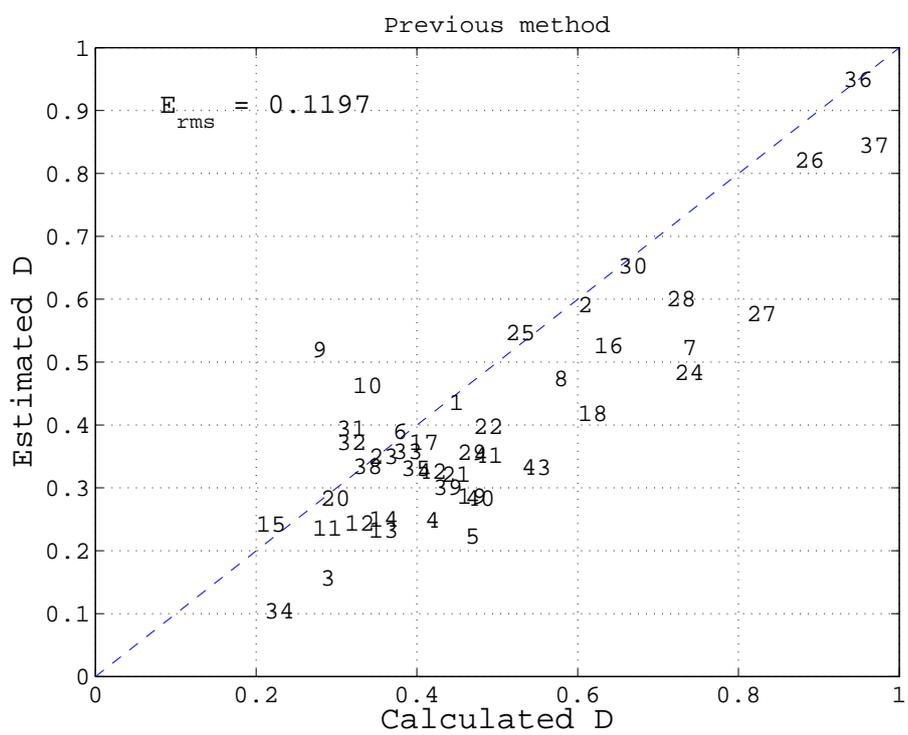


図 2.8: 従来法による D 値の推定結果

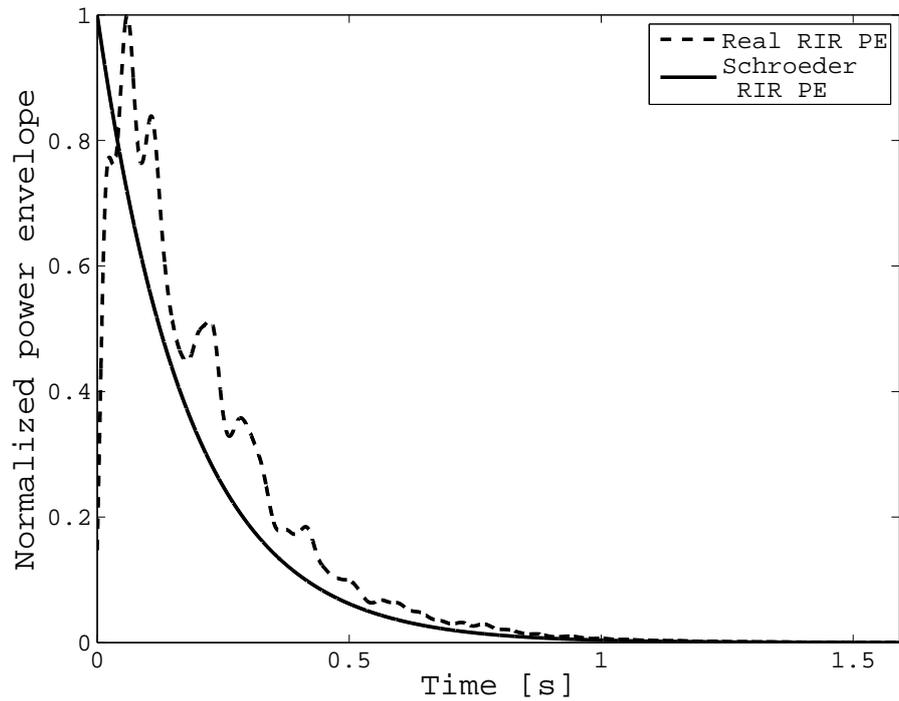


図 2.9: 実測 RIR とそれを推定した Schroeder の RIR のパワーエンベロープ

規格された STI と室内音響特性を，その実測値と同程度の精度で推定できることをフィールドテストにより実証し（推定精度の担保），人のいる音環境での STI と室内音響特性の推定が可能であることの実証を行う．

第3章 提案法

3.1 室内インパルス応答モデルの一般化

従来法での 1. の問題点を解決するために，Schroeder の RIR モデルを次式のように拡張する（以後，これを一般化 RIR と呼ぶ）．

$$h(t) = t^{(n-1)} \exp\left(-\frac{6.9t}{T_R}\right) c(t) \quad (3.1)$$

ここで， n は 1.0 以上の実数をとる次数である．一般化 RIR モデルは Schroeder の RIR モデルよりも自由度が高い．また， n が 1.0 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．このときの MTF は，式 (3.1) を式 (2.6) に代入することで求められ，次式で表される．

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

一般化 RIR の MTF と Schroeder の RIR の MTF の違いは，指数部の $-(2n-1)/2$ のみである．

実測の RIR に対して，一般化 RIR で詳細に表現できるかを調べるため，実測の RIR に対して近似を行った．結果の一例を図 3.1，図 3.2 に示す．RIR のパワーエンベロープに関して，図 3.1(a) から，立ち上がりとその傾きが一般化 RIR によって正確に表現できている．また，MTF についても，図 3.1(b) その特性を適切に表現できている．したがって，一般化 RIR モデルは Schroeder の RIR モデルよりも，実測の RIR のエンベロープ形状を正確に近似でき，かつ MTF も正確に表現できることがわかる．

3.2 一般化 RIR を利用した変調伝達関数の推定アルゴリズム

従来法での MTF 推定法をそのまま一般化 RIR を利用した MTF 推定に適用すると，一般化 RIR には残響時間 T_R と次数 n の二つの変動するパラメータが存在するため，主要な変調周波数を一つと仮定している従来法の推定方法では，一意に推定値を決定することができない．そこで，一般化 RIR を利用した MTF 推定では，推定法を適用するための仮定について検討した．入力信号のエンベロープが主要な 2 つの変調周波数を持ち，かつそれぞれの変調度が 1.0 であるものと仮定する．この仮定を表したものを図 3.2 に示す．一般化

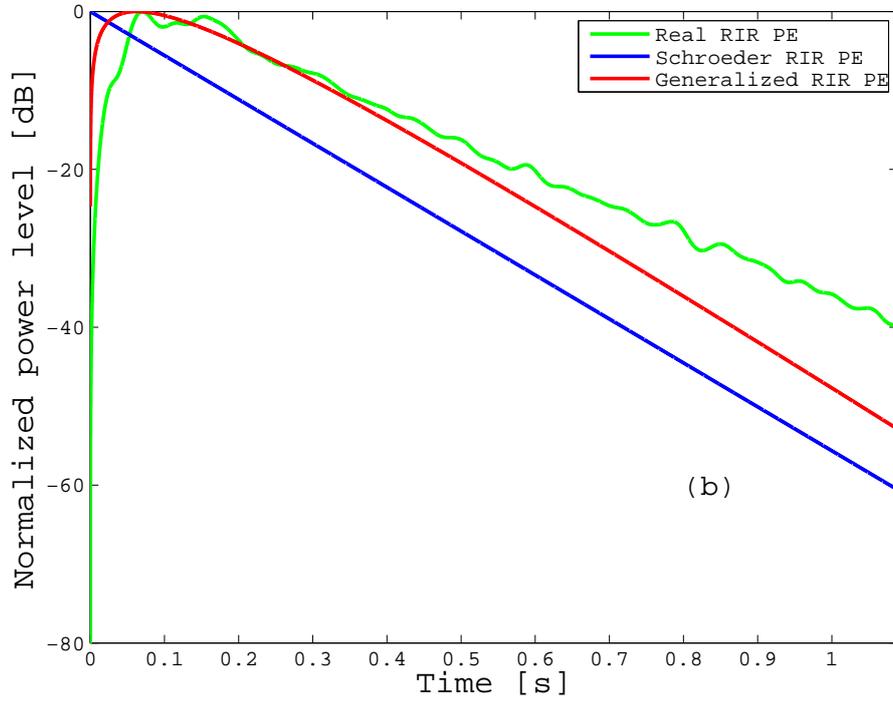
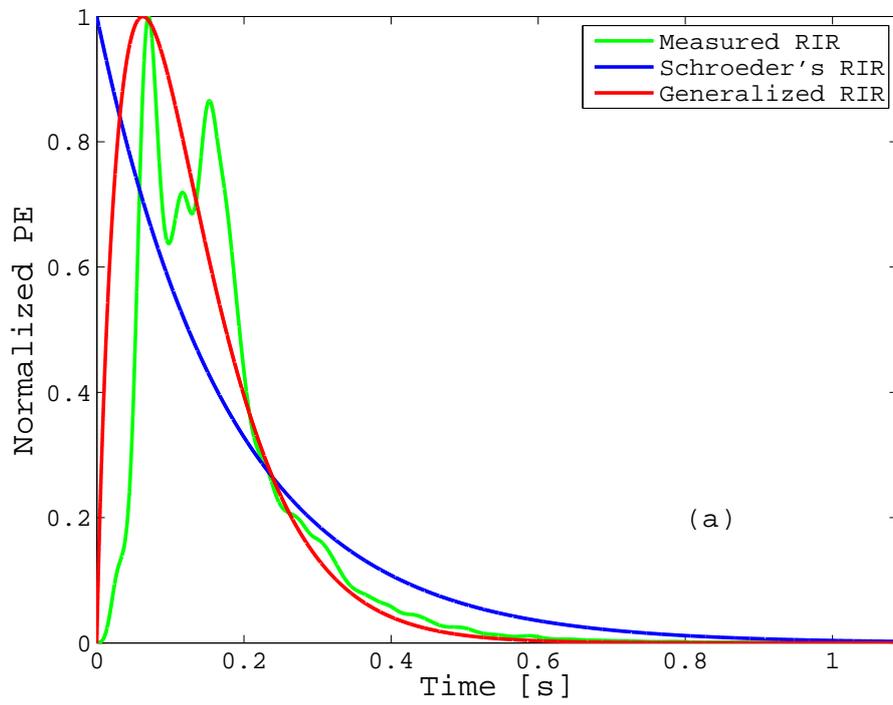


図 3.1: 2つの RIR モデルによる実測 RIR へ近似例 ((a) 正規化されたパワーエンベロップ, (b) dB 変換されたパワーエンベロップ)

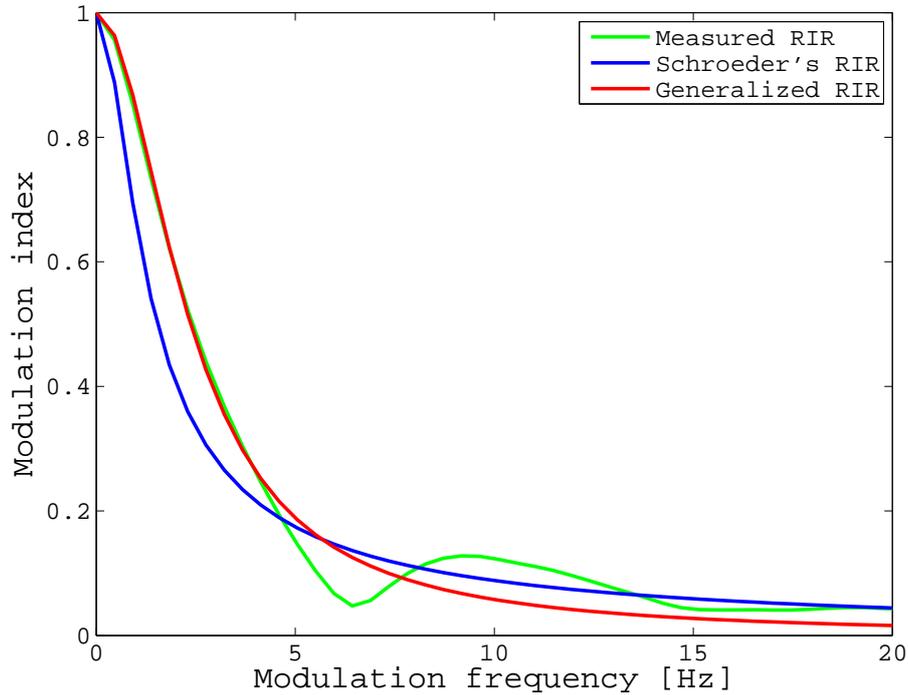


図 3.2: 2 つの RIR モデルによる実測 RIR の MTF への近似例

RIR にある 2 つのパラメータ (T_R と n) を求めるために, 2 つの変調周波数での残響信号の変調度と式 (3.2) の MTF 間の二乗平均平方根 (RMS) 誤差を求める. RMS 誤差は次式で定義される.

$$\text{RMS}(T_R, n) = \sqrt{\frac{1}{L} \sum_{\ell=1}^L [E_y(f_{m\ell}) - m(f_{m\ell}, T_R, n)]^2} \quad (3.3)$$

ここで, $E_y(f_{m\ell})$ は残響信号の主要な変調周波数 $f_{m\ell}$ での変調スペクトル, $m(f_{m\ell}, T_R, n)$ は, 式 (3.2) より算出される MTF, L は 2 である. RMS 誤差が最小のとき, 最適な T_R と n が得られる. 推定される T_R と n は次式で表される.

$$\{\hat{T}_R, \hat{n}\} = \arg \min_{T_R, n} \text{RMS}(T_R, n) \quad (3.4)$$

3.3 音声伝達指標のブラインド推定法

図 3.4 に推定手順のブロックダイアグラムを示す. 一般化 RIR モデルを用いた MTF/STI のブラインド推定法 (以下, 提案法と呼ぶ) の処理手順は, 従来法と同様であるが, 主な改良点は 1 番目と 2 番目のブロックである.

STI 推定は, 以下のように行われる. まず, 残響信号 $y(t)$ から, 式 (3.3) と式 (3.4) を用いて \hat{T}_R と \hat{n} を推定する. \hat{T}_R と \hat{n} は, 主要な 2 つの変調周波数 $f_{m\ell}$ における MTF $m(f_{m\ell})$

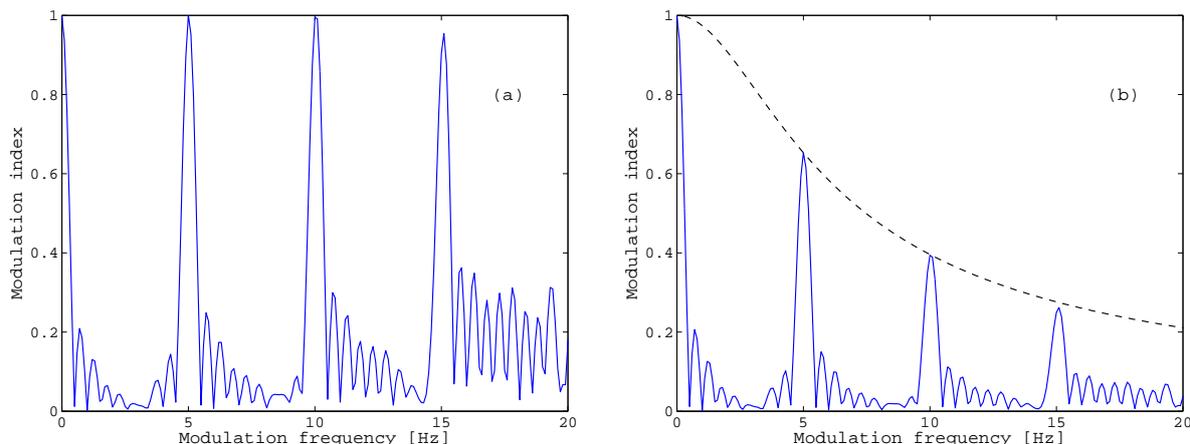


図 3.3: 一般化 RIR を利用した MTF 推定のコセプト ((a) 入力信号の変調スペクトル, (b) 出力信号の変調スペクトル)

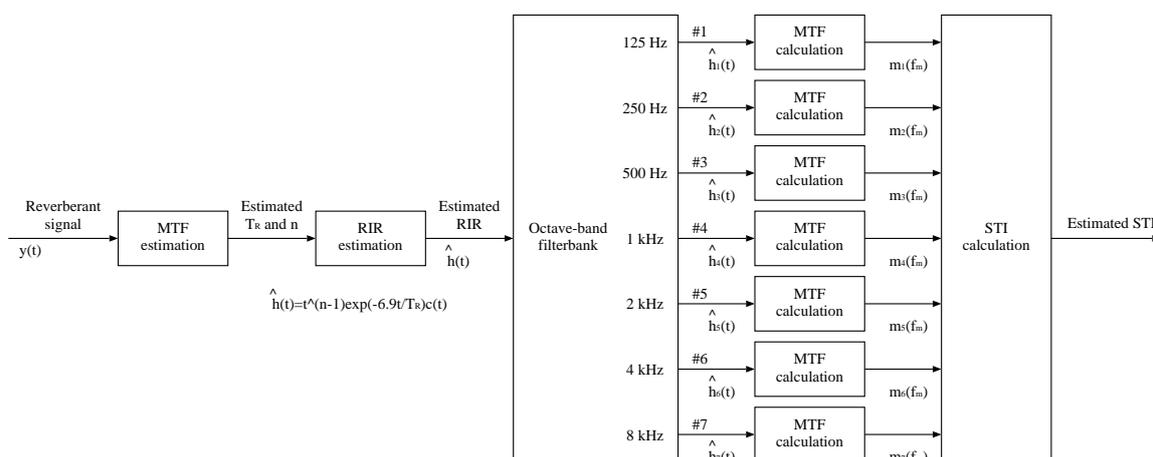


図 3.4: 提案法の処理手順

を 1.0 であるとする事で決定できる。このとき、一般化 RIR モデルに推定された \hat{T}_R と \hat{n} を代入することで、RIR $\hat{h}(t)$ が推定される。次に、推定された RIR を、中心周波数が 125 Hz から 8 kHz の範囲とするオクターブフィルタバンクを利用して 7 つの帯域に分割し、各帯域内の MTF を求める。最後に、推定された 7 つの MTF から、IEC60268-16 の計算方法を用いて STI を算出する。

3.4 室内音響特性の推定アルゴリズム

提案法では、STI 以外にも、残響時間 T_R と D 値の推定を行う。残響時間 T_R の推定値は、MTF 推定の際に推定されるものをそのまま用いる。また、D 値の推定は、推定され

た RIR を用いて行う。D 値は、次式を利用して推定される。

$$\hat{D} = \frac{\int_0^{50ms} [\hat{h}(t)]^2 dt}{\int_0^{\infty} [\hat{h}(t)]^2 dt} \quad (3.5)$$

3.5 音声信号を利用した音声伝達指標の推定

従来法での (2) の問題を解決するため、身の回りの音を利用した STI 推定法を提案する。ここでは、身の回りの音として音声に着目した。パワーエンベロープと MTF の関係を探り、音声を STI 推定に利用する方法を検討する。

3.5.1 音声信号の変調スペクトル

提案法によって MTF/STI を推定するには、 T_R と n の推定のために 2 つの変調周波数上の MTF が 1.0 でなければならない。残響音声信号は、この条件を満たすか調査した。図 3.5 は 5 Hz の正弦波 1 周期を 2 つ配置して作成したパワーエンベロープとその MTF を、図 3.6 はオクターブフィルタバンクによって帯域分割された、音声信号のパワーエンベロープとその MTF を表している。また、図 3.5 は 2 つの正弦波の間隔を徐々に広げたものとなっている。この図を見ると、パワーエンベロープにある 2 つの正弦波の間隔が長いほど、変調スペクトルが最大値に近い周波数ピークが、多くなっていることがわかる。このことから、残響音声信号へ MTF/STI 推定法を適用する場合、パワーエンベロープにある 2 つのピークの間隔を広くとった残響音声信号を用いれば、推定が可能であることがわかる。

上述の状態は音声信号にも見られた。図 3.6 での 2 ~ 5 s がそれである。その範囲のみを取り出し、再度 MTF を算出した結果が図 3.7 になる。音声の包絡線上に周期性のある 2 つ以上のピークがある場合、図 3.7 のように MTF 上に所望の特性が現れ、かつ MTF が 1.0 であるピークが複数現れる。これより、残響音声信号を利用して MTF/STI を推定することが可能であることがわかる。

3.5.2 音声信号の音声伝達指標の推定への適用法

音声のパワーエンベロープとその MTF を調査した結果、音声のパワーエンベロープ上で十分な時間間隔のある 2 つのピークを取り出す必要がある。したがって、推定を行う前に残響音声信号に対して、以下の操作を行う。

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. 残響音声信号のパワーエンベロープ内で最も理想的な 2 つのピークを取り出す。

ここで、(2) に関しては、パワーエンベロープ内にある各ピーク間隔の中で、最も長いものを取り出すことを行った。MTF/STI 推定は図 3.7 の状態の信号に対して行う。

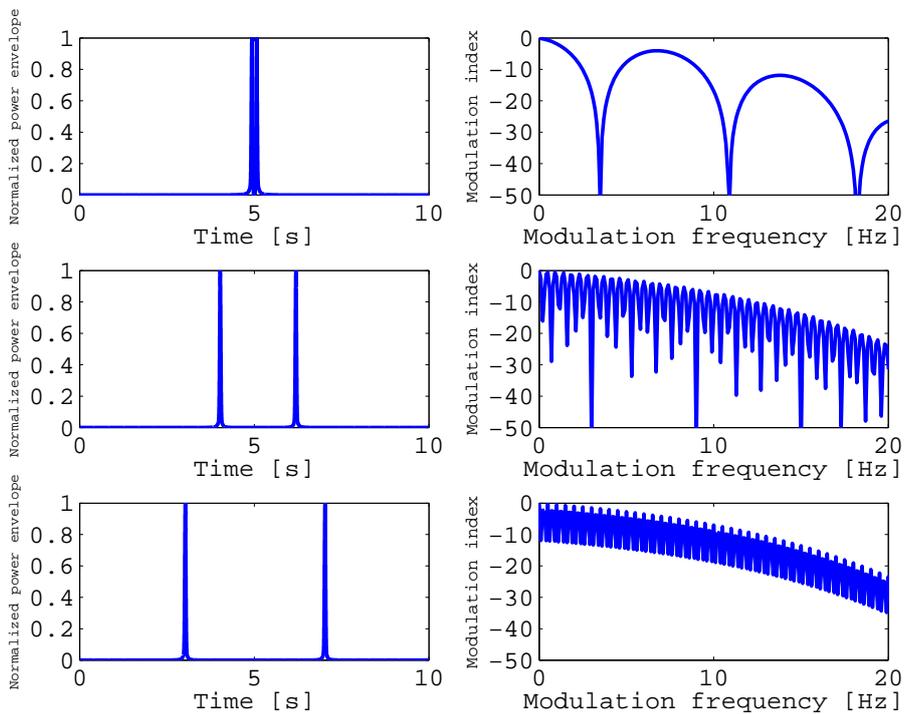


図 3.5: パワーエンベロープと変調スペクトルの形状の関係

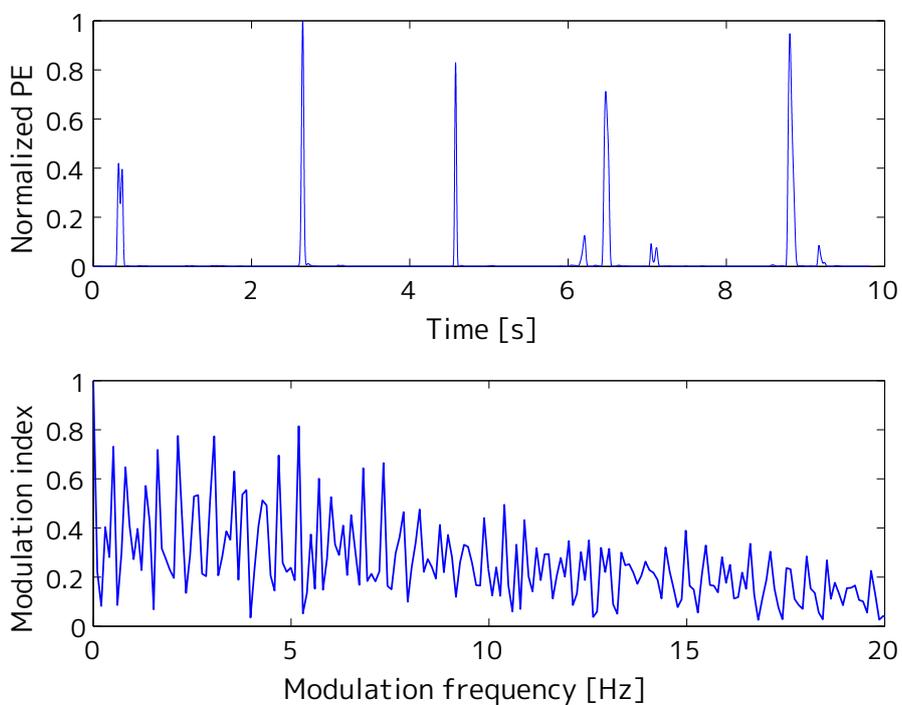


図 3.6: オクターブフィルタバンクで帯域分割した音声信号の変調スペクトル

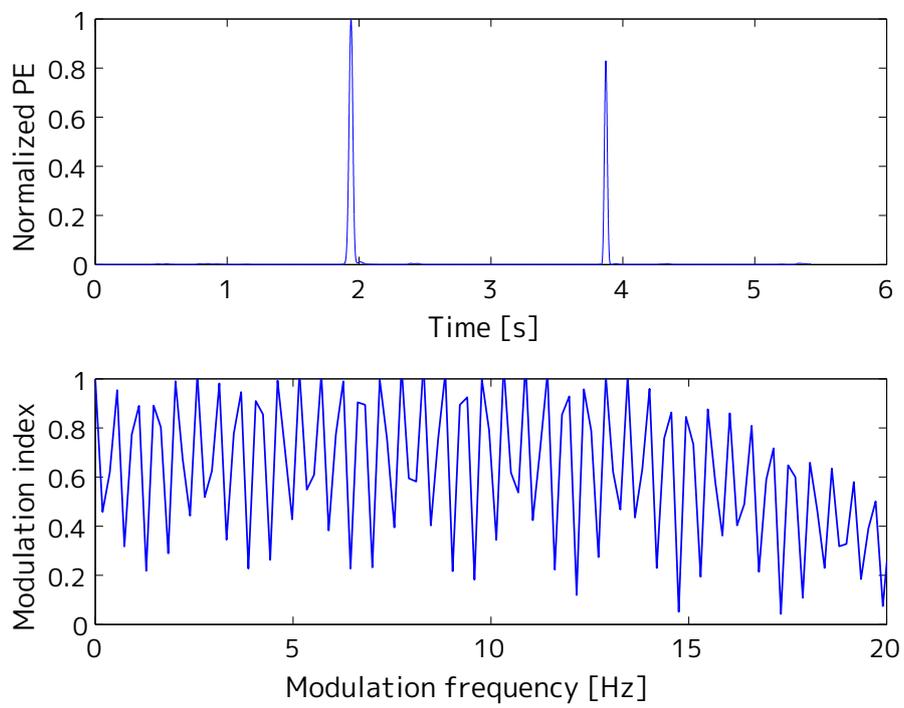


図 3.7: STI 推定に利用する音声区間

第4章 評価実験

4.1 実験の目的

従来法での問題を，提案法によって解決されたか評価するために，様々な信号に対して STI と室内音響特性を推定する実験を行った．評価 1 では，従来法と提案法の推定結果を，評価 2 では，AM 信号を利用した推定の結果と，残響音声信号を利用した推定の結果を，評価 3 では，測定結果と推定結果および室に人が存在する環境と存在しない環境での，推定結果を比較することで，三つの問題それぞれについて解決されたか，評価している．

4.2 評価 1：従来法との比較評価

ここでは，従来法における，推定精度の問題について解決されたかを，従来法と提案法の推定結果を比較することで，評価した．

4.2.1 実験条件

実際の残響環境の室内インパルス応答を用いて，MTF/STI を推定し，推定精度の評価実験を行った．ここでは，入力信号を提案法の仮定を満たすような，2 Hz と 18 Hz に変調スペクトルのピークを持つ正弦波の AM 信号を用いた．残響信号はその入力信号と SMILE2004 データベース [27] に収録されている 43 個の RIR とをそれぞれ畳み込んだ信号とした．表 4.1 にその詳細を示す．これらの RIR は従来法の評価でも用いられたものであり，これを MTF/STI 推定に用いることで従来法と同様の評価が出来る．評価は従来法と提案法の残響時間 T_R ，次数 n ，STI と D 値の推定値を比較することで行った．

4.2.2 実験結果

残響時間 T_R と次数 n ，STI の推定結果を図 4.1，図 4.2，図 4.3 に示す．まず，一般化 RIR のパラメータである，残響時間 T_R と次数 n の推定結果を見る．従来法では，残響時間 T_R の算出値が大きくなるにつれて，推定値との差が大きくなっており，RMS 誤差も 0.8852 と高い値を示している．これに対して，提案法では，RMS 誤差が 0.1760 と従来法よりも低く，推定値と算出値が一致していることがわかる．このことから，算出値と比較

した場合、提案法での推定は算出値通りにできているといえる。また、次数 n の推定結果についても、RMS 誤差は 0.0562 であり、算出値から大きく外れる推定値が見られないことから、精度よく推定できている。

次に STI の推定結果を見る。従来法の推定結果では算出値と大きく外れた箇所が見られるが、提案法では従来法での外れた箇所が、正しく推定されている。また、推定値を全体的に見ても、従来法の RMS 誤差は 0.0589 であるのに対し、提案法は 0.0487 であり、従来法よりも精度よく推定できていることがわかる。これらより、提案法の方が、すべての RIR に対して算出値通りに、高精度に STI を推定できているといえる。

最後に D 値の推定結果を見る。結果の図を見ると、算出値から大きく外れたものはいくつか見られる。また、全体的な推定結果について、誤差の二乗平均平方根 (RMS) をとったが、 $E_{rms} = 0.1413$ と従来法の $E_{rms} = 0.1197$ よりも精度の低いものとなった。高精度に D 値を推定できなかった原因として、二つのことが考えられる。一つは D 値の計算範囲である。従来法では D 値の計算は RIR 信号の最大値を始点にして行なっていたが、提案法では RIR 信号が発生する点を始点として、D 値を算出した。どちらが正しい計算法かは不明であるが、RIR 信号の立ち上がりからピークまでの誤差が比較的大きいことから、この計算範囲の差によって、従来法での D 値よりも誤差が大きくなってしまったと考えられる。もう一つが実測 RIR に対するエンベロップの表現の限度である。MTF は RIR のパワーエンベロップより算出されるが、そのパワーエンベロップはカットオフ周波数が 20 Hz のローパスフィルタがかけられた、RIR のエンベロップによって算出される。これでは、特に RIR の立ち上がりについて、RIR 信号を詳細に表現することができないため、RIR のエンベロップを求めた段階で D 値に差が生じる。そして、これを基に推定された MTF では、D 値の推定が正確に行われないと考えられる。

4.3 評価 2：音声信号を利用した音声伝達指標の推定の評価

ここでは、音声信号を利用した STI 推定について、従来法と提案法の推定値を比較することで、評価を行った。また、音声信号を利用した場合の STI の推定精度が AM 信号を利用したものと同程度であるか、評価 1 での推定結果を比較することで、評価を行った。

4.3.1 実験条件

実際の残響環境の室内インパルス応答を用いて、MTF/STI を推定し、推定精度の評価実験を行った。残響信号は音声信号と RIR とを畳み込んだ信号とした。このとき、音声信号は、ATR 音声データベース [28] に収録されている音声信号を用いた。表 4.2 にその詳細を示す。また、RIR は SMILE2004 データベース [27] に収録されている 43 個の RIR を用いた。表 4.2 にその詳細を示す。従来法と提案法、また評価 1 と評価 2 の提案法の残響時間 T_R 、次数 n 、STI、D 値の推定値を比較することで評価を行った。

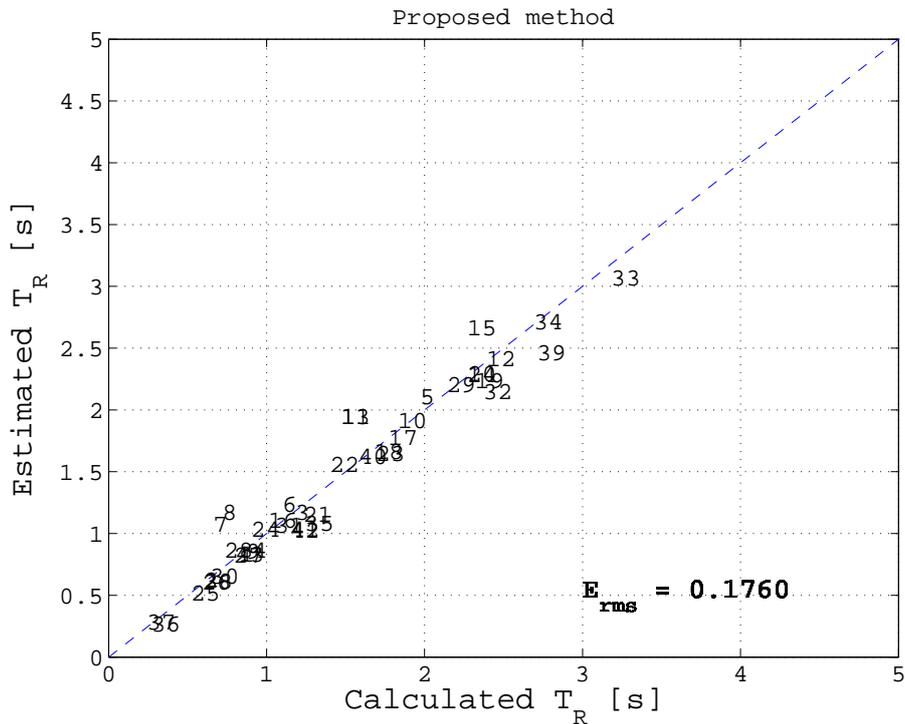


図 4.1: 提案法による残響時間 T_R の推定結果

4.3.2 実験結果

評価 2 での実験結果の図は RIR の No. 毎の推定値にける平均の、3 次の回帰曲線と、RIR の No. 毎の標準偏差を表している。また、RMS 誤差については、RIR の No. 毎の推定値の平均と、算出値を用いて算出した。

まず、従来法と提案法の比較評価を行う。図 4.5~ 図 4.7 に T_R と n の推定結果を示す。 T_R の推定結果について、従来法では過大推定、提案法では過小推定の結果となった。過大推定については従来法での問題点に挙げた、RIR モデルによる誤差と、音声信号を推定対象にしたことによる誤差が原因である。過小推定については n の推定値が影響していると考えられる。 n が過大評価すると T_R が過小推定される傾向がある。また、 n の値の変化は T_R よりも MTF への影響が大きい。これらのことから、提案法での T_R は過大推定が行われた。しかしながら、推定精度を全体的に見ると、従来法の RMS 誤差が 0.9043 であるのに対して、提案法の RMS 誤差は 0.6508 である。また、従来法では標準偏差が比較的大きいものが多数見られるが、提案法ではそれが少なく、また全体的に標準偏差が小さく抑えられている。これらのことから、提案法は従来法よりも高精度に、また音声信号によらず安定して T_R の推定ができることが示された。図 4.8、図 4.9 に STI の推定結果を示す。従来法では全体的に過小推定になっているが、提案法では推定値が算出値に近づいており、推定結果が全体的に改善されている。また、推定値を全体的に見ても、従来法の RMS 誤差が 0.0946 であるのに対して、提案法の RMS 誤差は 0.0595 である。これらのことから、

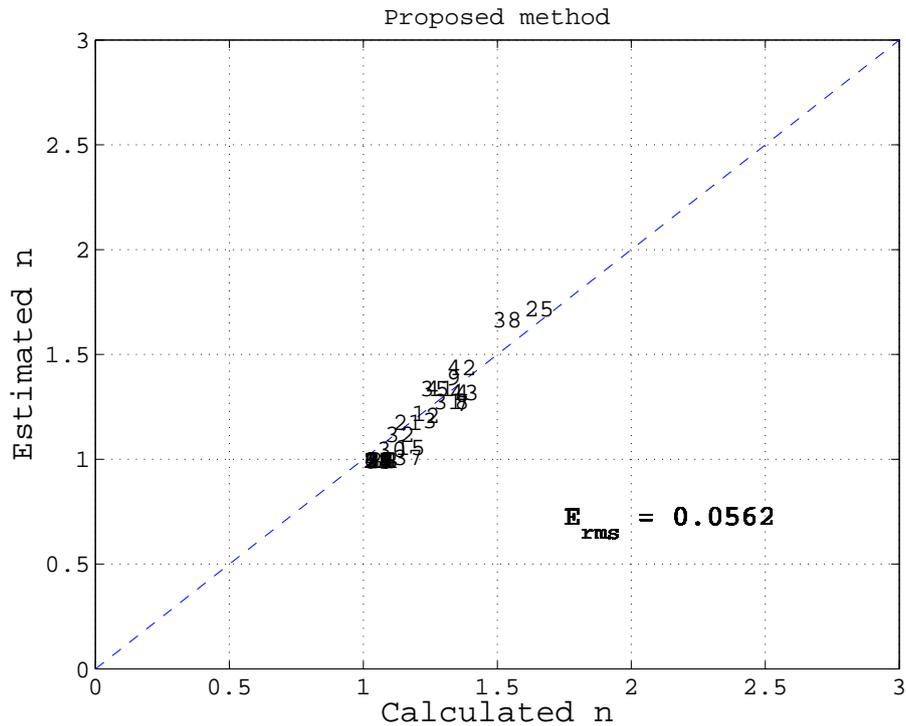


図 4.2: 提案法による次数 n の推定結果

提案法は従来法よりも高精度に STI の推定ができることが明らかになった．図 4.10，図 4.11 に D 値の推定結果を示す．提案法の方が，標準偏差が大きくなりがちだが，推定値を全体的に見ると，従来法の RMS 誤差が 0.2242 であるのに対して，提案法の RMS 誤差は 0.1800 である．これより提案法は，従来法よりも高精度に D 値を推定できているといえる．D 値の推定において，評価 1 では従来法が高精度に推定できたにもかかわらず，評価 2 では提案法が高精度に推定できた．この結果より，提案法は従来法よりも残響音声信号に対して特に高精度に推定できることがわかり，残響音声信号に対する提案法の有効性が示された．

次に，評価 1 と評価 2 の，提案法の推定結果を比較評価する．評価 1 と評価 2 の RMS 誤差はそれぞれ， T_R では 0.1760 と 0.6508， n では 0.0562 と 0.4278，STI では 0.0487 と 0.0595，D 値では 0.1413 と 0.1622，であった．全てにおいて，評価 2 の RMS 誤差が高くなっているが，特に STI と D 値の推定について，RMS 誤差の差が 0.01，0.02 程度と小さく，AM 信号での推定とほぼ変わらない推定を，残響音声信号を利用してできることが明らかになった．しかし， T_R と n の RMS 誤差については，差が大きかった．音声信号による減衰が含まれてしまい， T_R と n の最適な値が得られなかったことが挙げられる，また，正解として定めた値にも原因があるとも考えられる． T_R と n の正解値について，参考となる値がなかったため，推定の正解値を自身によって，実測の RIR に直接一般化 RIR モデルを近似することで決定していた．近似は MTF 上で行っており，RIR のパワーエンベロープについては，近似された MTF から求めている．STI と D 値の推定精度が高かった

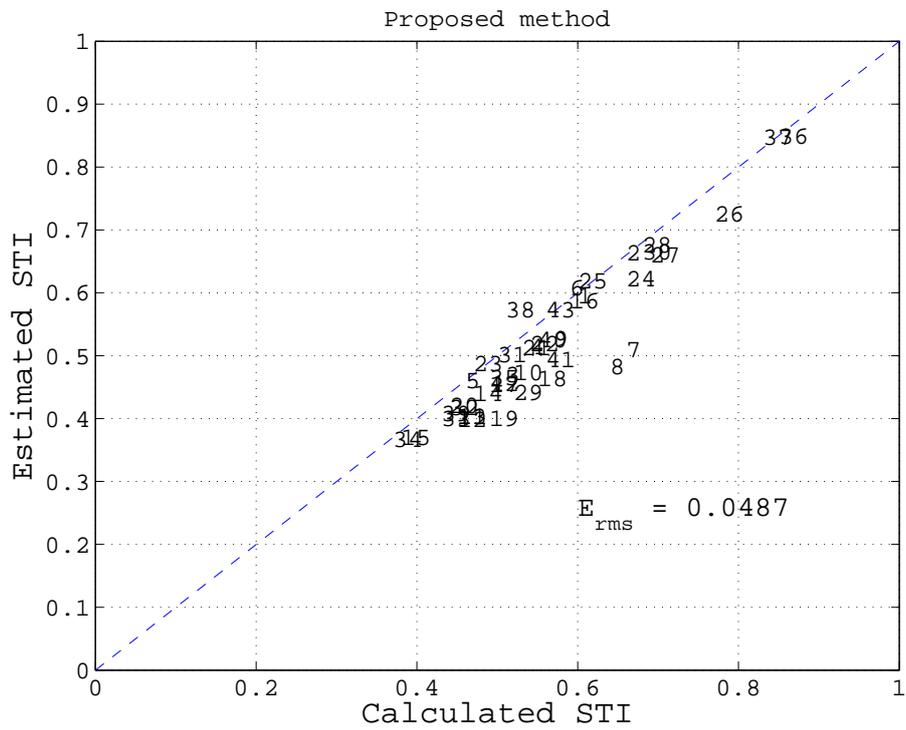


図 4.3: 提案法による STI の推定結果

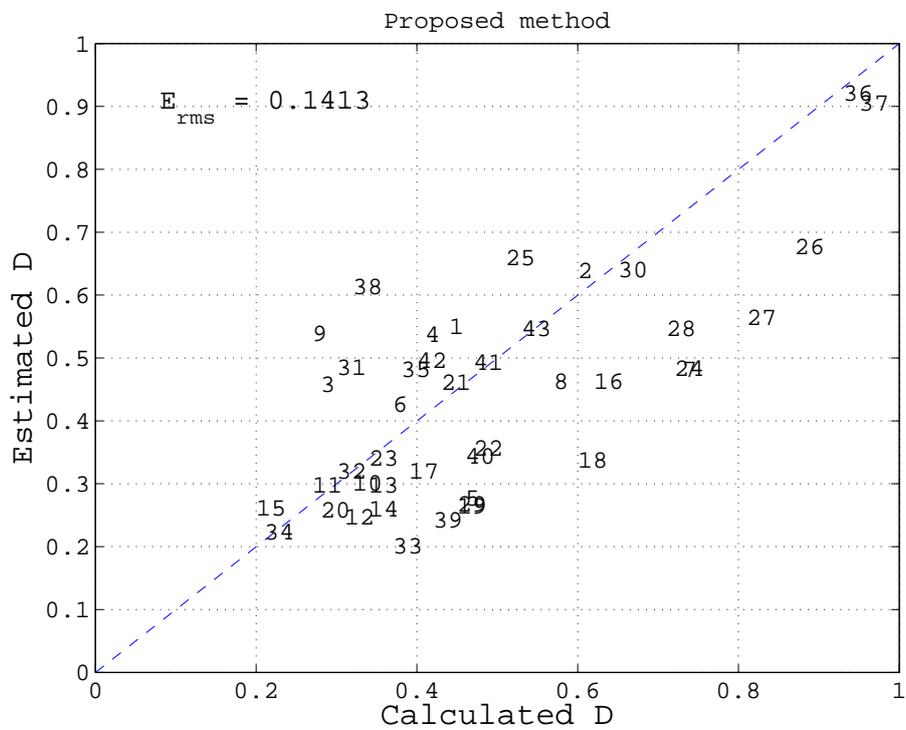


図 4.4: 提案法による D 値の推定結果

ことから，MTF 上ではよく近似されているものの，パワーエンベロープの近似ではズレが生じている，という可能性がある．これによって， T_R と n の正解値に誤りがあり，推定精度が低く見えていると考えられる．

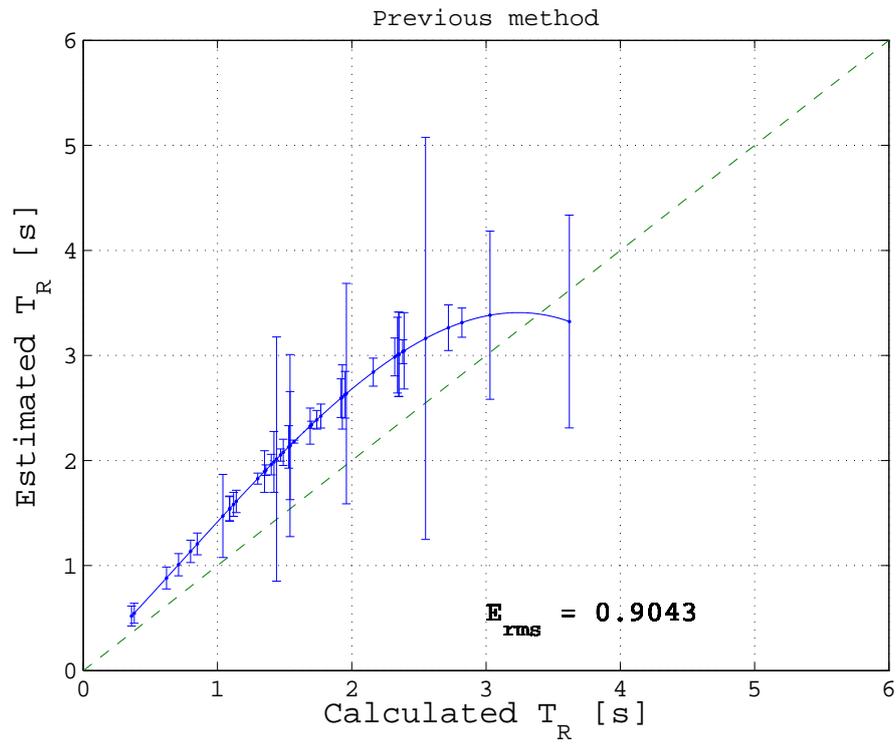


図 4.5: 従来法による音声を利用した残響時間 T_R の推定結果

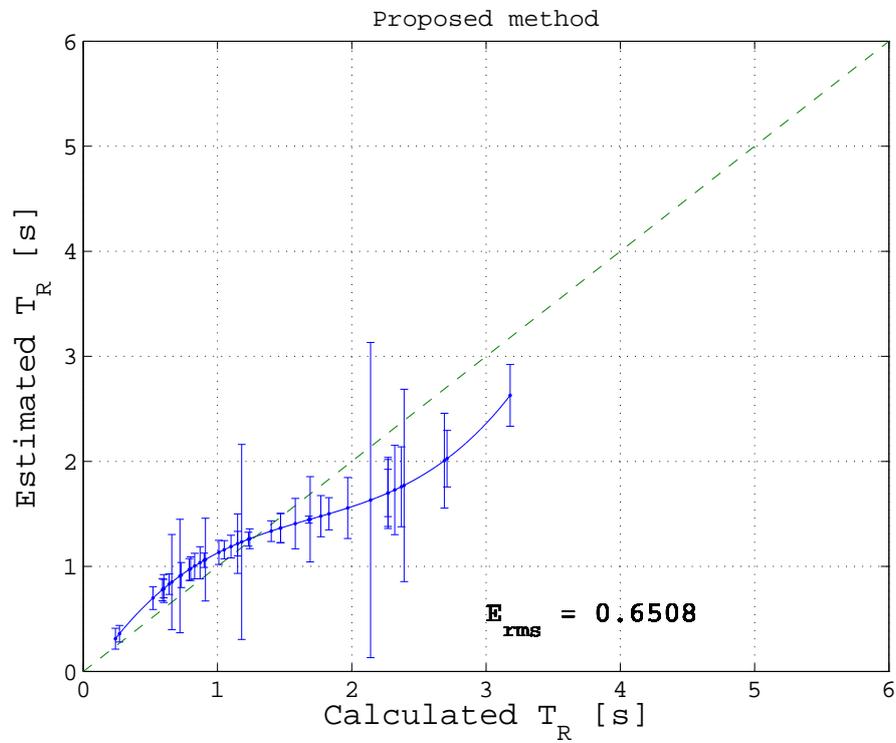


図 4.6: 提案法による音声を利用した残響時間 T_R の推定結果

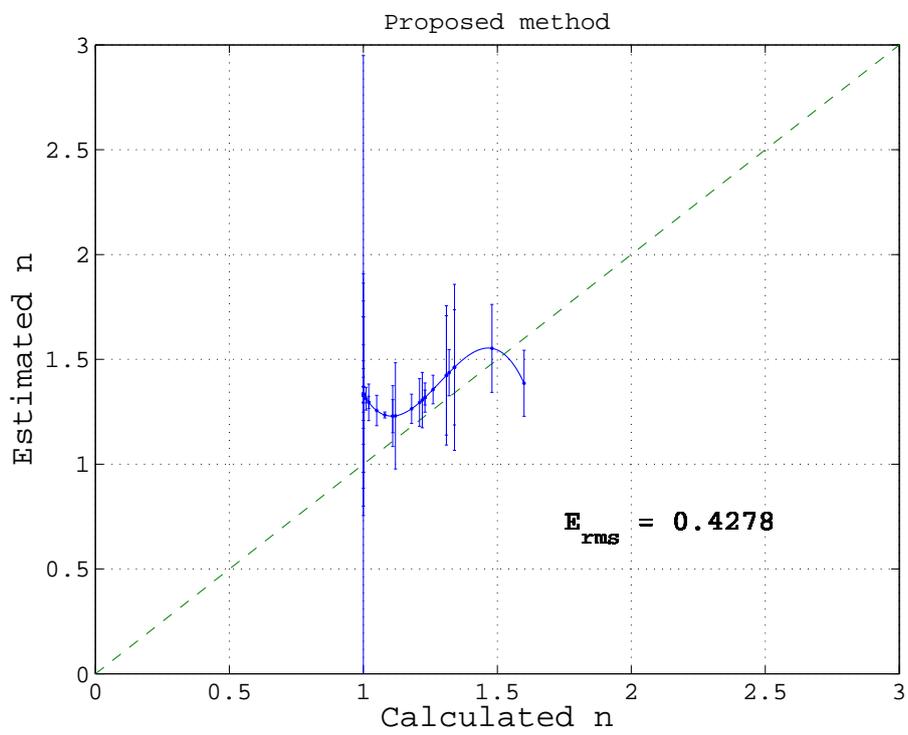


図 4.7: 提案法による音声を利用した次数 n の推定結果

表 4.1: SMILE2004 データベースの各 RIR 条件

No.	RIR No.	Room	T_{60} (s)
1	301	多目的ホール 1 反射板あり	1.09
2	302	多目的ホール 1 反射板なし	0.80
3	303	多目的ホール 2 反射板あり	1.44
4	304	多目的ホール 2 反射板なし	1.04
5	305	多目的ホール 3 反射板あり	1.93
6	306	多目的ホール 3 反射板なし	1.35
7	307	多目的ホール 4 吸音板あり	1.42
8	308	多目的ホール 4 吸音板なし	1.54
9	319	多目的ホール 5	1.47
10	320	多目的ホール 6	2.16
11	309	クラシックホール 1	2.35
12	310	クラシックホール 1 ($d = 6$ m)	2.34
13	311	クラシックホール 1 ($d = 11$ m)	2.35
14	312	クラシックホール 1 ($d = 15$ m)	2.39
15	313	クラシックホール 1 ($d = 19$ m)	2.38
16	314	クラシックホール 2	1.14
17	315	クラシックホール 3	1.96
18	316	クラシックホール 4 吸音カーテンあり	1.92
19	317	クラシックホール 4 吸音カーテンなし	2.55
20	323	クラシックホール 5	2.32
21	324	クラシックホール 6 (1F front)	1.77
22	325	クラシックホール 6 (2F side)	1.74
23	326	クラシックホール 6 (3F)	1.69
24	201	リビングルーム (自由音場)	1.36
25	318	劇場ホール	0.85
26	401	会議室	0.62
27	402	リビングルーム (400 m ³)	1.12
28	403	リビングルーム ($2,400$ m ³)	1.09
29	404	スピーチホール ($11,000$ m ³)	1.54
30	405	教会 1 ($1,200$ m ³)	0.71
31	406	教会 2 ($3,200$ m ³)	1.30
32	407	イベントホール 1 ($28,000$ m ³)	3.03
33	408	イベントホール 2 ($41,000$ m ³)	3.62
34	409	体育館 1 ($12,000$ m ³)	2.82
35	410	体育館 2 ($29,000$ m ³)	1.70
36	411	リビングルーム	0.36
37	412	映画館	0.38
38	413	アントリウム	1.57
39	414	トンネル	2.72
40	415	コンコース	1.95
41	416	スピーチホール 2 (1F 前列席)	1.53
42	417	スピーチホール 2 (1F 中央席)	1.49
43	418	スピーチホール 2 (1F バルコニー席)	1.40

表 4.2: 評価 2 に用いた音声信号

File name	Sentence
FYNOSA433.ad	もちろん発表のときも日本語でよろしいですね .
FKNSA209.ad	私は文化系の研究者なんですが , 言語学の分野での発表はどれだけありますか .
MTTSA103.ad	その要項についてちょっとお尋ねしたいんですけども よろしいですか .

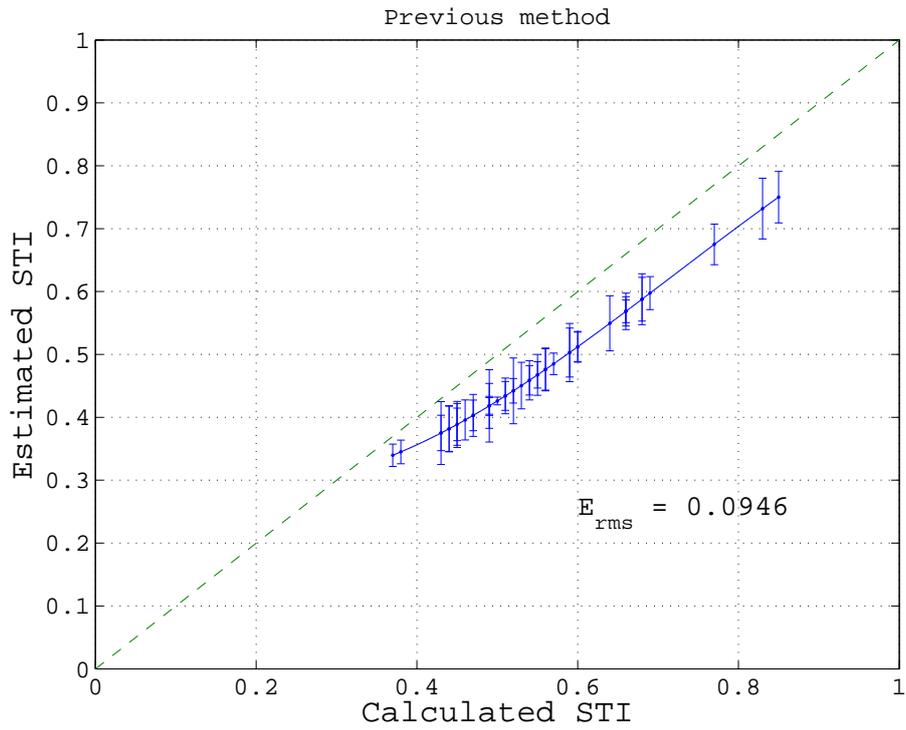


図 4.8: 従来法による音声を利用した STI の推定結果

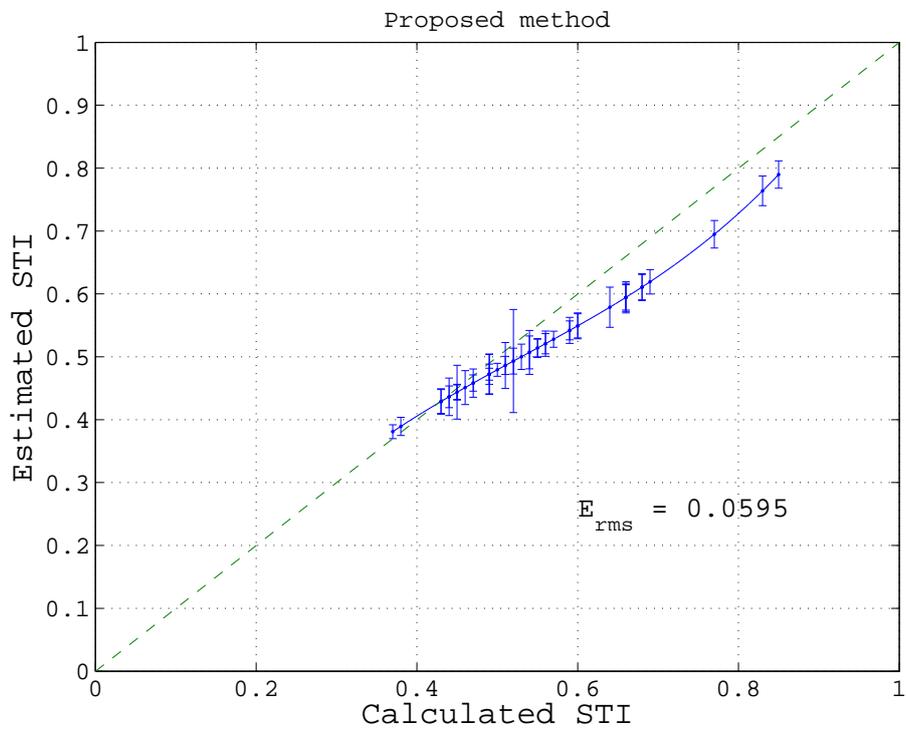


図 4.9: 提案法による音声を利用した STI の推定結果

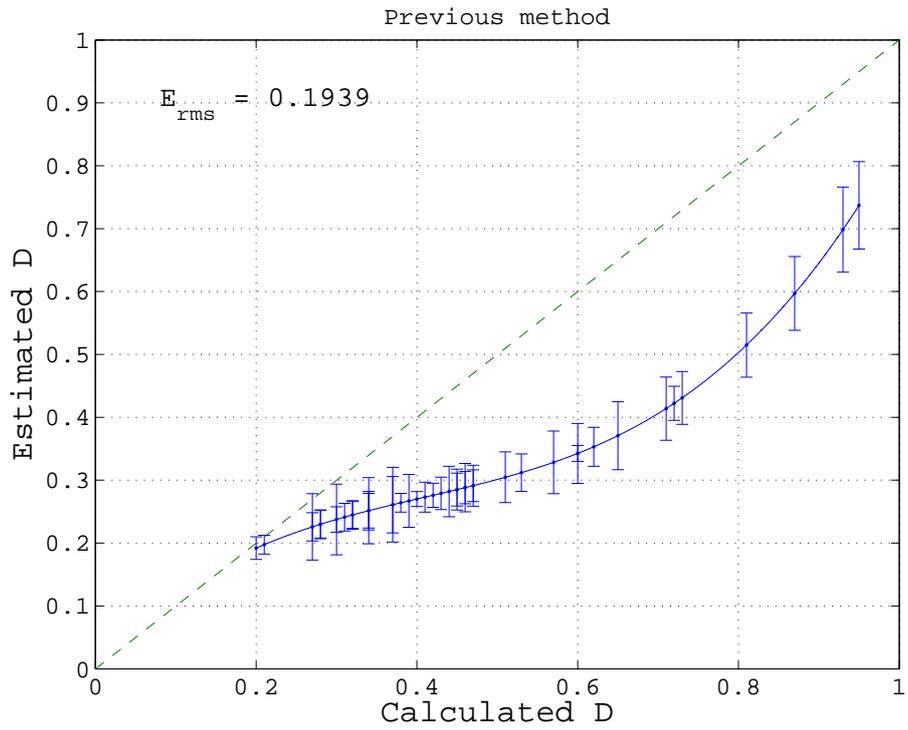


図 4.10: 従来法による音声を利用した D 値の推定結果

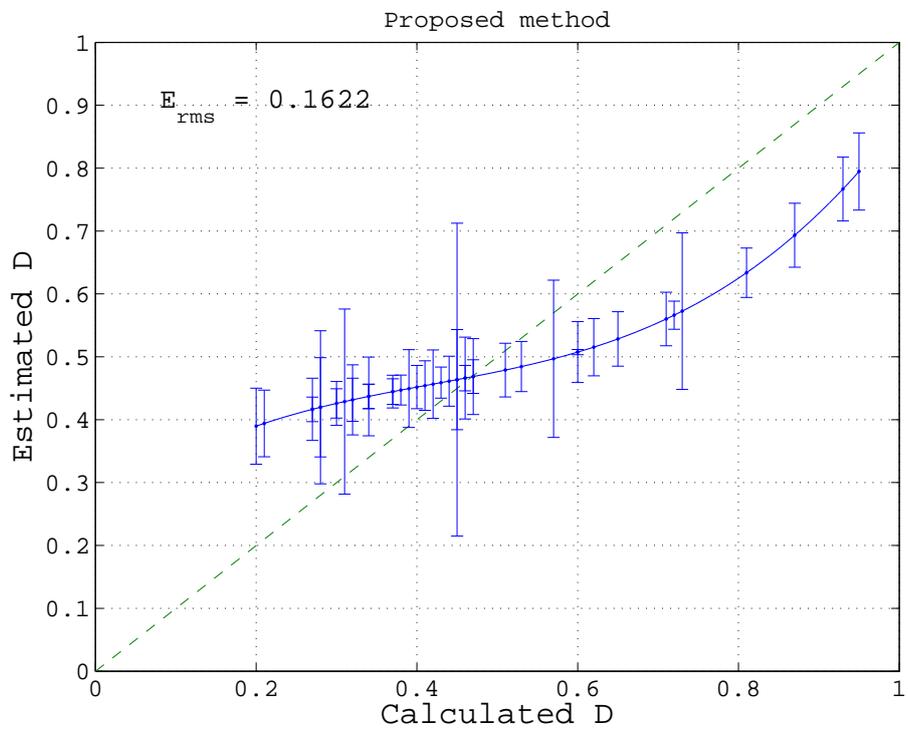


図 4.11: 提案法による音声を利用した D 値の推定結果

表 4.3: 評価 3 に用いた室

Room	Temperature [°C]	Humidity [%]
9 階輪講室	15.9	43
9 階実験室	21.0	39
I34 講義室	12.7	50
大講義室	12.3	49

4.4 評価 3：実測した残響音声信号を利用した評価

ここでは、実環境で STI と室内音響特性が推定可能であるかどうかを明らかにするため、実測値と推定値を比較し、評価を行った。さらに、従来法での問題であった、人がいる音環境でも STI と室内音響特性の推定が可能であるかの実証について、人のいる状態の室と、人を排除した状態の室にて推定を行い、評価を行った。

4.4.1 実験条件

自ら音響測定を行い、残響時間 T_{60} 、D 値と STI を測定した。測定装置を図 4.12 に示す。測定機器には Brüel & Kjær 社製の室内音響特性の測定装置 (2250 Based Building Acoustics System, DIRAC v5.0) を用いた。測定はインパルス応答を測定する手法 [29, 30, 31] をとり、測定された RIR から、残響時間 T_{60} 、D 値と STI を得ることで行った。室内にて、DIRAC v5.0 に収録されているインパルス信号を、アンプをとおして、12 面体スピーカから放射させる。その放射音を騒音計により収録し、DIRAC v5.0 にて信号の解析を行い、残響時間 T_{60} 、D 値と STI を得た。また、推定については、12 面体スピーカから表 4.2 の音声信号を放射させ、その放射音を騒音計により収録し、収録した残響音声信号に推定法を適用した。評価は、測定結果と推定結果を比較することで行った。さらに、人が存在する環境でも推定を行った。室内に人を 2 名配置し、残響音声信号を収録した。そして、その残響音声信号に推定法を適用した。評価は、測定結果、人の存在しない環境での推定結果と比較することで行った。

4.4.2 実験結果

まず、測定結果と推定結果について評価する。残響時間 T_R と次数 n の測定結果と推定結果を図 4.13, 図 4.14 に示す。両者の値とも、過大推定される傾向にあったが、RMS 誤差については低く抑えられており、特に残響時間 T_R の RMS については、測定値と大きく異なる推定値が存在しなかった。これらのことから、残響時間 T_R と次数 n 測定値に対して、有効に推定できていることが明らかになった。STI と D 値の測定結果と推定結果

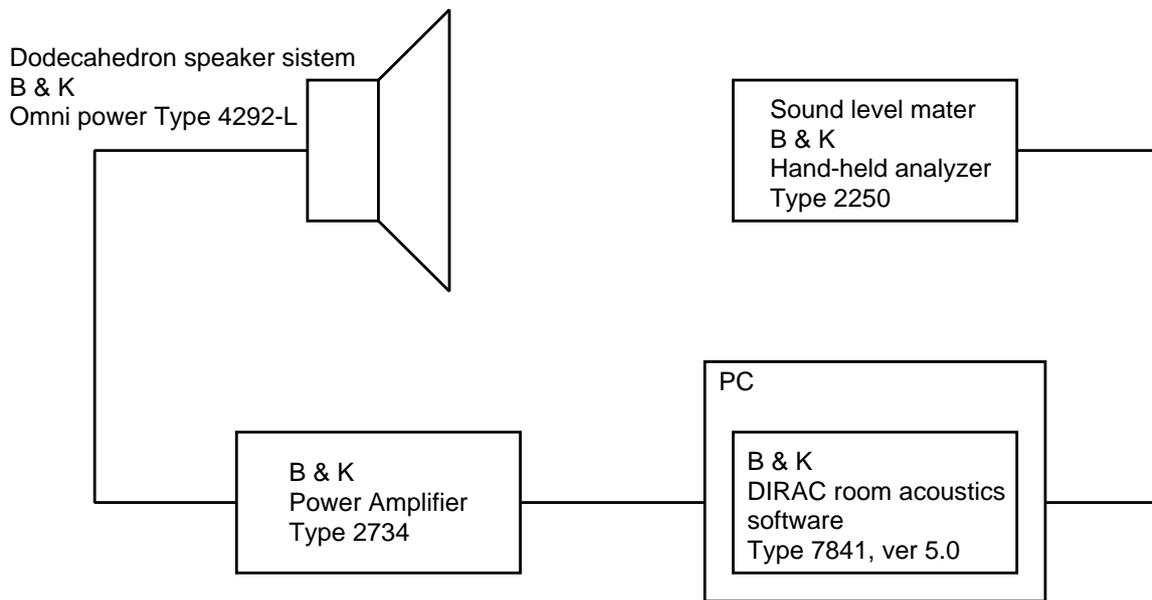


図 4.12: 評価 3 に用いた実験装置

を図 4.15, 図 4.16 に示す. これらの推定値は過小推定されていた. これは残響時間 T_R と次数 n が過大推定で合ったためである. しかしながら, こちらも RMS 誤差が低く, また, STI, D 値ともに測定値と大きく異なる推定値が存在しなかった. STI と D 値の推定についても, 有効に推定できていることが明らかになった. 以上から, 測定値と推定値の精度に大きな違いはないことが示された.

次に, 人が存在する環境と存在しない環境での推定結果について評価する. 推定結果を図 4.17~ 図 4.20 に示す. 推定結果としては, 全ての推定において, ばらつきが見られた. この結果から, 人の存在によって各値が変動することを確認することができる. さらに, 全ての推定において, 人が存在する環境での推定値が, 存在しない環境での推定値よりも大きいものと小さいものの両方が存在した. この結果によって, 人が音声信号を吸音, 反射の両方を行なっていることがわかり, 人の存在は, 室の音声伝送品質を高くおよび低くすることが明らかになった.

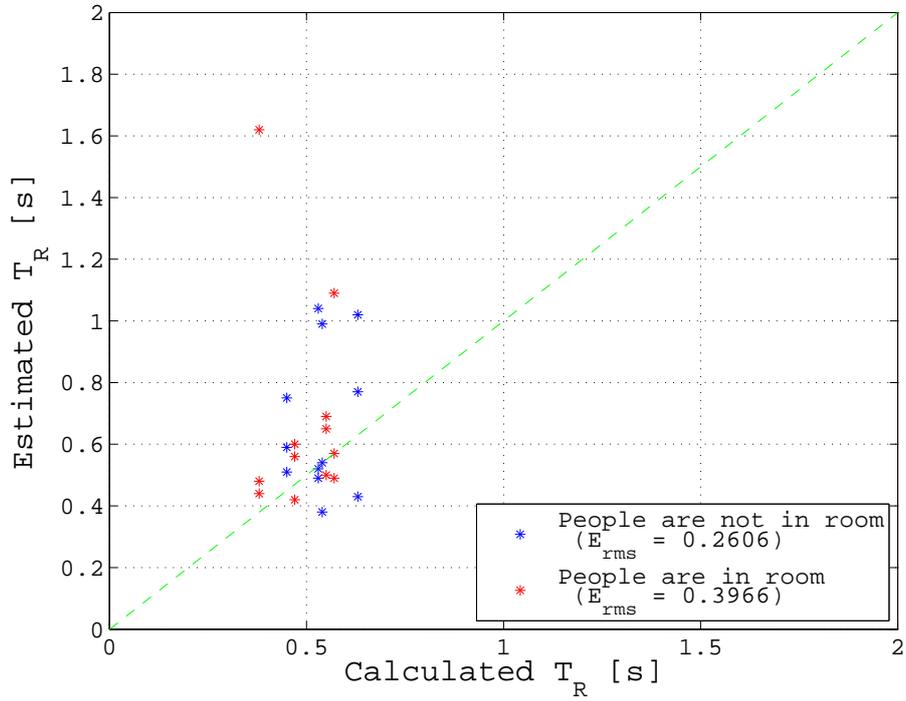


図 4.13: 残響時間 T_R の測定結果と推定結果

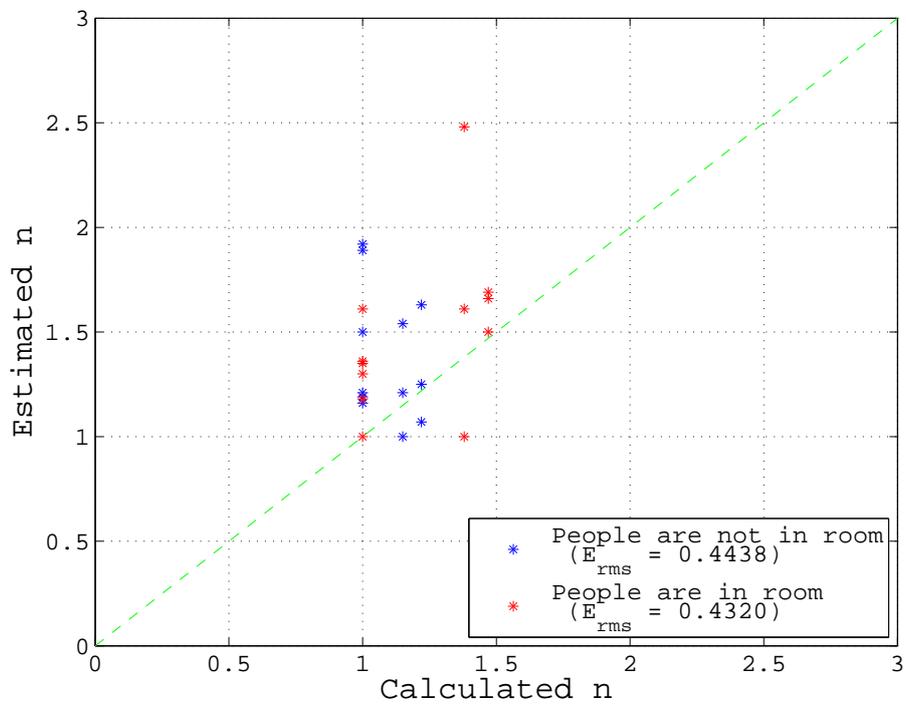


図 4.14: 次数 n の測定結果と推定結果

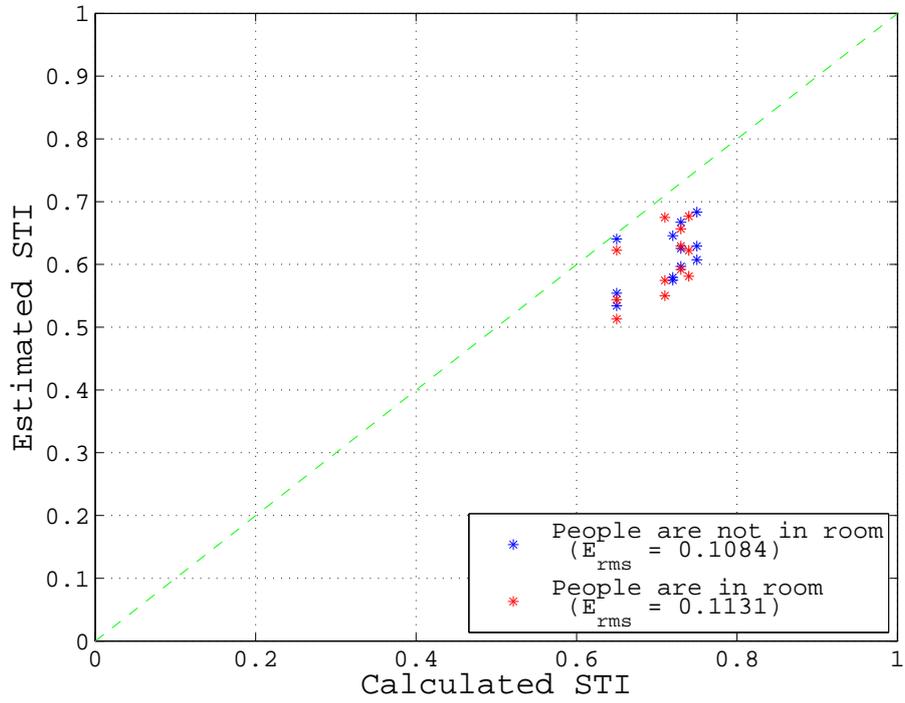


図 4.15: STI の測定結果と推定結果

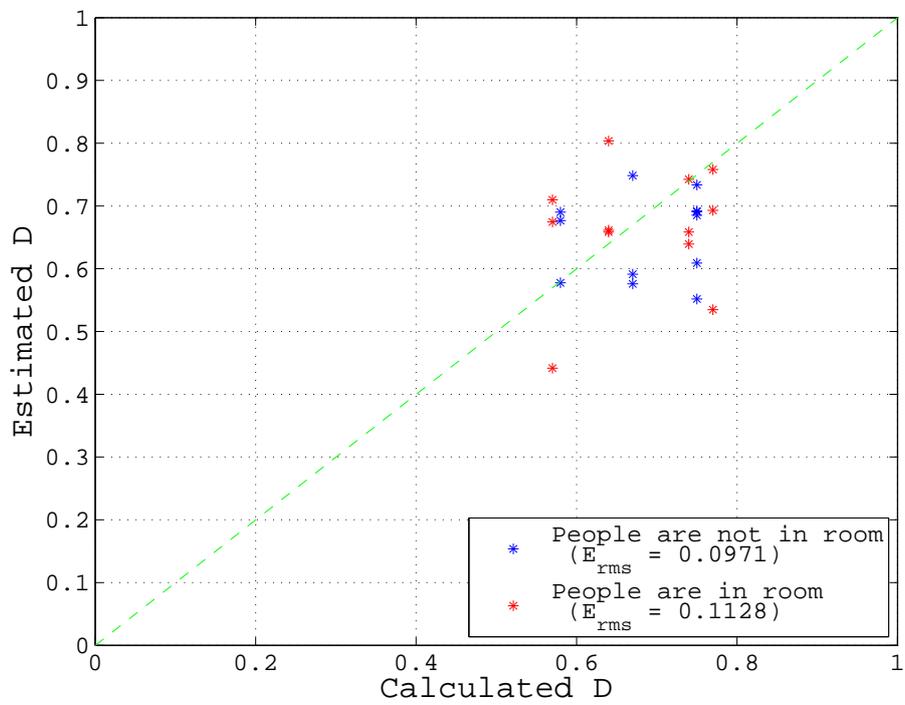


図 4.16: D 値の測定結果と推定結果

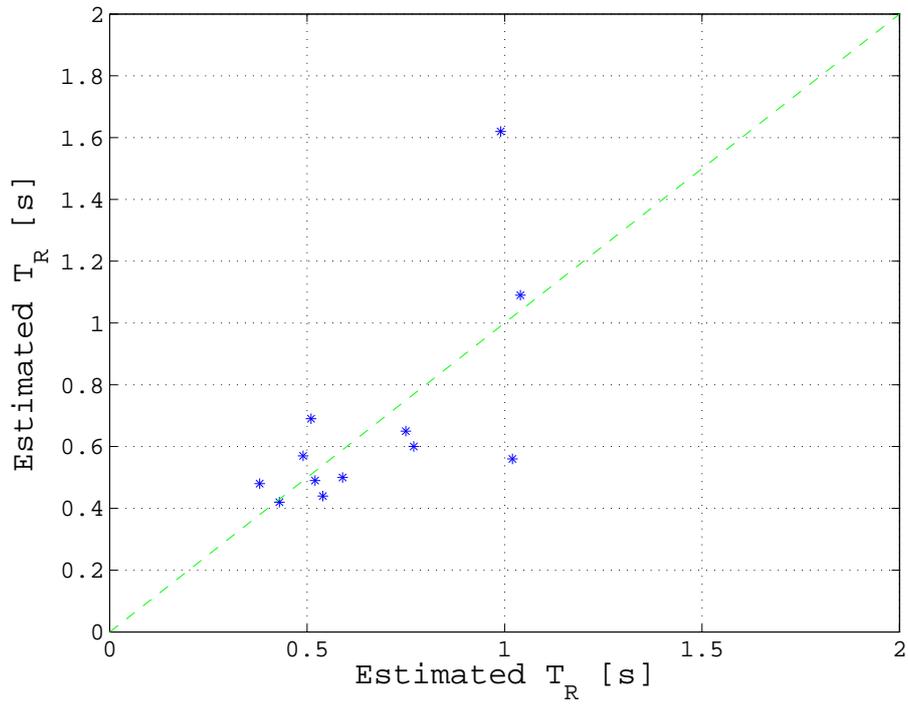


図 4.17: 残響時間 T_R の推定結果 (縦軸: 人が存在する環境, 横軸: 人が存在しない環境)

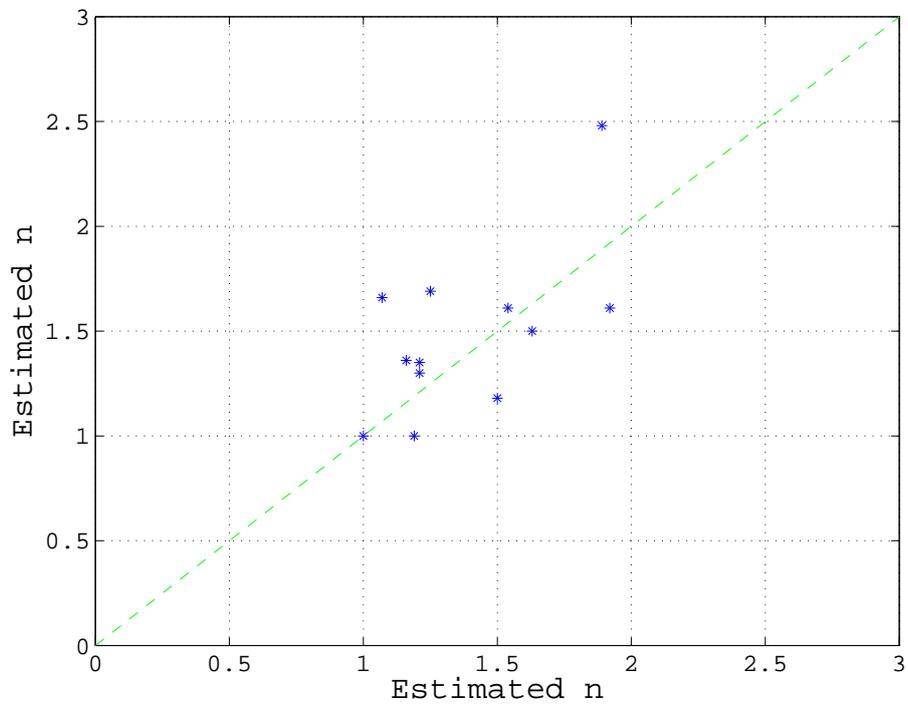


図 4.18: 次数 n の推定結果 (縦軸: 人が存在する環境, 横軸: 人が存在しない環境)

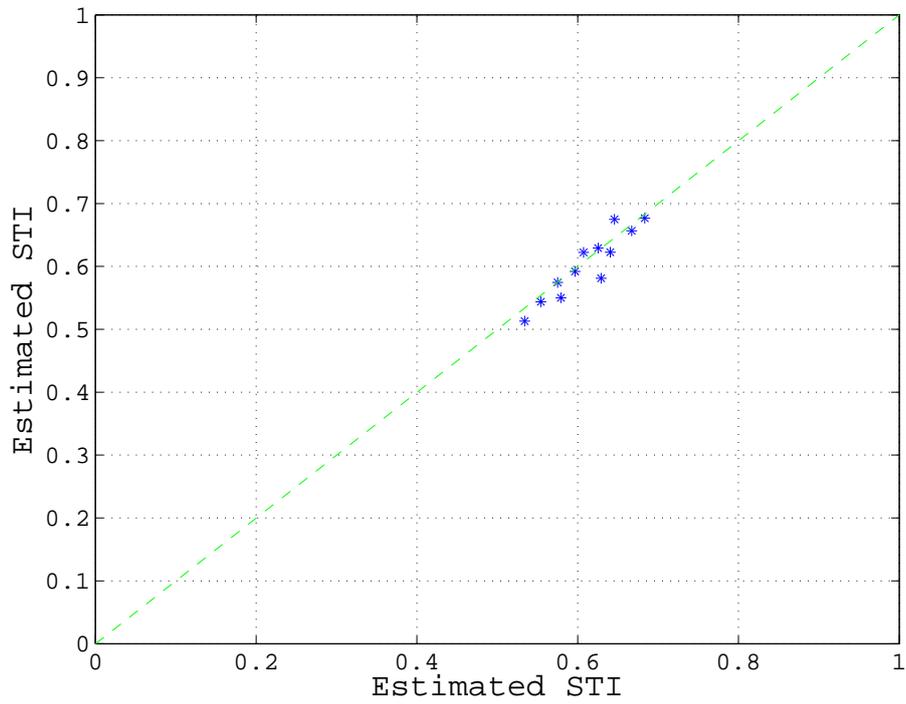


図 4.19: STI の推定結果 (縦軸: 人が存在する環境, 横軸: 人が存在しない環境)

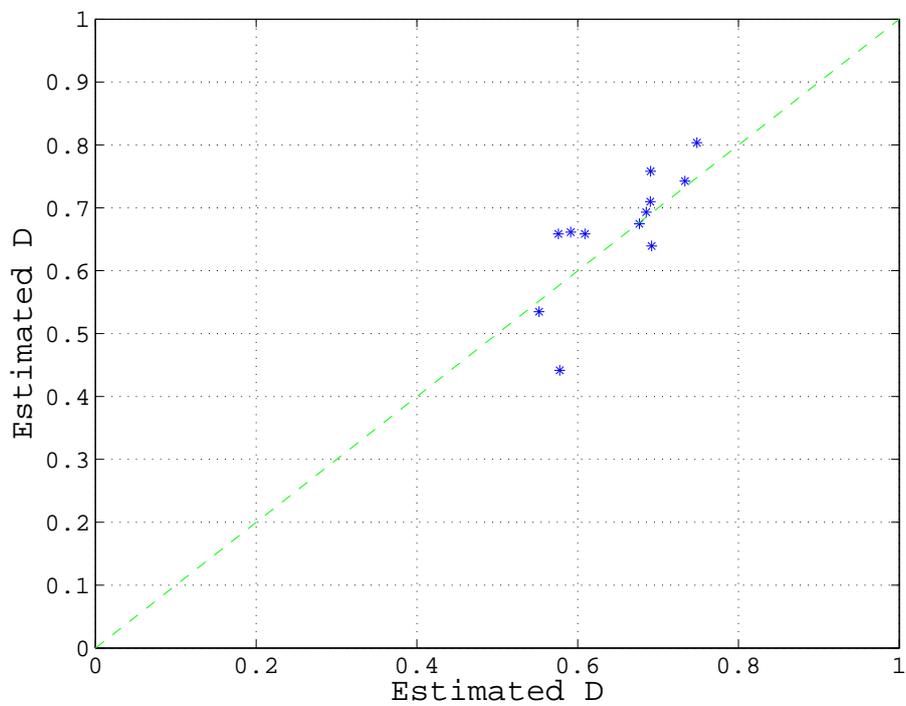


図 4.20: D 値の推定結果 (縦軸: 人が存在する環境, 横軸: 人が存在しない環境)

第5章 結論

5.1 本研究で明らかになったこと

本研究では、STI推定の従来法における三つの問題について解決を行い、室の音響測定を必要としないSTIと室内音響特性の推定法を提案した。一つ目の問題に対して、一般化RIRモデルを用いたMTF/STI推定法を提案した。AM信号を用いた評価実験を行った結果、残響時間 T_R とSTIにおいて、提案法は従来法よりも高精度に推定できることを示し、一般化RIRモデルはSchroederのRIRモデルよりも推定に適していることを明らかにした。しかしながら、D値の推定では提案法は従来法よりも推定精度が低かった。D値の計算範囲と、実測RIRに対するエンベロープの表現の限度にあると考えられた。

二つ目の問題に対して、音声信号を利用したSTIの推定法を提案した。評価実験として残響音声信号を利用してSTIと室内音響特性の推定を行った。結果、全ての推定において、提案法は従来法よりも高精度に推定できることを示した。特にD値の推定で、AM信号を利用した推定では従来法が高精度に推定されたことに対して、残響音声信号を利用した推定では提案法が高精度に推定できたことから、残響音声信号を利用した場合の提案法の有用性を示すことができた。また、AM信号を利用した推定の結果と比較した結果について、全体的にRMS誤差が大きく、特に T_R と n の推定値での差が大きかった。これについては、 T_R と n の最適な値が得られなかったことと、正解として定めた値に原因があると考えられた。しかしながら、STIとD値の推定について、AM信号を利用した場合と残響音声信号を利用した場合とのRMS誤差の差が小さかった。この結果から残響音声信号を利用したSTIとD値の推定では、AM信号を利用した推定と同程度に推定できることが明らかになった。

三つ目の問題に対して、様々な室にて音響測定を行い、実測値と、実測された残響音声信号から提案法によって推定された値とを比較し、評価を行った。結果から、 T_R 、STI、D値について、大きな誤差がなく、全体的に、実測値と推定値の精度に大きな違いはないことが示された。その後、人が存在する環境と存在しない環境で各値の推定値を比較したところ、全ての値において、人の存在による変化が見られた。これにより、人の存在によって音声伝送品質の変化を、提案法によって確認できることが明らかになった。

以上から、室の音響測定を必要としない音声伝達指標と室内音響特性の推定が実現できたことを示した。これより、人の存在する実際の室の環境において、STIと室内音響特性の推定が可能であることを明らかにした。

5.2 残された課題

本研究の今後の展望として残された課題は、二つある。一つ目は、残響音声信号を利用した場合の、 T_R と n の推定値の改善である。残響音声信号を利用した場合の、 T_R と n の推定値の改善については二つ考えがあり、一つは音声信号の音声伝達指標の推定への適用の工夫である。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していたが、帯域幅を変更することで、より推定に適した信号が見つかる可能性がある。実際、Hiramatsu & Unoki による残響音声信号を利用した残響時間の推定 [19] では帯域幅を 100 Hz とする別の分割方法をとっており、最適な分割方法は決まっていない。よって、分割方法の工夫によって、 T_R と n の推定値の改善が考えられる。もう一つは推定での正解値の決定法の変更である。本研究では、正解値の決定について、実測 RIR の MTF に対して、一般化 RIR モデルの MTF を近似することで、正解値を定めていた。しかし、実測 RIR 信号のパワーエンベロープに対して RIR モデルの近似を行うことで MTF での近似とはまた別の結果が出る。このため、実測 RIR 信号のパワーエンベロープを利用した正解値の決定によって、 T_R と n の推定値の改善ができるのではないかと考えられる。しかしながら一般化 RIR モデルを利用したとしても実測 RIR のパワーエンベロープは表現しきれないため、減衰の傾きやピークの位置など、部分的に重点をおいて近似を行い、正解値を定める必要がある。

二つ目は多様な音環境での実験である。本研究にて実測を行った室は、実験室、講義室のみであり、室の数も4箇所であった。使用する目的が同じ室は、STIなどの値にばらつきがなく、推定精度の担保をとるには、実験する室の種類が少なく、十分ではない。また、人の存在する環境での推定も、配置した人が2名と、少ない人数で行った。特に大講義室では、何十人も人が存在することを想定して設計されているため、実際の環境とはかけ離れたものである。これらのことから、室の種類や人の配置を多様化し、より実際の環境に近い室での実験が必要であると考えられる。

謝辞

本研究を進める上で、指導教官である鷓木祐史准教授には、懇切丁寧かつ熱烈峻厳なご指導を賜りました。その言伝身教に甚大なる感謝の意を表します。また、研究室会議など、様々な機会にて、有益な助言を賜った赤木正人教授、研究室の諸氏にも感謝の意を表します。

参考文献

- [1] 戸井田 義徳: 小特集-音声の明瞭度と認識率-, “空間内における音声情報伝達,” 日本音響学会誌, vol. 51, no. 4, pp. 312-316, 1995.
- [2] 小林 陽太郎, 石井 聖光, “教室に於ける聴取明瞭度試験に就いて,” 日本音響学会誌, vol. 7, No. 2, pp. 59-63, 1951.
- [3] 久我 新一, “建築分野への明瞭度試験の応用について,” 日本建築学会研究報告, vol. 27, pp. 255-256, 1954.
- [4] 高岡 知康, 翁長 博, 池田 哲朗, “音声伝達性能を評価できる主観評価方法に関する研究,” 平成 15 年度 日本建築学会近畿支部 研究報告集, pp. 17-20, 2003.
- [5] 近藤 和弘, 泉 良, 藤森 雅也, 加賀 類, 中川 清司, “二者択一型日本語音声了解度試験方法の検討,” 日本音響学会誌, vol. 63, No. 4, pp. 196-205, 2007.
- [6] Morimoto, M., Sato, H. and Kobayashi, M., “ Listening difficulty as a subjective measure for evaluation of speech transmission performance in public spaces, ”J. Acoust. Soc. Am. 116, pp. 1607-1613, 2004.
- [7] 佐藤 逸人, 森本 政之, 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, vol. 63, no. 5, pp.275-280, 2007.
- [8] 佐藤 洋, “音声伝達研究の新しい動き”, 日本音響学会誌, vol. 60, no. 3, pp.141-146, 2004.
- [9] ハイブリット・クットルフ, 室内音響学-建築の響きとその理論-, 藤原恭司 (訳), 日高孝之 (訳), 市ヶ谷出版社, 東京, 2003.
- [10] 佐藤 逸人, 森本 政之, 佐藤 洋, “主観評価に基づいた音声伝送品質のランク,” 第 61 回音シンポジウム, pp. 19-26, 2007.
- [11] 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, vol. 40, no. 3, pp. 181-191, 1984.
- [12] Houtgast, T. and Steeneken, H. J. M., “The modulation transfer function in room acoustics as a predictor of speech intelligibility,” *Acustica.*, vol. 28, pp. 66-73, 1973.

- [13] Steeneken, H. J. M. and Houtgast, T., “A physical method for measuring speech-transmission quality,” *J. Acoust. Soc. Am.*, vol. 67, pp. 318-326, 1980.
- [14] Houtgast, T. and Steeneken, H. J. M., “A review of the MTF concept in room acoustics and its use for estimating speech intelligibility in auditoria,” *J. Acoust. Soc. Am.*, vol.77, no.3, pp. 1069-1077, 1985.
- [15] IEC 60268-16:2003.Sound system equipment Part 16:Objective rating of speech intelligibility by speech transmission index.
- [16] 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.
- [17] 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,” *INTERSPEECH 2007*, pp. 514-517, 2007.
- [18] Wen, J. Y. C., Habets, E. A. P. and Naylor, P. A., “Blind estimation of reverberation time based on the distribution of signal decay rates,” *Proc. ICASSP 2008*, pp. 329-332, 2008.
- [19] Unoki, M. and Hiramatsu, S. “MTF-based method of blind estimation of reverberation time in room acoustics,” *Proc. EUSIPCO2008, Lausanne, Switzerland, 2008(CDROM)*.
- [20] Li, F. F. and Cox, T. J., “A neural network for blind identification of speech transmission index,” *Proc. ICASSP 2003*, vol. 2, pp. 757-760, 2003.
- [21] Kendrick, P., Cox, T. J., Zhang, Y., Chambers, J. A. and Li, F. F., “Room Acoustic Parameter Extraction From Music Signals,” *Proc. ICASSP 2006*, pp. 801-804, 2006.
- [22] Schlesinger, A., “Transient-based speech transmission index for predicting intelligibility in nonlinear speech enhancement processors,” *Proc. ICASSP 2012*, pp. 3993-3996, 2012.
- [23] 鷓木 祐史, 池田 友洋, 宮内 良太, 赤木 正人, “変調伝達関数の概念に基づいた音声伝達指標のブラインド推定法の検討,” *信学技報*, vol. 111, no. 136, pp. 27-32, 2011.
- [24] Unoki, M., Ikeda, T. and Akagi, M., “Blind Estimation Method of Speech Transmission Index in Room Acoustics,” *Proc. Forum Acusticum 2011, Aalborg, Denmark*, pp. 1973-1978, 2011.

- [25] Schroeder, M. R., “Modulation Transfer Function: Definition and Measurement,” *Acustica*, vol. 49, pp. 179-182, 1981.
- [26] Lu, X. and Unoki, M., “A study on temporal envelope restoration methods for reverberant signals based on the modulation transfer function concept,” *Proc.JCA2007*, 2007(CDROM).
- [27] Architectural Institute of Japan, *Sound library of architecture and environment*, Gihodo Shuppan Co., Ltd., Tokyo, 2004.
- [28] Takeda, K., Sagisaka, Y., Katagiri, S., Abe, M. and Kuwabara, H., *Speech Database, ATR Interpreting telephony Research Laboratories*, Kyoto, 1988.
- [29] Schroeder, M. R., “New Method of Measuring Reverberation Time,” *J. Acoust. Soc. Am.*, vol. 37, no. 3, pp. 409-412, 1965.
- [30] 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*, vol. 55, No. 5, pp. 364-369, 1999.
- [31] 横山 栄, “残響時間の測定における注意点,” *日本音響学会誌*, vol. 68, No. 8, pp. 403-408, 2012.

学会発表リスト

1. Kyohei Sasaki, Masashi Unoki, “Study on Estimation Method of MTF/STI Using Generalized Model of Room Impulse Response,” 2013 RISP International Workshop on Nonlinear Circuits and Signal Processing, 2013.
2. 佐々木恭平, 鷗木祐史, “室の音響測定を必要としない音声伝達指標と室内音響特性の推定法の研究,” 2013年日本音響学会春季研究発表会.
3. 佐々木恭平, 鷗木祐史, “室内インパルス応答のモデル化と MTF/STI 推定法の検討,” 信学技報, vol. 112, no. 266, pp. 79-84, 2012.
4. Kyohei Sasaki, Masashi Unoki, “ Study on method for blindly estimating speech transmission index based on concept of modulation transfer function,” 2012 IEEE Workshop on Nonlinear Signal Processing, 2012.
5. 佐々木 恭平, 鷗木 祐史, “室内インパルス応答のモデル化とその MTF の推定,” 平成 24 年度電気関係学会北陸支部連合大会, 2012.

付録

SMILE2004 データベースに収録されている RIR および MTF の形状の近似結果を以下に示す。近似は MTF 上で行い，近似された MTF から，近似されたパワーエンベロープを求めている。それぞれの図の上段にはパワーエンベロープを，中段には上段を dB 変換したものを，下段には MTF を表している。また，緑線は SMILE2004 データベースに収録されている RIR のパワーエンベロープおよび MTF，青線は緑線の RIR に対して Schroder の RIR モデルで近似を行ったもの，赤線は緑線の RIR に対して一般化 RIR で近似を行ったものである。

