

Title	音声伝送指標を基準としたスピーチプライバシー保護の研究
Author(s)	柏原, 佑太
Citation	
Issue Date	2017-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/14139
Rights	
Description	Supervisor: 鶴木 祐史, 情報科学研究科, 修士

修 士 論 文

音声伝送指標を基準とした
スピーチプライバシー保護の研究

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

柏原 佑太

2016年3月

修士論文

音声伝送指標を基準とした
スピーチプライバシー保護の研究

指導教官 鷗木 祐史 教授

審査委員主査 鷗木 祐史 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

1510010 柏原 佑太

提出年月: 2016年2月

概要

音声伝送指標 (STI) は、室における音声伝送品質の評価に利用される客観評価尺度であり、主観評価尺度である聴き取りにくさと高い相関を持つ指標である。STI は、変調伝達関数 (MTF) の概念に基づいており、室内インパルス応答 (RIR) から算出される。本論文の目的は、この STI を基準として、STI を操作し、漏えい音声を聴き取りにくくすることによって、スピーチプライバシー保護を行うことである。

実環境において、室の伝送経路を直接操作して STI を制御することは難しい。しかし、直接音に対して、後部残響に対応する音刺激を遅延和として加算呈示することで見かけ上、RIR を畳み込むことと等価な音響処理が可能である。さらにはこの方法によって STI を制御により音声の聴き取りを制御することもできるかもしれない。まず、その室の RIR を残響音声で操作する。すると STI は、ターゲットとなる音声と RIR を畳み込んだ残響音声によって減少する。したがって、音声プライバシー保護は、第三者の近くの STI を減少させることによって行われる。本研究で提案した方法は、音声通信を安全に行うことができる安全なエリアを作り出すことができる。本論文では、この仕組みに基づき、直接音と後部残響で構成した RIR を用いて、そのパラメータを操作し、STI を制御することで、音声の聴き取りを制御し、スピーチプライバシーが保護されたかどうか検討する。

ある室の音声の聴き取りの状況を予測するには、その室の STI、つまり、その室の RIR を事前に知っておかなければならない。RIR は、一般的に、直接音、初期反射、後部残響の三つの成分で構成される。これらの成分のうち後部残響がもっとも音声の聴き取りに影響を与えるとされている。直接音と後部残響からなる統計的 RIR モデルを利用して、そのモデルパラメータの変化に伴う STI の変化を系統的に予測できることを示した報告がある。これらの結果に基づけば、後部残響を構成するモデルパラメータを操作することで STI を制御できるだけでなく、さらには音声の聴き取りも制御できるし、スピーチプライバシー保護を行うことができる。スピーチプライバシー保護を行うにあたり、まず、拡張型 RIR を後部残響とした後部残響モデルを定義した。拡張型 RIR を後部残響とした後部残響モデルは、立ち上がりを制御するパラメータ T_h や立ち下がり制御するパラメータ T_t 、振幅項 a を容易に制御する事ができる。それゆえに、STI を操作することが容易であった。

提案法の性能評価を行うために、STI を変化させたときに単語了解度試験、聴き取りにくさ、わずらわしさがどのように変化するかを検討した。その結果、単語了解度試験においては、単語親密度に注意すれば、STI の操作によって、単語了解度を操作できることがわかった。また、聴き取りにくさとわずらわしさは、単語親密度にかかわらず STI の操作によって、制御されることがわかった。この結果より、STI を操作し、主観評価尺度である、単語了解度、聴き取りにくさ、わずらわしさを制御することができたといえる。この結果から、スピーチプライバシー保護ができることがわかった。

最後に、提案法が他手法より優れているかどうか検討するために比較実験を行った。STIが0.23のときに、提案法と残響音声とピンク雑音について比較するため、単語了解度、聞き取りにくさ、わずらわしさについて聴取実験をした。その結果、提案法は、他の2手法とほとんど同等の性能を発揮した。さらに、提案法は、他の2手法よりもSNRが低くなったため、効率よくSTIが制御できたといえる。これにより、他の2つのスピーチプライバシー保護の技術よりもスピーチプライバシー保護が容易にできることがわかった。

目次

第1章	序論	1
1.1	はじめに	1
1.2	関連研究	3
1.3	問題点	5
1.4	本研究の目的	5
1.5	本論文の構成	6
第2章	スピーチプライバシーを評価するための尺度と提案法の着目点	7
2.1	主観評価尺度	7
2.2	客観評価尺度	8
2.3	本研究の着目点	9
第3章	本研究における提案法の原理	11
3.1	変調伝達関数の概念	11
3.1.1	変調伝達関数の算出	12
3.1.2	STIの算出方法	12
3.2	雑音のMTFとSTI	13
3.3	RIRモデルのMTFとSTI	15
3.3.1	SchroederのRIRのMTFとSTI	15
3.3.2	一般化RIRのMTFとSTI	19
3.3.3	拡張型RIRのMTFとSTI	23
3.3.4	単一エコーのMTFとSTI	27
3.3.5	SchroederのRIRを後部残響としたモデルのMTFとSTI	31
3.3.6	後部残響を一般化RIRとしたモデルのMTFとSTI	35
3.4	提案法	39
第4章	提案法の評価	44
4.1	評価方法	44
4.2	提案法の原理の評価	49
4.3	提案法と従来法の比較評価	53
4.4	考察・まとめ	54

第 5 章 結論	61
5.1 本研究で明らかになったこと	61
5.2 残された課題	62
付 録 A MTF の導出方法の一例	68
A.1 はじめに	68
A.2 MTF の定義	68
A.3 導出方法及び導出した MTF	68
A.3.1 単一エコーに対応する MTF	68
A.3.2 Schroeder の RIR に対応する MTF	69
A.3.3 後部残響を Schroeder の RIR としたものに対応する MTF	69
A.3.4 一般化 RIR に対応する MTF	70
A.3.5 後部残響を一般化 RIR としたものに対応する MTF	71
A.4 おわりに	71

目次

1.1	スピーチプライバシーの概念図	2
1.2	パワーマスキングでのスピーチプライバシー保護	3
1.3	残響音を用いたスピーチプライバシー保護	4
2.1	本研究の概要図	10
3.1	変調伝達関数の概念図	11
3.2	雑音の MTF と STI	14
3.3	Schroeder の RIR のパワーエンベロープ	16
3.4	Schroeder の RIR の MTF	17
3.5	Schroeder の RIR の STI	18
3.6	一般化 RIR のパワーエンベロープ	20
3.7	一般化 RIR の MTF	21
3.8	一般化 RIR の STI	22
3.9	拡張型 RIR のパワーエンベロープ	24
3.10	拡張型 RIR の MTF	25
3.11	拡張型 RIR の STI	26
3.12	単一エコーのパワーエンベロープ	28
3.13	単一エコーの MTF	29
3.14	単一エコーの STI	30
3.15	Schroeder の RIR を後部残響としたモデルのパワーエンベロープ	32
3.16	Schroeder の RIR を後部残響としたモデルの MTF	33
3.17	Schroeder の RIR を後部残響としたモデルの STI	34
3.18	一般化 RIR を後部残響としたモデルのパワーエンベロープ	36
3.19	一般化 RIR を後部残響としたモデルの MTF	37
3.20	一般化 RIR を後部残響としたモデルの STI	38
3.21	提案法の概念図	40
3.22	拡張型 RIR を後部残響としたモデルのパワーエンベロープ	41
3.23	拡張型 RIR を後部残響としたモデルの MTF	42
3.24	拡張型 RIR を後部残響としたモデルの STI	43
4.1	実験機構の図	46

4.2	SNR の決め方	47
4.3	単語了解度の結果	50
4.4	聴き取りにくさの結果	51
4.5	わずらわしさの結果	52
4.6	他手法の比較：単語了解度の結果（高親密度の場合）	55
4.7	他手法の比較：単語了解度の結果（低親密度の場合）	55
4.8	他手法の比較：聴き取りにくさの結果（高親密度の場合）	56
4.9	他手法の比較：聴き取りにくさの結果（低親密度の場合）	56
4.10	他手法の比較：わずらわしさの結果（高親密度の場合）	57
4.11	他手法の比較：わずらわしさの結果（低親密度の場合）	57
4.12	他手法の比較：単語了解度と SNR の関係（高親密度の場合）	58
4.13	他手法の比較：単語了解度と SNR の関係（低親密度の場合）	58
4.14	他手法の比較：聴き取りにくさと SNR の関係（高親密度の場合）	59
4.15	他手法の比較：聴き取りにくさと SNR の関係（低親密度の場合）	59
4.16	他手法の比較：わずらわしさと SNR の関係（高親密度の場合）	60
4.17	他手法の比較：わずらわしさと SNR の関係（低親密度の場合）	60

第1章 序論

1.1 はじめに

音声コミュニケーションでは、多くの情報が伝送される。近年、病院や薬局、銀行などが、個人情報保護法などの強化に基づいて、個人情報における情報セキュリティ対策が重要視されている。それに伴い、個人情報が伝送される音声コミュニケーションでは、話者やそれとは関係なく盗聴する第三者も音声コミュニケーションの秘話性について、細心の注意を払う必要があるとされる。例えば、話者同士が話した内容（言語情報）だけでなく、話ただけで、話者個人の特定が可能な情報（非言語情報）が伝送される。このような情報は、病院や薬局、銀行、ホテルのロビー、重要な会議が行われているオープンオフィスなどのオープンスペースで伝送されることが多々ある。話者同士で情報の伝送をやり取りすることは問題にならない。しかし、図1.1のように、音声コミュニケーションに全く関係ない第三者に漏えいすることにより、個人情報の漏えいや重要な機密情報の漏えいにつながった場合、プライバシー保護の観点から見ると非常に重大な問題となる。このような問題に対して、音声コミュニケーションにおけるプライバシー（スピーチプライバシー）保護は、実現しているとは言いがたい [1]。

一方、日本と欧米では、スピーチプライバシー保護に対する考え方が異なっている。日本では、上記のような話者の声が第三者に聞こえない状況を作ることを指すが、欧米では、話者ではなく第三者の執務などに邪魔にならないような環境を構成法のことを指す。これは、日本と欧米の文化や考え方の相違から発生する違いである。本論文では、オープンスペースでのスピーチプライバシー保護を行うため、前者のスピーチプライバシー保護に焦点を合わせて研究を行う。スピーチプライバシーが現在保護されていない理由は、例えば、聴知覚の機能としてあげられるカクテルパーティー効果 [2] やマガーク効果 [3]、聴覚情景解析 [4, 5] などがある。これらの現象によって、ヒトは、雑音環境下や情報が少ない状況であっても、目的音を知覚することができる。カクテルパーティー効果とは、聴取者が様々な音（雑音や人の話し声）の中から目的音だけを選択的に聴取できる現象で、マガーク効果とは、話者の口の形だけで、聴取者が何を話しているか会話の内容を予測する現象である。聴覚情景解析は、音声の立ち上がりや立ち下がりや周波数成分の類似によって、音声を分凝するような現象である。ヒトが音声を頑健に聴取するために必要な効果であるが、情報セキュリティが重要視されるなかこのような効果は、スピーチプライバシー保護に対して妨害となるような効果である。様々な音環境において、スピーチプライバシー保護を行うには、上記のような様々な問題を解決する必要がある。このような音声の

聴取に頑健となるような効果を逆手に取ったようなスピーチプライバシー保護に関する研究や技術は存在するが、どれもスピーチプライバシー保護を行うには、決定的な効果を得ていない。スピーチプライバシー保護の度合いを評価するためには、ヒトによる聴取実験において主観評価が不可欠になっており、評価には大規模かつ、多大な時間を要する。また、聴取実験でしかスピーチプライバシー保護の度合いを評価していないため、ある音環境においてスピーチプライバシー保護ができたからと言って、他の音環境でスピーチプライバシー保護ができるかわからない。そのため、聴取実験において主観評価尺度を用いることは別に、物理測定値から算出される客観評価尺度を基準にすることで主観評価に頼らず、どのような音環境にも適応できるスピーチプライバシー保護の技術は、必要不可欠である。

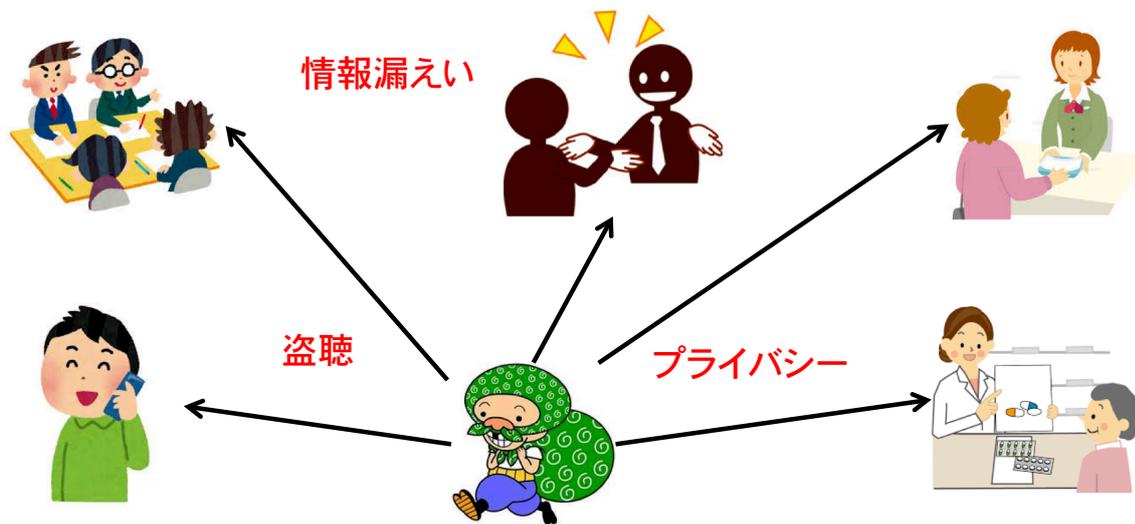


図 1.1: スピーチプライバシーの概念図

1.2 関連研究

スピーチプライバシー保護の方法は、山口等による防音壁を用いて音声の漏えいを防ぐ方法 [6] やアクティブノイズコントロール (ANC) [7] を用いて音を物理的に遮断する方法がある。前者は、防音壁によるオープンスペースの利便性の制限という欠点を持つ。話者同士を、密閉した室内に閉じ込めてしまえば良いが、これだと銀行や病院の店舗におけるオープンスペースを活用しているとは言いがたい。後者は、目的音の逆位相の音を話者が話している目的音に付加することによって、音を消す事ができるが、オープンスペースを想定した場合、すべての音を消すことは現実的ではない。そこで、佐伯らはパワーマスクングによって、目的音をかき消す方法を提案した [8]。ここでは、ヒトの聴覚特性に着目して、帯域制限したピンク雑音を使用してマスクングを試みた。帯域制限を行ったピンク雑音は、他の様々な雑音と比較したとき、単語の了解性を示す単語了解度が低いことがわかった。また、多種類の音声を混合して作成されるバブル雑音は、単語了解度が低い結果が得られた [9]。しかし、この方法では、図 1.2 のように目的音をかき消すために目的音よりも多大な音圧の定常雑音を放出するため、騒音暴露の発生や、不快感を発生させる。また、常時定常雑音を放出するため、銀行などの衆人環境下では、使用することができない。

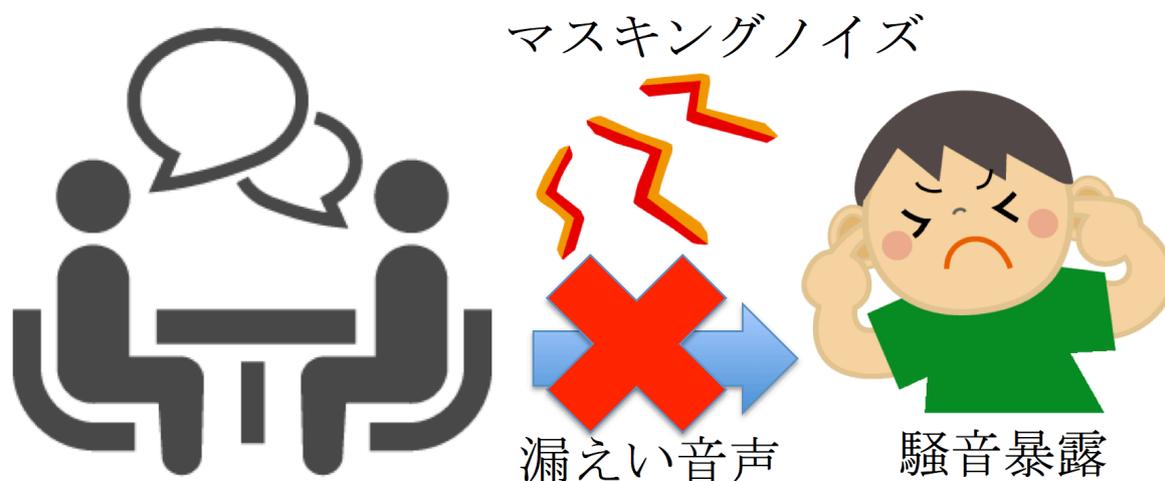


図 1.2: パワーマスクングでのスピーチプライバシー保護

これらの問題に対して、赤木と入江は、聴覚情景解析に基づいた方法を提案した [10]. これは、音の知覚的融合を利用したカクテルパーティー効果を逆手に取った方法である. 知覚的融合とは、通常、分凝されて聴こえてくるふたつ音（目的音と防聴音）に対して、音声の周波数に着目して、音声の重要な情報を消す消す処理を行い、全く違う一つの音として知覚する現象である. その結果、マスキング法と同程度の聴き取りにくさを持ちながら、不快感が低いことがわかった. しかし、目的音と防聴音が融合された漏えい音声は、多々、不快感を与える場合がある. 音環境によっては、カクテルパーティー効果によって目的音が聴き取れてしまうこともある. このことより、オープンスペース内の音環境に十分適応できているとは言い切れない. また、話者の声からマスクを作成する方法 [11, 12] もある. この方法は、ヒトの音声からマスクを作成するため、個人の音声をあらかじめ収録しなければならないため、スピーチプライバシー保護には向いているとは言いがたい.

一方で、佐藤らによる残響音声を用いる方法 [13, 14] は、室内音響指標のひとつである音声伝送指標 (STI) [15] と聴き取りにくさに着目したものである. 残響音声とは、会話音声に室内インパルス応答 (RIR) を畳み込んだ音声のことである. 残響音声は、我々が普段よく聴く音で、風呂場や協会などで、音を発したときに連続的に複雑な反射を起こしながら聴こえてくる音声のことである. ここでは、図 1.3 のように、音声にある一定の RIR を畳み込むことによって、残響音声を作成し第三者に漏えいさせるため、オープンスペースの利用は制限されず、騒音暴露は起こらず、その残響における聴き取りにくさによって、スピーチプライバシーが保護される. また、聴覚情景解析の観点から、漏えい音声は原音声と残響のふたつに分凝されず、常に知覚的融合が行われ、残響音声そのものの不快感は抱かないものと考えられる. しかしこの方法では、漏えい環境に依存せずに残響を付与するような処理を行うため、場所によっては、スピーチプライバシー保護が十分でなかったり、過剰に残響が付与される状況を生み出したりする可能性がある.

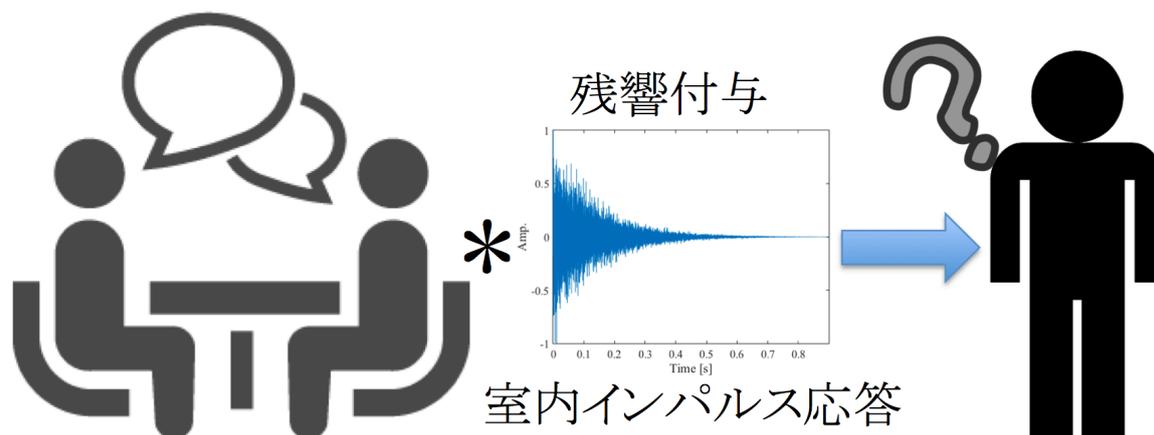


図 1.3: 残響音を用いたスピーチプライバシー保護

1.3 問題点

従来の研究では、スピーチプライバシー保護を評価するために主観評価尺度を使用している。主観評価尺度は、聴取者が聴取者個人の実験環境によって変動する評価である。主観評価尺度を用いてスピーチプライバシー保護を評価することは問題にはならない。しかし、スピーチプライバシー保護を評価するうえで主観評価尺度は、実験環境における音環境しか考慮していないため、様々な音環境に対応しているとはいえない。さらに、目的の場所に適応したスピーチプライバシー保護の技術が完成したとしても、オープンスペースを考慮したとき、ヒトの往来があることが想定され、常に音環境が変化すると考えるべきである。そのときに、従来の研究では音環境が変化することまでは、踏み込んでいない。パワーマスキングや、残響音声を用いる方法は、その最たる例である。そこで、主観評価尺度ではなく、物理測定値から算出される客観評価尺度を用いてスピーチプライバシー保護を遂行することが重要な課題である。

1.4 本研究の目的

スピーチプライバシー保護を行うにあたり、主観評価尺度でスピーチプライバシー保護を評価することは、様々な音環境に対応できないことに直接関わってくる。そこで、本研究は、漏れ音声において、主観評価尺度ではなく、客観評価尺度の一つである STI を基準として漏れ音声の聴き取りにくさに着目することで、スピーチプライバシー保護の実現を目的とする。宮崎らや鷗木ら [16, 17] による STI 推定により会話者の周辺に漏れる音声の聴き取りにくさを常に予測し、STI を能動的にかつ正確に制御することで、その場における聴き取りにくさを担保することである。RIR は、一般的に、直接音、初期反射、後部残響の三つの成分で構成される。その中でも、後部残響が音の聞き取りに最も影響を与えているという事は、知られている。後部残響を変化させるだけで、STI を能動的に操作ができるような機構を提案できれば、防音壁などを使わずとも、安全に音声コミュニケーションを行えるエリアをそうではないエリアに分割することができる。また、適切な聴き取りにくさを保持するために、単に残響時間を長くするのではなく、聴き取りにくさに直接寄与する RIR における概形の特徴（後部残響）を変形させ、不快感のないスピーチプライバシー保護を実現することである。提案法を実現できれば、漏れ音を聞こうとする第三者のいる場の STI を能動的に制御ができる。STI の能動的制御が可能となることで、様々な音環境にも対応したスピーチプライバシー保護技術を提供できる。さらに、本研究は、客観評価尺度を基準にすることにより、スピーチプライバシー保護の技術の中に、新風を吹かすことができる。

1.5 本論文の構成

本論文は，5章で構成される．

第1章 本研究の背景とスピーチプライバシー保護の関連研究，その問題点，本研究の目的を述べる．

第2章 スピーチプライバシーを評価するための尺度と提案法の着目点，その背景について述べる．

第3章 変調伝達関数（MTF）の概念やSTIの算出方法，様々なRIRモデルのMTFとSTI，本研究における提案法の原理を述べる．

第4章 提案法に対して，主観評価尺度において評価を行う．また，他の手法と比較を行う．

第5章 本研究で明らかになったことと残された課題についてを述べる．

第2章 スピーチプライバシーを評価するための尺度と提案法の着目点

2.1 主観評価尺度

主観評価尺度とは、人間の被験者による試験のことである。電話系や音響再生系などの人間を対象とするシステムは、最終的に人間の主観評価により決定する必要があるため、主観評価尺度が用いられる。主観評価尺度は、人間の被験者の感覚によって定められたもののため、被験者個人の特徴によって結果が大きく影響されることがある。また、被験者を収集しなければならないため、経費や稼働、時間が膨大にかかる。

明瞭度

単音や無意味な音節を単位として構成された試験刺激を用いて行う主観評価である [19, 20]。刺激を聴いたときに正しく聴き取れた割合を百分率で表現したものである。明瞭度は、試験信号と被験者の特性に依存して変動するため、評価を正しくするには、十分に訓練を必要とする。スピーチプライバシーを評価にあたり、無意味な音節をどれだけ正しく聴き取れたか、明瞭度で試験をすることもある。

了解度

了解度には、単語了解度と文章了解度がある [20]。単語了解度は、有意味な単語を音声刺激としたとき、正しく聴き取れた単語の割合を百分率で表現されている。単語リストには、特に定まったものはないが、単語の親密度、音節数などは揃えなければならない。文章了解度は、単文刺激を用意して、正しく聴き取れた単語の割合を百分率で表しており、単語了解度と高い相関を持つ。また、最近の研究では、2者択一で単語了解度試験を行うこともされている [9, 22]。

うるささ

「うるささ」は、音そのものに付随した不快感や音色の悪さなどを示す指標である [21]。一般的に、騒音レベルが高くなると、音の大きさも「うるささ」とともに増加するが、継続時間が長くなると音の大きさは変化しないが「うるささ」は増加する。また、音色が異

なるときに音の大きさは同じでも「うるささ」は変化するため、音の大きさと「うるささ」は別の指標としてあげられる。

わずらわしさ

「わずらわしさ」は、音の大きさ、「うるささ」とともに騒音の心理的な特性の一つである [21]. 「うるささ」が音そのもののもつ不快感を表した指標にたいして、「わずらわしさ」は、それ以外の要因により評価される指標である。例えば、オフィスで執務中に話し声が聞こえてきて、仕事が妨害されるなどがこれに当たる。

聞き取りにくさ

聞き取りにくさは、森本ら [23, 24, 25, 26, 27] によって提案された音声の「聞き取りにくさ」を評価する指標である。明瞭度や了解度は、音声刺激が変化すると正答率が変わり、「再現性」がないことが問題点としてあった。特に、了解度は、2度3度同じ音声を重複して聴いたとき、スコアが上がってしまう傾向にある。そこで、「音声伝送性能の評価には「再現性」が必要とされるため、森本らは「聞き取りにくさ」を提案した。聞き取りにくさは、「聞き取りにくくはない」、「やや聞き取りにくい」、「かなり聞き取りにくい」、「非常に聞き取りにくい」の4段階の評価尺度で、被験者に矯正選択をさせ評価を行う。「聞き取りにくさ」の評価によって、目的の音声を聴こうとしたときに評価できる了解度の他に、受動的に音声が届いてくるときに聞き取りにくい評価できるようになった。また、この主観評価尺度は、客観評価尺度である STI と高い相関を持つことが報告されている [23, 25, ?].

2.2 客観評価尺度

客観評価尺度とは、測定された、物理測定値から、主観品質を推定する尺度である。人間の被験者を介して行う客観評価尺度に対して、物理測定さえ行えば、経費や稼働、時間がかからない点において、客観評価尺度は優れている。

明瞭度指数 (AI : Articulation Index)

明瞭度指数 A は、明瞭度と関係があり、French と Steinberg によって提案された客観評価尺度である [28]. 明瞭度指数は次の式で求められる。

$$A = \int_0^{\infty} W(f)D(f)df \quad (2.1)$$

ここで、 D は音声の周波数あたりの明瞭度貢献度である。 D は、周波数特性を変化させた系の明瞭度試験から算出される。 W は、音声信号のひずみや雑音などの妨害要因から求まる。

表 2.1: 音声伝送性能と STI の関係.

Quality	Bad	Poor	Fair	Good	Excellent
STI	0.00 ~0.29	0.30 ~0.44	0.45 ~0.59	0.60 ~0.74	0.75 ~1.00

プライバシー指数 (PI : Privacy Index)

プライバシー指数 PI は、次式で求められる [1].

$$PI = 100 \times (1 - A) \quad (\%) \quad (2.2)$$

ここでは、 A は式 (2.1) の A である。この資料は、どれだけ明瞭ではなかったかを示す指標であり、スピーチプライバシー保護を評価するための目安となっている。

音声伝送指標

音声伝送指標 (STI) [15] は、室における音声伝送品質を評価する客観評価尺度であり、主にスピーチホールや教室における音声明瞭度の予測や室の設計・評価に利用される。STI は、Houtgast & Steeneken によって提案され、変調伝達関数 (MTF) [30, 31] の概念に基づき、室内インパルス応答 (RIR) およびその MTF より算出される。STI は、表 2.1 に表すように 0 ~ 1 の範囲で表現され、Bad から Excellent までの 5 段階評価でランク分けられる。STI が 0 に近ければ音声伝送品質が悪く、1 に近ければ音声伝送品質が良いとなるような指標となっている。STI は主観評価尺度である聴き取りにくさ [24] と高い相関を持つことも知られている。

2.3 本研究の着目点

本研究の着目点は、STI を基準として漏えい音声の聴き取りにくさに着目することで、スピーチプライバシー保護を実現することである。図 2.1 は、本研究における提案法の概念図になっており、簡便な線形システムで表現できる。上図中の $x(t)$ は、音声コミュニケーションの音声信号となっており、 $h_0(t)$ は、会話の伝送路の RIR で、 $y(t)$ は、第三者に聞こえる漏えい音声となっている。* は畳み込み演算子である。この状況だと、音声信号 $x(t)$ は、その室の特性 $h_0(t)$ が畳み込まれることによって、 $y(t)$ として第三者に漏えいし聴き取れてしまう。提案法は、その漏えい音声に着目することによって、宮崎らや、鷗木らによって提案された STI 推定法 [16, 17] を用いて、常に $h_0(t)$ を推定し、適切な残響の付与によって、スピーチプライバシー保護を行う。下図中の $h(t)$ は、本来のその室の特性 $h_0(t)$ と $h_0(t)$ と後部残響である $h_1(t)$ を畳み込んだものである。図中のスピーカーから残響音声を出力することによって、室の RIR を擬似的ではあるが、変化させ、その RIR

を畳み込んだ音声を第三者に聴かせることによって、第三者に聴こえる音声の STI は減少されることによって、スピーチプライバシー保護が行われる。スピーカーから音を出力することによって、安全に音声コミュニケーションが行うことのできる安全エリアを作り出すことを可能とする。後部残響に着目した理由は、音声明瞭度に大きく影響を与えるとされる室内音響指標の一つであるドイットリヒカイト (D_{50}) にある [18]。 D_{50} は、RIR の全体と初期反射のパワー比で表され、RIR の初期反射である直接音から 50 ms の残響成分には明瞭度を上げる効果があることを示している。すなわち、スピーチプライバシー保護を行う上で、RIR の初期反射は、明瞭度や了解度を増加させてしまうため、不要であると考えられる。本研究では、まず、どの概形の後部残響が、STI に最も影響があるか検討するため、 $h_0(t) = \delta(t)$ 、つまり、無響室内を仮定し、提案法を確立させる。次に、その方法を用いて聴取実験による原理の確認を行い、提案法とスピーチプライバシー保護の技術である残響音声を用いた方法とピンク雑音とを性能比較を行う。

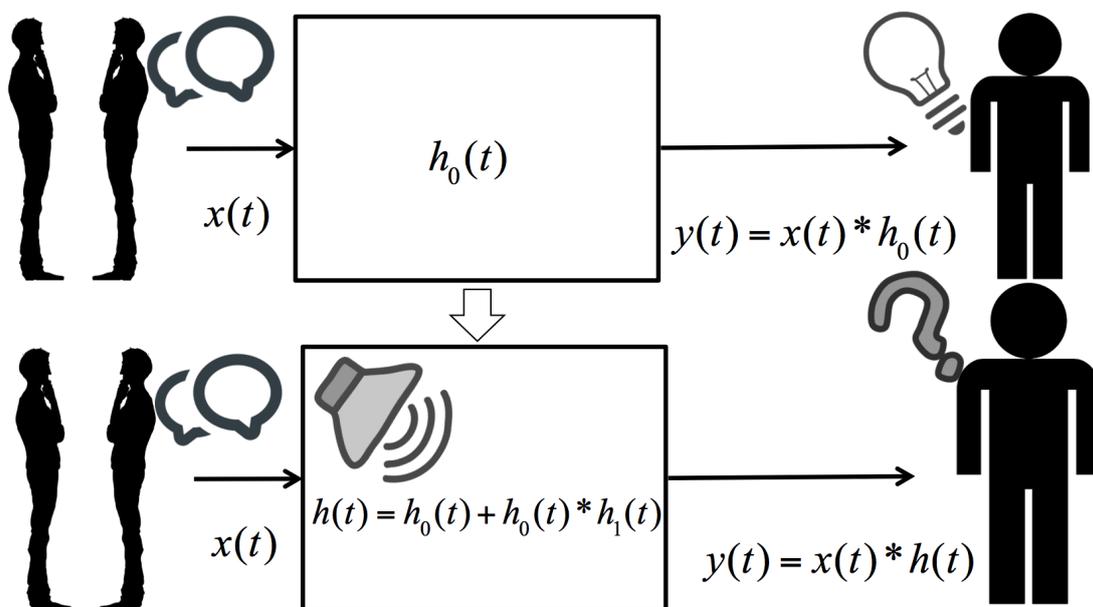


図 2.1: 本研究の概要図

第3章 本研究における提案法の原理

3.1 変調伝達関数の概念

図 3.1 は、変調伝達関数 (MTF) [30, 31] の概念に基づいた残響環境の室内音響における信号モデルである。入力信号, 出力信号, RIR をそれぞれ, $x(t)$, $y(t)$, $h(t)$ とする。出力信号は, 入力信号と RIR の畳み込みで表現され, $y(t) = x(t) * h(t)$ となる。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, 出力信号である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 RIR のそれぞれのパワーエンベロープの畳み込みとして表現できる。この関係から, 入力信号と出力信号のパワーエンベロープを再度表現すると $e_y^2(t) = e_x^2(t) * e_h^2(t)$, $e_x^2(t) = e_x^2(1 + \cos(2\pi f_m t))$ となる。ここで, t は時間, f_m は変調周波数, $*$ は畳み込みを表す。

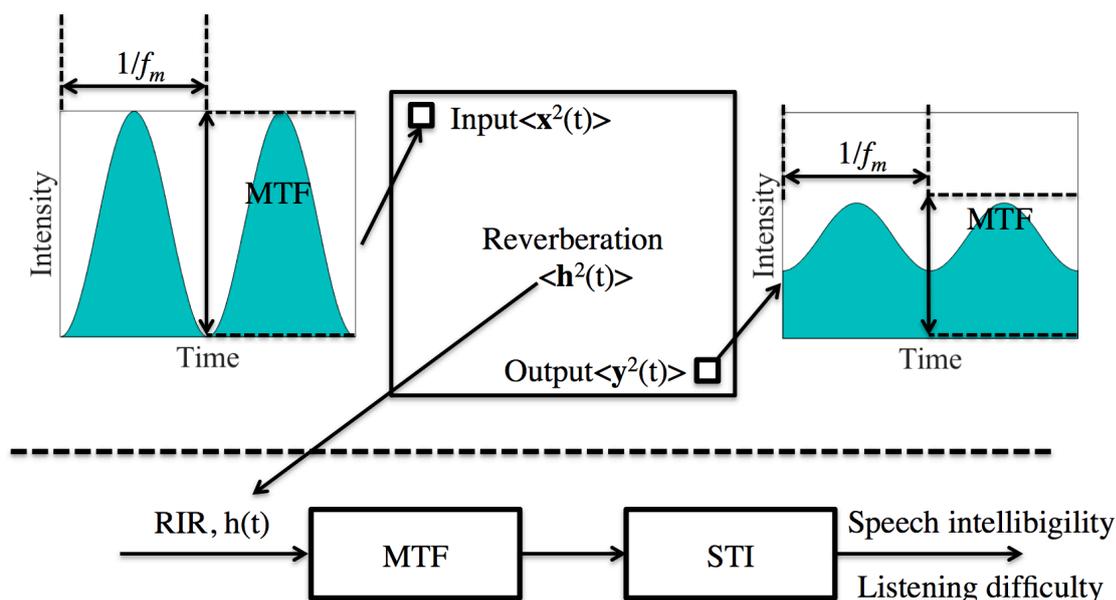


図 3.1: 変調伝達関数の概念図

表 3.1: 14 個の変調周波数の代表値 (Hz)

F_1	F_2	F_3	F_4	F_5	F_6	F_7
0.63	0.80	1.00	1.25	1.63	2.0	2.50
F_8	F_9	F_{10}	F_{11}	F_{12}	F_{13}	F_{14}
3.15	4.00	5.00	6.30	8.00	10.0	12.5

3.1.1 変調伝達関数の算出

図 3.1 に示す MTF $m(f_m)$ は、次式のように定義される。

$$m(f_m) = \frac{|\int_0^\infty h^2(t) \exp(-j2\pi f_m t) dt|}{\int_0^\infty h^2(t) dt} \quad (3.1)$$

式 (1) は、 $h^2(t)$ に対する一種のフーリエ変換をそのパワーで正規化したものである。また、 $h(t)$ のパワーエンベロープ $e_h^2(t)$ を利用して同様に MTF を導出できる。

3.1.2 STI の算出方法

STI は、次に示す 5 つの手順で求められる [31]。

7 つのオクターブ帯域での MTF の計算

まず、 $h(t)$ を $h_k(t)$ の 7 つのオクターブ帯域 (125 ~ 8000 Hz) に分割する。分割されたインパルス応答は $h_k(t)$ 、その MTF は $m_k(f_m)$ とし、 k は、オクターブ帯域の番号を指し、 $k = 1, 2, \dots, 7$ である。変調周波数 f_m は連続値を取るが、 $m_k(f_m)$ の重要な値を算出にあたり、 f_m を 14 個の代表値 F_i として求める。 F_i の詳細な値は表 3.1 に示した。なお、 $\mathbf{h}(t)$ はキャリアをランダム変数とした統計的なモデルで表現されるため、本論文での MTF は 7 つのオクターブ帯域で同様の形をもつことになる。

信号対雑音比 (SNR) の計算

次式のように $m_k(F_m)$ から SNR $N(k, i)$ を求める。

$$N(k, i) = 10 \log_{10} \frac{m_k(F_i)}{1 - m_k(F_i)} \quad (3.2)$$

伝送指標 (TI) の計算

次式により, SNR $N(k, i)$ を正規化することで, 伝送指標 $T(k, i)$ を求める.

$$T(k, i) = \begin{cases} 1, & (15 < N(k, i)) \\ \frac{N(k, i) + 15}{30}, & (-15 \leq N(k, i) \leq 15) \\ 0, & (N(k, i) < -15) \end{cases} \quad (3.3)$$

変調伝達指標の計算

次式に示すように, 各オクターブバンドの $T(k, i)$ の平均値より, 変調伝送指標 $M(k)$ を求める.

$$M(k) = \frac{1}{14} \sum_{i=1}^{14} T(k, i) \quad (3.4)$$

STI の計算

次式から STI を求める.

$$STI = \sum_{k=1}^7 W(k)M(k) \quad (3.5)$$

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.

3.2 雑音の MTF と STI

雑音の MTF は, 次の式で表せる.

$$m(f_m) = \frac{1}{1 + 10^{-\frac{SNR}{10}}} \quad (3.6)$$

ここで, SNR は信号体雑音比を表す. 図 3.2 から雑音の MTF は, その SNR にあわせて, フラットな特性を持つことがわかる. また, STI は, SNR が小さくなるほど減少し, 大きくなるほど増加する傾向が見られた.

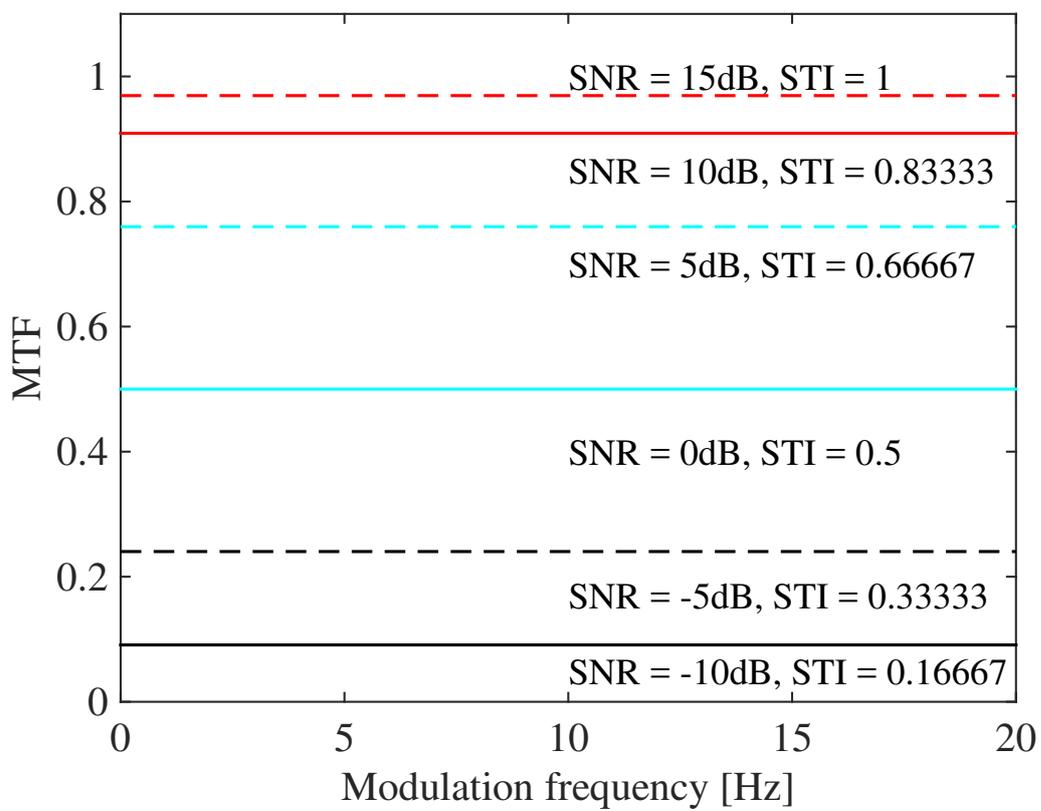


図 3.2: 雑音の MTF と STI

3.3 RIRモデルのMTFとSTI

本研究では、まず、STIを操作する仕組みを提案しなければならない。そこで、既存のRIRモデルおよび、直接音と後部残響で構成された後部残響モデルを定義し、そのMTFとSTIを算出して、各モデルのパラメータを変化させたときのMTFとSTIの振る舞いを確認する。

3.3.1 SchroederのRIRのMTFとSTI

残響におけるRIRでよく知られているものは、SchroederのRIRである[32]。SchroederのRIRは次式で定義される。

$$\mathbf{h}(t) = a \exp\left(\frac{-6.9t}{T_R}\right) \mathbf{c}_h(t) \quad (3.7)$$

SchroederのRIRモデルのパワーエンベロープの概形は、図3.3に示すようになる。ここで、 T_R は残響時間、 $\mathbf{c}_h(t)$ は白色雑音キャリアである。 T_R は、室内音響指標の一つである T_{60} と等しく、パワーエンベロープの減衰の傾きを制御する役割がある。SchroederのRIRに対応するMTFは、次式のようになる。

$$m(f_m) = \frac{1}{\sqrt{1 + (2\pi f_m \frac{T_R}{13.8})^2}} \quad (3.8)$$

MTFの概形は、図3.4のように低域通過特性を表し、 T_R の増加とともにその通過帯域が低域側に制限される。図3.5は、SchroederのRIRモデルのSTIの特性を示す。このモデルには、振幅項 a と残響時間 T_R のパラメータがある。振幅項 a はMTFの導出時に正規化されるため、 a は式3.8にはない。ここでは残響時間 T_R のパラメータのみの変化におけるSTIの変化を表示している。図3.5は、SchroederのRIRモデルのSTIの特性を示す。このモデルには、振幅項 a と残響時間 T_R のパラメータがある。図3.5に示すように、STIは T_R を増加すると減少する傾向が見られた。しかし、その値は0.2程度に収束し、それより小さくなることはなかった。

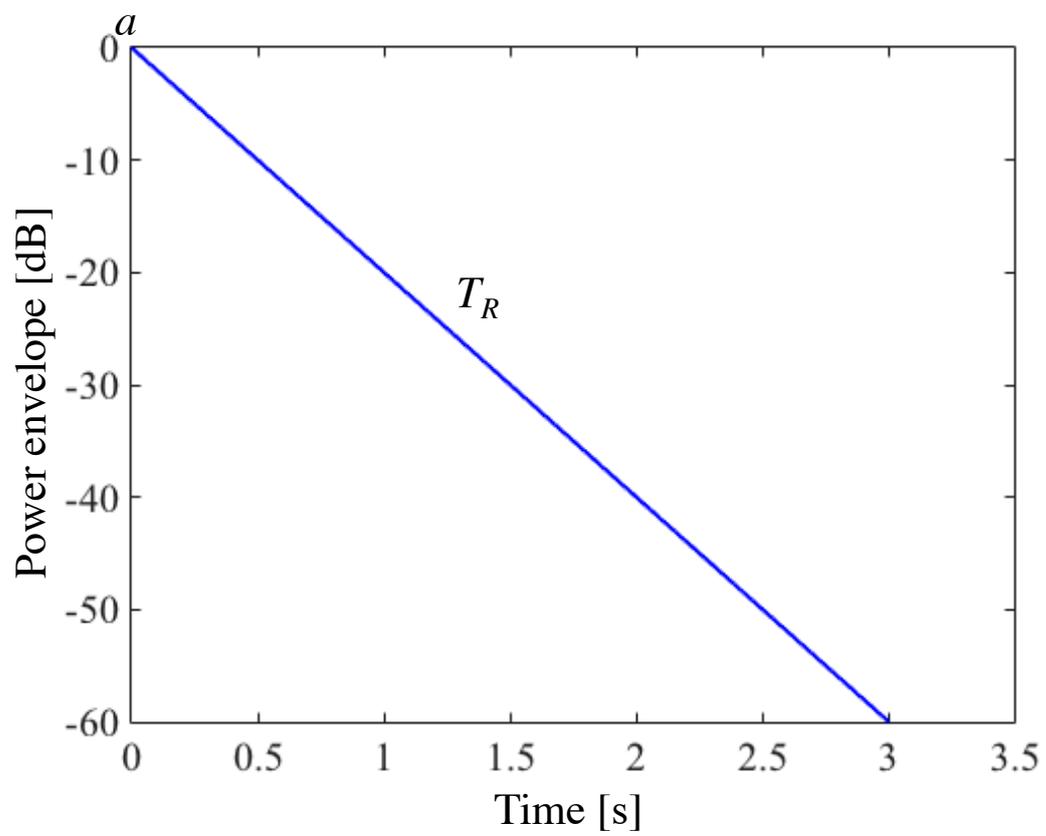


図 3.3: Schroeder の RIR のパワーエンベロープ

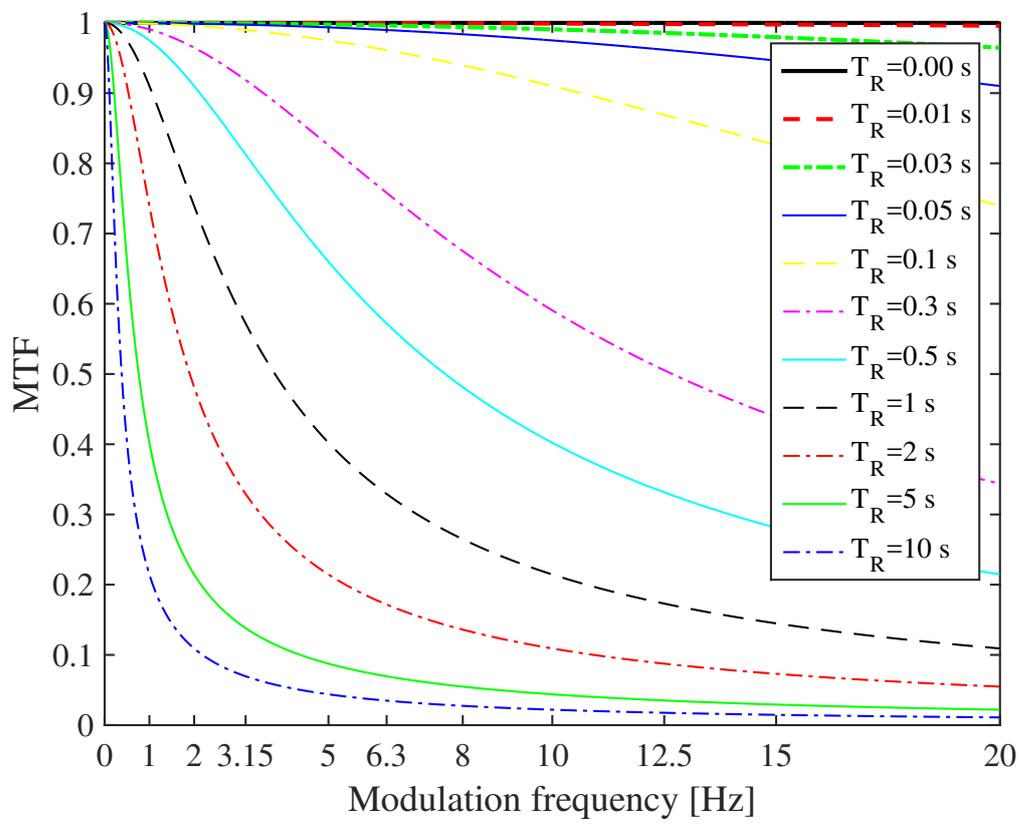


図 3.4: Schroeder の RIR の MTF

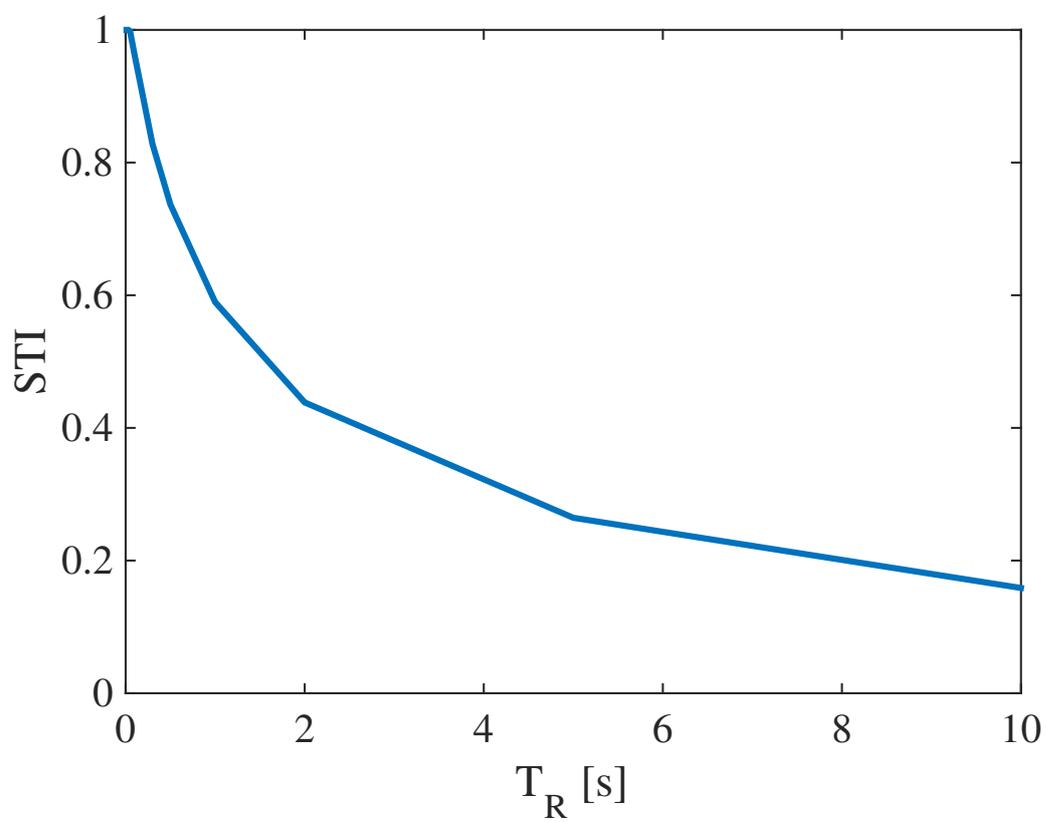


図 3.5: Schroeder の RIR の STI

3.3.2 一般化 RIR の MTF と STI

Schroeder の RIR より実測の RIR を近似したモデルとして一般化 RIR がある [33]. 一般化 RIR は, Schroeder の RIR を拡張したものであり, 次式で定義される.

$$\mathbf{h}(t) = at^{(b-1)} \exp\left(\frac{-6.9t}{T_R}\right) \mathbf{c}_h(t) \quad (3.9)$$

パワーエンベロープの概形は図 3.6 に表すようになる. このように, RIR をガンマ関数で表したモデルであり, パラメータを 2 つ持つ関数である. この式において, b は一般化 RIR の次数であり, $b \geq 1$ の値を取る. $b = 1$ のとき, 式 (3.9) は式 (3.7) と等しくなる. $\mathbf{h}(t)$ の MTF は, 式 (3.1) から得られ, 次のようになる.

$$m(f_m) = \frac{1}{(1 + (2\pi f_m \frac{T_R}{13.8})^2)^{\frac{2b-1}{2}}} \quad (3.10)$$

ただし, ここで, 振幅項 a は Schroeder の RIR のときと同様の理由で正規化される. MTF は, 図 3.7 のようになり, 図 3.4 と同じく低域通過特性を持つ. このときの b は 1.5 である. b を大きくすると Schroeder の RIR よりも MTF の通過帯域の傾きが急峻になる. これは, RIR モデルの形状が MTF に影響を与えることが原因である. 図 3.8 は, 一般化 RIR モデルの STI の特性を示す. このモデルには, 残響時間 T_R と次数 b のパラメータがあるため, これらを変化させたときの STI の変化を表示している. 図 3.8 に示すように, STI は T_R と b を大きくすると減少し, Schroeder の RIR とは違って, その値は 0 に収束した. T_R のみを大きくしたときは図 3.5 のような特性がみられ, b のみを大きくしたときは STI が劇的に減少する. MTF と同様に, RIR の形状は, STI に影響を与えることがわかる.

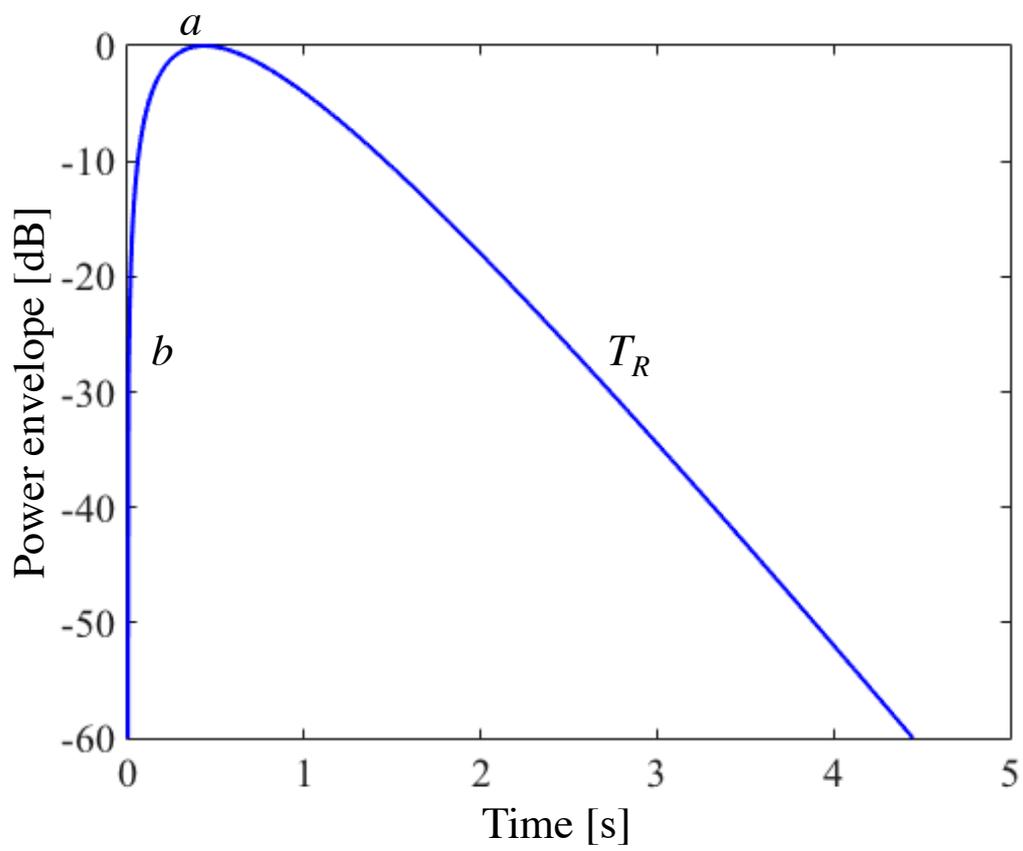


図 3.6: 一般化 RIR のパワーエンベロープ

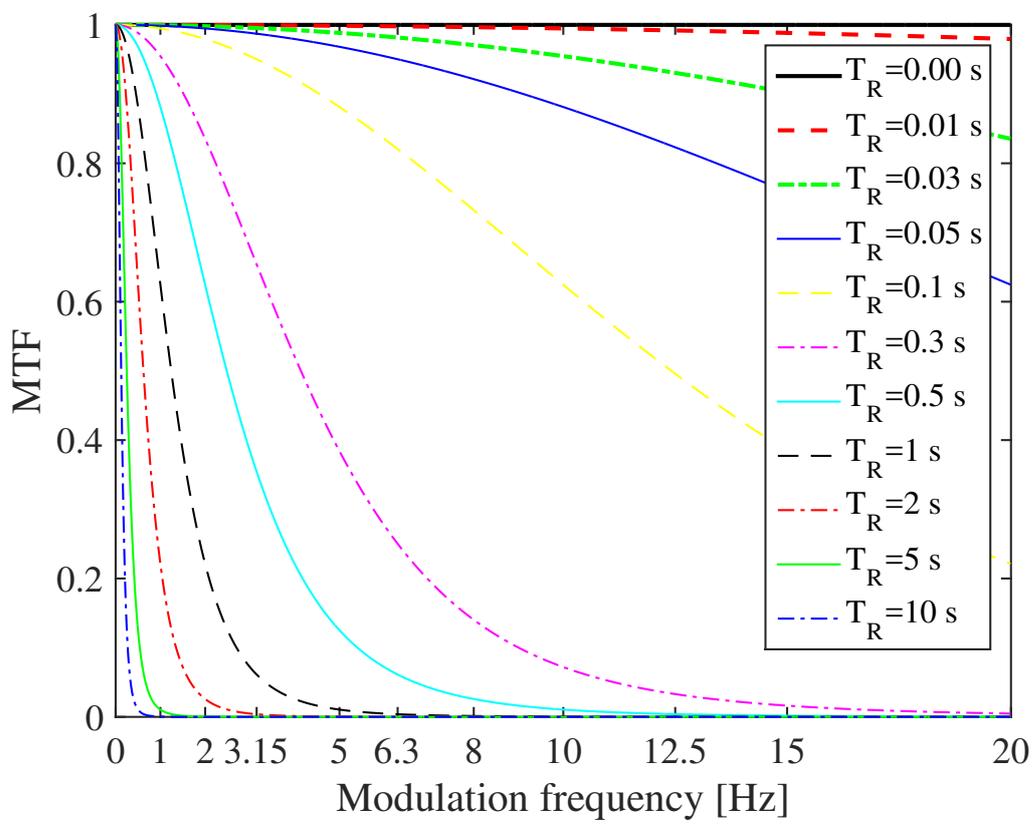


図 3.7: 一般化 RIR の MTF

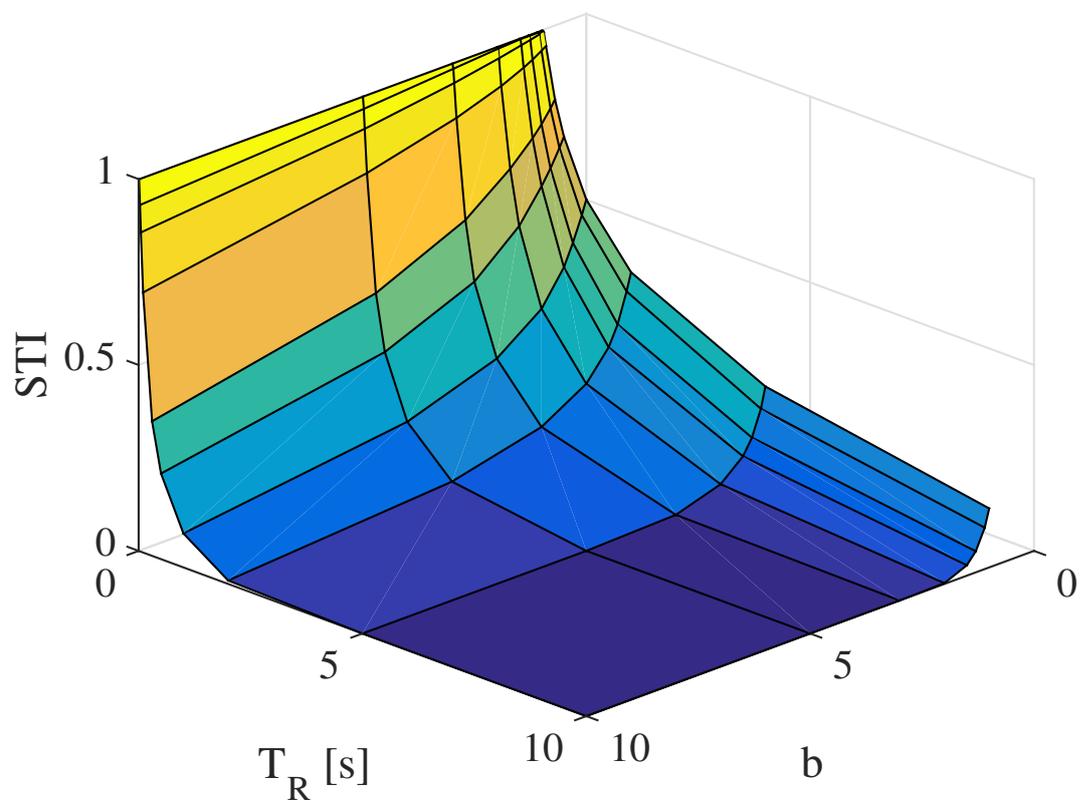


図 3.8: 一般化 RIR の STI

3.3.3 拡張型 RIR の MTF と STI

一般化 RIR よりも実測の RIR に近似できるモデルとして拡張型 RIR がある。拡張型 RIR は次の式で定義される。

$$\mathbf{h}(t) = \mathbf{h}_0(t - t_0), \quad t_0 \geq 0 \quad (3.11)$$

$$\mathbf{h}_0(t) = \begin{cases} a \exp(6.9t/T_h) \mathbf{c}_h(t), & t < 0 \\ a \exp(-6.9t/T_t) \mathbf{c}_h(t), & t \geq 0 \end{cases} \quad (3.12)$$

パワーエンベロープの概形は図 3.9 のようになっている。この式で、 T_h は RIR の立ち上がりを制御するパラメータで、 T_t は立ち下がり制御するパラメータである。 t_0 は、RIR のピークの時間を示すパラメータである。また、 T_t は、室内音響指標の一つである T_{60} と等しいという特性を持つ。そのため、一般化 RIR よりも室の特性を近似できているモデルだといえる。さらに、拡張型 RIR は、立ち上がりと立ち下がり指数関数で表現していることにより、ガンマ関数により表現される一般化 RIR よりも制御が容易である。 $\mathbf{h}(t)$ の MTF は、式 (3.1) から得られ、次のようになる。

$$m(f_m, T_h, T_t) = \frac{1}{\sqrt{\left(1 + \left(2\pi f_m \frac{T_h}{13.8}\right)^2\right) \left(1 + \left(2\pi f_m \frac{T_t}{13.8}\right)^2\right)}} \quad (3.13)$$

ただし、振幅項 a は Schroeder の RIR のときと同様の理由で正規化されるため、ここでは省略する。MTF は、図 3.10 のようになり、低域通過特性を持つ。このときの T_h は 0.1 である。 T_h を大きくすると Schroeder の RIR よりも MTF の通過帯域の傾きが急峻になる。図 3.11 は、拡張型 RIR モデルの STI の特性を示す。このモデルには、 T_h と T_t のパラメータがあるため、これらを変化させたときの STI の変化を表示しており、図 3.11 に示すように、STI は T_h と T_t を大きくすると減少し、一般化 RIR と同じようにその値は 0 に収束した。 T_t のみを大きくしたときは図 3.5 のような特性がみられ、 T_h と T_t の相互作用によって STI が劇的に減少する傾向がみられた。

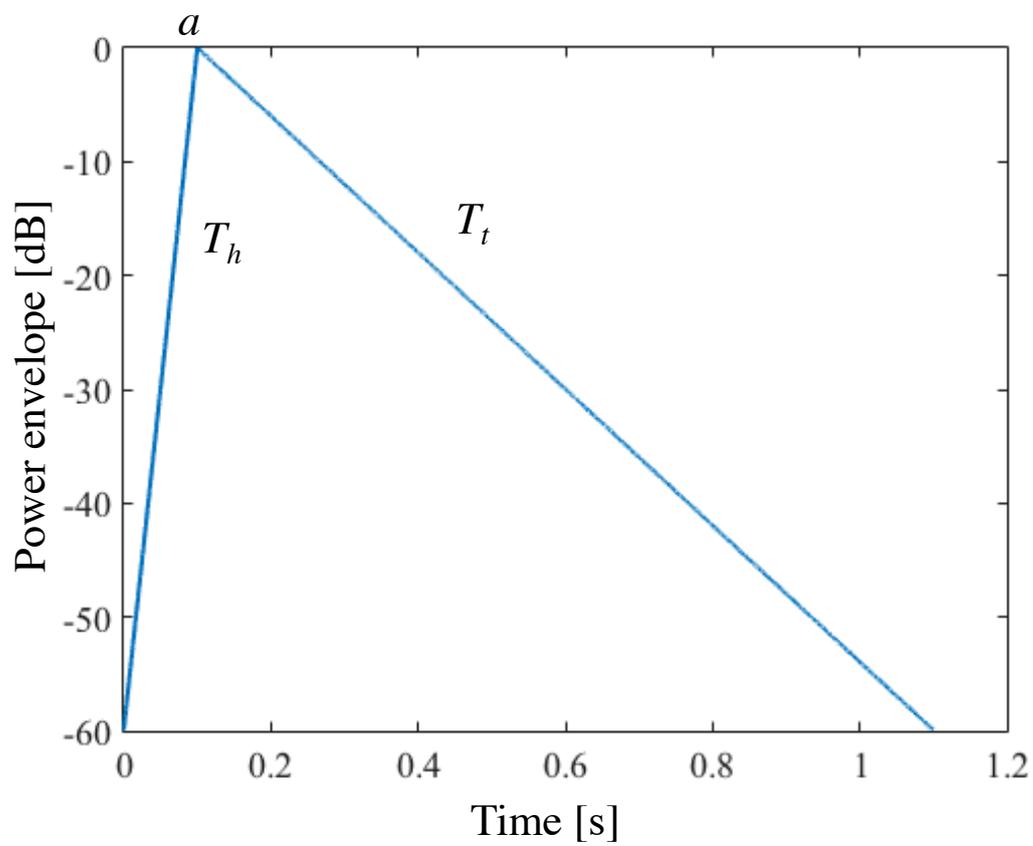


図 3.9: 拡張型 RIR のパワーエンベロープ

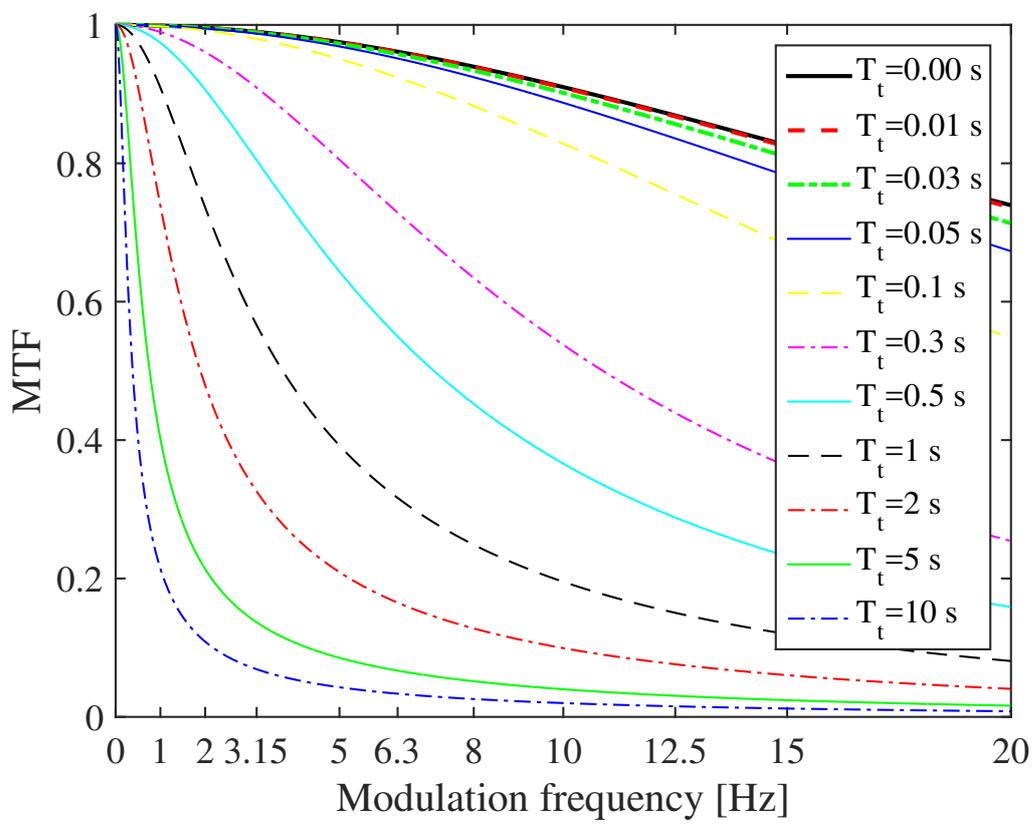


図 3.10: 拡張型 RIR の MTF

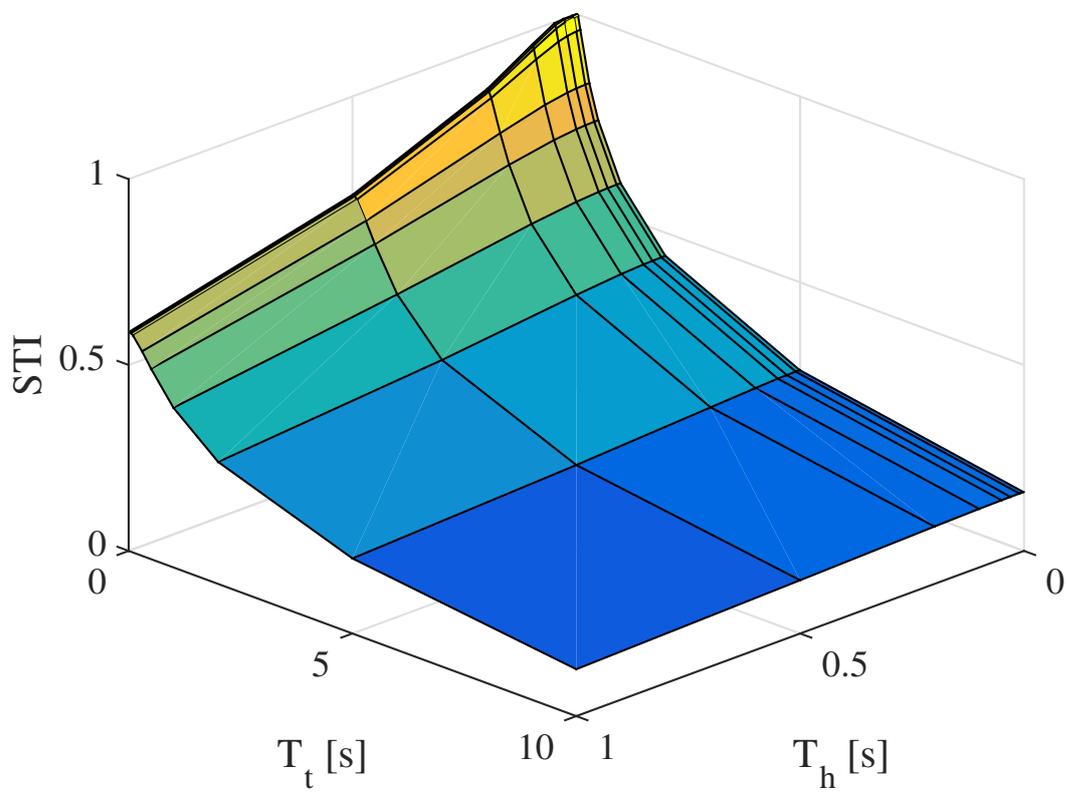


図 3.11: 拡張型 RIR の STI

3.3.4 単一エコーの MTF と STI

本研究では、前述の通り、後部残響の変化によって、STIの操作を行い、スピーチプライバシー保護を行うことが目的である。そこで、直接音と後部残響で構成したモデルを定義し、そのモデルのパラメータの変化にともなう MTF と STI の特性を知ることによって、後部残響を変化させることで目的とする STI となるような予測を立てる。直接音と後部残響で構成する最も簡易なモデルは、 δ 関数のみで構成される単一エコーだと考えられる。単一エコーの特性を知ることにより、直接音と後部残響の関係の一端が見えてくるはずである。単一エコーは、次式で定義される。

$$\mathbf{h}(t) = \delta(t) + a\delta(t - \tau)u(t - \tau) \quad (3.14)$$

単一エコーのパワーエンベロープは、図 3.12 のような概形を取る。ここで、 $\delta(t)$ は、直接音に対応するインパルス、 a は振幅項、 τ は遅延時間を示している。 $\mathbf{h}(t)$ の MTF は、式 (3.1) から得られ、次式のようになる。

$$m(f_m) = \frac{\sqrt{1 + 2a^2 \cos(2\pi f_m \tau) + a^4}}{1 + a^2} \quad (3.15)$$

MTF の概形は、図 3.13 に示されたように変調周波数領域で $1/\tau$ を基本周波数とする調波性をもつ。図 3.13 では、 $a = 0$ dB としたが、これを減少させると、MTF の概形は変化し、おおよそ 1 になる。 $\tau = 0.00$ s のとき、直接音とエコーが同時刻に発生するため、結果としてインパルスしか観測されないため、どの周波数においても 1 でフラットな特性をとる。図 3.14 は、単一エコーモデルの STI の特性を示す。このモデルには、振幅 a と遅延時間 τ のパラメータがあるため、これらを変化させたときの STI の変化を表示している。図 3.14 のように、STI は $\tau = 0.2$ s、 $a = 0$ dB のときに最も減少し、その値は 0.6 程度に収束することがわかった。

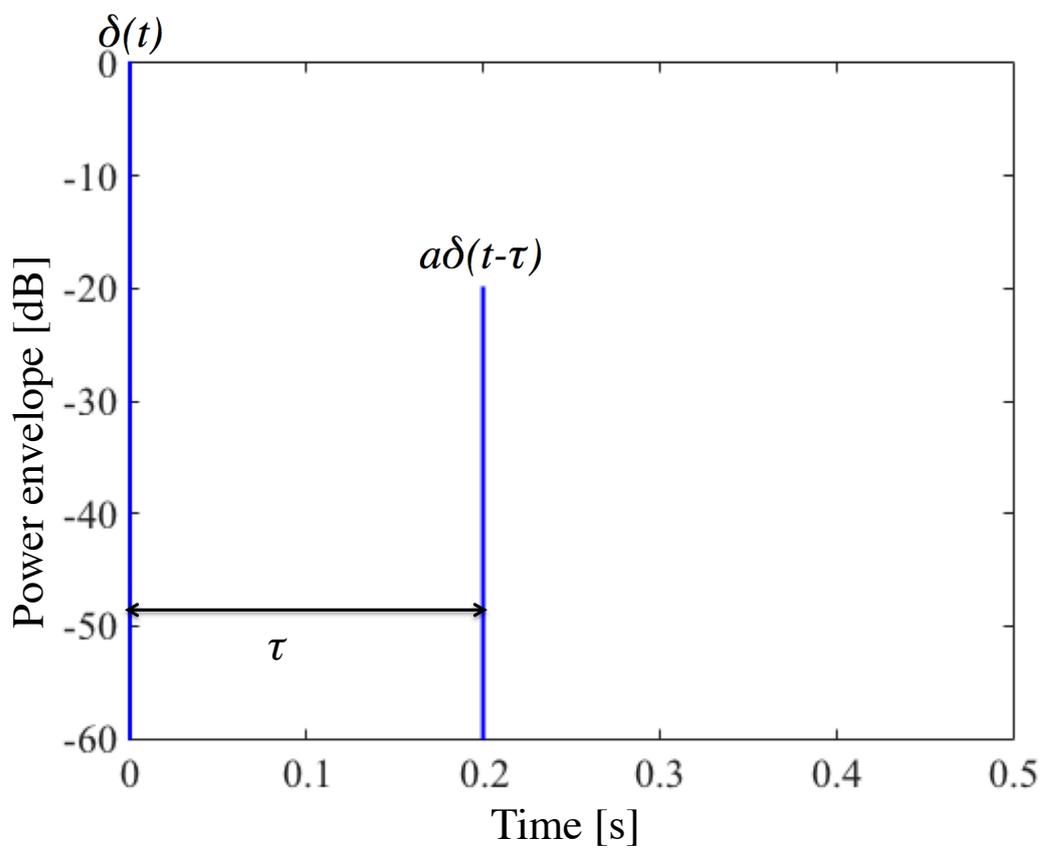


図 3.12: 単一エコーのパワーエンベロープ

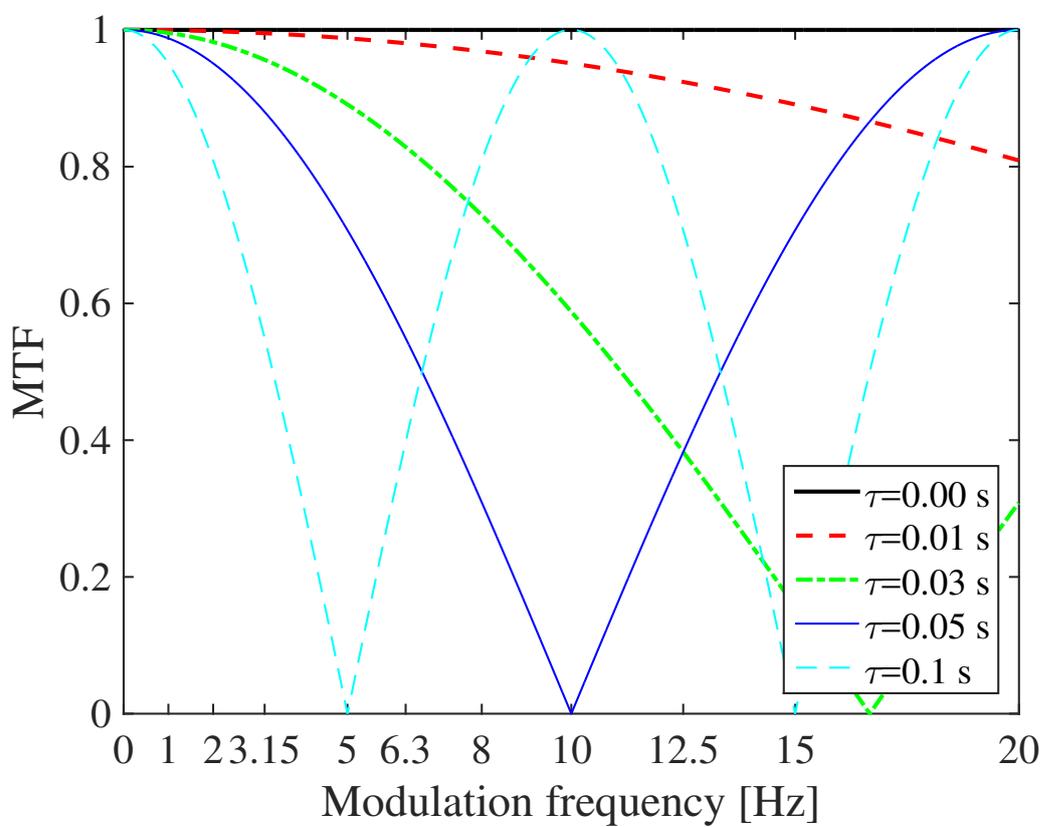


図 3.13: 単一エコーの MTF

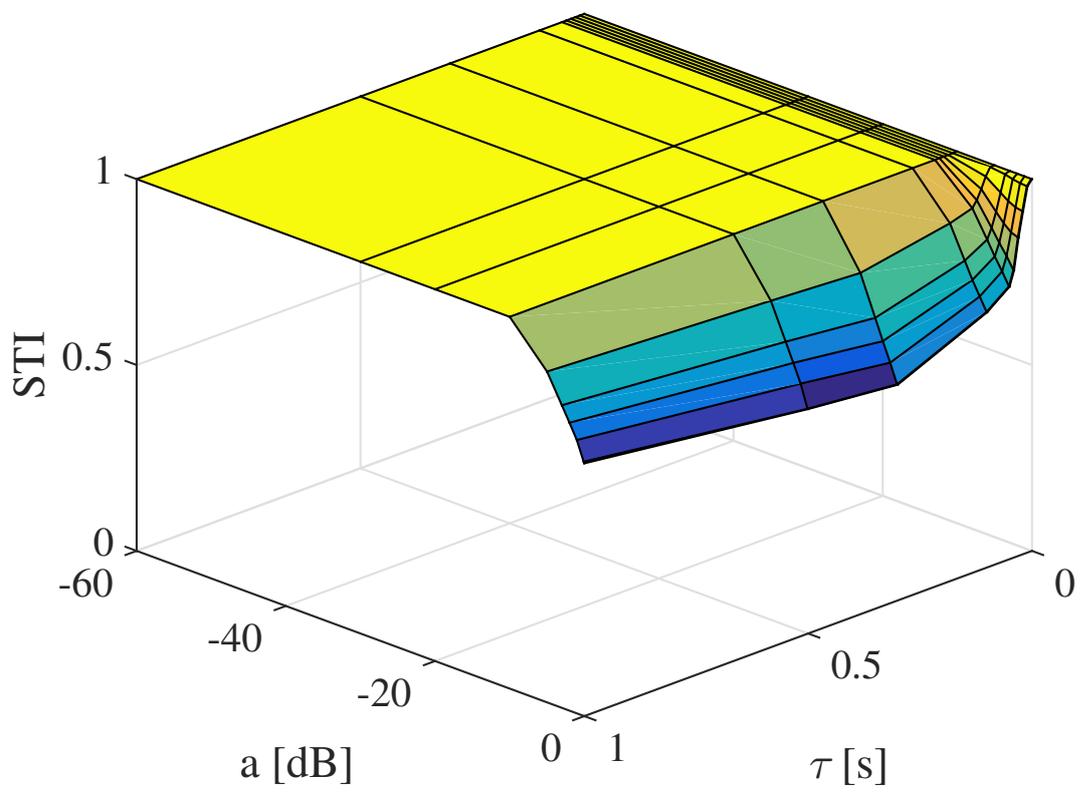


図 3.14: 単一エコーの STI

3.3.5 Schroeder の RIR を後部残響としたモデルの MTF と STI

先に上げた Schroeder の RIR を後部残響としたモデルを定義することで、直接音と残響の関係を紐解ける。Schroeder の RIR を後部残響としたモデルは、次式で定義される。

$$\mathbf{h}(t) = \delta(t) + a \exp\left(\frac{-6.9(t - \tau)}{T_R}\right) u(t - \tau) \mathbf{c}_h(t) \quad (3.16)$$

Schroeder の RIR を後部残響としたモデルのパワーエンベロープは、図 3.15 のような概形を取る。ここで、 $\delta(t)$ は、直接音に対応するインパルス、 a は振幅項、 τ は遅延時間であり、 T_R は残響時間を表すパラメータ、 $u(t - \tau)$ はステップ関数である。これまで、MTF は式 (3.1) から得られたが、 $\delta(t)$ はディラックのデルタ関数と呼ばれるもので、連続時間で定義され、 $t = 0$ s のとき、無限大になり、それ以外では 0 になる特性を持つ。式 (3.21) は直接音と後部残響の大きさを比べることができないため、式 (3.1) で MTF に変換できない。そこで、式 (3.21) をこれまで定義してきた連続時間ではなく、離散時間で再定義し、 z 変換することによって、IIR フィルタを作成し、そのフィルタの周波数特性を用いて MTF を求める。Schroeder の RIR を後部残響としたモデルは、次のように離散時間で再定義される。

$$\mathbf{h}[n] = \delta[n] + a \exp\left(\frac{-6.9 \cdot (n - \tau f_s)}{T_R \cdot f_s}\right) u[n/f_s - \tau] \mathbf{c}[n] \quad (3.17)$$

この $\delta[n]$ はクロネッカーのデルタ関数と呼ばれるものである。 z 変換は、以下の式で定義される。

$$\mathbf{H}(z) = \sum_{n=-\infty}^{\infty} \mathbf{h}^2[n] z^{-n} \quad (3.18)$$

式 (3.18) を次式な RIR モデルのパワーエンベロープの面積で除算する。

$$S = \int_0^T \mathbf{h}(t) dt \quad (3.19)$$

ここで、 T は RIR の信号長である。この正規化された IIR フィルタの周波数特性は、MTF として扱える。式 (3.17) を式 (3.18) で z 変換すると次式が得られる。

$$E_{\mathbf{h}}(z) = \frac{a^2 z^{-\tau f_s}}{1 - \exp\left(\frac{-13.8}{T_R \cdot f_s}\right) z^{-1}} + 1 \quad (3.20)$$

式 (3.20) を式 (3.19) で除算したものの周波数特性 (MTF) は、図 3.15 のようになる。 T_R を大きくすると Schroeder の RIR のように MTF が急峻になる。また、 T_R が小さいとき、 τ の影響を受けやすくなる。図 3.17 は、Schroeder の RIR を後部残響としたモデルの τ と T_R を変化させたときの STI である。 τ と T_R によって、STI を 0.4 程度まで減少させることができた。

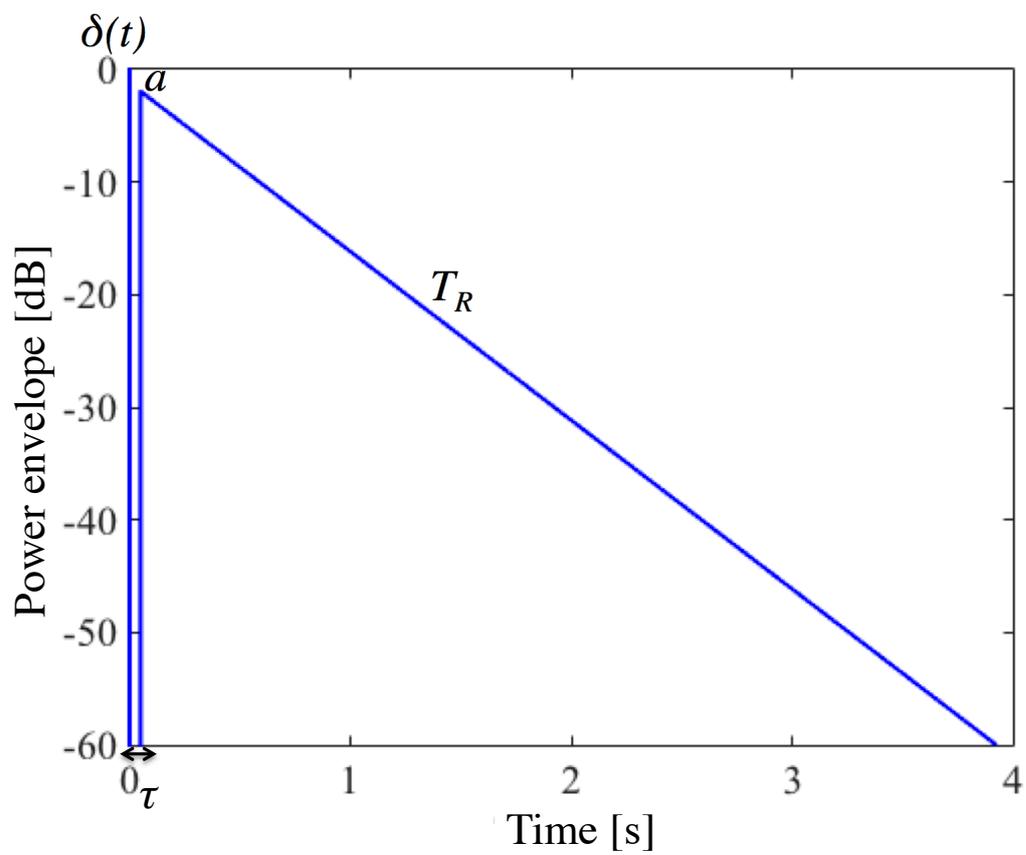


図 3.15: Schroeder の RIR を後部残響としたモデルのパワーエンベロープ

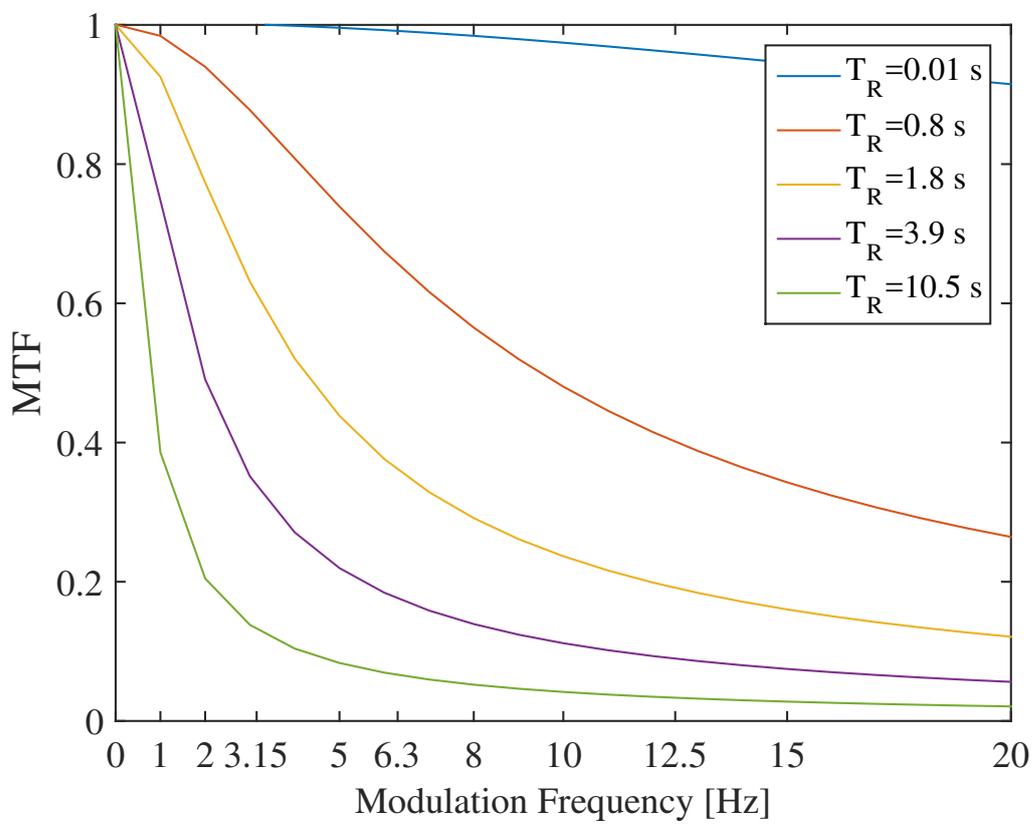


図 3.16: Schroeder の RIR を後部残響としたモデルの MTF

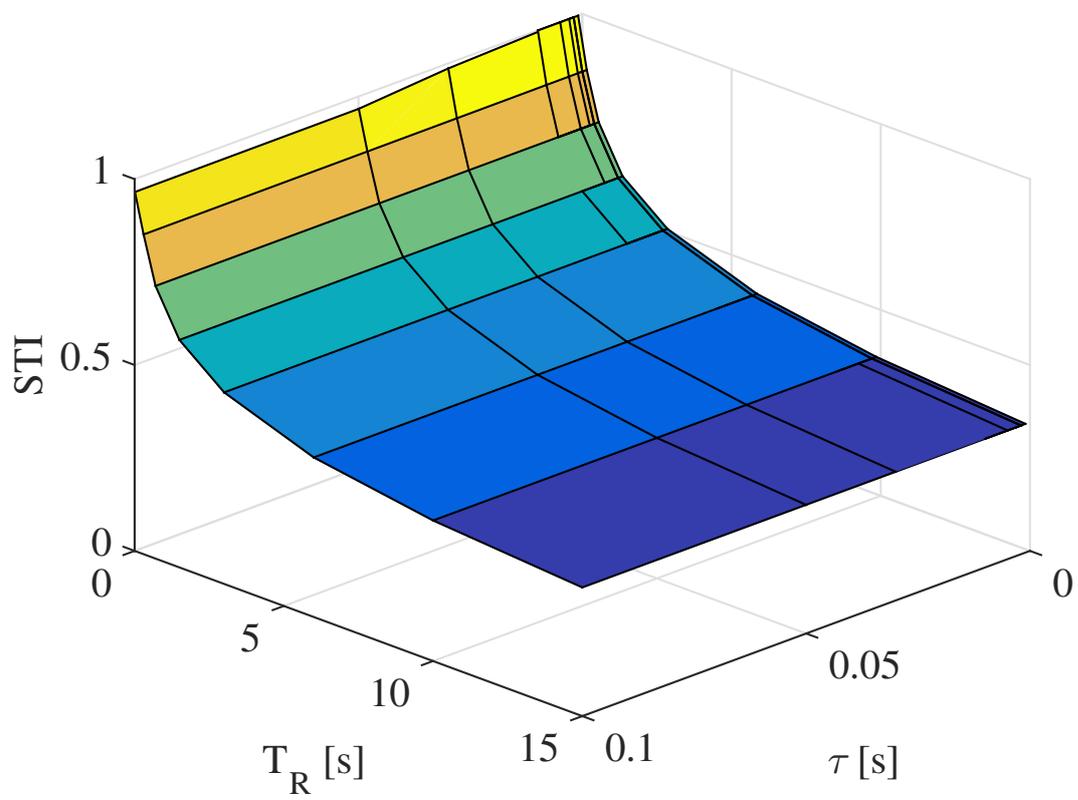


図 3.17: Schroeder の RIR を後部残響としたモデルの STI

3.3.6 後部残響を一般化 RIR としたモデルの MTF と STI

一般化 RIR を後部残響としたモデルは、次式で定義される。

$$\mathbf{h}(t) = \delta(t) + a(t - \tau)^{(b-1)} \exp\left(\frac{-6.9(t - \tau)}{T_R}\right) u(t - \tau) \mathbf{c}_h(t) \quad (3.21)$$

一般化 RIR を後部残響としたモデルのパワーエンベロープは、図 3.15 のような概形を取る。ここで、 $\delta(t)$ は、直接音に対応するインパルス、 a は振幅項、 τ は遅延時間であり、 T_R は残響時間を表すパラメータ、 b は一般化の次数で $b \geq 1$ の範囲で値を取る。また、 $u(t - \tau)$ はステップ関数である。Schroeder の RIR を後部残響としたモデルのときと同様に、一般化 RIR を後部残響としたモデルを次式のように離散時間で再定義する。

$$\mathbf{h}[n] = \delta[n] + a(n/f_s - \tau)^{(b-1)} \exp\left(\frac{-6.9 \cdot (n - \tau f_s)}{T_R \cdot f_s}\right) u[n/f_s - \tau] \mathbf{c}[n] \quad (3.22)$$

式 (3.22) を式 (3.18) で z 変換すると式 (3.23) や式 (3.24) が算出される。

$$E_{\mathbf{h}(1.5)}(z) = \frac{a^2 \exp\left(\frac{-13.8}{T_R \cdot f_s}\right) z^{-1} z^{-\tau f_s}}{f_s \left(1 - \exp\left(\frac{-13.8}{T_R \cdot f_s}\right) z^{-1}\right)^2} + 1 \quad (3.23)$$

$$E_{\mathbf{h}(2.0)}(z) = \frac{a^2 \left\{ \exp\left(\frac{-13.8}{T_R \cdot f_s}\right) z^{-1} + \exp\left(\frac{-27.6}{T_R \cdot f_s}\right) z^{-2} \right\} z^{-\tau f_s}}{f_s^2 \left(1 - \exp\left(\frac{-13.8}{T_R \cdot f_s}\right) z^{-1}\right)^3} + 1 \quad (3.24)$$

また、 $b = 1$ のときの $E_{\mathbf{h}(1.0)}(z)$ は、式 (3.20) となる。ここで、 b に一般性を見つけることはできず、 $b = 1.0, 1.5, 2.0, \dots$ のように 0.5 刻みでしか IIR フィルタを解くことができなかった。式 (3.23) を式 (3.19) で除算したものの周波数特性は、図 3.18 に示す。Schroeder の RIR を後部残響としたモデルよりものように MTF が急峻になる。また、 T_R が小さいとき、 τ の影響を受け、振動を励起させる。図 3.20 は、一般化 RIR を後部残響としたモデルの b と T_R を変化させたときの STI である。 b と T_R によって、STI が 0.1 程度まで減少させることができた。

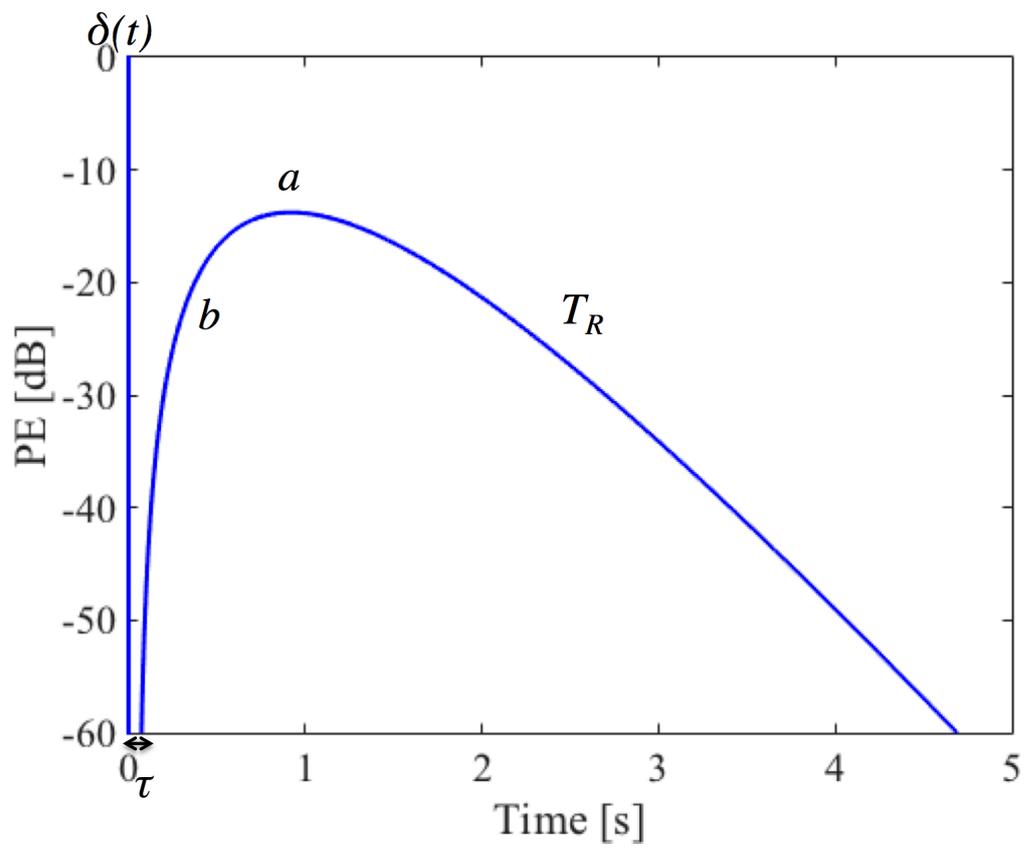


図 3.18: 一般化 RIR を後部残響としたモデルのパワーエンベロープ

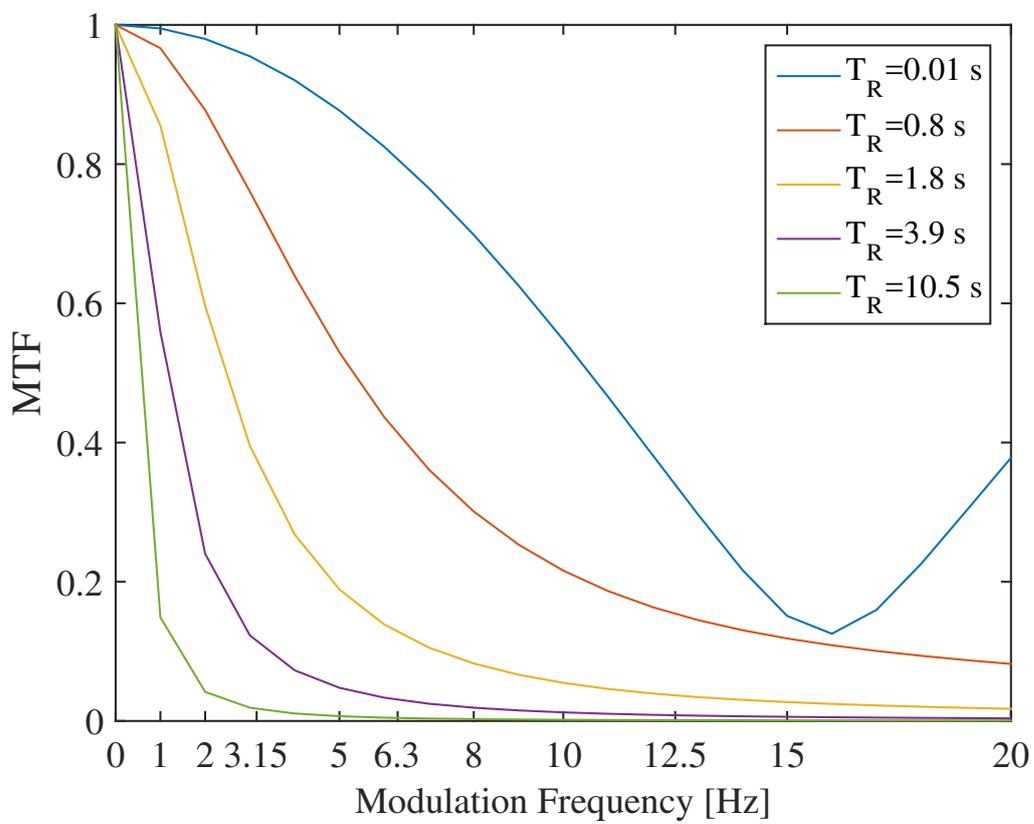


図 3.19: 一般化 RIR を後部残響としたモデルの MTF

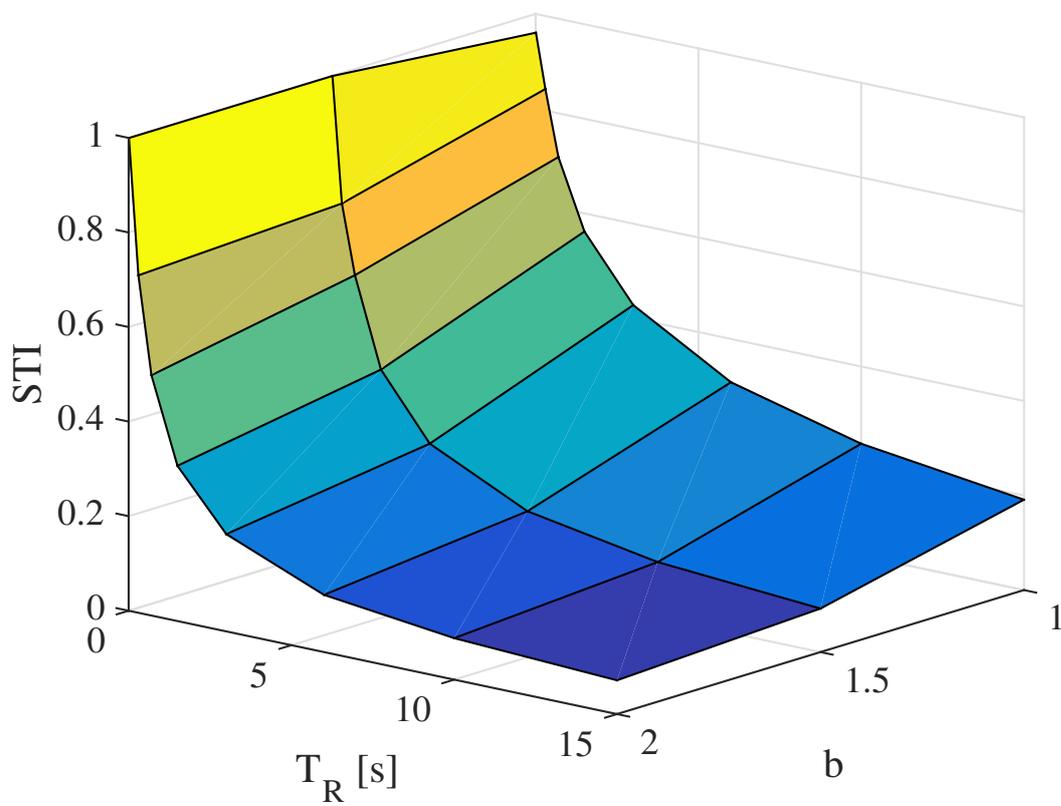


図 3.20: 一般化 RIR を後部残響としたモデルの STI

3.4 提案法

本研究の提案法は、図 3.21 のように漏えい音声をその残響音声により聞き取りにくくすることで、スピーチプライバシー保護を行うことをコンセプトとしている。ここで、 $x(t)$ は話者自身の声、 $h(t)$ は後部残響モデルである RIR、 $\delta(t)$ は直接音、 $h_L(t)$ は、後部残響である RIR モデル、 $y(t)$ は、漏えい音声である。提案法は、漏えい音声を $y(t) = x(t) * (\delta(t) + h_L(t))$ となることを目指す。これによって、自然で不快感のないスピーチプライバシー保護ができると考えられる。直接音と後部残響で構成されたモデルの中で最も STI を減少させることができたモデルは、一般化 RIR を後部残響としたモデルであった。しかし、2 点問題が残る。まず、前述のように $b = 1.0, 1.5, 2.0, \dots$ のみでしか MTF が計算できず、一般性も見いだせていない。そのため、目的であるあらゆる環境において最適な残響を付加するということができない。次に、ガンマ関数の特性上、振幅項 a の制御が非常に困難である。 b と T_R によって、 a が変動するため、自然な残響が付加されないおそれがあり、これでは、わずらわしくなることが予想される。そこで、立ち下がり制御するパラメータ T_h と立ち下がり制御するパラメータ T_t で構成された拡張型 RIR であれば、上記のような問題は解決されるため、本研究の提案法として、後部残響を拡張型 RIR としたモデルを採用する。拡張型 RIR を後部残響としたモデルは、次式で定義される。

$$\mathbf{h}(t) = \delta(t) + \mathbf{h}_L(t - \tau)u(t - \tau) \quad (3.25)$$

$$\mathbf{h}_L(t) = \mathbf{h}_0(t - t_0), \quad t_0 \geq 0 \quad (3.26)$$

$$\mathbf{h}_0(t) = \begin{cases} a \exp(6.9t/T_h) \mathbf{c}_h(t), & t < 0 \\ a \exp(-6.9t/T_t) \mathbf{c}_h(t), & t \geq 0 \end{cases} \quad (3.27)$$

拡張型 RIR を後部残響としたモデルのパワーエンベロープは、図 3.22 のような概形を取る。ここで、 $\delta(t)$ は、直接音に対応するインパルス、 a は振幅項、 τ は遅延時間、 T_h は RIR の立ち上りを制御するパラメータで、 T_t は立ち下がり制御するパラメータ、 t_0 は、RIR のピークの時間を示すパラメータ、 $u(t - \tau)$ はステップ関数である。遅延時間 τ は、室内音響指標の一つである D_{50} より、50 ms とし、自然な残響を付加することを想定するため、振幅項 a は距離減衰を表し、 $\exp(0.05\tau/T_t)$ とした。Schroeder の RIR を後部残響としたモデルのときと同様に、拡張型 RIR を後部残響としたモデルを次式のように離散時間で再定義する。

$$\mathbf{h}[n] = \delta[n] + \mathbf{h}_L[n]u[n/fs - \tau] \quad (3.28)$$

式 (3.28) を式 (3.18) で z 変換すると式 (3.29) が算出される。

$$E_h(z) = 1 + \frac{\left(\left(-\frac{13.8}{T_t f_s} \right) - \left(\frac{13.8}{T_t f_s} \right) \right) z^{-(t_0 + \tau) f_s}}{\left(1 - a^2 \exp\left(-\frac{13.8}{T_t f_s} \right) z^{-1} \right) \left(1 - \left(\frac{13.8}{T_t f_s} \right) z^{-1} \right)} \quad (3.29)$$

この式を RIR の全面積で除算したものの周波数特性が拡張型 RIR を後部残響としたモデルの MTF となる。MTF の概形は、図 3.23 に示す。このとき、 T_h は 0.1 s であった。 T_h ,

T_t がともに小さい時, MTF は振動を励起されることがわかった. また, T_t を大きくすると, MTF の傾きが急峻になることがわかった. 音声は, 変調スペクトルの 2 ~ 16 Hz にピークがあることが知られている [34, 35]. 図 3.23 のように低域通過特性を持つ MTF を音声の変調スペクトルに掛け合わせることで, 音声の情報がなくなるのではないかと考えられる. 図 3.24 は, 拡張型 RIR を後部残響としたモデルの T_h と T_t を変化させたときの STI である. このように, T_h と T_t を変化させることによって, STI は 0.2 程度まで現象が確認できた. これより, このモデルにより, STI を自在に操作することができる. このモデルは, 一般化 RIR を後部残響としたモデルのように, 振幅項 a が様々なパラメータによって変化することなく, a を自在に決定することができるため, 立ち上がりと立ち下がり制限されることがない. そのため, 実際の環境に近似した残響を付加することに向いているといえる.

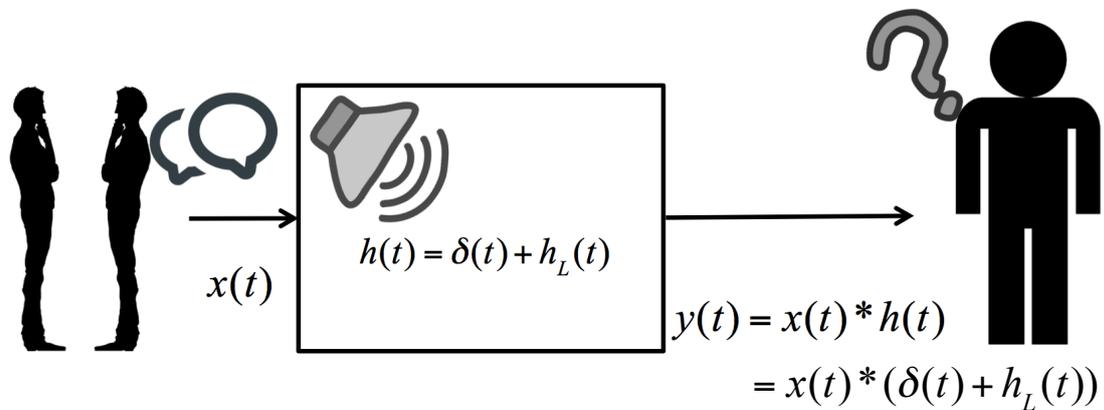


図 3.21: 提案法 の 概念 図

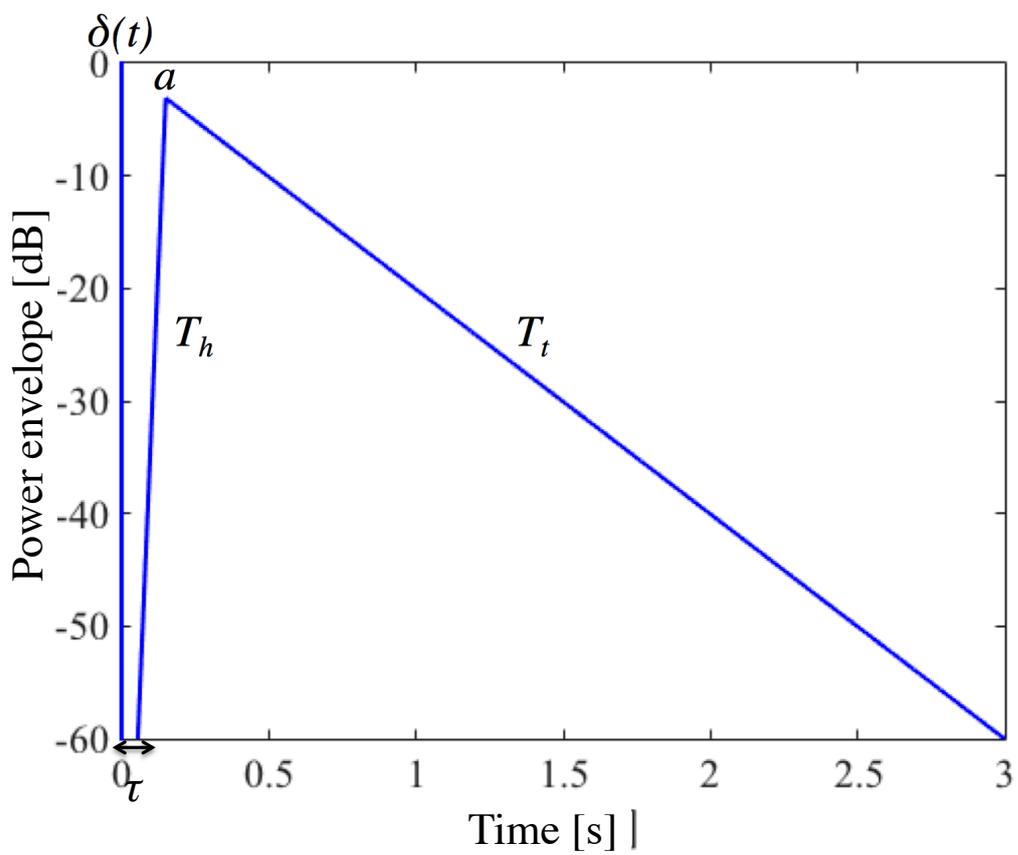


図 3.22: 拡張型 RIR を後部残響としたモデルのパワーエンベロープ

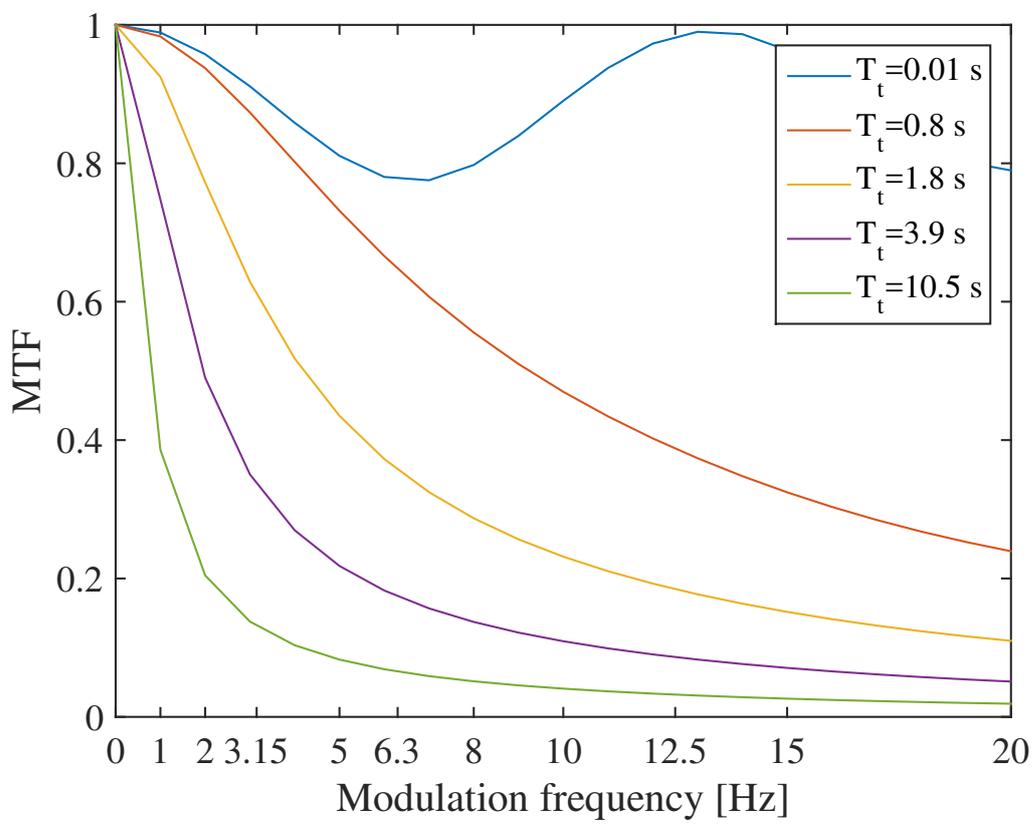


図 3.23: 拡張型 RIR を後部残響としたモデルの MTF

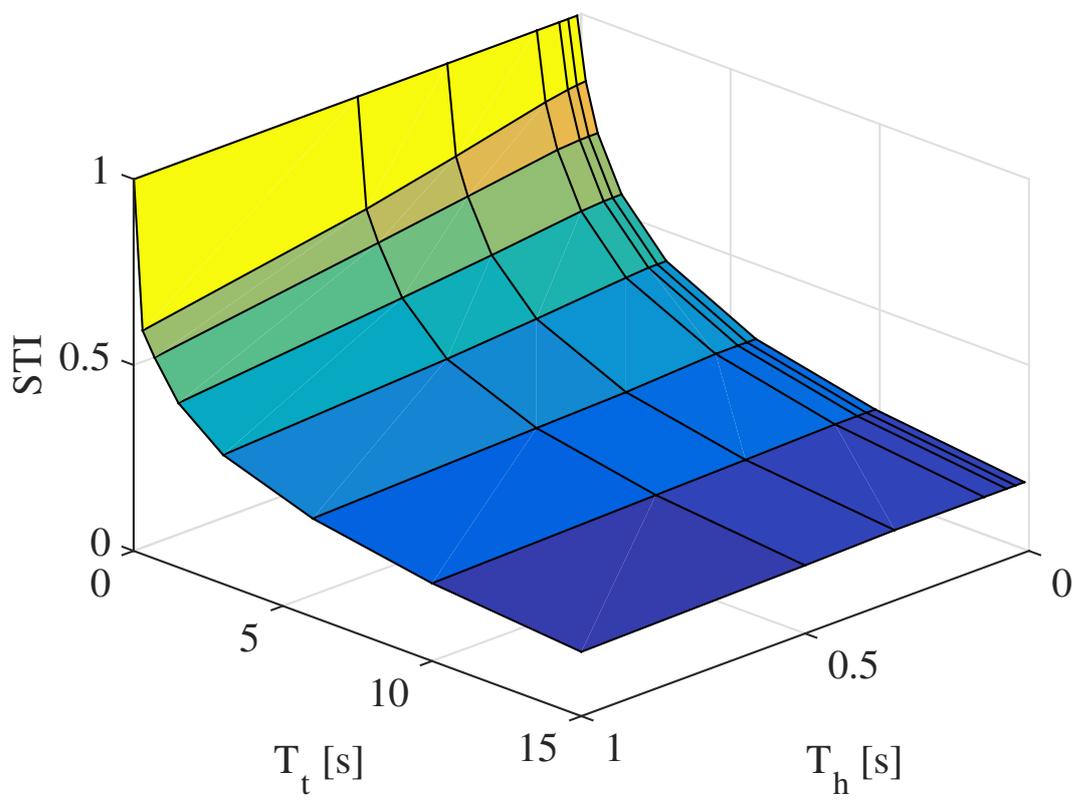


図 3.24: 拡張型 RIR を後部残響としたモデルの STI

第4章 提案法の評価

まず、提案法の原理を確認するため、提案法においてSTIを操作したとき、単語親密度と聞き取りにくさ、わずらわしさがどのような変化するか検討を行う。次に、提案法が、従来のスピーチプライバシー保護の技術と比べるため、提案法と残響音声とピンク雑音を用いて、性能比較を行う。従来法は、スピーチプライバシー保護を行うために、音環境に一切配慮せず一定の処理を行っていたことに対し、提案法は、室のSTIを基準にしているため、最低限のコストでスピーチプライバシー保護を行うことができる。

4.1 評価方法

RIRの後部残響を拡張型RIRとしたとき、そのパラメータを操作してSTIを制御することにより、スピーチプライバシー保護できるか検討するために、単語了解度試験と聞き取りにくさの聴取実験を行った。単語了解度は、聴取者が能動的に聴こうとすることを想定し、実験を行った。聞き取りにくさは、能動的に聴取者が効いてしまうことを想定したときに、音が聴こえてくるけれど聞き取りにくいという状況が、スピーチプライバシー保護には重要であるため実験を行った。また、STIを制御したときに不快感があるか検討するために、わずらわしさによる聴取実験を行った。

評価項目・条件

単語了解度試験では、MATLABのGUIを用い、聞き取った単語をカタカナでキーボード入力させ、その単語正答率を集計した。聞き取りにくさの聴取実験では、実験参加者に、聴取音声の印象を、表4.1のように「聞き取りにくくはない」、「やや聞き取りにくい」、「かなり聞き取りにくい」、「非常に聞き取りにくい」の4つの評価から1つ強制選択させた。その集計の結果、次の式のように、「聞き取りにくくはない」以外の選択数を刺激の総数で除算した値を集計した。

$$\text{LDR} = \frac{N - \text{Count}(I)}{N} \quad (4.1)$$

ここで、LDRは、音声がどの程度「聞き取りにくい」かを示している。また、 N は実験刺激数、 $\text{Count}(I)$ は「聞き取りにくくはない」の数を数える関数となっている。この式は、全体の刺激から、聞き取りにくかった音声はいくつあるかを計算する式である。聞き取りにくさの聴取実験では、実験参加者に、聴取音声の印象を、表4.2のように「わずら

わしくはない」,「ややわずらわしい」,「かなりわずらわしい」,「非常にわずらわしい」の4つの評価から1つ強制選択させた。その集計の結果, 次の式のように, 「わずらわしくはない」以外の選択数を刺激の総数で除算した値を集計した。

$$\text{ANR} = \frac{N - \text{Count}(i)}{N} \quad (4.2)$$

ここで, ANR は, 音声かどの程度「わずらわしい」かを示している。また, N は実験刺激数, $\text{Count}(i)$ は「わずらわしくはない」の数を数える関数となっている。この式は, 全体の刺激から, わずらわしかった音声はいくつあるかを計算する式である。

刺激

試験用の音声刺激として, 親密度別単語了解度試験用データベース (FW07) [36, 37, 38] の男性発話者 (mya) の単語親密度 1.0~7.0 の単語を用いた。STI が 0.875, 0.675, 0.525, 0.375, 0.230 になるように拡張型 RIR の後部残響モデルを作成し, 音声刺激に重畳し残響音声を作成した。このとき, 遅延時間である τ は, 室内音響指標の一つである D_{50} を基準にして, 50 ms とし, 振幅項である a は自然な残響を付加することを想定し, $\exp(-6.9\tau/T_t)$ とした。STI が 0.875, 0.675, 0.525, 0.375 とした理由は, 音声伝送性能の Excellent から Poor の中央値を取ったためである。また, STI を 0.23 にした理由は, スピーチプライバシー保護するためには, STI が 0.23 以下にならなければならないという報告されたものを基準としたためである [29]。刺激の総数は, STI 5 条件 \times 単語親密度 4 条件 \times 20 単語 \times 実験数 3 の 1200 単語であった。1200 単語中, 単語の単語親密度を上げるおそれがあるため, 刺激の重複呈示を許さなかった。

比較実験では, 刺激の総数は, 親密度 2 条件 \times 20 単語 \times 実験数 3 \times 3 手法の 360 単語であった。この 3 手法は, 残響音声とピンク雑音である。

実験機器

図 4.1 は, 実験機構を表す。PC (LG Sharkoon, with OS Windows 8.1) により音声刺激を出力した。出力した刺激は, オーディオインタフェース (FIREFACE UCX) を経由してアンプ (audio-technica AT-HA5000) で増幅された後, ヘッドフォン (Sennheiser HD280) から呈示された。ここで, ヘッドフォンから出力された信号は, 被験者の聴力に影響が出ないように全刺激中最も SNR が高いもので, A 特性音圧レベル 65 dB とした。この SNR は, 図 4.2 原信号の信号がある部分を信号 (S) とし, 後部残響を雑音 (N) としたときの信号体雑音比である。

参加者

提案法の原理の評価における単語了解度と聴き取りにくさの実験には, 正常聴力を有すると自己申告をした, 日本語を母語とする 23~25 歳の男性 6 名が参加した。わずらわし

さの実験には、正常聴力を有すると自己申告をした、日本語を母語とする 23~25 歳の男性 3 名が参加した。

比較実験には、正常聴力を有すると自己申告をした、日本語を母語とする 23~31 歳の男性 8 名が参加した。

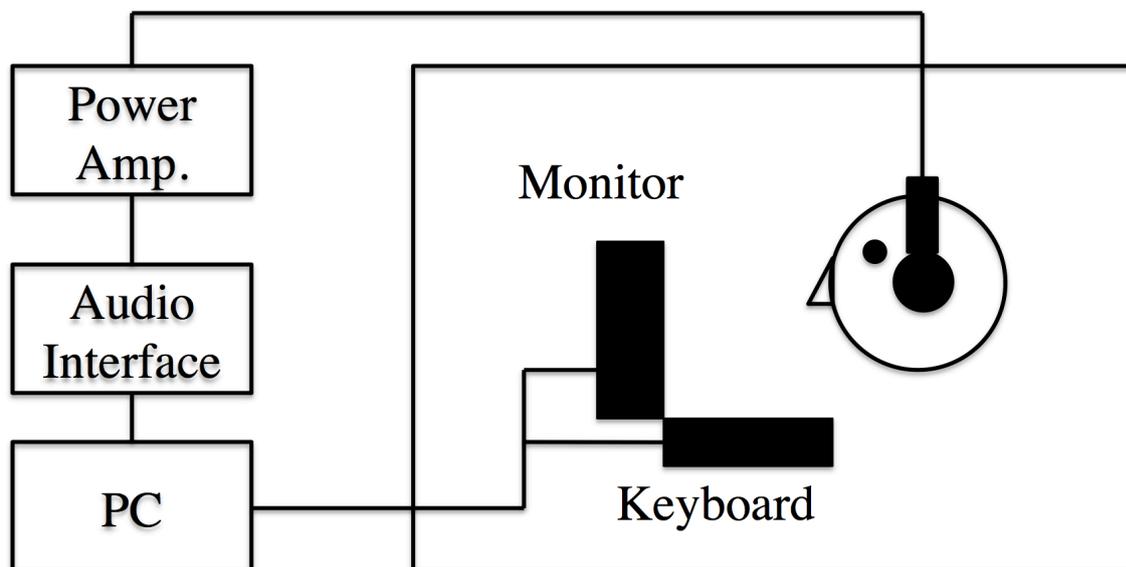


図 4.1: 実験機構の図

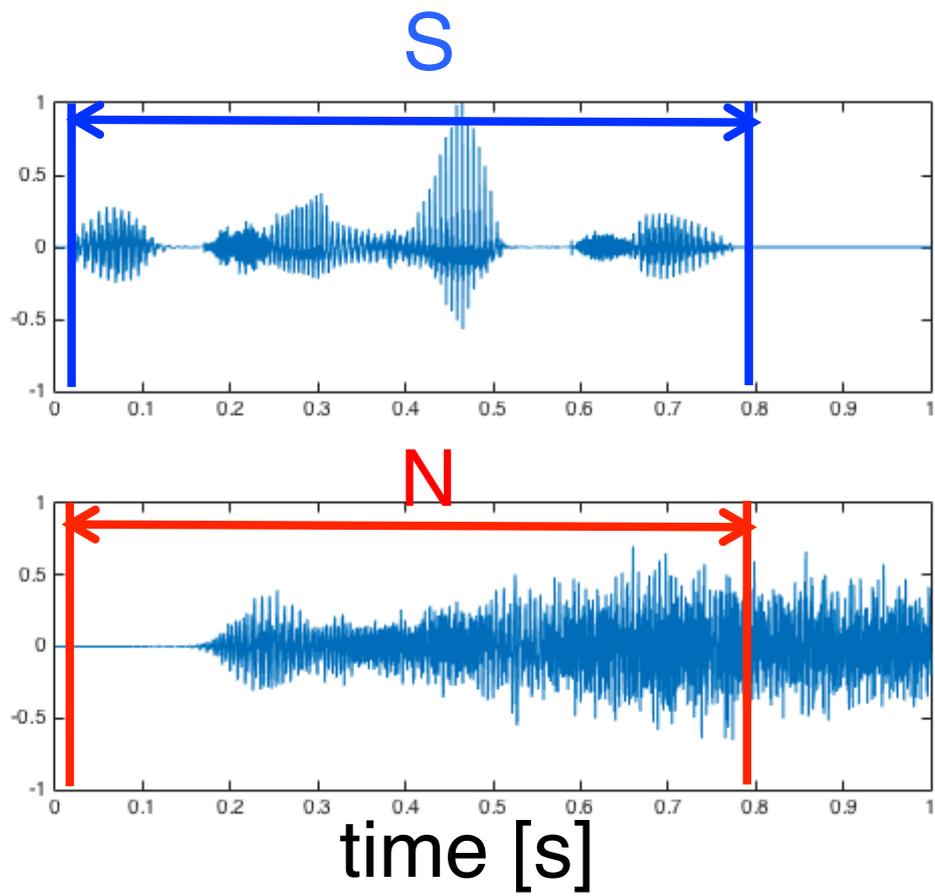


図 4.2: SNR の決め方

表 4.1: 聴き取りにくさの評価尺度

	評価尺度
(I)	聴き取りにくくはない
(II)	やや聴き取りにくい
(III)	かなり聴き取りにくい
(IV)	非常に聴き取りにくい

表 4.2: わずらわしさの評価尺度

	評価尺度
(i)	わずらわしくはない
(ii)	ややわずらわしい
(iii)	かなりわずらわしい
(iv)	非常にわずらわしい

4.2 提案法の原理の評価

図 4.3 に各 STI における親密度別単語了解度試験の結果を示す。図中の WF は単語親密度を示している。エラーバーは標準誤差である。おおむね、STI の低下によって単語了解度も低下した。また、STI が単語了解度に及ぼす影響は、単語親密度によって異なる傾向があった。単語了解度を従属変数、単語親密度と STI を要因とした 2 要因分散分析を行った。その結果、STI が 0.525 のとき、 $F(3, 75) = 8.03$, $p < 0.001$ で有意差があった。STI が 0.375 のとき、 $F(3, 75) = 14.64$, $p < 0.001$ で有意差があることがわかった。STI が 0.23 のとき、 $F(3, 75) = 13.95$, $p < 0.001$ で有意差があることが確認できた。一方、STI が、0.675 及び 0.875 の時は、単語親密度の単純主効果は認められなかった。また、交互作用は、単語親密度にかかわらず、STI が高い値を取ると、有意差は確認できなかった。このことから、単語親密度に注意すれば STI を操作することで単語了解度を制御できたといえる。

図 4.4 に、各 STI における聞き取りにくさの聴取実験の結果を示す。STI を低くすると、聞き取りにくさが著しく増加し、1.0 に飽和した。また、この傾向は単語親密度に依存しなかった。この結果を分散分析した。STI が 0.875~0.23 のとき、単語了解度は、単語親密度別に有意差がないことが確認された。このことから、聞き取りにくさは、単語親密度にかかわらず、STI で制御できたといえる。

図 4.5 に、各 STI における聞き取りにくさの聴取実験の結果を示す。STI を低くすると、わずらわしさが著しく増加し、1.0 に飽和した。また、この傾向は単語親密度に依存しなかった。この結果を分散分析した。STI が 0.875~0.23 のとき、単語了解度は、単語親密度別に有意差がないことが確認された。このことから、聞き取りにくさは、単語親密度にかかわらず、STI で制御できたといえる。

これらの結果より、STI を制御した結果、単語親密度、聞き取りにくさ、わずらわしさを制御できることがわかった。また、提案法の原理通り、STI を後部残響を用いて、制御できることがわかった。

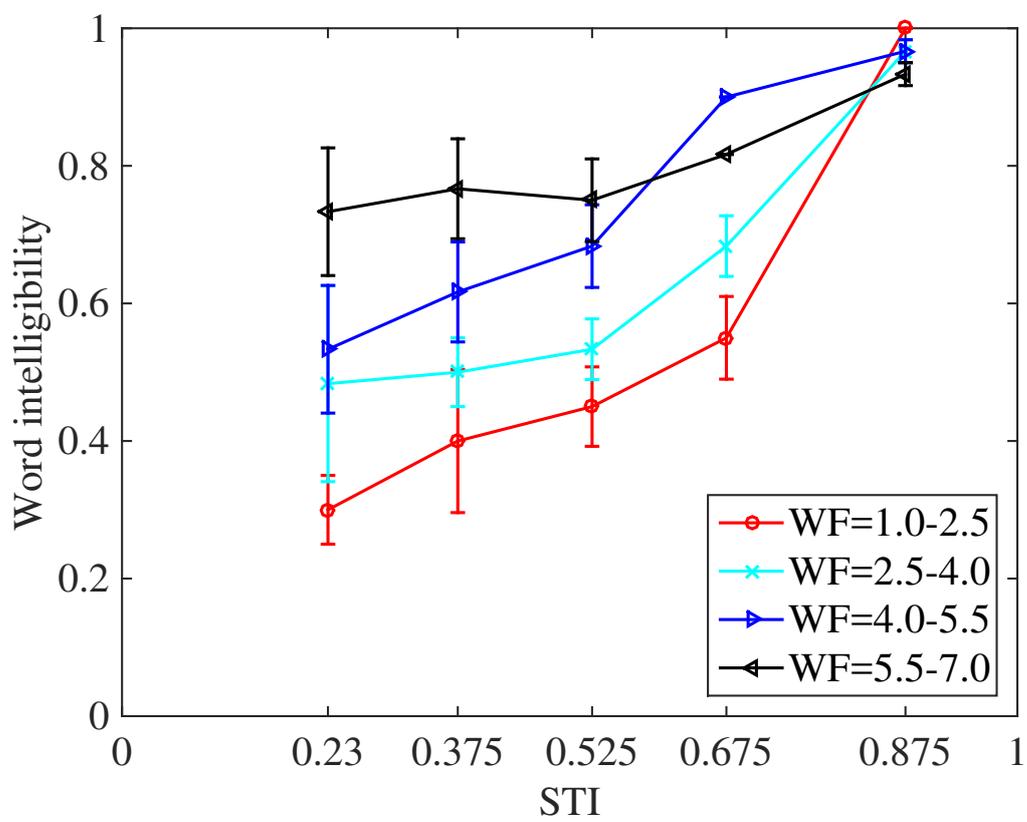


図 4.3: 単語了解度の結果

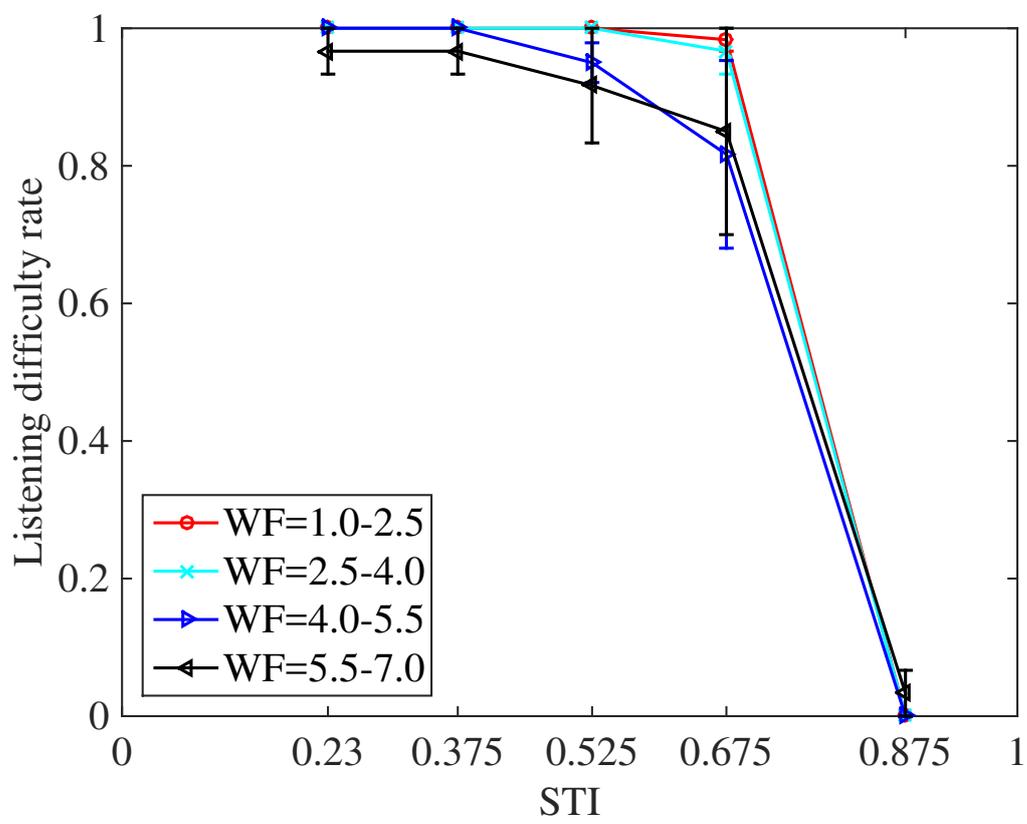


図 4.4: 聴き取りにくさの結果

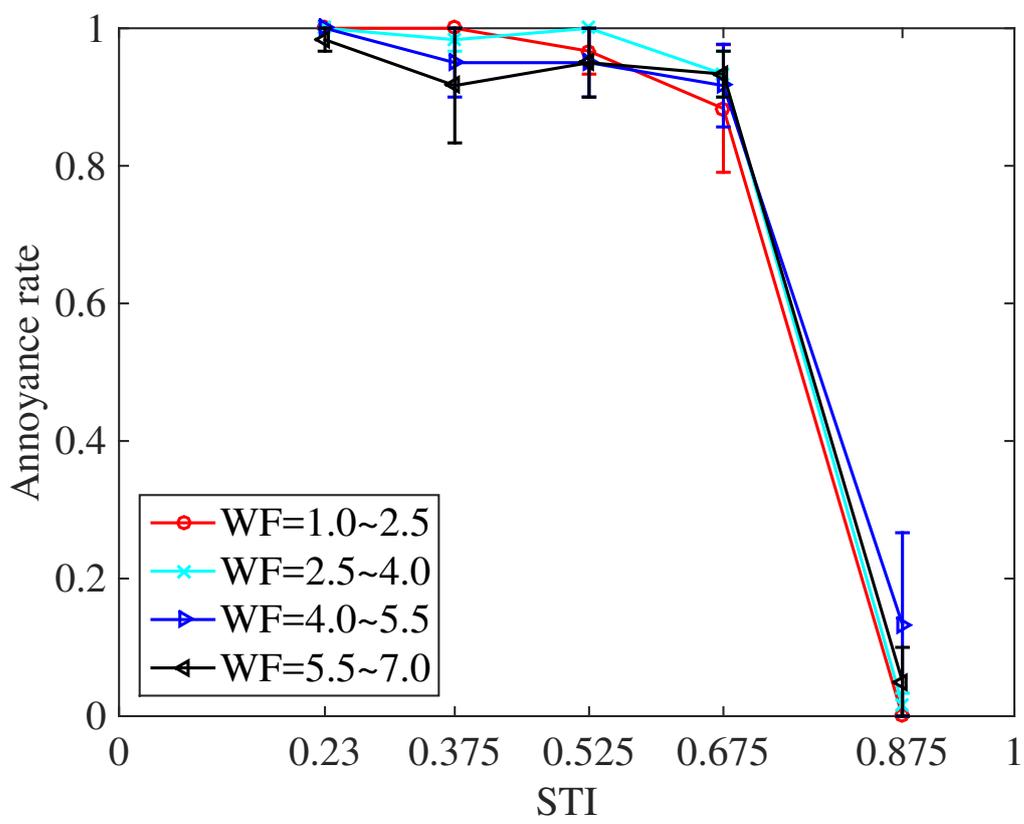


図 4.5: わずらわしさの結果

4.3 提案法と従来法の比較評価

これまでの検討より、単語了解度は、単語親密度別ではあるが、STIの操作によって、制御ができた。また、単語親密度に関係なく、聴き取りにくさ、わずらわしさは制御できた。この結果から、スピーチプライバシー保護に関して新たな方法が提案できた。ここからは、現存するスピーチプライバシー保護の技術との比較を行う。その比較対象は、SchroederのRIRを畳み込み残響を付加しただけの残響音声と、帯域制限されたピンク雑音である。残響音声は、最も一般的なRIRモデルであるSchroederのRIRとの性能比較を行うため、帯域制限されたピンク雑音は、定常雑音の中で最も了解度が低いという結果であったため、これらの手法を比較対象とした[8]。

図4.6に高親密度における提案法と残響音声、ピンク雑音の単語了解度試験の結果を示す。エラーバーは標準誤差を表している。ここで、単語了解度は、3手法においてほとんど変化しない結果となった。図4.7に低親密度における提案法と残響音声、ピンク雑音の単語了解度試験の結果を示す。高親密度のときと同じく、3手法ともほとんど同じ値となった。1要因の分散分析を行った結果、単語親密度にかかわらず、有意差がないことがわかった。これは、単語親密度に注意すれば、3手法ともSTIを操作によって、単語了解度を制御可能であることの裏付けとなる。

図4.8に高親密度における提案法と残響音声、ピンク雑音の聴き取りにくさの結果を示す。高親密度において、提案法は、ピンク雑音と同等の聴き取りにくさを持つことがわかった。提案法は、残響の恩恵を受け、パワーマスキングであるピンク雑音と同等の効果が発揮された。また、提案法は、残響音声よりも聴き取りにくさが高くなることがわかる。残響音声より提案法の残響時間が長いため、聴き取りにくいという印象を被験者に与えたためであると考えられる。図4.9に低親密度における提案法と残響音声、ピンク雑音の聴き取りにくさの結果を示す。低親密度において、3手法には、ほぼ変化がないことがわかり、どれも同等の聴き取りにくさを保っていたことが確認された。図4.4の結果のように、単語親密度にかかわらないという結果にならなかった理由は、主観評価尺度である聴き取りにくさが、ピンク雑音によって他の結果に影響を与えたためである。1要因の分散分析を行った結果、単語親密度にかかわらず、有意差がないことが確認できた。

図4.10に高親密度における提案法と残響音声、ピンク雑音のわずらわしさの結果を示す。高親密度において、提案法は、ピンク雑音よりもわずらわしくないことがわかった。その原因として、ピンク雑音のようなパワーマスキングとは違い、後部残響として残響音声を呈示したためにわずらわしくないと判断されたと考えられる。一方で、残響音声よりも提案法は、わずらわしいと判断された。これは、直接音と後部残響からなる拡張型RIRの後部残響モデルを使用したために、直接音と残響音声との間に空隙が発生し、そこで、不快感があったためであるといえる。図4.11に低親密度における提案法と残響音声、ピンク雑音のわずらわしさの結果を示す。高親密度のときと同じような結果となった。1要因の分散分析を行った結果、単語親密度にかかわらず、有意差がなかった。しかし、提案法は、従来法と類似する傾向が得られた。これより、単語親密度に関係なく、わずらわしさを制御することができることの裏付けになった。

図 4.12, 4.13, 4.14, 4.15, 4.16, 4.17, を見ると, STI が 0.23 において提案法の SNR が最も低いことがわかる. それにもかかわらず, 提案法が他の方法とほとんど性能が変わらなかったため, 提案法が他の手法よりも優れていることを表している.

4.4 考察・まとめ

本章では, 提案法の性能評価および, 提案法と残響音声とピンク雑音との比較を行った. 性能評価の結果, 提案法を用いて, STI を操作することで, 単語理解度, 聴き取りにくさ, わずらわしさを制御することができ, 提案法は, スピーチプライバシー保護ができることがわかった. また, 単語理解度の制御は, 単語親密度に注意して行う必要があることがわかり, 聴き取りにくさ, わずらわしさを制御は, 単語親密度に関係しないことがわかった. 提案法と残響音声とピンク雑音を比較した結果, 提案法の SNR が最も低かったにも関わらず, おおむね同等の性能があった. これより, 他の 2 手法と比較すると効率よく STI を制御できた.

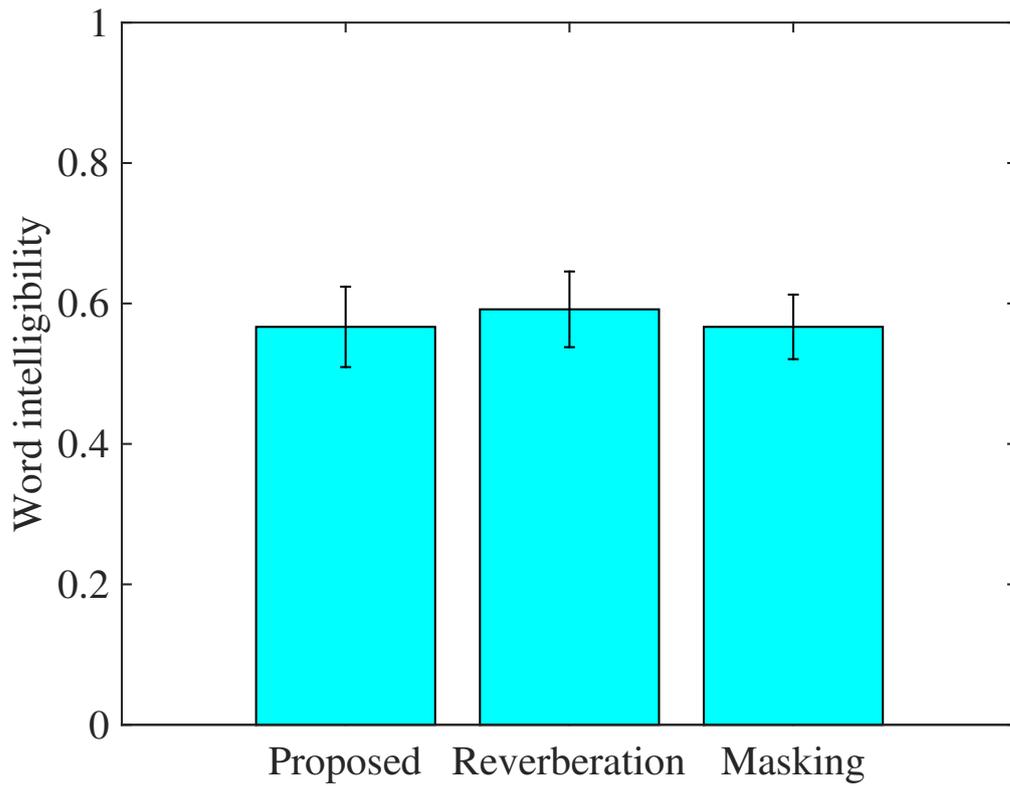


図 4.6: 他手法の比較：単語了解度の結果（高親密度の場合）

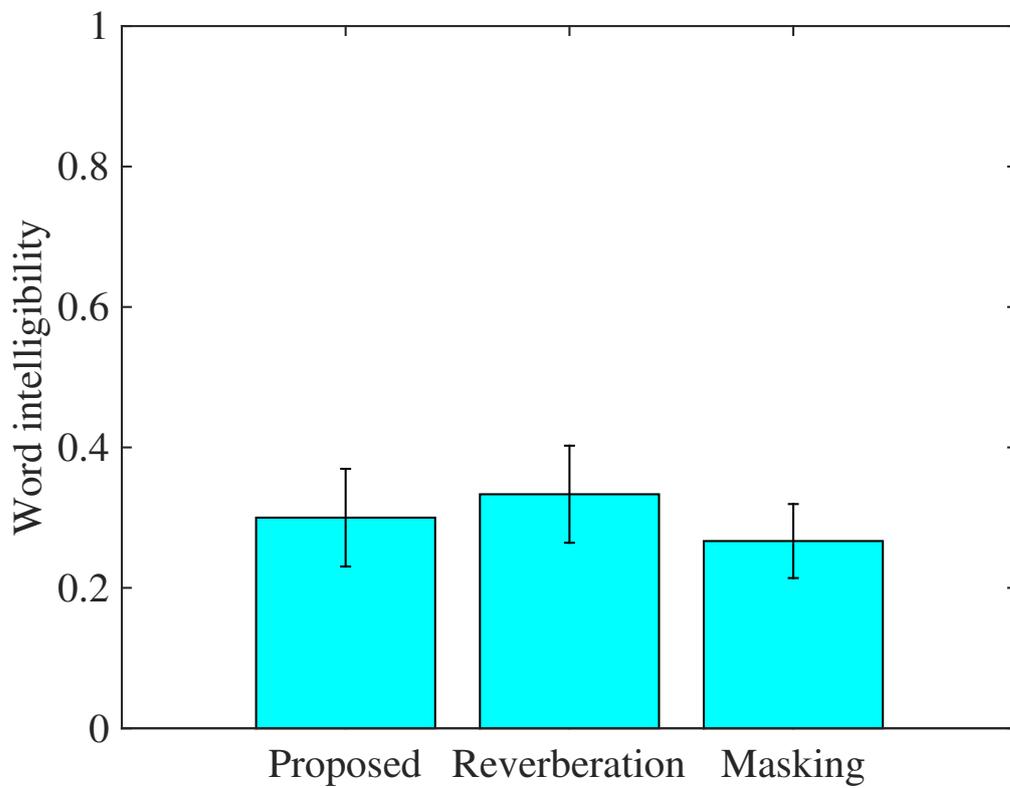


図 4.7: 他手法の比較：単語了解度の結果（低親密度の場合）

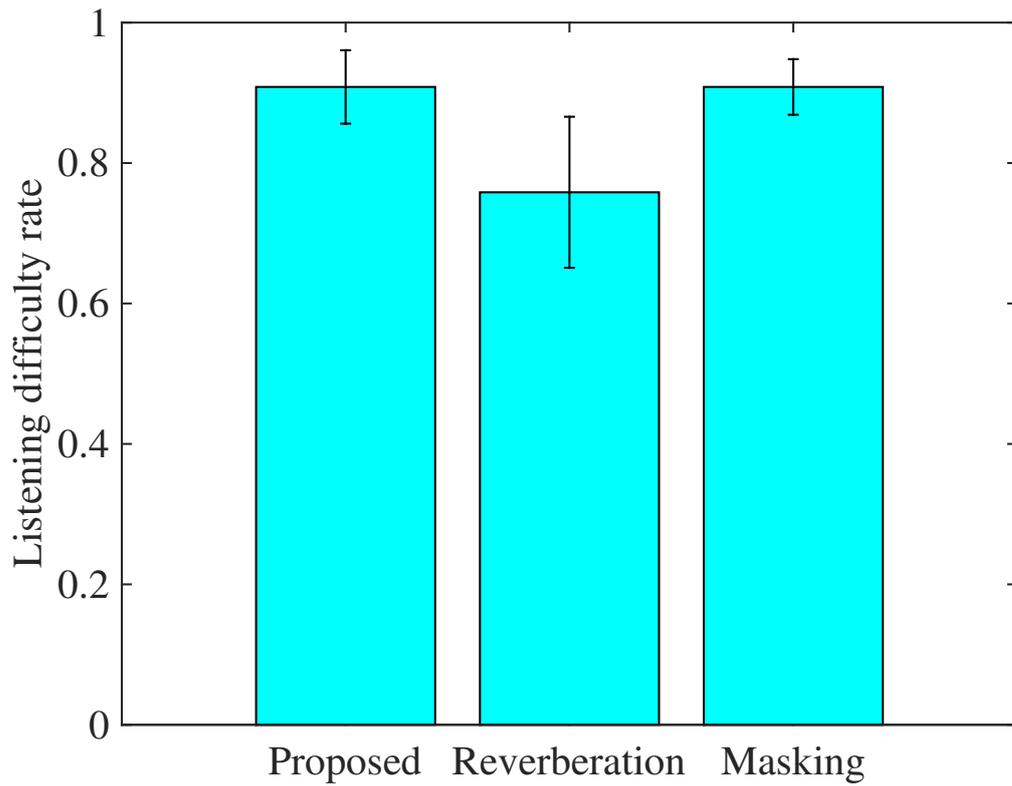


図 4.8: 他手法の比較 : 聴き取りにくさの結果 (高親密度の場合)

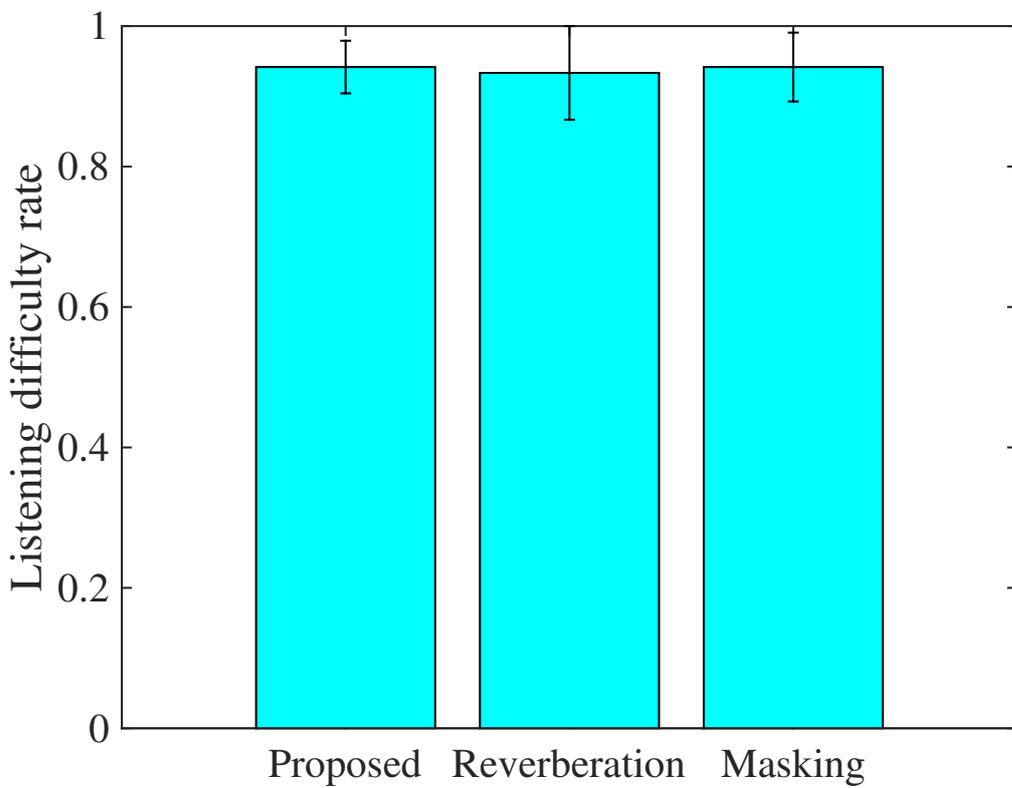


図 4.9: 他手法の比較 : 聴き取りにくさの結果 (低親密度の場合)

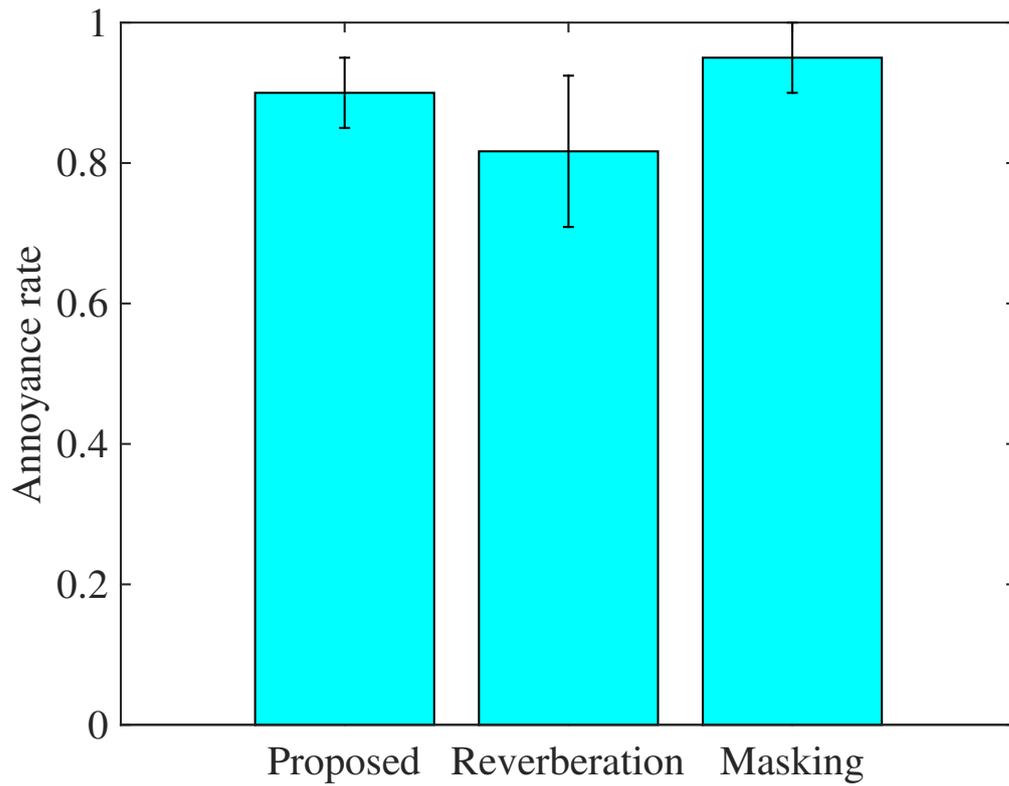


図 4.10: 他手法の比較：わずらわしさの結果（高親密度の場合）

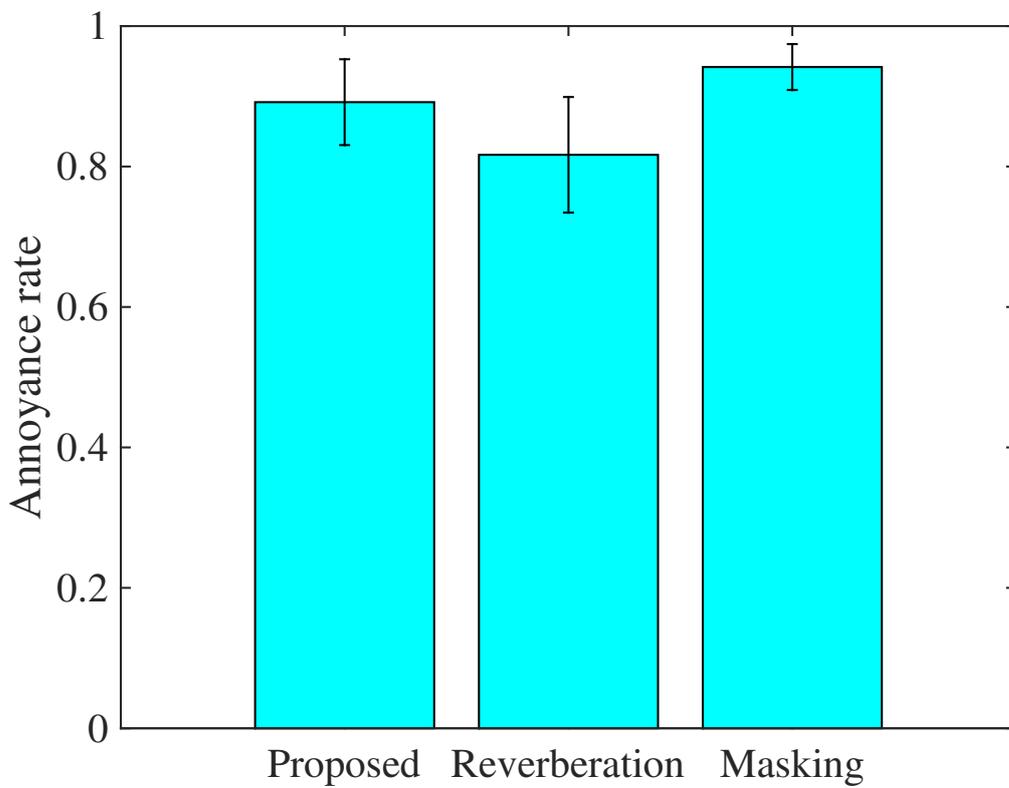


図 4.11: 他手法の比較：わずらわしさの結果（低親密度の場合）

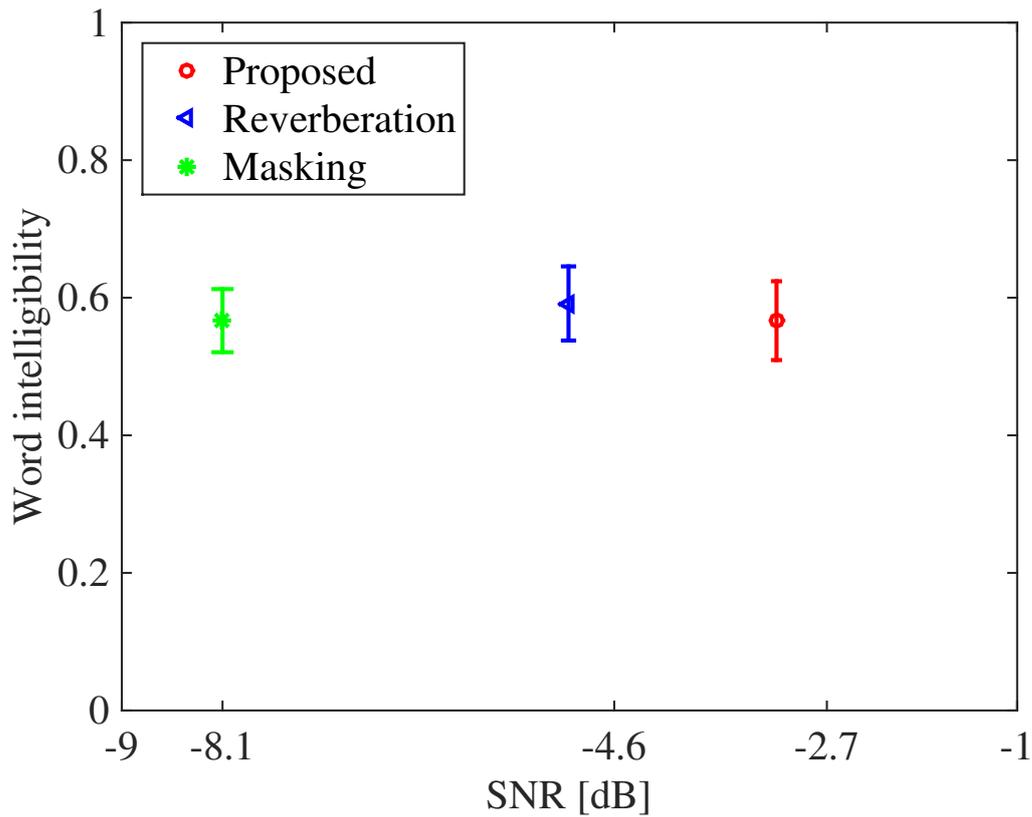


図 4.12: 他手法の比較：単語理解度と SNR の関係（高親密度の場合）

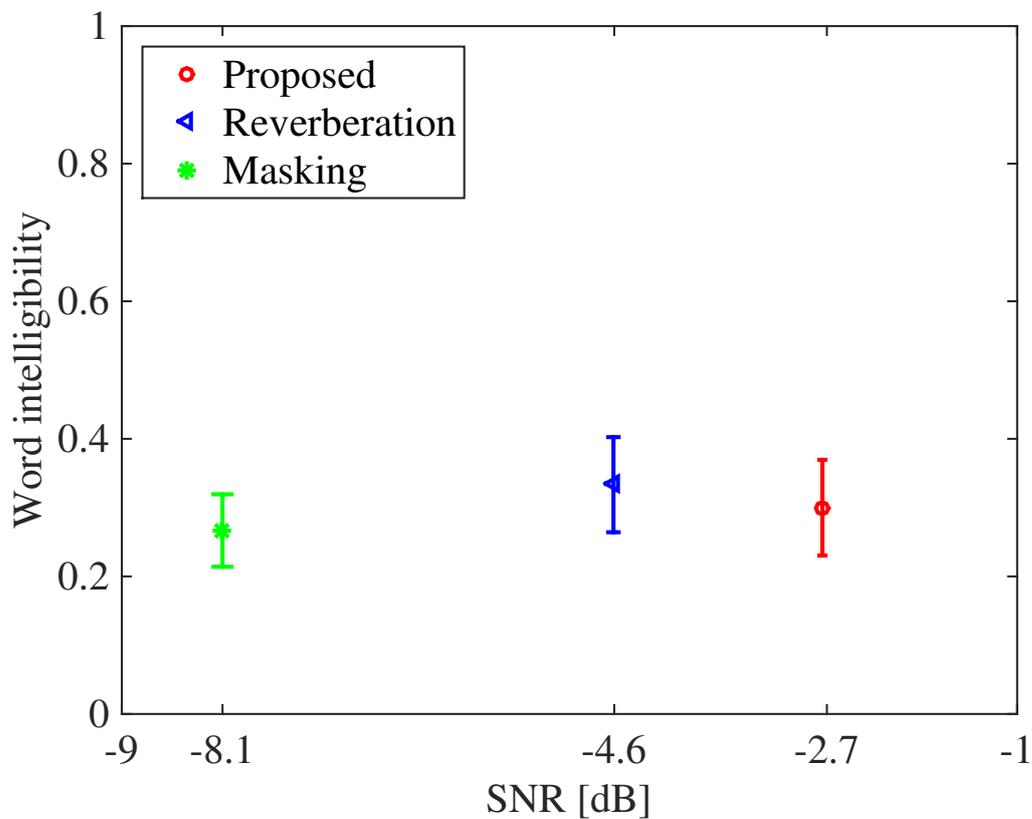


図 4.13: 他手法の比較：単語理解度と SNR の関係（低親密度の場合）

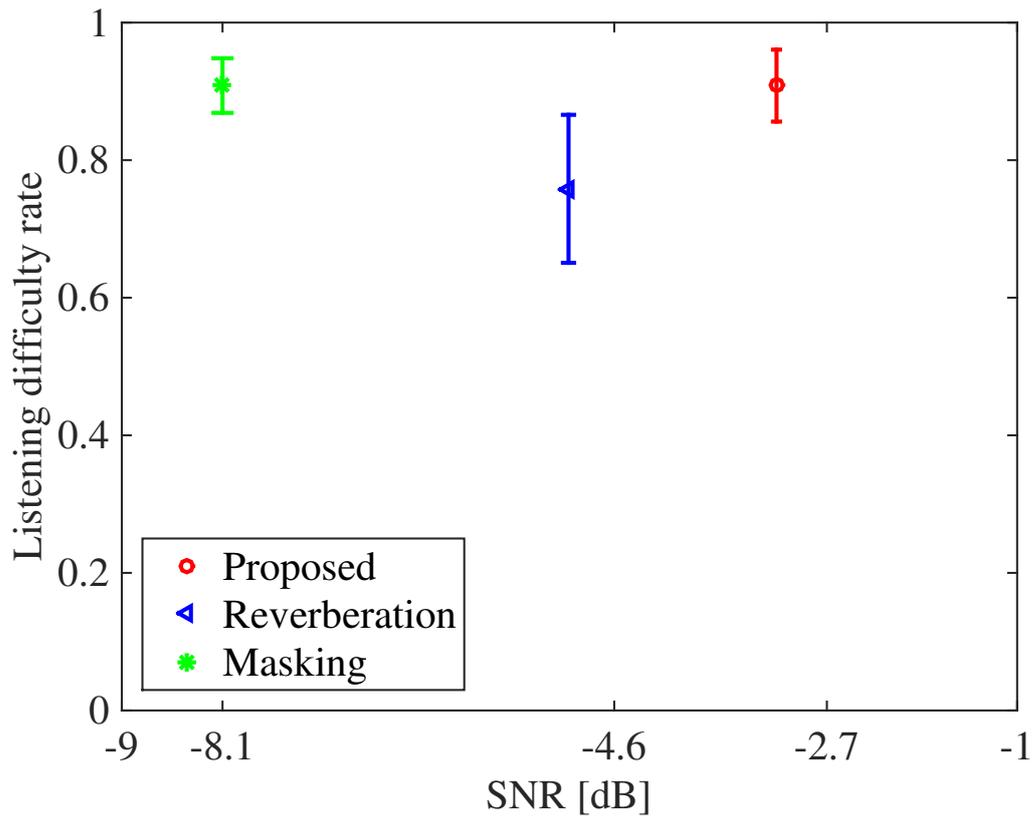


図 4.14: 他手法の比較：聴き取りにくさと SNR の関係（高親密度の場合）

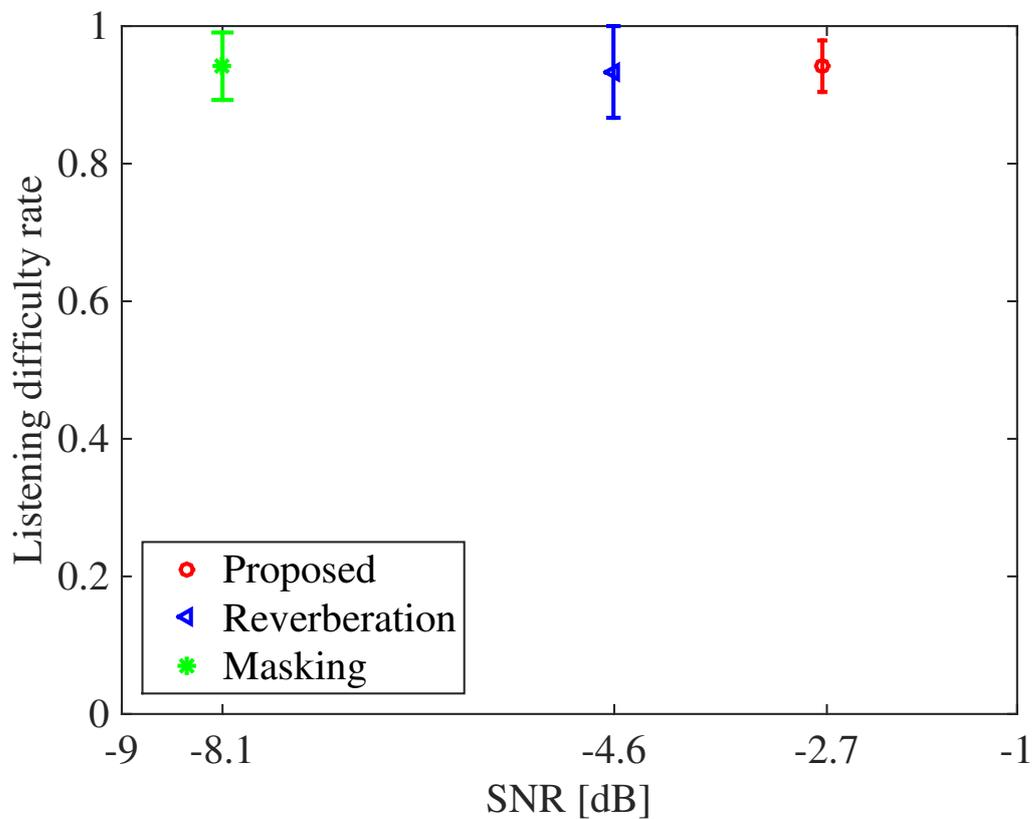


図 4.15: 他手法の比較：聴き取りにくさと SNR の関係（低親密度の場合）

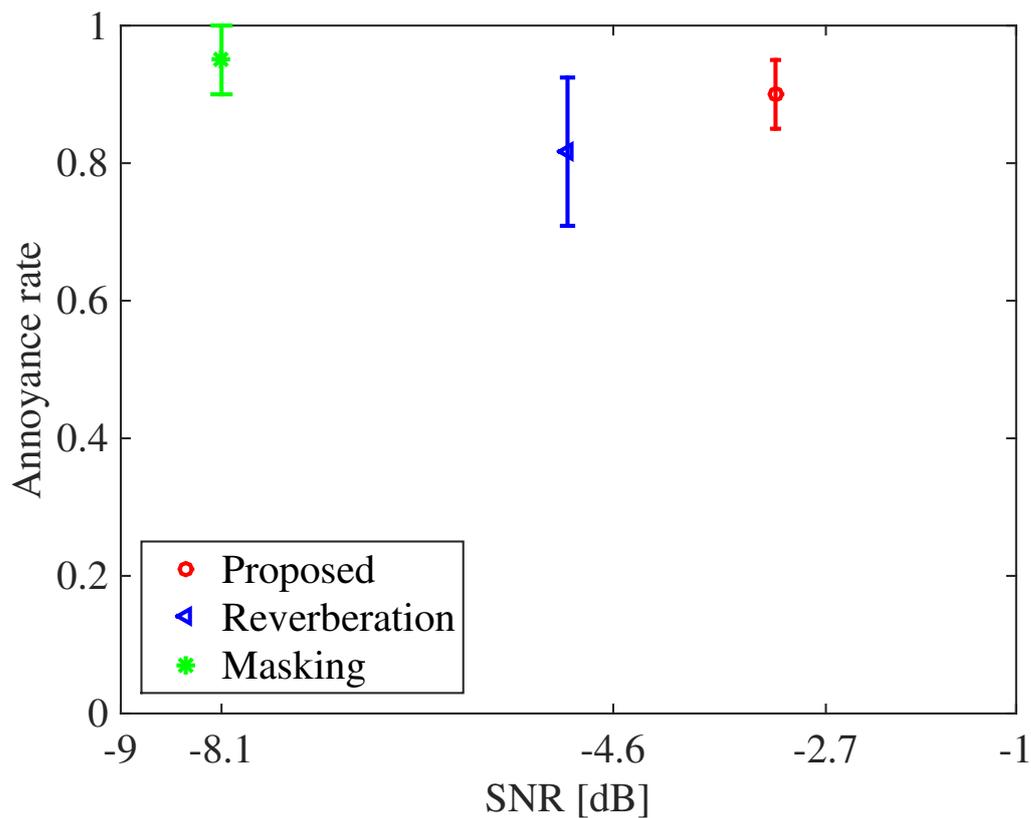


図 4.16: 他手法の比較：わずらわしさと SNR の関係（高親密度の場合）

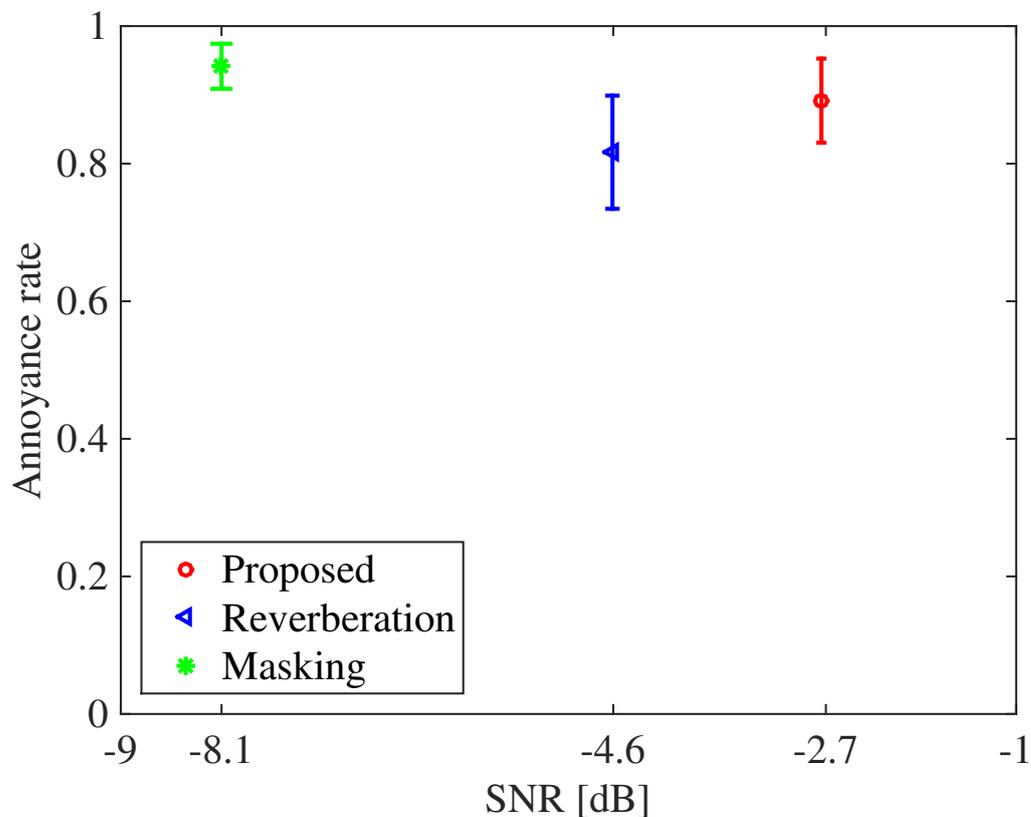


図 4.17: 他手法の比較：わずらわしさと SNR の関係（低親密度の場合）

第5章 結論

5.1 本研究で明らかになったこと

本研究は、後部残響を変化によって、STIを自在に操作することで、第三者に漏洩する音声の聴き取りに影響を与え、スピーチプライバシー保護を目的とした。そのために、以下の3つのことを行った。

- 本研究における原理を確立するために後部残響モデルを定義し、どのような後部残響が、STIに影響があるか検討した。
- 提案法の性能評価を行うために、STIを変化させたときに単語理解度試験、聴き取りにくさ、わずらわしさがどのように変化するか検討を行った。
- 他の方法と比較し、提案法が他の方法よりも良いかどうか検討した。

一つ目は、拡張型RIRを後部残響とした後部残響モデルを定義した。拡張型RIRを後部残響とした後部残響モデルは、立ち上がりを制御するパラメータである T_h や立ち下がり制御するパラメータである T_t 、振幅項 a を制御することが容易で、それゆえに、STIを操作することが容易であった。これは、直接音と後部残響である拡張型RIRで構成されたモデルであり、他手法と比べてSNRが低いことから、効率よくSTIを減少させられるモデルであったといえる。明らかになったことは、拡張型RIRの立ち上がり立ち下がり制御するパラメータ T_h と T_t がSTIに大きく影響を与えるということである。

二つ目は、提案法の性能評価を行うために、STIを変化させたときに単語理解度試験、聴き取りにくさ、わずらわしさがどのような変化をするか確かめた。このことから、単語理解度試験においては、単語親密度に注意すれば、STIの操作によって、単語理解度を操作できることがわかった。また、聴き取りにくさとわずらわしさは、単語親密度にかかわらずSTIの操作によって、制御されることがわかった。これより、STIを操作することで、主観評価尺度である、単語理解度、聴き取りにくさ、わずらわしさを制御することができたといえる。これにより、STIを操作することによって、スピーチプライバシー保護ができることを明らかにした。

三つ目に、STIが0.23のときに、提案法と残響音声とピンク雑音を単語理解度、聴き取りにくさ、わずらわしさについて比較した。その結果、提案法は、ほとんど同等の性能を発揮できた。さらに、提案法は、他の2手法よりもSNRが低くなったため、効率よくSTIが制御できたといえる。これにより、従来のスピーチプライバシー保護の技術よりもスピーチプライバシー保護が容易に可能なことを明らかにした。

5.2 残された課題

本研究において、残された課題は、以下のような3つが挙げられる。

- 直接音と後部残響で構成された簡易なモデルを使用したため、実際の室の環境に適したモデルではなかったこと。
- 実際の部屋の環境を無視したこと。
- スピーチプライバシー保護を行う上で、音声の特徴を考慮しなければならないが、まずは、問題を簡便化するため、直接音と後部残響で構成されたモデルを作成したため、音声そのものの特徴を考慮していないこと。

一つ目は、直接音と後部残響で構成された簡易なモデルを使用したため、実際の室の環境に適したモデルではなかった事が考えられる。本研究では、直接音と後部残響のみで構成したモデルを使用した。実際の室の環境では、直接音ではなく、SchroederのRIRのような指数減衰するようなモデルを直接音に置き換える必要があると考えられる。

二つ目は、実際の室の環境を無視したことがあげられる。本研究の最終目的は、銀行や病院、薬局などのオープンスペースにおいて、スピーチプライバシーを保護することである。しかし、実際のオープンスペースを用いて、提案法は、評価されていない。そこで、提案法が実環境にどの程度有効か確かめるために、実際のオープンスペースにおいて、提案法の評価が必要となる。本研究では、図2.1における $h_0(t) = \delta(t)$ としていたが、実際の室の特性にあわせた $h_0(t)$ を考慮する必要がある。そのためには、宮崎らや鷓木らによるSTI推定法を利用して実際の室のSTIを常に推定し、どのような残響を音声に畳み込めば良いかを常に予測することが必要となる。

三つ目は、音声の特徴を一切無視した方法の提案となったことである。音声は、変調スペクトルの2～16 Hzにピークがあることが知られている。しかし、本研究における提案法は、STIを減少させることにしか着目していない。音声の特徴に着目すれば、本研究の提案法よりも効率の良いスピーチプライバシー保護ができる。

謝辞

本研究を遂行する上で、主指導教官である鵜木祐史教授には、研究のことから生活態度、人との付き合い方などにいたるまで、厳しくも優しく非常に丁寧な指導を賜り、深く感謝いたします。また、研究室会議やミーティングなどの場において、数多くのご助言を頂いた赤木正人教授をはじめとする鵜木研究室・赤木研究室内のメンバーの皆様には感謝いたします。公私共に、私が問題に直面するとともに悩み考えてくれた本学の仲間、両親に深く感謝いたします。

参考文献

- [1] 佐藤 洋, 清水 寧, “スピーチプライバシ研究の歴史と近年の動向,” 日本音響学会誌, Vol. 64, No. 8, pp. 475–480, 2008.
- [2] Cherry, E. C., “Some experiments on the recognition of speech, with one and with two ears,” J. Acoust. Soc. Am., Vol. 25, No. 5, pp. 975–979, 1953.
- [3] 積山 薫, “顔と声による音声知覚–視聴覚情報の相互作用–,” 信学技報, HIP97–21, pp. 83–90, 1997.
- [4] 安部 素嗣, 安藤 繁, “共有 FM-AM の時間周波数統合に基づく聴覚情景解析 (I) : Lagrange 微分特徴量とその周波数軸統合,” 電子情報通信学会論文誌, Vol. 83, No. 2, pp. 458–467, 2000.
- [5] 安部 素嗣, 安藤 繁, “共有 FM-AM の時間周波数統合に基づく聴覚情景解析 (II) : 最適な時間軸統合とストリーム音の再合成,” 電子情報通信学会論文誌, Vol. 83, No. 2, pp. 468–477, 2000.
- [6] 山口 晃治, 増田 潔, 買手 正浩, “スピーチプライバシ評価のための RASTI の予測に関する検討,” 大成建築技術センター報, vol. 44, no. 49, pp. 1–4, 2011.
- [7] 西村 正治, “アクティブノイズコントロール (<連載企画> 音響学の温故知新),” 日本音響学会誌, Vol. 69, No. 8, pp. 433–434, 2013.
- [8] 佐伯 徹郎, 山口 静馬, 為末 隆弘, “マスキングノイズによるスピーチプライバシ保護に関する一考察,” 信学技報, EA2003-93, pp. 43–48, 2003.
- [9] 近藤 和弘, 泉 良, 中川 清司, “新しい日本語了解度試験方法の評価,” 信学技法, SP2000-163, pp. 25–32, 2001.
- [10] 赤木 正人, 入江 佳洋, “音情景解析の概念にもとづいた音声プライバシ保護,” 電子情報通信学会論文誌 A, J97-A(4), pp. 247–255, 2014.
- [11] 小林 洋介, 近藤 和弘, “主観・客観評価値を用いたスピーチマスキングシステムの品質推定,” 信学技報, EA2016-59, pp. 67–72, 2016.

- [12] 三戸 武大, 荒井 隆行, 安 啓一, “サウンドマスキングシステムにおけるデータベースを用いた音声マスク作成法の提案,” 日本音響学会誌, Vol. 71, No. 8, pp. 382–389, 2015.
- [13] 渡邊 康平, 森本 政之, 佐藤 逸人, “残響音の単語了解度,” 日本建築学会近畿支部研究報告集, 環境系 (52), pp. 49–52, 2012.
- [14] 星野 康, 森本 政之, 佐藤 逸人, “遮音性能とスピーチプライバシーの関係,” 日本建築学会学会公園論文集, D-1, pp. 343–344, 2010.
- [15] IEC 60268-16:2003. “Sound system equipment - Part 16: Objective rating of speech intelligibility by speech transmission index”.
- [16] 宮崎 晃和, 森田 翔太, 鷗木 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,” 信学技法, EA2013–88, pp. 1–6, 2013.
- [17] 鷗木 祐史, 佐々木 恭平, 宮内 良太, 赤木 正人, “残響音声からの音声伝達指標のブラインド推定法の検討,” 信学技法, EA2013–44, pp. 63–68, 2013.
- [18] Kuttruff, H., “Room Acoustics”, 4th ed. (Elsevier Science Publishers Ltd., Lindin), 1991.
- [19] 飯田 茂隆, “明瞭度試験法について,” 日本音響学会誌, Vol. 43, No. 7, pp. 532–536, 1987.
- [20] 戸井田 義徳, “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.
- [21] 難波 精一郎, “ノイジネス・アノイヤンスについて,” 日本音響学会誌, Vol. 44, No.10, pp. 775–780, 1988.
- [22] 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.
- [23] 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., Vol. 116, No. 3, pp. 1607–1613, 2004.
- [24] Sato, H., Morimoto, M., Sato, H., and Wada, M., “Relationship between listening difficulty and acoustical objective measures in reverberation fields,” J. Acoust. Soc. Am., Vol. 123, No. 4, pp. 2087–2093, 2008.
- [25] 佐藤 逸人, 森本 正之, 佐藤 洋, “聴き取りにくさによる音声伝送性能の評価,” 日本音響学会誌, Vol. 63, No.5, pp. 275–280, 2007.

- [26] 佐藤 逸人, “「聴き取りにくさ」の測定方法,” 騒音制御, Vol. 35, No. 1, pp. 69–74, 2011.
- [27] 佐藤 洋, “音声情報伝達 : 「聴き取りにくさ」からスピーチプライバシーまで,” 騒音制御, Vol. 33, No. 4, pp. 261–264, 2008.
- [28] French, N. R. and Steinberg, J. C., “Factors governing the intelligibility of speech sounds,” J. Acoust. Soc. Am., Vol. 19, pp. 90–119, 1947.
- [29] Joint Subcommittee on Speech Privacy of the ASA/INCE/NCAC, “Interim Sound and Vibration Design Guidelines for Hospital and Healthcare Facilities,” Public draft dated Nov.28, 2007.
- [30] 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.
- [31] Steeneken, H. J. M. and Houtgast, T., “A physical method for measuring speech-transmission quality,” J. Acoust. Soc. Am., Vol. 67, No. 1, pp. 318–326, 1980.
- [32] Schroeder, M. R., “Modulation transfer functions: definition and measurement,” Acustica, Vol. 49, pp. 179–182, 1981.
- [33] 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).
- [34] Drullman, R., Festen. J. M., and Plomp, R., “Effect of temporal envelope smearing on speech reception,” J. Acoust. Soc. Am., Vol. 95, No. 2, pp. 1053–1064, 1994.
- [35] Drullman, R., Festen. J. M., and Plomp, R., “Effect of reducing slow temporal modulations on speech reception,” J. Acoust. Soc. Am., Vol. 95, No. 5, pp. 2670–2680, 1994.
- [36] 近藤 公久, 坂本 修一, 天野 成昭, 鈴木 陽一, “信号対雑音比調整による単語リスト間の単語了解度差補正 –親密度別単語了解度試験用音声データセット (FW07) を用いた検証–,” 日本音響学会誌, Vol. 69, No. 5, pp. 224–231, 2013.
- [37] 近藤 公久, 坂本 修一, 天野 成昭, 鈴木 陽一, “親密度別単語了解度試験用音声データセット 2007(FW07) の作成,” 信学技法, SP2007-157, pp.43–48, 2008.
- [38] 近藤 公久, 天野 成昭, 坂本 修一, 鈴木陽一, “親密度別単語了解度試験用音声データセット (FW07),” NII 音声資源コンソーシアム, 2007 年.

研究業績一覧

柏原 佑太, 森田 翔太, 鷗木 祐史, “インパルス応答の後部残響成分が音の聴取に与える影響の調査,” 平成 27 年度電気関係学会北陸支部連合大会, G-6, 2015.

Kashihara, Y. and Unoki, M., “Study on IIR Implementation for Modulation Transfer Function of Room Impulse Response,” Proc. NCSP ’ 16, pp.165-168, 2016.

Kashihara, Y. and Unoki, M., “Study on IIR Implementation for Modulation Transfer Function of Room Impulse Response,” Journal of Signal Processing, Vol. 20, No. 4, pp. 201-204, 2016

柏原 佑太, 鷗木 祐史, “変調伝達関数モデルを利用した音声伝送指標の検討,” 平成 28 年度電気関係学会北陸支部連合大会, G-1, 2016.

石川 大介, 柏原 佑太, 鷗木 祐史, “室内インパルス応答のパラメータ逆推定の検討,” 平成 28 年度電気関係学会北陸支部連合大会, G-2, 2016.

柏原 佑太, 鷗木 祐史, “変調伝達関数モデルを利用した音声伝送指標の検討,” 信学技報, EA 2016-33 (2016-10), pp.13-18, 2016.

鷗木祐史, 石川大介, 柏原佑太, 小林まおり, 赤木正人, “室内インパルス応答のモデル化とその室内音響特性の検討,” 信学技報, EA 2016-61 (2016-11), pp.79-84, 2016.

Kashihara, Y. and Unoki, M., “Study on controlling listening comprehension by manipulating speech transmission index,” Proc. NCSP ’ 17, 2017.

柏原 佑太, 鷗木 祐史, “音声伝送指標に着目した音声の聴き取り制御に関する検討,” 日本音響学会 2017 年春季研究発表会, 3-Q-18, 2017.

付録A MTFの導出方法の一例

A.1 はじめに

STIを算出するためには、変調伝達関数（MTF）を正しく算出する必要がある。MTFは、室内インパルス応答（RIR）から求められる。RIRを正しくモデル化することにより、正しいMTFが算出可能となり、正しいSTIが求められるようになる。算出したMTFがどのような形状のとき、STIがどのような変化をするかを検討する必要がある。そのため、MTFを算出することは非常に重要な課題である。本稿では、実際に算出したMTFを紹介する。

A.2 MTFの定義

MTFを表す $m(f_m)$ は、次式のように定義される。

$$m(f_m) = \frac{|\int_0^\infty h^2(t) \exp(-j2\pi f_m t) dt|}{\int_0^\infty h^2(t) dt} \quad (\text{A.1})$$

ここで、 $h(t)$ はインパルス応答を表す。MTFの定義は、 $h^2(t)$ に対する一種のフーリエ変換を表している。

A.3 導出方法及び導出したMTF

A.3.1 単一エコーに対応するMTF

単一エコーのモデルは次式のように表される。

$$h(t) = \delta(t) + a\delta(t - \tau) \quad (\text{A.2})$$

ここで、 $\delta(t)$ はインパルス、 a は振幅項、 τ は遅延時間とする。式(A.2)を式(A.1)に代入すると次のような式になる。

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

$$\begin{aligned}
&= \frac{|\exp(-j2\pi f_m t)|_{t=0} + a^2 \exp(-j2\pi f_m t)|_{t=\tau}|}{1 + a^2} \\
&= \frac{\sqrt{1 + 2a^2 \cos(2\pi f_m \tau) + a^4}}{1 + a^2}
\end{aligned} \tag{A.3}$$

A.3.2 Schroeder の RIR に対応する MTF

Schroeder の RIR は、統計的に RIR を近似した室内インパルス応答のモデルであり、次式のように表される。

$$h(t) = e_h(t)c_h(t) = a \exp\left(\frac{-6.9t}{T_R}\right)c_h(t) \tag{A.4}$$

ここで、 a は振幅項、 t は時間、 T_R は残響時間、 $c_h(t)$ は白色雑音である。式 (A.4) を式 (A.1) に代入すると次のような式になる。

$$\begin{aligned}
m(f_m) &= \frac{|\int_0^\infty h^2(t) \exp(-j2\pi f_m t) dt|}{\int_0^\infty h^2(t) dt} \\
&= \left| \frac{a^2}{\frac{13.8}{T_R} + j2\pi f_m} \right| \cdot \frac{\frac{13.8}{T_R}}{a^2} \\
&= \frac{a^2}{\sqrt{\left(\frac{13.8}{T_R}\right)^2 + (2\pi f_m)^2}} \cdot \frac{\frac{13.8}{T_R}}{a^2} \\
&= \frac{1}{\sqrt{1 + (2\pi f_m \frac{T_R}{13.8})^2}}
\end{aligned} \tag{A.5}$$

A.3.3 後部残響を Schroeder の RIR としたものに対応する MTF

先行研究より、後部残響が STI に影響を与えることが示唆された。そこで、次式のようなモデルを考案した。

$$h(t) = \delta(t) + g(t - \tau) \tag{A.6}$$

$g(t - \tau)$ は、次のような式である。

$$g(t - \tau) = a \exp\left(\frac{-6.9(t - \tau)}{T_R}\right)c_h(t) \tag{A.7}$$

式 (A.6) を式 (A.1) に代入すると次のような式になる。

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

$$\begin{aligned}
&= \frac{|\exp(-j2\pi f_m t)|_{t=0} + \int_{\tau}^{\infty} g^2(t-\tau) \exp(-j2\pi f_m t) dt}{1 + \frac{a^2}{\frac{13.8}{T_R}}} \\
&= \left| 1 + \frac{a^2 \exp(-j2\pi f_m \tau)}{\frac{13.8}{T_R} + j2\pi f_m} \right| \cdot \frac{\frac{13.8}{T_R}}{\frac{13.8}{T_R} + a^2} \\
&= \frac{\frac{13.8}{T_R} \sqrt{(\frac{13.8}{T_R})^2 + (2\pi f_m)^2} + a^4 + 2a^2 \frac{13.8}{T_R} \cos(2\pi f_m \tau) - 4a^2 \pi f_m \sin(2\pi f_m \tau)}{(\frac{13.8}{T_R} + a^2) \sqrt{(\frac{13.8}{T_R})^2 + (2\pi f_m)^2}}
\end{aligned} \tag{A.8}$$

A.3.4 一般化 RIR に対応する MTF

一般化 RIR は次式のようにになる。

$$h(t) = e_h(t)c_h(t) = at^{(b-1)} \exp\left(-\frac{13.8t}{T_R}\right) c_h(t) \tag{A.9}$$

ここで、 b は一般化 RIR の次数であり、 $b \geq 1$ の値を取る。以下に式 (??) の MTF を算出するための準備を行う。

$$\int_0^{\infty} \exp(-\alpha t) dt = \frac{1}{\alpha} \tag{A.10}$$

ここで、 α は変数とし、 α を n 回微分すると次式のようにになる。

$$\int_0^{\infty} t^n \exp(-\alpha t) dt = \frac{n!}{\alpha^{n+1}} \tag{A.11}$$

次に複素数の絶対値に関して準備を行う。

$$\begin{aligned}
\left| \frac{1}{(\alpha + j\omega)^n} \right| &= \left| \frac{1}{\alpha + j\omega} \right| \cdot \left| \frac{1}{\alpha + j\omega} \right| \cdots \left| \frac{1}{\alpha + j\omega} \right| \\
&= \frac{1}{\sqrt{\alpha^2 + \omega^2}} \cdot \frac{1}{\sqrt{\alpha^2 + \omega^2}} \cdots \frac{1}{\sqrt{\alpha^2 + \omega^2}} \\
&= \frac{1}{(\alpha^2 + \omega^2)^{\frac{n}{2}}}
\end{aligned} \tag{A.12}$$

式 (A.10)、式 (A.12) を利用し、式 (A.9) を式 (A.1) に代入すると次のような式になる。

$$\begin{aligned}
m(f_m) &= \frac{|\int_0^{\infty} h^2(t) \exp(-j2\pi f_m t) dt|}{\int_0^{\infty} h^2(t) dt} \\
&= \frac{a^2 (2b-2)!}{((\frac{13.8}{T_R})^2 + (2\pi f_m)^2)^{2b-1}} \cdot \frac{(\frac{13.8}{T_R})^{2b-1}}{a^2 (2b-2)!} \\
&= \frac{1}{(1 + (2\pi f_m \frac{T_R}{13.8})^2)^{\frac{2b-1}{2}}}
\end{aligned} \tag{A.13}$$

$b = 1$ のとき、式 (A.5) 式と同様の式が得られる。

A.3.5 後部残響を一般化 RIR としたものに対応する MTF

直接音に後部残響が付加しているものを次のモデルで定義した.

$$h(t) = \delta(t) + g(t - \tau) \quad (\text{A.14})$$

$g(t - \tau)$ は, 次のような式である.

$$g(t - \tau) = a(t - \tau)^{(b-1)} \exp\left(\frac{-6.9(t - \tau)}{T_R}\right) c_h(t) \quad (\text{A.15})$$

式 (A.13) の結果を踏まえ, 式 (A.14) を式 (A.1) に代入すると次のような式になる.

$$\begin{aligned} m(f_m) &= \frac{|\int_0^\infty h^2(t) \exp(-j2\pi f_m t) dt|}{\int_0^\infty h^2(t) dt} \\ &= \frac{|\exp(-j2\pi f_m t)|_{t=0} + \int_\tau^\infty g^2(t - \tau) \exp(-j2\pi f_m t) dt|}{1 + \frac{(\frac{13.8}{T_R})^{2b-1}}{a^2(2b-2)!}} \\ &= \left| 1 + \frac{a^2(2b-2)! \exp(-j2\pi f_m \tau)}{(\frac{13.8}{T_R} + j2\pi f_m)^{2b-1}} \right| \cdot \frac{(\frac{13.8}{T_R})^{2b-1} + a^2(2b-2)!}{(\frac{13.8}{T_R})^{2b-1}} \end{aligned} \quad (\text{A.16})$$

$b = 1$ のとき, 式 (A.8) と同様の式になる.

A.4 おわりに

本報告では, 定義通りに MTF の算出を行ったことにより, MTF を求めることができた. しかし, 式 (A.8) と式 (A.16) は本論文に記載したとおり, 本来の MTF とは違う結果になったため, z 変換にて IIR フィルタを設計し, MTF を導出した.