課題研究報告書

音声明瞭度回復を目的とした雑音・残響除去に関する調査研究

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

森田 翔太

2010 年 3 月

課題研究報告書

音声明瞭度回復を目的とした雑音・残響除去に関する調査研究

指導教官　鵜木祐史 准教授

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

北陸先端科学技術大学院大学
情報科学研究科情報科学専攻

0810062 森田 翔太

提出年月: 2010 年 2 月

概 要

音声コミュニケーションは，人にとって欠くことのできない情報伝達方法である．しかし，雑音や残響などの環境によって音声コミュニケーションが阻害されることがある．例えば，音声アプリケーションのハンズフリー機能を使う時に，マイクロホンから離れていると会話が阻害され，話者の発話内容がうまく相手に伝わらないことがある．このような問題を解決するために，音声の明瞭性を回復する雑音残響除去法が必要となるが，これまでにそのような手法は提案されていない．

本研究では，雑音残響環境での円滑な音声コミュニケーションの実現を最終ゴールとして，音声明瞭度の回復を目的とした雑音・残響除去法に関する調査をした．まず，円滑な音声コミュニケーションを評価するのに最善な伝達性能の評価方法を調査した．次に，これまでに提案されてきた雑音・残響除去法の調査をした．最後に，どのようなアプローチにより音声明瞭度を回復する雑音残響除去法の実現ができるかを調査をした．

音声伝達性能の評価方法を調査した結果，音声コミュニケーションを評価するためには，単語の親密度の統制を取りながら単語了解度による評価を行う必要があることがわかった．また，この評価と同時に「聴き取りにくさ」の評価尺度を行うことにより，音声コミュニケーションをより的確に評価を行えることがわかった．これらの主観評価と相関の高い客観評価に Speech Transmission Index（STI）があり，この物理指標を回復することで音声明瞭度と聴き取りにくさを回復できると考えた．

従来の雑音・残響除去法の調査を行った結果，雑音除去法は，雑音環境のみでは雑音をよく除去できるが，残響の特性と雑音の特性は全く異なることから雑音除去法で残響を取り除くのは難しい．また，音声明瞭度や聴き取りにくさを回復するような物理指標を使っていないことから，的確に音声明瞭度や聴き取りにくさを回復できないと考える．一方，多くの残響除去法は，事前に室内インパルス応答の測定を必要とするため実用的でない．また，音声に関係する特徴パラメータを回復している処理ではないことから音声明瞭度の回復を効率よく出来るとは考え難い．これらの手法を用いて雑音残響除去を実現するには，雑音除去法と残響除去法を組合わせる手法でしか実現できない．このような逐次的な処理を行う手法が近年提案されているが，雑音成分を減算し，残響成分を線形フィルタリングする処理である．これらの手法は，雑音と残響除去を行う手法であり，音声強調を行う手法ではないことから音声明瞭度回復には限界があると考えられる．

一方，音声伝達性能の評価方法である STI と相互関係にある Modulation Transfer Function（MTF）に基づく雑音・残響除去法が提案されている．この手法は，音声明瞭度を保つのに重要な 20 Hz 以下の信号のエンベロープが，雑音と残響の影響により振幅と位相が影響を受けるため，MTF に基づきエンベロープを回復するという手法である．MTF は，理論的に雑音残響を同時に扱うことができることから，MTF により雑音残響を除去することで STI が回復し，音声明瞭度及び聴き取りにくさの回復処理を行えると考える．

音声伝達の評価方法及び従来の雑音・残響除去法の調査を行った結果，MTF に基づいて雑音残響除去を行い音声明瞭度及び聴き取りにくさを回復するというアプローチが最善であると考え，どのような課題があるのかを調査した．その結果，MTF に基づいた雑音残響除去の実現には，音声のエンベロープ回復だけでは，音声の明瞭度を向上させることができないが，キャリアの回復を行うことで音声の明瞭度が向上することがわかった．キャリア回復の方法として，キャリア再生成処理があるが，雑音残響に頑健な基本周波数推定や音声区間推定などが必要である．今後，これらの推定法の提案を行い，最終的に雑音残響環境において MTF に基づいた音声明瞭度及び聴き取りにくさの回復を実現する．

目 次

図 目 次

第1章　序論

1.1　背景

音声コミュニケーションは，人にとって欠くことのできない情報伝達方法である．近年，経済産業省や文部科学省主導の下，音声によるユニバーサル音声コミュニケーション「いつでも・どこでも・誰とでも・安心安全な音声会話」の実現に向けた様々な研究が取り組まれている．ユニバーサル音声コミュニケーションのツールとしては，遠隔で音声コミュニケーションを行う遠隔会議システムや音声認識技術を用いて翻訳を行い音声合成により他言語にリアルタイムで翻訳するリアルタイム自動通訳器，情報の秘匿性を守る暗号化などがある．

本研究では，ユニバーサル音声コミュニケーションの「どこでも」という環境に焦点をあてる．音声コミュニケーションは，周囲の環境に影響されるという問題がある．例えば，飛行機の機内では騒音，電車のホームでは騒音や残響，ホールや教会では残響や他の人の話し声，レストランや空港では機械音や他の人の話し声などの騒音によって，音声コミュニケーションが阻害されてしまう．これは，実環境において話し声や機械音などの雑音と室特有の残響による影響を受けるためである．この雑音や残響の影響により，音声の明瞭性や了解性が低下し，音声コミュニケーションに影響を与えている（図1.1）．従って，音声の明瞭性や了解性を改善する雑音残響除去法が必要である．

これまでに，雑音や残響を除去する数多くの手法が提案されている．雑音除去法として，適応フィルタなどで雑音を取り除くフィルタを設計するアプローチなどが提案され，残響環境では，マイクロホンアレーを用いるアプローチや調波構造に基づくアプローチなどが提案されてきた．しかし，これらの手法では，雑音と残響の両方を除去できない．雑音と残響の両者を除去する手法として，雑音除去法と残響除去法を逐次的に処理する雑音残響除去法が提案されている．雑音と残響の除去は，処理体系が複雑となるため，ほとんど提案されてこなかったが，近年，雑音残響環境下で音声認識の性能向上や音質改善を目的として提案された．しかし，この手法は知覚ベース（明瞭性や了解性）でのアプローチではなく，また，雑音と残響を同時に除去することができないことから，音声の明瞭性を主観的に評価する音声明瞭度を最適に向上させることが難しいと考えられる．従って，雑音と残響を逐次的に除去するのではなく，同時に除去し，円滑な音声コミュニケーションのために音声明瞭度を向上させる手法が必要である．このような手法が提案できると，環境の影響を受けづらいオンデマンド音声アプリケーション（図1.2）が実現可能となる．

図 1.1: 雑音や残響により音声コミュニケーションが阻害される場面

図 1.2: 環境の影響を受け難いオンデマンド音声アプリケーション

1.2 本論文の目的

本研究の目標は雑音残響環境下での円滑な音声コミュニケーションの実現であり，音声明瞭度を回復する雑音残響除去法の提案を目指す．そこで，まず，円滑な音声コミュニケーションを評価するのにどのような評価尺度を用いれば良いのか，これらの評価尺度を回復する物理指標には何があるのかを明らかにする．その結果をふまえ，これまでに提案されてきた雑音・残響除去法の調査することで，どのようなアプローチにより音声明瞭度を回復する雑音残響除去法を実現できるのか調査を行う．

1.3 本論文の構成

本論文は，全6章で構成されている．以下に各章の概要を述べる．第1章では，問題点等を示すために本研究の背景と目的について述べる．第2章では，雑音・残響環境下での音声コミュニケーションを考えた時にどのような主観評価尺度により評価することで的確に評価が行えるのか，また，音声明瞭度を回復するために有効な物理指標は何なのか，を明らかにする．第3章では，2章の結果をふまえて，これまでに提案されている雑音・残響除去法を調べることにより，音声明瞭度を回復する雑音残響除去法のアプローチを考える．第4章では3章の結論より導き出されたMTFの逆フィルタ処理の概念を説明する．そして，MTFの逆フィルタのこれまでに提案されている手法のコンセプトと実例を挙げ，雑音残響環境での処理に発展させていくための課題を示す．第6章では，本論文をまとめ，今後の課題と展望について述べる．

第2章　音声伝達の評価法

本研究では，雑音残響環境での円滑な音声コミュニケーションの実現を最終ゴールとしているため,（i）どのような評価を行うことで円滑な音声コミュニケーションを評価できるのか，また,（ii）どのような評価尺度を用いることで雑音残響環境下においての評価ができるのか,（iii）どのような物理指標を回復することで音声の明瞭性や了解性を回復する雑音残響除去が実現できるのかを明らかにすることを目的とした．(i) を明らかにすることで音声コミュニケーションに必要な評価尺度や特徴を明らかにできる．また,（i）と（ii）より雑音残響環境での音声コミュニケーションを最善な方法で評価を行うことができる.（iii）を明らかにすることでこれまでに提案されている雑音・残響除去法を評価する時に，音声明瞭度を回復するのに最善なアプローチを見つけることができる．

音声コミュニケーションを行うとき，お互いの意思を相手に伝えることが重要である．このような音声伝達の評価方法に関して，建築音響の分野ならびに音声通信の分野などで活発に研究されてきた．建築分野では，音声明瞭度が，空間内での音声伝達性能を主観的に評価する方法として示され，また，この主観評価尺度を Speech Transmission Index（STI）で客観的に評価してきた．これらの評価尺度は，室が雑音や残響の影響によりどの程度音声伝達に影響を及ぼすのかを評価するために提案されてきた．一方，電話での受話音声の品質等の評価のために，音声明瞭度や Mean Opinion Score（MOS）値が用いられてきた．これは，音声伝送路での雑音や歪み等により音声劣化した音声の品質を評価することを目的として提案された．この MOS の客観評価方法として，PESQ が提案された．これらの詳細な説明を次に述べる．

2.1　主観評価尺度

主観評価は，人間の主観に基づいて評価する方法である．音声の主観評価は，室内や伝送路等における，雑音やエコー，残響等によって，劣化した音声を直接人が評価する方法である．客観評価とは異なり人間が直接聴いて行う試験であるので，最善の評価法ではあるが，被験者の体調等のコンディションに影響を受ける等の問題がある．この評価方法として，音声明瞭度や MOS，聴き取りにくさなどがある．音声コミュニケーションを評価する場合には，人と人とのコミュニケーションであることから主観評価で行うことが最善である．注意点として，主観評価では年齢により受聴能力が異なることから，被験者を選ぶときには，年齢や正常な聴力を有しているのかという試験をする必要がある．

2.1.1 MOS

MOS は，ITU-T 勧告 P.800 に規定されるオピニオン評価法によって得られる評価値である．オピニオン評価とは，通話したときに感じる音声品質を定量的に評価する主観評価である．受聴者により 5 段階の採点を行ない，統計的処理がなされたのが MOS 値である．これは，伝送路にける歪みや雑音の影響を受けた受話音声の品質を評価する．雑音等の評価方法ではあるが，室内音響特性等を評価するために提案されたものではなく，また，評価量が多い為に被験者への負担が大きいという問題がある [1]．しかし，電話などの音声の品質を評価するものであり目的は異なるものの，音声の品質を評価する評価尺度として重要である．

2.1.2 音声明瞭度

次に，話者の意志が相手に正確に伝わっているのかどうかを評価するのに重要な評価尺度である音声明瞭度について説明を行う．一般的に述べられている音声明瞭度は，単音節に対する聴取者側の正答率である明瞭度と，言語として意味のある単語あるいは短文に対する聴取者側の正答率である了解度に分けられる．

明瞭度試験は発話者と聴取者によって行われる．明瞭度試験には，単音節明瞭度試験，2 連音節明瞭度試験，3 連音節試験がある．単音節明瞭度は，試験音に文字通り単音節を用いて評価する方法である．音節の数は直音，濁音，半濁音，拗音を合計して 100 音節である．この 100 音節を 1 行に 10 音節で 10 行にランダムに配置し，1 行ずつ約 2 秒おきに 1 音節を発声する．次に，2 連音節明瞭度試験であるが，この方法も，文字通り 2 音節による試験である．これは，二つの音節を一組にして，ある一定の時間間隔で発声し，単音節明瞭度試験と同様に聴き取れた順序に記録をさせて行う．一組の二つの音節は単音節試験に用いた 100 音節の中からランダムに選び，組合わせたものが無意味になる必要がある．2 連音節明瞭度では単音節明瞭度より明瞭度は低い値を示す．これは，前の音節の母音が後の音節の子音をマスクすることにより，2 音目の音節が違聴され易いためと考えられる．3 連音節明瞭度試験は，三つの音節を連続で並べた，無意味の 3 連音を用いた試験である．これまでの報告より，単音節による試験よりも部屋の音響特性に良く合うとの報告がなされている．飯田らの報告によると，明瞭度試験の結果と部屋の音響特性が比例的に一致しない現象がしばしば観測されている [2]．特に単音節明瞭度試験でこの傾向が顕著である．残響の影響を評価する場合には，2 連音節明瞭度試験や 3 連音節明瞭度試験を用いる方が有効であると言えるが，こちらの試験は，無意味音声による試験である．実際の音声コミュニケーションを考えると，普段の生活においては意味のある音声しか用いていない．一方で，雑音環境を考えた時には，単音節明瞭度試験では，訓練された受聴者でも困難な作業であるとされている．これが，2 連や 3 連音節明瞭度試験では，組合わせが多いことから，試験自体が難しい．これらのことを考えると，明瞭度試験が音声コミュニケーションに適した評価方法であるとは考え難い．

空間で話を聴いたり会話を行った場合，その内容をどれだけ理解したかが，音声コミュニケーションでは重要であり，その評価方法として了解度試験がある．了解度は，発声・伝送された章句又は単語の正しく聴取された割合である．明瞭度との違いは，有意味の単語で文章フレーズであることである．単語の場合は単語了解度，文章の場合は文章了解度である．単語了解度は，発声された単語の正しく聴き取られた単語の割合である．文章了解度は質問又は命令文を読み上げ，正しい解答が得られた文章数の割合で表す．この，単語了解度試験には，2 音節単語，3 音節単語等の音表を用いる．また，質問文や命令文を用いたものと普通文を用いる方法とがある．この了解度試験では，単語の親密度により了解度の正答率が変わることが報告されている．坂本らは，臨床現場への応用を目的とした，4 段階の親密度に分類した単語 [3] から FW03 という試験用音声データベースを提案している．この試験用音声を用いた単語了解度試験の結果，単語の新密度が高いほど正答率が高くなることが明らかにされている [4]．この試験用音声データベースの単音節では，音声レベルが等価騒音レベルと等しくなるように校正されているため，聴感レベ ルが単音節ごとに異なるという問題があった．そこで，長谷らは，この単音節音声のらラウドネス校正を行った [5]．その結果，補正量が大きく，FW03 の単音節音声が不十分であることを示した．また，近藤らは，FW03 を簡略化し，日常の音声聴取能力を測定するための試験用音声データセット（FW07）を提案した [6]．こちらの方が，試験用単語リスト間の了解度の差が小さく，より精度の高いものである．また，建築音響の分野では，佐藤らが雑音・残響環境においての単語親密度と単語了解度の関係の検討を行っている [7]．この結果，親密度が異なると音場の単語了解度に影響を与える場合があることから，雑音・残響環境での単語了解度の評価を行う際は，単語親密度の統制が必要である．従って，どのような親密度の単語を使うのかというのは，どのような評価を行うのかという目的別に検討が必要である．実環境は，雑音と残響が共存することから音声コミュニケーションを評価する場合には，親密度別に評価を行うことが最善と考えられる．従って，単語了解度試験を親密度を統制しながら行うことが，音声コミュニケーションの評価において重要である．

ここまで，日本語についての音声明瞭度について述べた．他言語でも様々な方法が提案されている．語頭または語尾の 1 音素のみ異なるミニマルペア 6 単語を 1 セット，計 50 個からなる単語リストを用意し，そのうちの 1 単語を聴かせ，セット内の 6 単語から選らばせる Modified Rhyme Test (MRT) が House *et al.* によって提案している．ミニマルペアとは，語の意味を弁別する最小の単位である音素の範囲を認定するために用いられる言語形式の二つの単語のことをいう．この評価試験は，6 単語から選ぶため，比較的簡単である．更に，語頭のみ異なるミニマルペア 2 単語を 1 セット，計 96 セットからなる，Diagnostic Rhyme Test (DRT) が Voiter によって提案された．DRT は，各セットで対比させる語頭の音素が音素特徴空間内の特定の要素のみ異なるように吟味されている．そのため，音素特徴別の了解度を評価することができる．MRT，DRT の両者は米国において標準化されている [8]．今後，他言語による音声コミュニケーションを評価する場合には，他言語の評価方法を更に調べる必要があるが，現時点では日本語を想定しているので，以上の代表

的な評価方法の紹介に留める．

2.1.3　聴き取りにくさ

先に述べた音声明瞭度の了解度であるが，親密度が高い単語による了解度試験を行う場合には，伝達経路による劣化の影響が評価結果に現れにくくなることがわかっている．これは，親密度が高いと単語の類推が可能になるためであるが，本来ならば了解度に多少の差があると考えられる．この場合，単語を認識できても「聴き取りやすい」場合と「聴き取りにくい」場合とがある．要するに，音声伝達性能として最善ではないにも関わらず，単語了解度試験では最高な結果を導いてしまうのである．これは，音声コミュニケーションを考えた時の最善な評価方法を考えると，単語了解度試験のみでは不十分であることを示唆している．そこで，佐藤らは正答率で求める単語了解度試験とは異なる評価方法として,「聴き取りやすさ」という主観評価尺度を提案した．これは，了解度試験の評価結果には差が出ない場合でも，聴感的に差があると感じる音声伝達性能の違いを評価することができる評価尺度である [9]．しかし,「聴き取りやすさ」は評価結果にばらつきがあり，また，音声明瞭度との関係も明確にされなかった．そこで,「聴き取りにくさ」が森本らによって提案された [10, 11, 12]．この評価方法は,「聴き取りにくい」と判断された割合で音声伝達性能を評価する方法で，高い親密度の単語においても，聴き取りにくさの評価結果では差が生じることがわかっている．

この主観評価の調査を行った結果，円滑な音声コミュニケーションを評価するためには，単語了解度又は文章了解度による評価を行うと同時に聴き取りにくさの評価を行う必要があることがわかった．これにより音声コミュニケーションを的確に評価できることがわかった．

2.2　客観評価尺度

客観評価は，主観評価を客観的に評価するために考えられた評価方法である．客観評価は，人を用いて評価する必要がなく，計算機等により求めることができるので，主観評価に比べ容易に音声の伝達性能を評価できる．しかし，客観評価は物理指標を用いて主観評価の結果を予測していることから，主観評価結果との誤差が多少生じることが問題である．これまでの多くの研究における音声伝達性能の評価は，はじめは客観評価を行い傾向を掴み，最終的には主観評価により評価がなされている．この主観評価を予測する際に用いる物理指標は，音声回復等に使うことができるため，直接物理指標を回復することで，精度よく主観評価を回復することができる．そのため，音声明瞭度及び聴き取りにくさを最善な方法で回復することを考え，音声明瞭度や聴き取りにくさとの高い相関を持つ物理指標についても調査する．

2.2.1 明瞭度指数（AI）

French&Steinberg による電話受聴や Kryter によるスピーカー受聴を対象とする明瞭度指数という客観評価方法 [13] がある．この手法は，元々音声通信における伝送路での雑音等の影響による明瞭性を評価するためのものであるので，残響音場での適用ができないという問題がある．そこで，Latam が反射音と暗騒音を有害として S/N から明瞭度を求める手法を提案した．

2.2.2 Perceptual Evaluation of Speech Quality（PESQ）

主観評価値である MOS 値を客観的に評価する尺度である Perceptual Evaluation of Speech Quality（PESQ）は，ITU-T P.862 として勧告されている．近年，Voice over IP（VoIP）や携帯電話での音声品質の評価によく用いられている．PESQ の特徴として，VoIP などのパケット損失等の影響により発生する歪みを扱えることである [14]．しかし，PESQ では背景雑音として雑音が重畳み込みされている場合には，MOS の評価特性が反映されていないことも報告されている [15]．また，VoIP での MOS と PESQ の相関の調査も行われており，パケット損失等の影響を受けにくく相関が高いことも報告されている [16]．また，同様に MOS を客観的に評価する評価尺度として，PAMS や ITU-T P.861 として勧告されている Perceptual Speech Quality Measurement（PSQM）があるが，PESQ の方が音声品質評価として優れているためよく用いられている．PESQ は，音声品質を客観的に評価するのに有効な評価方法である．

2.2.3 D 値

D 値は，室において初期エネルギーが全エネルギーに占める割合として計算される音響指標である．音響品質に対応する物理指標である．50 ms 以上の遅延成分を含まなければ，D 値は 100%となる．単音節了解度と D 値には良好な相関関係が得られている [17]．この評価尺度は，残響の影響を評価するのによく用いられる尺度であるが，雑音がある環境で評価できるかどうかはこれまで検討されていないため，雑音残響環境において評価尺度として用いる際には検討を行う必要がある．

2.2.4 Speech Transmission Index（STI）

音声明瞭度や聴き取りにくさとの相関が高い物理指標として Speech Transmission Index (STI) がある．この評価尺度は，Houtgast&Steeneken によって音声明瞭度予測理論 [18, 19, 20] として提唱された．評価尺度は，音場内では音声波形の時間包絡（エンベロープ）が雑音や残響の影響により低下することに着目している．この評価尺度は，理論的に明快であり，雑音と残響の両方が同時に存在する音場を評価することができる．STI は建築音響

における現場での音声伝達性能の評価に用いられている．戸井田の報告に基づくと明瞭度や了解度との相関関係には，それぞれの適応限界が原因で，相関関係が高くない時があることが報告されている [21]．STI の測定方法については中島が詳細を解説しており [22]，STI は MTF から計算により求めることができる．一方で，佐藤らの「聴き取りにくさ」とは相関関係が高い [10, 11] ことから，STI は音声の伝達を評価するのに重要な物理指標であると考える．

従って，雑音残響環境下で主観評価尺度である音声明瞭度及び聴き取りにくさを回復する物理指標として STI がある．この STI を回復することにより円滑な音声コミュニケーションを実現できると考える．

第3章　雑音・残響除去法

これまでに，音声通信や音声認識などの音声アプリケーションにおいて耐雑音性を向上させるために，様々な特徴等に基づき音声処理技術 [23, 24] を用いて雑音・残響除去の取り組みが行われてきた．雑音除去法は，振幅スペクトルやパワースペクトルなどにおいて適応フィルタなどが用いられ，残響除去法は，逆フィルタ処理を中心として発展してきた．そして，単一マイクロホンを用いた手法だけでなく，複数のマイクロホンを用いることで室内の音響特性を推定するマイクロホンアレー技術 [25] を用いた手法が多く提案されてきている．近年，雑音残響除去に対する取り組みも始められてきており，実環境により近い処理が検討されはじめている．

本章では，円滑な音声コミュニケーションを実現するために主観評価である音声明瞭度や聴き取りにくさを改善するような手法が，従来の雑音・残響除去法にはないのかどうか，また，どのようなアプローチにより音声回復を行っているのかを調査する．そして，物理指標である STI を回復できるような手法があれば，直接音声明瞭度及び聴き取りにくさを回復できることから最善の手法ではないかと考える．

3.1　雑音除去

3.1.1　Spectral Subtraction 法

Spectral Subtraction（SS）法は Boll によって，音声圧縮や音声認識，音声認証などの音声処理装置の精度向上を目的として提案された [26]．この手法は，観測された信号の振幅スペクトルから雑音の振幅スペクトルの推定平均値を減算することで，原音声の振幅スペクトルを得る方法である．Boll らの手法では，雑音の振幅スペクトルの推定平均値を音声の無音区間から推定を行う．この手法は，マイクロホン 1 本から利用でき，処理が簡単かつ良い回復結果が得られることから，現在でもよく使われている．しかし，この手法は，雑音の推定誤差などの原因により回復音声にミュージカルノイズが生じ，雑音に定常雑音を想定しているために雑音の時間変化に弱いという問題がある．ミュージカルノイズを取り除く手法として，異なるサブトラクション係数で二つの SS の処理を行い，その差から音声成分を残しミュージカルノイズを取り除く手法 [27] が提案されているが，計算量が多いという問題がある．これらの手法でもミュージカルノイズは，軽減する程度であり完全に取り除くことができていないため，聴き取りにくさが残ると考えられる．また，雑音の時間変化に頑健な手法として，マイクロホンアレーを用いる手法などが提案されている

[28]．この手法は，マイクロホンアレーを用いて，信号の到達時間差の推定を行い，短時間フレームごとに雑音を推定するため，非定常雑音及び突発性雑音を除去できるSS法であると言える．しかし，位相情報については処理を施していないことからミュージカルノイズが生じる．これらのSS法においては，音声強調を目的とした手法ではなく，雑音残響環境で音声明瞭度や聴き取りにくさを回復するような物理指標は使われていないことから，音声明瞭度や聴き取りにくさを的確に回復することができる手法ではないと考える．

3.1.2 Adaptive Noise Canceling（ANC）法

他のアプローチとしてAdaputive Noise Cancelling（ANC）は，LMSを用いて適応フィルタの係数を推定することでフィルタを設計し，これを用いて雑音が付加された信号から雑音を取り除くという概念で，Samburによって提案されている[29]．ここでは，雑音を白色雑音としており，定常雑音にしか対応できていないという問題がある．また，適応フィルタの係数の推定精度を向上させるための改良法 [30, 31] などが提案されている．しかし，音声強調を目的とした手法ではなく，音声認識などのための雑音除去を目的としており，雑音残響環境で音声明瞭度や聴き取りにくさを回復するような物理指標は使っていないことから，的確に音声明瞭度や聴き取りにくさを的確に回復することができない手法であると考える．

3.1.3 MMSE-STSA

MMSE (Minimum Mean Square Error)-STSA(Short-Time Spectral Amplitude) は，音声のフーリエ係数をガウス分布に従うと仮定し，推定短時間振幅スペクトルの平均2乗誤差を最小にする方法 [32] で，Ephraim&Malahによって提案された．この手法は，音声強調を目的とした手法である．手順は，短時間フーリエ分析を行い，雑音音声のフーリエ変換を行うことで振幅スペクトルと位相を得る．劣化音声の振幅スペクトルにスペクトルゲインを乗算することで強調された音声の振幅スペクトルが得られ，短時間フーリエ合成で強調された音声の振幅スペクトルと位相情報を補正していない雑音音声の位相の積に対して逆フーリエ変換を求める．位相を的確に補正することにより音声ミュージカルノイズを発生しないが， 非音声区間から雑音推定を行っているため，非定常雑音に対して弱く音声品質の低下が避けられない．強調音声の歪みを低減する手法として加藤らの雑音推定の時に重み付けを行うMMSE-STSA法などが提案されている [33]．しかし，この手法は音声強調を行っているものの，雑音残響環境で音声明瞭度や聴き取りにくさを回復するような物理指標は使っていないことから，的確に音声明瞭度や聴き取りにくさを回復できる手法ではないと考える．

3.1.4 Winner filter 法

Winner によって提案された Winner filtering を音声に適用した手法は，最適フィルタを周波数領域での平均 2 乗誤差（MSE）の最小化により導出する手法 [34] として Lim&Oppenheim によって提案された．最小平均 2 乗誤差を振幅スペクトルで取る点では MMSE-STSA と共通する点もある．この手法は，クリーンな音声のパワースペクトルと雑音のパワースペクトルから Winner filter は設計する．LPC 分析を用いてクリーンな音声のパワーエンベロープ推定を行い，Winner filter によって音声強調された音声に対して LPC 分析を行い，フィルタの再設計を行い，音声強調を繰り返す方法が取られている．この方法では，繰り返し処理を行うことで推定音声はクリーンな音声に近づくものの，反復回数が多いとスペクトル歪みが生じる問題があり，反復回数の決定が難しいことが知られている．ミュージカルノイズは発生しない．音声強調を行っているものの，雑音残響環境で音声明瞭度や聴き取りにくさを回復するような物理指標は使っていないことから，的確に音声明瞭度や聴き取りにくさを回復できる手法ではないと考える．

3.1.5 最大尤度に基づくフィルタ法

この手法は，最大尤度法からパラメータを推定しフィルタを設計する手法であり，McAulay&Malpass によって提案された [35]．この手法は，ウィナーフィルタ同様に，パワースペクトル上での減算処理を行う．こちらでは，評価実験は行われていないが，雑音を軽減することができている．この手法は，雑音除去であり音声強調を行うものでなく，雑音残響環境で音声明瞭度や聴き取りにくさを回復するような物理指標は使っていないことから，的確に音声明瞭度や聴き取りにくさを回復できる手法ではないと考える．

3.1.6 RASTA 法

RelAtive SpecTrAl processing: RASTA は，変調スペクトルの約 1-12 Hz の変調周波数のみを通過させるフィルタを用いた雑音除去法で，Hermansky &Morgan によって提案された [36]．RASTA は，変調スペクトルの重要な周波数成分のみを通過させることで，音声認識性能を向上させる手法である．RASTA での重要な点は，音声認識における重要な特徴がどの変調スペクトル成分に存在しているかであり，これに基づきなフィルタを設計することである．Hermansky&Morgan の手法においても雑音に頑健な手法となり，RASTA の先駆け的手法となった．更に，音声認識に重要な変調周波数を Kanedera らが調べ [37]，2 Hz 以下と 16 Hz 以下の変調周波数成分が音声認識性能を低下させることを示し，RASTA のフィルタ形状の再設計を行った．しかし，この手法は，雑音残響環境で音声明瞭度や聴き取りにくさを回復するような物理指標は使っていないことから，的確に音声明瞭度や聴き取りにくさを回復できる手法ではないと考える．

3.2 残響除去

3.2.1 最小位相逆フィルタ法

この手法は，Neely&Allen によって提案された残響除去法 [38] である．この手法は，室内音場が最小位相特性を有している時に室内インパルス応答の逆フィルタをかけることにより残響除去できる．しかし，実際の室内音場では，最小位相特性であることはほとんどなく，非最小位相特性であることが多くを占める．また，事前に室内インパルス応答を測定しておく必要があり，時間変化による環境の変化に追従できないことから，回復精度を常に高く保つことはできない．また，雑音残響環境で音声明瞭度や聴き取りにくさを回復するような物理指標は使っていないことから，的確に音声明瞭度や聴き取りにくさを回復できる手法ではないと考える．

3.2.2 MINT 法

Miyoshi&Kaneda は，音源から受音点までの室内インパルス応答を事前に測定しておき，その逆フィルタをマイクロホンに畳み込む，音場逆フィルタ処理（Multiple-input/output inversee theorem: MINT）[39] を提案した．この手法は，音場を 1 入力多出力の線形システムでモデル化し，単一音源から複数マイクロホンまでの多チャンネル線形システムの逆フィルタ問題として定式化を行っている．非最小位相特性であっても残響除去を可能とした．しかし，MINT 法では事前に室内インパルス応答を測定しておく必要があり，最小位相逆フィルタ処理同様にインパルス応答の時間変化による回復精度の低下は免れない．また，残響の影響が小さい環境においては，あまり良い結果が得られない．MINT 法の改良法として，事前にインパルス応答を測定しなくても残響除去可能な Semi-blind MINT 法 [40] がある．この手法は，マイクロホンに一番近いマイクロホンを既知とし，各入力マイクロホン間の相関行列からインパルス応答を推定して逆フィルタ処理を行っている．また，音声信号は有色信号であるため，MINT 法では性能が低下する問題があった．残響の影響を受けた音声に音声の平均スペクトルの逆特性をもつ白色化フィルタを用いることで，この問題の解決に取り組んでいる．その結果，室内インパルス応答を事前に測定せずに，残響除去が実現されている．ただし，音源に近いマイクロホンを既知としており，この情報がなければうまく残響除去を行えことから，完全なブラインド処理ではない．また，Semi-blind MINT 法で取りきれなかった残響を SS 法を組合わせることにより取り除く手法 [40, 41] なども提案されている．しかしながら，MINT 法では，ブラインド処理を実現できていないように見受けられ，複数のマイクロホンを用いることからシステムが大掛りになってしまうという問題が残る．また，雑音残響環境で音声明瞭度や聴き取りにくさを回復するような物理指標は使っていないことから，的確に音声明瞭度や聴き取りにくさを回復できる手法ではないと考える．

3.2.3 帯域分割逆フィルタ処理

MINT 法と同様に複数のマイクロホンと帯域分割処理理を用いた手法を Wnag& Itakura が提案している [42]．この手法は，各マイクロホンの入力に対して，それぞれの帯域毎に最小 2 乗誤差を計算し，各帯域毎に最適なマイクロホンの入力を選び，各帯域毎に逆フィルタ処理を行い，各帯域の回復信号を合成することにより音源波形を復元する方法である．広帯域の音声を回復することができる．雑音残響環境で音声明瞭度や聴き取りにくさを回復するような物理指標は使っていないことから，的確に音声明瞭度や聴き取りにくさを回復できる手法ではないと考える．

3.2.4 調波構造に基づく処理

音声の調波構造に着目した，Hermonic-based dEReverBeration（HERB）が Nakatani らによって提案されている [43]．この手法は，残響を含む音声信号の調波構造を回復する逆フィルタが，近似的に室内伝達関数の逆フィルタになることを用いて，ブラインドでの残響除去を実現している．単一マイクロホンで残響除去できるが，残響時間 1.0 s 程度までしかその有用性は得られていない．この改良法として，逆フィルタの設計に平均伝達関数（ATF）や最小平均 2 乗誤差（MMSE）を用いた HERB の改良法が提案された [44]．従来の HERB では，残響時間 1.0 s の時の音声認識率には課題を抱えていたものの，改良法の HERB では，90 %以上の音声認識率が得られている．しかし，音声品質等の評価がなされていない為，どの程度音声が回復していえるのかわからない．また，雑音残響環境で音声明瞭度や聴き取りにくさを回復するような物理指標は使っていないことから，的確に音声明瞭度や聴き取りにくさを回復できる手法ではないと考える．

3.3 雑音・残響除去

3.3.1 音情報分析（CASA）のアプローチ

Bregman は，カクテルパーティ効果に代表される人間の聴覚による音の分離である聴覚情景解析（Auditory Scene Analysis）において，聴覚が利用している制約条件を心理的規則として述べた．これらの問題を計算モデルとして実現する試みが，音環境解析（Computational Auditory Scene Analysis）のアプローチである [45]．CASA で重要となるのが，音響ストリーム分離であり，混合音から個々の音を分離するための統一的な計算モデルが求められる．分離を行うためには，音クラス，各音クラス属性，それら関係が階層的に定義される．最上階では，音源グループに分類され音源が，音声，音楽機械音などに分類される．そして音声クラスには，調波構造（周波数成分，フォルマント），音色，ラウドネス，変調，パワースペクトラム，LPC ケプストラムがある．このように音響ストリームは，属性を入力である混合音から抽出することであると説明さている．これに基づいて，

音楽からの音声の抽出 [46] が試みられている．この考えに基づけば，雑音環境下での音声は，混合音と見なせ，CASA のアプローチから音声を抽出することも可能であるため，雑音から純音を抽出する手法が [47] 提案され，更に音声の抽出へと発展させるべく調波複合音を抽出する提案がなされている [48, 49]．この手法は，音声などの特徴には基づいているものの，雑音残響環境で音声明瞭度や聴き取りにくさを回復するような物理指標は使っていないことから，的確に音声明瞭度や聴き取りにくさを回復できないと考える．

3.3.2 独立成分分析（ICA）のアプローチ

CASA とは異なり，独立成分分析（Independent Component Analysis: ICA）に基づくブラインド音源分離（Blind Source Separation: BSS）が提案されている．典型的なアルゴリズムでは，複雑さを減らすための前処理として白色化や中心化，次元削減などの処理を行う．また，ブラインド音源分離における ICA では，時間領域において FIR フィルタを推定する時間領域 ICA と周波数領域で周波数毎のフィルタを推定する周波数領域 ICA とがある [50]．周波数領域 ICA を用いて雑音抑圧を行った手法 [51] がある．この手法は，ICA を用いてブラインド信号分離を行うことで雑音除去を行う手法で，ノンブラインドな信号分離の精度と同等の精度が得られている．また，マイクロホンアレーを用いて ICA を行う手法 [52] がある．この手法は，ICA と SS 法を用いてパワースペクトル上で処理を行い，音声認識の対雑音性を向上させることに特化した手法となっている．また，耐残響についての検討もなされている [53]．しかし，三つの手法を組み合わせた手法で，処理が複雑化しているなどの問題がある．また，雑音残響環境で音声明瞭度や聴き取りにくさを回復するような物理指標は使っていないことから，的確に音声明瞭度や聴き取りにくさを回復できないと考える．

3.3.3 MTF に基づく逆フィルタ処理

MTF の概念が提案されてから，事前に室内のインパルス応答を測定を行わないブラインドな残響除去法が提案されてきた．音声明瞭度の客観評価尺度である STI は MTF から計算されることから，この手法は音声明瞭度や聴き取りにくさを直接改善するような手法であると考えられる．

Langhans&Strube は，パワーエンベロープを STFT 上の変調スペクトル上で回復する方法を提案した [54]．この手法は，パワーエンベロープの対数を取り，逆フィルタ処理を行っている．その結果，雑音環境と残響環境で従来の方法より音声明瞭度がわずかながらに向上したことが報告された．また，Avendano&Hermansky は，変調周波数 8 Hz 以上の強調を抑圧する MTF と高域通過フィルタを組合わせた逆フィルタ処理を提案した [55]．これにより，変調スペクトル上で原音声に近づくような回復処理が得られている．これらは，変調スペクトル上での回復処理に着目した手法である．しかし，音声明瞭度の回復は得られなかった．一方で，Moujopoulos&Hammond は，MTF の逆フィルタを設計す

る際の近似的なインパルス応答の推定方法について検討を行い，エンベロープの回復を行った [56]．この手法においても音声明瞭度の回復は得られていない．広林らは，音声信号の時間包絡（エンベロープ）のパワーを取ったパワーエンベロープに着目し，回復処理を行った [57, 58]．しかし，これまでに提案されてきた手法では，MTF の逆フィルタを設計する際に必要なパラメータをブラインド推定できておらず，ノンブラインドな手法であるために実用的でないという問題点があった．そこで，Unoki らは，ブラインド残響除去を行うためにパワーエンベロープ抽出法，残響時間の推定法，振幅項の推定法を提案し [59]，フィルタバンクを用いて帯域分割した手法 [59] を提案し，ブラインドでのパワーエンベロープの回復を実現した．しかし，回復パワーエンベロープと残響の影響を受けたキャリアを合成して回復音声を求めたために，異音が生じた [60]．これまでに提案されてきた MTF に基づく逆フィルタ法の音声明瞭度が回復しなかったのもそこに起因していると考えられる．そのため，分析合成器を用いてキャリアの再生成処理を行った手法が提案された [61]．その結果，明瞭度等の回復が得られら [62]．一方，音声認識を目的として，パワーエンベロープ逆フィルタ処理の改良法 [63] が提案されているが，あまり精度の向上に至っていないように思える．

雑音環境でのパワーエンベロープ回復につていは Yamasaki&Unoki [64] によって提案されており，音声認識に対する同様の手法 [65] も提案されている．雑音残響環境でのパワーエンベロープ回復について Unoki&Yamasaki [66] によって提案されているが，キャリア回復が行われていないことなどから人工的な異音が生じるなどの問題がある．しかし，雑音と残響を同時に除去でき，音声明瞭度を回復できる手法であることから有用な手法であると考える．

3.3.4 雑音除去と残響除去を組み合わせた雑音残響除去

Kinoshita らは，マイクロホンアレーを用いた雑音残響除去法を提案している [67]．この手法は，SS 法により雑音を抑圧し，多段線形予測を用いることでパワースペクトル上で残響を除去する方法である．この手法は，雑音残響を除去し音声認識率を向上させることができる手法ではあるが，高域においては雑音が残っているように見受けられ，音声コミュニケーションを目的とした場合には，有効な手法であるかどうかはわからない．

吉岡らは，雑音除去にはウィナーフィルタなどの非線形フィルタと残響除去には線形フィルタを用いる雑音残響除去法を提案した [68, 69]．それぞれのフィルタに用いるパラメータを単一の最尤推定から推定し，フィルタを設計して，雑音除去，残響除去の順に逐次処理を行っている．

しかし，これらの手法は単純に雑音の成分を減算し，残響成分を逆フィルタする手法であり，音声に重要な特徴パラメータを回復するような処理は取られていない．そのため，音声明瞭度の回復には限界があるものと考えられる．

3.4 まとめ

雑音除去法は，様々なアプローチから提案され，一定の回復精度が得られている．また，多く用いられている手法としては，SS 法が簡単かつ回復精度が高いことから一般的に良く利用されており，問題となるミュージカルノイズの低減法も多く提案されていることから雑音除去法として優れた手法である．しかし，全く特性のことなる残響の存在する環境では，このような減算処理等では抑圧できるとは到底考え難い．また，RASTA は，音声認識のために重要でない変調スペクトルを削ぎ落とす処理であり，音声強調処理ではないこと，そして，雑音残響環境で音声明瞭度や聴き取りにくさを回復するような物理指標は使っていないことから，的確に音声明瞭度や聴き取りにくさを回復できないと考える．しかし，RASTA における音声認識に重要な変調周波数成分は，音声の明瞭性とも関係があり，MTF に基づく雑音・残響除去法に生かすことが可能である．一方で，残響環境においては，MINT 法に代表されるマイクロホンを複数用いる手法により，非最小位相特性においても残響除去が行われてきた．また，HERB は，残響を単一マイクロホンによりブラインド残響除去を行えている点で有用であるが，音声認識を目的として回復が行われており，音声回復をするような手法ではない．音情報解析のアプローチや独立成分分析のアプローチでは雑音や残響を除去することができるが，音声を抜き出すことなどにはまだ課題が残っているようにも見える．また，音声の特徴を用いているが，音声明瞭度や聴き取りにくさを回復するような物理指標は使っていない．また，雑音残響環境においては取り組みがないようで，課題が残っている．近年，雑音と残響の両方を除去する手法が提案されはじめているが，どちらも音声認識を目的としており，雑音除去と残響除去を組合わせた逐次的な処理であり，音声強調を行う手法でないことから音声コミュニケーションにおける雑音残響除去法として有用であるようには考え難い．MTF に基づく手法は，雑音残響環境で音声明瞭度や聴き取りにくさを回復する物理指標の STI と相互関係にある MTF を使っていることから，雑音残響環境において的確に音声明瞭度や聴き取りにくさを回復できると考える．しかし，この手法においても，雑音残響環境において音声明瞭度を回復するには課題が多く残っているようにも見受けられるが，この手法は音声明瞭度との関係が高いことから音声明瞭度を飛躍的に回復できるものと考えられる．従って，本研究において MTF に基づく逆フィルタ処理を用いて音声明瞭度及び聴き取りにくさを回復し，円滑な音声コミュニケーションの実現を目指す．

雑音・残響除去法の調査結果を一覧表として付録に示す．評価項目は，雑音残響環境下で円滑な音声コミュニケーションを簡易的に行うことためには，どのような事が満たされていることが重要なのかを考えて決めた．音声明瞭度や了解度，聴き取りにくさを回復する手法である必要があるため，知覚ベースの回復処理の手法なのかが重要である．対応環境としては，雑音残響環境を目指しているので，雑音残響環境に対応できる手法が ◯ となっている．また，マイクロホンアレーの技術を用いるとシステムが大きくなるなどの問題があることから，単一マイクロホンを想定している手法を ◯ としている．実用的であるためにはブラインド処理である必要がある．そして，音声コミュニケーションには音声明瞭度の了解度の評価が絶対必要であると考え，了解度試験が行われて回復精度が良かっ

た場合に ◯ とする．著者らは，全ての評価項目で ◯ となるような雑音残響除去法の提案を目指す．

第4章　変調伝達関数（MTF）

4.1　MTFの概念

変調伝達関数（Modulation Transfer Function: MTF）は，MTFの逆フィルタ処理の基礎となる概念である．この概念は，Houtgast&Steenekenによって音声明瞭度予測理論[18, 19, 20]として提案された．MTFの概念では，室内を伝達系と見たときの入力・出力の強度変化に着目し，この強度変化をMTFと定義している．室内音響における入出力の強度変化を余弦波を用いて定式化すると次のように示すことができる．

$$\text{Input} = \overline{I_i^2}(1 + \cos(2\pi f_m t)) \tag{4.1}$$

$$\text{Output} = \overline{I_o^2}\{1 + m(f_m)\cos(2\pi f_m(t-\theta))\} \tag{4.2}$$

$\overline{I_i^2}$は入力の強度，$\overline{I_o^2}$は出力の強度であり，f_mは変調周波数，θは位相情報である．

例えば，入力パワーエンベロープが100 %振幅変調（変調度が1）であるとき，室内の残響の影響を受けることで，出力パワーエンベロープの変調度が$m(f_m)$だけ（1未満）減少する．残響時間とエンベロープの周波数の関数として変調度が変化することから，この関係がMTFと呼ばれる所以である．

次に，変調度である$m(f_m)$の導出のために，信号のエンベロープ（時間包絡線）のパワー（2乗）を取ったパワーエンベロープを定義する．入力信号$x(t)$のパワーエンベロープと出力信号$y(t)$のパワーエンベロープを

$$e_x^2(t) = \overline{e_x^2}(1 + \cos(2\pi f_m t)) \tag{4.3}$$

$$e_y^2(t) = \overline{e_y^2}(1 + m(f_m)\cos(2\pi f_m t)) \tag{4.4}$$

と定義する．ここでは，簡単のため，式4.2の$\theta = 0$とした．入力信号と出力信号を一般化すると，

$$x(t) = e_x(t)n_x(t) \tag{4.5}$$

$$y(t) = e_y(t)n_y(t) \tag{4.6}$$

と表現できる．$e_x(t)$は入力信号のエンベロープ，$e_y(t)$は出力信号のエンベロープ，$n_x(t)$及び$n_y(t)$は白色ガウス雑音をの特性を有するランダム変数であり，音信号を想定するとキャリアc_x，c_yに該当する．ここでは，白色雑音を用いるので，

$$< n(t)n(\tau) >= \delta(t-\tau) \tag{4.7}$$

という特性がある．ここで，$< \cdot >$は集合平均を表す．

4.2 雑音・残響環境でのMTF

雑音・残響環境でのMTFを先ほどの概念を用いて説明する．まず，雑音環境でのMTFについて説明する．雑音環境での出力$y(t)$は，入力信号$x(t)$と$w(t)$の加算で求まる．

$$y(t) = x(t) + w(t) \tag{4.8}$$

雑音環境でのMTFは，式4.3に雑音が加算された観測パワーエンベロープは次式で表現される．

$$\begin{aligned} e_{yN}^2(t) &= e_x^2(t) + e_n^2(t) \\ &= \overline{e_x^2}(1 + \cos(2\pi f_m t)) + e_n^2(t) \\ &= \left(\overline{e_x^2} + \overline{e_n^2}\right) 1 + m_N(f_m)\cos(2\pi f_m t) \end{aligned} \tag{4.9}$$

ただし，雑音パワーエンベロープ$\overline{e_n^2} = \frac{1}{T}\int_0^T e_n^2(t)dt$，$T$は信号の時間長である．ここで，$e_n^2(t)$を時間領域一定であると仮定すると，雑音環境におけるMTFは，次式で表現できる．

$$m_N(f_m) = \frac{\overline{e_x^2}}{\overline{e_x^2} + \overline{e_n^2}} = \frac{1}{1 + 10^{-(SNR)/10}} \tag{4.10}$$

ただし，$SNR = 10\log_{10}(\overline{e_x^2}/\overline{e_y^2})$ dBである．式4.10の特性を図4.1に示す．雑音環境でのMTFは，変調周波数f_mには依存せずに，SNRの関数として減少する．

次に，残響環境でのMTFについて説明する．残響環境での出力信号$y(t)$は，入力信号$x(t)$と室内インパルス応答$h(t)$の畳み込みから得られる．

$$y(t) = \int_0^\infty h(\tau)x(t-\tau)d\tau \tag{4.11}$$

そして，複素表現のMTFは次式で表現できる．

$$m_R(f_m) = \frac{|\int_0^\infty h^2(t)\exp(-j2\pi f_m t)dt|}{\infty_0^\infty h^2(t)dt} \tag{4.12}$$

インパルス応答$h(t)$を，室内音響特性の統計的近似として知られているSchroederの室内インパルス応答 (RIR) [70]を用いて定義する．

$$h(t) = e_h(t)n_h(t) = a\exp\left(-\frac{6.9t}{T_R}\right)n_h(t) \tag{4.13}$$

ただし，$e_h(t)$はインパルス応答のエンベロープ，$n_h(t)$はキャリアとして白色雑音，aは振幅項，T_Rは残響時間である．式4.12に式4.13を代入することで，次式の残響環境におけるMTFが得られる．

$$m_R(f_m) = \left[1 + \left(2\pi f_m \frac{T_R}{13.8}\right)^2\right]^{-\frac{1}{2}} \tag{4.14}$$

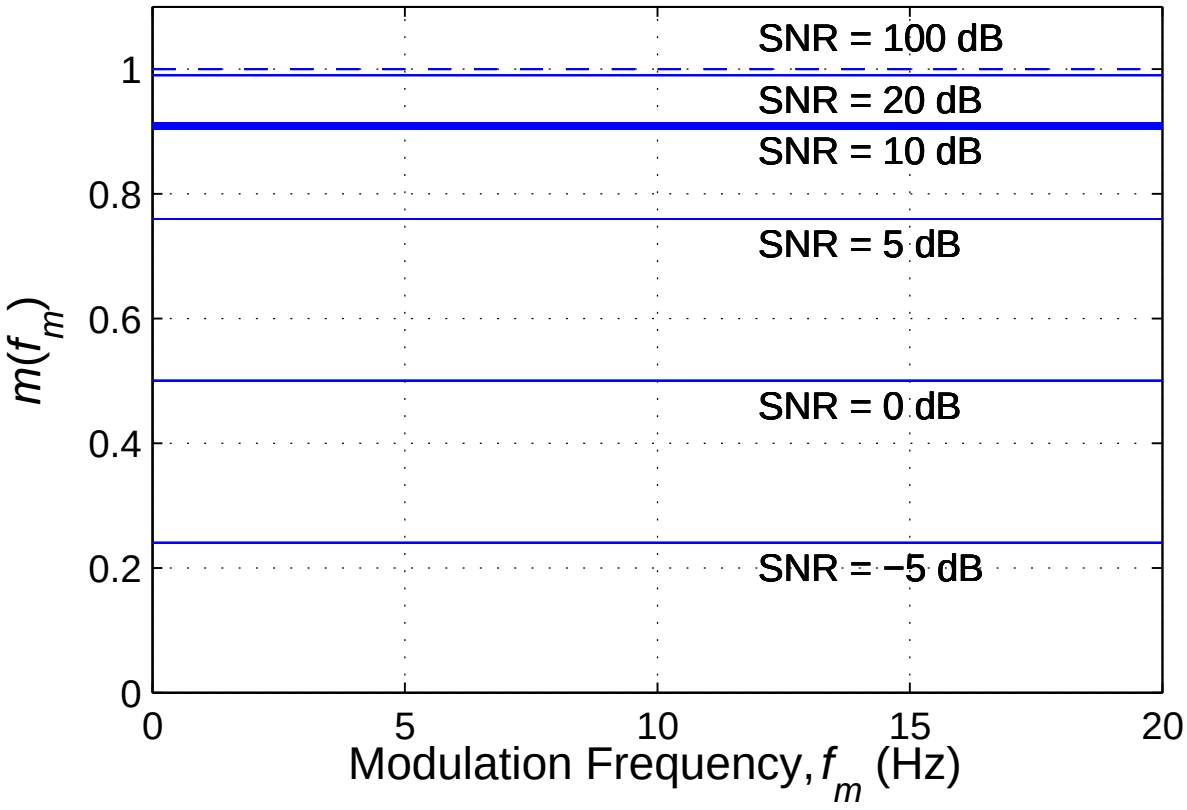

図 4.1: 雑音環境での MTF $m(f_m)$ の特性

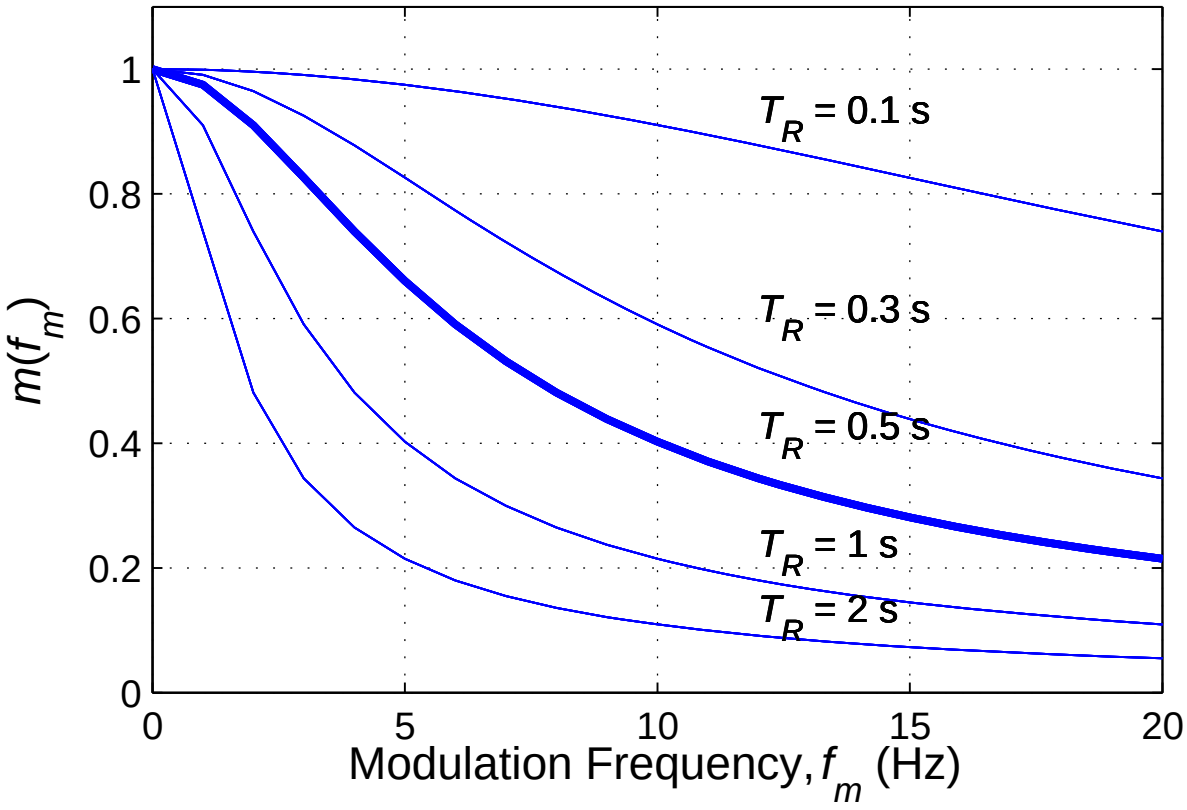

図 4.2: 残響環境での MTF $m(f_m)$ の特性

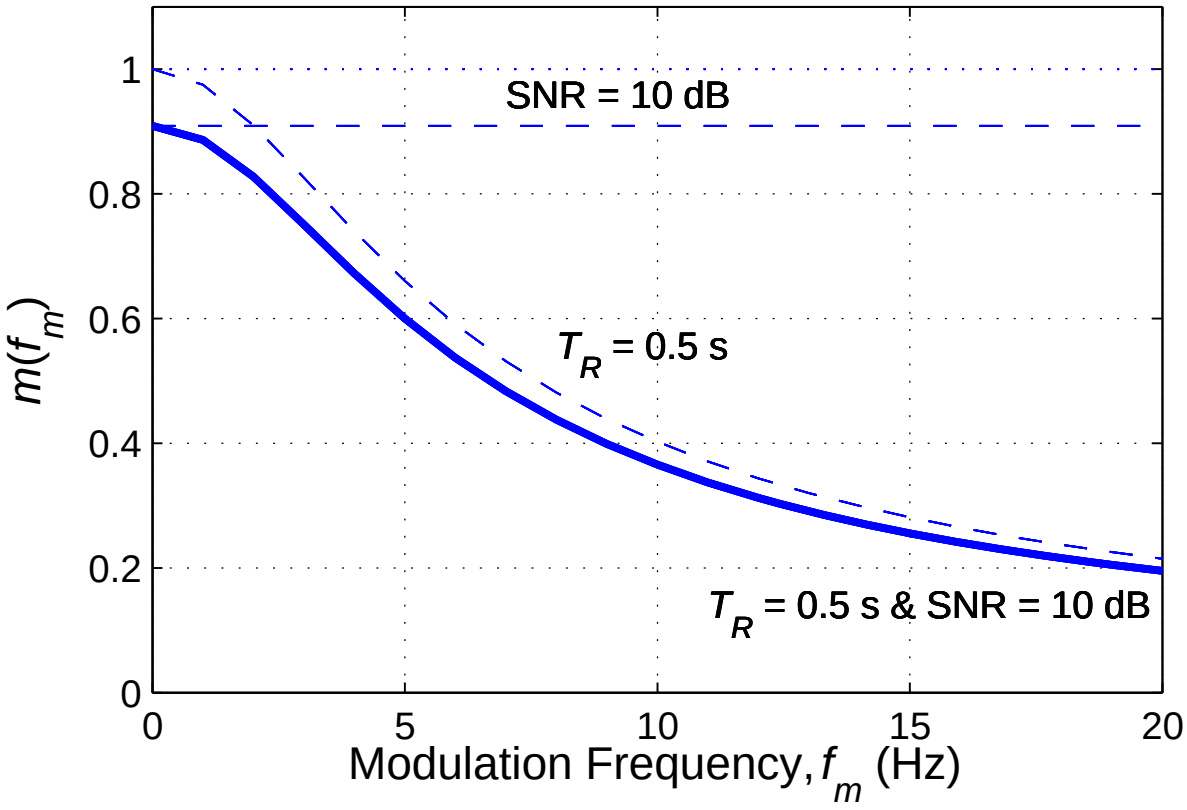

図 4.3: 雑音残響環境での MTF $m(f_m)$ の特性

式 4.14 の特性を図 4.2 に示す．これより，MTF は残響時間 T_R と変調周波数 f_m に依存した，一種の低域通過フィルタ特性を示していることがわかる．

次に，雑音と残響の両方を考慮した場合の MTF を考える．雑音残響環境の出力信号 y(t) は，入力信号 $x(t)$ と室内インパルス応答 $h(t)$ の畳み込みに，雑音 $n(t)$ を加算することで求まる．

$$y(t) = \int_0^\infty h(\tau)x(t-\tau)d\tau + w(t) \tag{4.15}$$

式 4.10 と式 4.14 より雑音残響環境による MTF は次のように導出される．

$$m(f_m) = m_N(f_m) \cdot m_R(f_m) = \left(1 + 10^{-(SNR)/10} \cdot \sqrt{1 + \left(2\pi f_m \frac{T_R}{13.8}\right)^2}\right)^{-1} \tag{4.16}$$

雑音残響環境での MTF は，f_m に依存する．式 4.16 の特性を図 4.3 に示す．これより，残響環境での MTF の低域通過特性と雑音環境での MTF の一定な減衰特性が見て取れる．この概念を用いることにより，雑音残響除去を MTF に基づく逆フィルタ処理で実現する．

4.3 音声の変調スペクトル

ここで，エンベロープをフーリエ変換した時の変調スペクトルが持つ，音声特有の特徴について述べる．入力信号の変調スペクトルと出力信号の変調スペクトルの差が MTF で表現される．Drullman らは，低域の周波数に限定したオランダ語音声 (CVC や VCV) を用いて音声明瞭度の試験を行い，変調周波数 4 Hz 以上かつ 16 Hz 以下は音声明瞭度に重要な帯域であることを示した．最終的に変調周波数 2-16 Hz が重要な特徴であることを示した [71, 72]．Arai らは，Drullman の研究をケプストラムに対応する対数領域に拡張し，日本語においても同様の実験を行い，音声エンベロープの変調周波数 1-16 Hz に音声の重要な特性があることを示した [73, 74]．この時，4 Hz 付近の変調スペクトル成分が最も重要であることが示された．

一方，金寺らは音声認識での重要な変調スペクトル成分を調査している [75, 37]．この結果も音声明瞭度と同様 1-16 Hz の変調スペクトル成分が音声認識において重要であるということが示されている．また，雑音環境において，2 Hz-16 Hz 以外の変調スペクトル成分が音声認識を低下させ，1 Hz 以下の変調スペクトル成分は認識性能を著しく低下させるという結果が示された．

これらの結果より，変調スペクトルでは 1-16 Hz の変調周波数成分が重要であり，雑音や残響により劣化したエンベロープを回復することにより，音声明瞭度が回復すると考えられる．また，変調スペクトル上の約 4 Hz 付近にピークが現れることがわかった．

4.4 MTF と STI の関係

Houtgast&Steeneke は，室内を伝達系とした時の入力と出力の強度変化に着目し，その変化を MTF として定義した．そして，この MTF に基づいて，音声明瞭度を予測する客

観評価尺度であるSTIが音声予測理論として示された．建築音響におけるSTIの測定方法は，既知の入力信号と室内の伝達特性の影響を受けた出力信号からMTFを求め，そこから計算によりSTIを求める．従って，MTFとSTIは対の関係になっており，MTFの変調度が低ければSTIも悪く，MTFの変調度が高ければSTIも良いということになる．また，STIは音声明瞭度との相関関係があり，聴き取りにくさとは相関が高いことがわかっていることから，総合的に考えるとMTFの変調度が高いということは，音声明瞭度及び聴き取りにくさも良いということである．従って，理論的には，室内伝達特性で減衰した出力信号のMTFを推定して逆フィルタ処理して室の影響を受けた信号を回復することで，入力信号を得ることができる．逆フィルタ処理の詳細を次の章で述べる．

第5章　MTFに基づく逆フィルタ処理

5.1　パワーエンベロープ逆フィルタ処理

ここでは，前の章で説明したMTFの概念を用いた，残響・雑音環境でのパワーエンベロープ逆フィルタ処理の説明を行う．まず，パワーエンベロープ逆フィルタ法では信号を以下のようにモデル化している．入力信号，インパルス応答，出力信号は次式のように示される．

$$x(t) = e_x(t)c_x(t) \tag{5.1}$$

$$h(t) = e_h(t)c_h(t) \tag{5.2}$$

$$n(t) = e_n(t)c_n(t) \tag{5.3}$$

$$y_R(t) = x(t) * h(t) \tag{5.4}$$

$$y_N(t) = x(t) + n(t) \tag{5.5}$$

$$y(t) = x(t) * h(t) + n(t) \tag{5.6}$$

$e_x(t)$，$e_h(t)$，$e_n(t)$ は，$x(t)$，$y(t)$，$n(t)$ のエンベロープであり，$c_x(t)$，$c_h(t)$，$c_n(t)$ はそれぞれのキャリアである．MTFの概念は $c_x(t)$，$c_h(t)$，$c_n(t)$ が無相関の時に成り立つ．

まず，残響環境でのパワーエンベロープ逆フィルタ処理 [60, 59, 76, 62, 61, 77, 78] について説明する．入力信号，室内インパルス応答，出力信号の各パワーエンベロープ $e_x^2(t)$，$e_h^2(t)$，$e_y^2(t)$ の間には，

$$\begin{aligned} \left\langle y^2(t) \right\rangle &= \left\langle \left\{ \int_{-\infty}^{\infty} x(\tau)h(t-\tau)d\tau \right\}^2 \right\rangle \\ &= \int_{-\infty}^{\infty} e_x^2(\tau)e_h^2(t-\tau)d\tau = e_y^2(t) \end{aligned} \tag{5.7}$$

の関係があり，式4.14と式5.7のMTFの関係として $m(f_m)$ は $e_h^2(t)$ に関するパワーを正規化した一種のFourier変換であることがわかる．

これらを，計算機で取り扱うために，離散信号に変換する．各パワーエンベロープ $e_x^2(n)$，$e_y^2(n)$，$e_h^2(n)$ のz変換をそれぞれ $E_x(z)$，$E_y(z)$，$E_h(z)$ と表す．インパルス応答のパワーエンベロープのz変換を考えると，式5.7の $e_h^2(t)$ は単純な指数減衰の関数であり，$t < 0$

では $e_h^2(t) = 0$ と仮定していることから，最小位相特性を有している．この特性をふまえ，z 変換した $E_h(z)$ を求めると

$$E_h(z) = \frac{a^2}{1 - \exp\left(-\frac{13.8}{T_R \cdot f_s}\right) z^{-1}} \tag{5.8}$$

を得る．ただし，f_s はサンプリング周波数である．式 (3) の畳み込みの関係から，$e_x^2(t)$ の変調スペクトル $E_x(z)$ は，$E_y(z)$ と MTF の逆数（逆数フィルタ：IMTF）$1/E_h(z)$ の積で求められる．この関係を次式に示す．

$$E_x(z) = \frac{E_y(z)}{a^2} \left\{ 1 - \exp\left(-\frac{13.8}{T_R \cdot f_s}\right) z^{-1} \right\} \tag{5.9}$$

このようにして，室内インパルス応答を推定して逆フィルタ処理を行うことにより入力信号のパワーエンベロープを得ることができる．

次に，雑音環境でのパワーエンベロープ逆フィルタ処理 [64] を説明する．雑音の MTF における変調度と平均パワーは雑音の影響のみを受ける．従って，雑音の平均パワーを求め，減算を行うことで雑音成分を取り除くことができ，この関係を次式に示す..

$$\hat{e}_x^2(t) = \overline{e_x^2} \left(1 + m_N(f_m) \cos(2\pi f_m t) \times \frac{1}{m_N(f_m)} \right) = e_y^2 - \overline{e_n^2} \tag{5.10}$$

ここで，雑音でのブラインド逆フィルタの実現には，$\overline{e_n^2}$ を求める必要がある．$\overline{e_n^2}$ は無音区間での雑音の平均パワーを取ることで求められている．この無音区間の推定は，頑健な音声区間推定（Voice Activity Detection: VAD）を用いることで可能となる．雑音に頑健な VAD として変調スペクトルに基づいた手法がある [79]．最終的に式 5.10 からわかる通り，$\overline{e_x^2}$ は，観測信号のパワーエンベロープ e_y^2 から雑音の平均パワー $\overline{e_x^2}$ を差し引くことで，雑音の除去ができることがわかる．

次に，雑音残響環境でのパワーエンベロープ逆フィルタ処理 [66] を説明する．この方法は，すでに説明している残響環境と雑音環境でのパワーエンベロープ逆フィルタ処理を同時に行うものである．処理体系として，式 5.10 により雑音成分を除去し，その後式 5.9 により残響除去を行う．式の上では，式 4.16 の逆フィルタである．MTF の概念に基づいたパワーエンベロープ回復の概念を図 5.1 に示す．図中（b）は正弦波信号（入力信号），(a）はそのパワーエンベロープ, (d）は $T_R = 0.5$ s の時のインパルス応答, (c）はそのパワーエンベロープ, (f）は SNR=3 dB の白色雑音, (e）はそのパワーエンベロープ, (h）は (b) , (d) , (f）から成る雑音残響入力信号, (g）はそのパワーエンベロープ, (i）は回復パワーエンベロープである．回復パワーエンベロープは適切な残響時間を推定できれば，入力信号と同じ残響時間が得られることを示しており，それ以外の残響時間では，過少推定や過剰推定となってしまうことがわかる．このように，パワーエンベロープ逆フィルタを用いることで雑音残響環境でのパワーエンベロープ逆フィルタ処理が行える．

残響環境において，回復したエンベロープにキャリアを再合成することにより回復音声を得ることができるが，キャリアも残響の影響を受けているにも関わらず回復処理を行っ

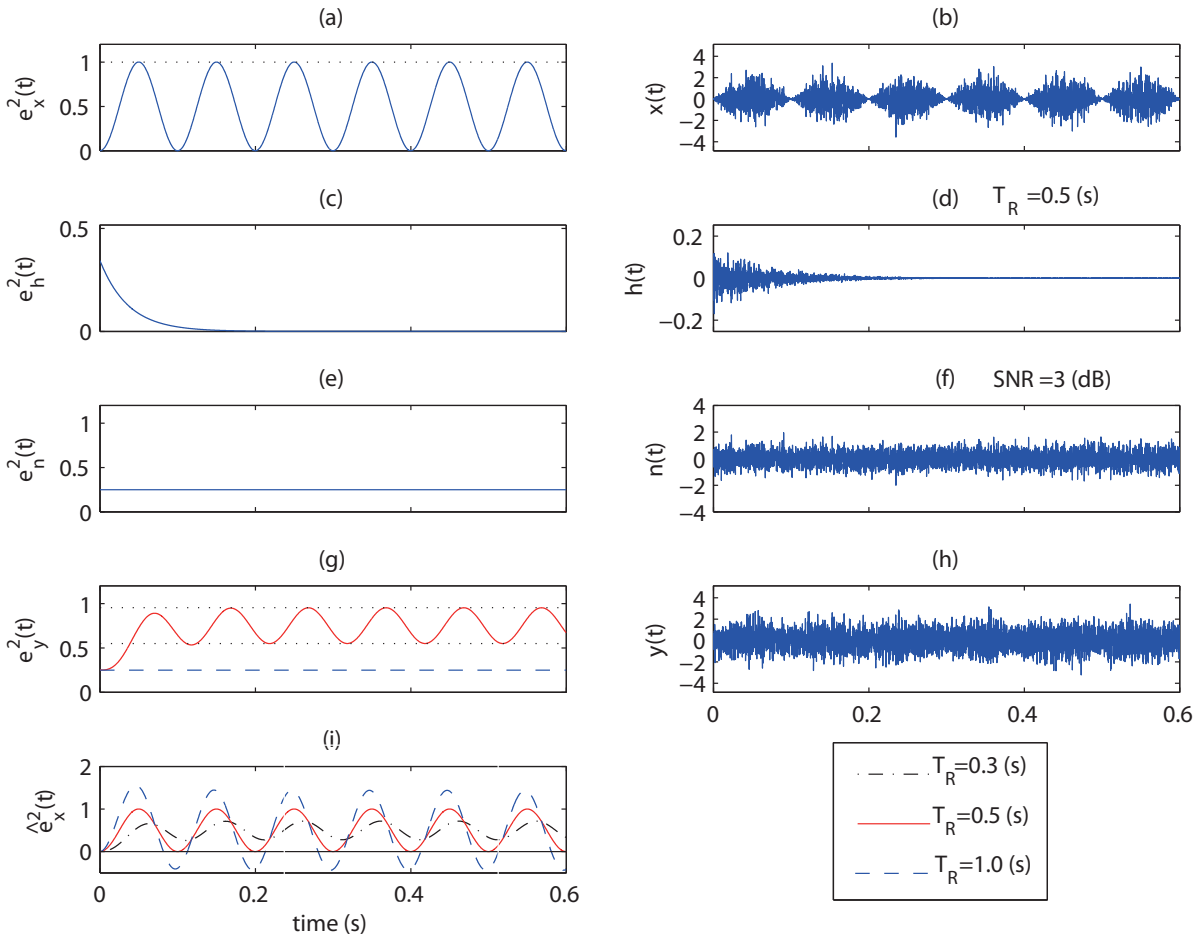

図 5.1: MTF に基づいたパワーエンベロープ回復の概念

ていないために，キャリアと回復したエンベロープを合成すると人工的な異音が生じることが Unoki らによって報告されている．この問題を解決すべくキャリアを再生成する試みが Unoki らによって行われている [61]．これは，残響の影響を受けたキャリアの位相情報を分析合成器の STRAIGHT を用いることで制御する方法である．キャリア再生成の時には，基本周波数情報を必要とするため，基本周波数推定法が必要となる．この結果，音声明瞭度が向上することが示されているが，一方で，原音声の品質と比較したときにキャリアの再合成を行っているために音色が変化することが報告されている．

5.2 エンベロープ抽出法

パワーエンベロープの抽出法について説明する．エンベロープの抽出方法は，Unoki らが二つの方法を提案している [59, 78]．一つは，次式の集合平均を取る方法である．

$$e_y^2(t) = \mathbf{LPF}\left[\left\langle \hat{y}^2(t) \right\rangle\right] = \mathbf{LPF}\left[\left\langle (y(t)\hat{n}(t))^2 \right\rangle\right] \tag{5.11}$$

但し，**LPF** は低域通過フィルタ（LPF）でカットオフ周波数 20 Hz である．もう一つの方法には，次式で示す．Hilbert 変換を利用する方法がある [59]．

$$e_y^2(t) = \mathbf{LPF}\left[|y(t) + j \cdot \mathbf{Hilbert}(y(t))|^2\right] \tag{5.12}$$

但し，**Hilbert** は Hilbert 変換である．これは，Hilbert 変換により瞬時振幅を求め，LPF を施すことによりパワーエンベロープを抽出する．二つの方法で，LPF を用いているのは，音声の主要な変調周波数が 1 ～ 16 Hz であるという報告に基づき，LPF の遮断周波数を 20 Hz とし，高い変調周波数成分を取り除いているためである．Unoki らの報告に基づくと，抽出精度はほとんど差がなく，後者の抽出方法の方が計算が簡単かつ計算量が少ない．実際，後者は精度よくエンベロープ抽出が可能なため，今後も後者の手法によりエンベロープ抽出を行うことを考える．

5.3 残響時間・振幅項の推定方法

まず，残響時間の推定方法について説明する．これまで，時間領域においてパワーエンベロープの回復量から残響時間を推定する方法が Unoki ら [59, 80] によって提案されている．パワーエンベロープは，少なくとも 1 つの 0 の値を取るディップもしくは無音区間をもつため，$e_x^2(t)$ の変調度が 1 だと仮定する．この仮定に基づき原信号と残響信号のパワーエンベロープの間の一致条件が，残響の影響によって減少した変調度 $m(f_m)$ を回復することであると定義する．この定義式は次式のように表現され，回復されたパワーエンベロープ $\hat{e}_x^2(t)$ の最大のディップが 0 である所，あるいは $\hat{e}_x^2(t)$ がちょうど 0 であるところを検出することで残響時間 T_R を推定できる．

$$\hat{T}_R = \max\left(\underset{T_{R,\min} \le T_R \le T_{R,\max}}{\arg\min} \int_0^T \left|\min\left(\hat{e}_{x,T_R}^2(t), 0\right)\right| dt\right) \tag{5.13}$$

ただし，T は $y(t)$ の信号長，$\hat{e}^2_{x,T_R}(t)$ は T_R を変数として得られたパワーエンベロープである．“$\max(\arg\min\{\cdot\})$” は，これらの候補 $\hat{e}_{x,T_R}(t)$ の中から，0 になるときの T_R の最大値を求めることを意味する．従って，回復されたパワーエンベロープが負の値をもつ前の $\hat{T}_R$ で制約を受けている．また，問題点として，パワーエンベロープ抽出の際に取りきれなかった 20 Hz 以上の周波数成分が，パワーエンベロープ逆フィルタ処理によって過剰に強調される．そのため，パワーエンベロープ逆フィルタ処理では後処理として LPF に通しているものの，回復パワーエンベロープの振幅に LPF で取りきれなかった 20 Hz 以上の周波数成分の影響を受けている．これはまた，残響時間の推定精度を左右する谷の形成にも影響を与えている．この影響の検討を変調スペクトル上で逆フィルタ処理を行うことにより検討した．この結果は，付録の変調スペクトル逆フィルタ処理の検討に示している．

そして，室の残響時間を正確に推定する方法を Hiramatsu&Unoki が提案している [81]．この手法は，変調スペクトル上で主要な変調周波数成分を変調度 1 に回復する手法で，ブラインドで推定が可能である．ブラインドでの推定を可能とするために，残響のない状態の変調スペクトルの主要な変調周波数のパワーは 0 Hz での値と同じであり，変調周波数 0 Hz の成分が残響の影響を受けないこと，変調スペクトルは残響が付加されると MTF に従って減少するという特性に基づいている．

$$\log|E_x(f_{dm})| = \log|E_x(0)| \log|E_y(0)| = \log|E_x(0)| \tag{5.14}$$

ただし，f_{dm} は変調スペクトルの主要な成分であり，E_x 及び E_y は，それぞれ e^2_x と e^2_y の変調スペクトルである．そして，残響時間 T_R を次式のように推定している．

$$\hat{T}_R = \arg\min_{T_R}(|\log|E_y(f_{dm})| - log|E_y(0)| - log\hat{m}(f_{dm}, T_R)|) \tag{5.15}$$

ただし，$\log|E_y(f_{dm})| - log|E_y(0)$ は，変調スペクトルの主要な成分 f_{dm} での減少した変調スペクトル，$\hat{m}(f_{dm}, T_R)$ は，T_R の関数として f_{dm} での MTF である．これは，変調スペクトル上で変調度 $m(f_{dm})$ が 1 に回復されるときの T_R に該当する．この方法では，他の周波数成分の影響を受けることもないためパワーエンベロープでの変調スペクトル推定より高精度に残響時間を推定できる．

次に振幅項 a の推定方法を説明する．この推定方法は，Unoki らによって提案された [59]．振幅項 a は，室内インパルス応答の増幅度に関係する．実際は残響の効果は信号の増幅よりも反射による遅延の重ね合わせの効果が強い．そこで，$e^2_h(t)$ の伝達特性の MTF の整合を取るためにパワーによる正規化を行い，次式で振幅項 a を推定する．

$$\hat{a} = \sqrt{1/\int_0^\infty \exp\left(-\frac{13.8t}{\hat{T}_R}\right) dt} \tag{5.16}$$

5.4 雑音残響環境でのMTFに基づく逆フィルタ処理実現に向けての課題

これまで，MTF に基づく逆フィルタ処理では，雑音・残響環境でのパワーエンベロープの回復を中心に研究が行われてきた．これは，Drullma らのエンベロープ情報に音声明瞭度に重要な成分があると示されたことに起因すると考えられる．しかし，彼らはキャリアの回復の必要がないと示した訳ではない．Unoki らの報告によるとキャリアを回復することにより音色が変わるが音声明瞭度が向上するという報告がなされている．そのため，雑音残響環境での音声明瞭度を向上する為には，エンベロープとキャリアの両方を回復する必要がある．

雑音残響環境でのエンベロープの回復は，すでに鵜木らによって検討されており，高い回復精度が得られているように見受けられる．この手法では，雑音に頑健な VAD [82, 83] を用いているが，雑音区間での音声区間の推定 [84] をうまく行えているように見受けられるが，推定精度は十分なものであるとは言いがたい．また，雑音残響の音声を用いて音声区間を推定しているため，残響の影響を受けている音声の音声区間を推定しており，本来ならば，原音声の区間で音声区間推定がなされるべきであり，回復精度の低下に結び付いていると考えられる．従って，雑音残響に頑健な VAD が必要となるが，残響環境での VAD などはこれまでに提案されていない．

一方で，キャリア回復を考えると残響環境でしか行われていない．雑音残響環境でのキャリアの再生成を考えると雑音残響環境に頑健な音声区間推定法と基本周波数の推定法が必要となる．残響環境での基本周波数推定法は，Unoki らが提案した複素ケプストラムを用いた推定法 [85] が提案されており，残響環境での頑健性などが調査されている [86]. 雑音環境での基本周波数推定法は，周期性・調波性を利用した方法 [87] を石本らが提案している．しかし，雑音残響環境での基本周波数推定法が必要となる．今後，検討を行う必要がある．また，キャリアの再生成処理においては，音色が変わるという報告がなされていることから，個人性を保つためにも音色が変わらないような再合成の処理を検討する必要がある．

これらの課題をまとめると次のようなモデル（図 5.2）を提案できれば，音声明瞭度及び聴き取りにくさを回復する雑音残響除去法の実現ができると考えている．

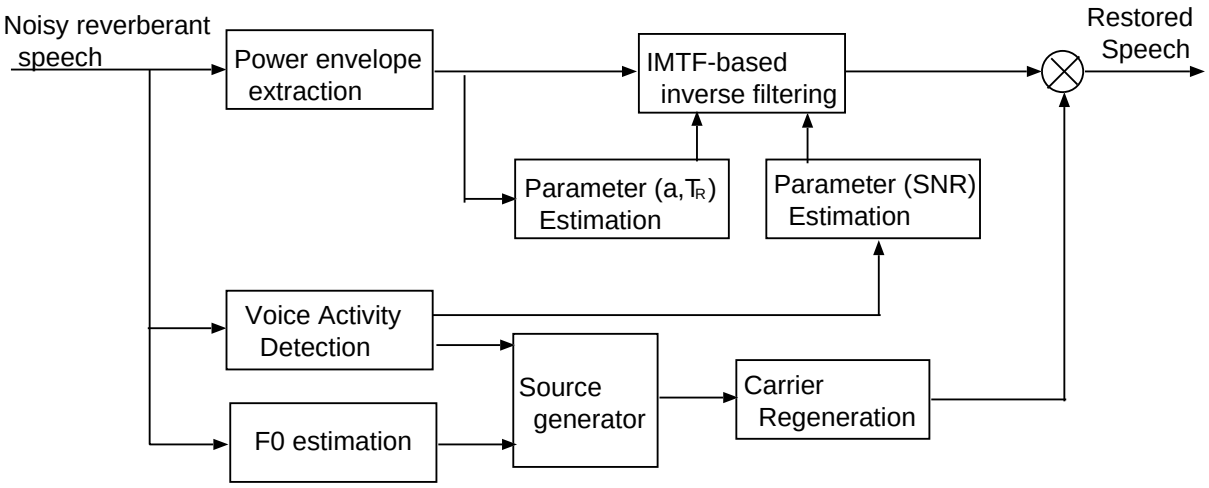

図 5.2: 音声明瞭度と聴き取りにくさを改善する雑音残響除去法の概要

第6章　結論

6.1　本報告書の要約

まず，音声コミュニケーションを的確に評価するために，音声伝達の評価方法の調査を行い，単語了解度及び聴き取りにくさにより評価を行うことが最善であることがわかった．また，これらの主観評価を客観的に評価する方法としてSTIを用いることが最善であるとわかった．この結果に基づき，これまでに報告されている雑音・残響除去法について調査を行った．多くの手法は，音声明瞭度や聴き取りにくさを回復する物理指標に着目して回復処理を行うっ手法ではなかった．雑音残響環境での音声明瞭度回復を行う手法としてMTFに基づくパワーエンベロープ逆フィルタ処理法が最適であると考えた．この手法は，音声明瞭度や聴き取りにくさと相関の高いSTIを求める時に用いるMTFを用いていることからMTFを回復することで音声明瞭度が回復すると考えた．しかし，この手法は，これまでパワーエンベロープのみの回復処理がなされており，残響の影響を受けたキャリアを用いて合成して回復音声を求めると人工的な異音が生じることが報告されている．キャリア回復に関する検討はほとんど行われておらず，音声明瞭度の回復にはキャリア回復は必要不可欠であるなどの課題を示した．

6.2　今後の展望

雑音残響環境での音声明瞭度回復のために，MTFに基づくパワーエンベロープ逆フィルタ処理を用いることを考えている．次の大きな課題を克服することで実現できると考えられる．(i）雑音残響に頑健な音声区間推定法,(ii）雑音残響に頑健な基本周波数推定法,(iii）個人性を低下させないキャリアの再生成法．これらの手法を提案し，パワーエンベロープ回復及びキャリアの回復を行うことができれば，雑音残響環境下での音声明瞭度の回復が実現できると考えている．

付録

雑音・残除去法一覧表

次に示す表は，雑音・残響除去法の調査結果を一覧表としてまとめたものである．評価項目の説明と記号の意味について示す．● 手法 … 各雑音・残響除去法の手法名
● 知覚ベース … 音声知覚に基づくアプローチであるかどうか（◯：基づいている，×：基づいていない）
● 雑音残響 … 手法がどの環境に対応しているのか（◯：雑音残響環境，△：雑音と残響環境，×：雑音又は残響環境）
● マイク … 単一マイクロホンで処理が可能なのかどうか（◯：可能， △：不可能）
● ブラインド … ブラインドによる処理が可能かどうか（◯：ブラインド処理，△：セミブラインド処理，×：ノンブラインド処理）
● 了解度 … 了解度が回復されているか（◯：回復できている，△：回復に課題が残る，–：了解度試験の結果が示されていないため不明）

表 6.1: 雑音・残響除去法の評価表

手法	知覚ベース	雑音残響		マイク	ブラインド	了解度
ICA に基づく処理 [51, 52, 53]	×	雑音・残響	△	×	○	−
Spectral Subtraction (SS) 法 [26, 27, 28]	×	雑音	×	○	○	−
Adaptive Noisee Canceling (ANC)[29, 30, 31]	×	雑音	×	○	○	−
MMSE-STSA [32, 33]	×	雑音	×	○	○	−
Winner filtering [34]	×	雑音	×	○	○	−
最小位相逆フィルタ処理 [38]	×	残響	×	○	×	−
MINT 法 [39, 41]	×	残響	×	×	△	−
帯域分割逆フィルタ処理 [42]	×	残響	×	×	△	−
逐次的雑音残響除去 [67, 68, 69]	×	雑音残響	○	×	○	−
RASTA [36]	○	雑音	×	○	○	−
調波構造に基づく処理 [43, 44]	○	残響	×	○	○	−
CASA [48, 49]	○	雑音・残響	△	○	○	△
MTF に基づく逆フィルタ処理 [64, 59, 76, 66]	○	雑音・残響	△	○	○	△

変調スペクトル逆フィルタ処理の検討

時間領域での残響時間推定法は，パワーエンベロープ回復では最善な残響時間として最適に求められている．しかし，パワーエンベロープを抽出する際に低域通過フィルタ（LPF）で取りきれなかった高調波成分 (20 Hz 以上) が，パワーエンベロープ逆フィルタ処理により過剰に強調される．そのため，後処理として回復処理後に LPF に通しているが，回復パワーエンベロープの振幅は，LPF で取りきれなかった 20 Hz 以上の変調周波数成分の影響を受けている．これはまた，残響時間推定の精度を左右する谷の形成にも影響を与えている．以上により，残響時間推定の精度とパワーエンベロープの回復精度が頭打ちの状態となっている．そのため，ここでは，変調スペクトル上での残響時間の推定と MTF ベースの逆フィルタ処理を行うことによって，これらの問題点を解決できるかどうか，その可能性を検討する．

提案法は，変調スペクトル上で残響時間を推定し，MTF ベースの逆フィルタ処理を行うことから変調スペクトル逆フィルタ処理と呼ぶ．提案法では，20 Hz 以上の周波数成分の影響を受けないようにするために，パワーエンベロープを 40 Hz にダウンサンプリングを行った．その上で，残響時間パラメータの推定を Hiramatsu&Unoki による変調スペクトルでのブラインド残響時間推定法 [81] のコンセプトに基づきノンブラインドで推定を行った．今回は，原信号の変調スペクトルの主要な周波数成分に残響信号の変調スペクトルが最も近い時の残響時間パラメータを，求めた推定パラメータとした．理論的には，この変調スペクトルで原信号の変調スペクトルで一致する場合にパワーエンベロープでの回復精度が向上するわけである．そして，推定パラメータを用いて，20 Hz までに帯域制限されている変調スペクトル逆フィルタ処理を行う．そして，FFT を行いパワーエンベロープに戻し，アップサンプリングすることにより，変調スペクトルでの回復パワーエンベロープが求まる．

今回は，シミュレーションには次に示す三つのパワーエンベロープと白色雑音の積で構成される原信号 $x(t)$ を利用する．

1. 正弦波で構成されるパワーエンベロープ：
   $e_x^2(t) = 1 - \cos(2\pi F t)$
2. 調波複合音で構成されるパワーエンベロープ：
   $e_x^2(t) = 1 + \dfrac{1}{K}\displaystyle\sum_{k=1}^{K} \sin(2\pi k F_0 t + \theta_k)$
3. 帯域制限雑音で構成されるパワーエンベロープ：
   $e_x^2(t) = \mathbf{LPF}[n_\omega(t)]$

但し，$F = 10$ Hz，$F_0 = 1$ Hz，$K = 2$，θ_k はランダム位相，$\mathbf{LPF}[\cdot]$ のカットオフ周波数は 20 Hz とした．評価シミュレーションでは，原理を確認する為，三つのパワーエンベロープに対して，一つのキャリア（白色雑音）を乗じて得た音源信号 $x(t)$ および，5 種類の残響時間（$T_R = 0.1, 0.3, 0.5, 1.0, 2.0$ s）に対して，$x(t)$ と $h(t)$ のキャリアが無相関性を

保っている 100 種類の室内インパルス応答 $h_t(t)$ を畳み込んで得た音源信号が得られた合計 1,500（$3 \times 5 \times 100$）個の残響信号 $y(t)$ を用意する．信号長は 1.0 s である．解析信号 $x(t)$ には，前に 0.5 s，後に 2.0 s の無音区間を入れて合計 3.5 s の信号にて解析を行った．

評価尺度は，パワーエンベロープに対する (i) 相関値と (ii) SNR（S をオリジナルのパワーエンベロープ，N を回復したパワーエンベロープ）とした．

$$\mathrm{Corr}(e_x^2, \hat{e}_x^2) = \frac{\int_0^T \left(e_x^2(t) - \overline{e_x^2(t)}\right)\left(\hat{e}_x^2(t) - \overline{\hat{e}_x^2(t)}\right) dt}{\sqrt{\left\{\int_0^T (e_x^2(t) - \overline{e_x^2(t)})^2 dt\right\}\left\{\int_0^T (\hat{e}_x^2(t) - \overline{\hat{e}_x^2(t)})^2 dt\right\}}} \tag{6.1}$$

$$\mathrm{SNR}(e_x^2, \hat{e}_x^2) = 10 \log_{10} \frac{\int_0^T (e_x^2(t))^2 dt}{\int_0^T (e_x^2(t) - \hat{e}_x^2(t))^2 dt} \tag{6.2}$$

但し，$\overline{e_x^2(t)}$ は $e_x^2(t)$ の時間平均である．

はじめに，サンプル例として残響時間 1.0 s の時の正弦波で構成されるパワーエンベロープを図 6.1 に示す．これより，パワーエンベロープ逆フィルタ処理では，原信号のパワーエンベロープに比べ振幅が小さく回復され形状も少し異なるのに対し，変調スペクトル逆フィルタ処理では原信号のパワーエンベロープに近い振幅が得られ，形状も原信号のパワーエンベロープに近いことがわかる．図 6.1 の時の変調スペクトルを図 6.2 に示す．この正弦波で構成されるパワーエンベロープは 10 Hz の成分を持っているため，変調スペクトル上で主要なスペクトルである 10 Hz のピークを見ると，変調スペクトル逆フィルタ処理で回復した変調スペクトルは原信号の変調スペクトルとよく一致している．これが，パワーエンベロープ上での一致につながっている．

次に図 6.3 に正弦波で構成されるパワーエンベロープの回復精度の結果を示す．図 6.3 (a) は相関値の改善度を，図 6.3 (b) は SNR の改善度を示す．この場合，相関値も SNR も共に大きく改善していることがわかる．

次に図 6.4 に調波複合音で構成されるパワーエンベロープの回復精度の結果を示す．この結果では相関値と SNR 共に改善がみられたが，図 6.3 の結果ほどではなかった．

図 6.5 に帯域制限雑音で構成されるパワーエンベロープの回復精度の結果を示す．この場合，図 6.4 と類似した結果となった．図 6.4 と図 6.5 の結果では，主要な周波数成分が 0-20 Hz の全体にあるために，変調スペクトルの外形構成が影響を受け，二つの処理での差がほとんど出なかったものと考えられる．

今回，パワーエンベロープ逆フィルタ処理と変調スペクトル逆フィルタ処理の回復精度について検討した．その結果，正弦波で構成されるパワーエンベロープでは大きな改善がみられた．残りの二つでも改善はみられたが予想よりは大きなものではなかった．変調スペクトル逆フィルタ処理の方が優位であることを示した．提案法の方が改善度が高かったことから，20 Hz 以上の周波数成分による影響がパワーエンベロープ逆フィルタの回復精度を頭打ちにしていた原因であったことが明らかとなった．

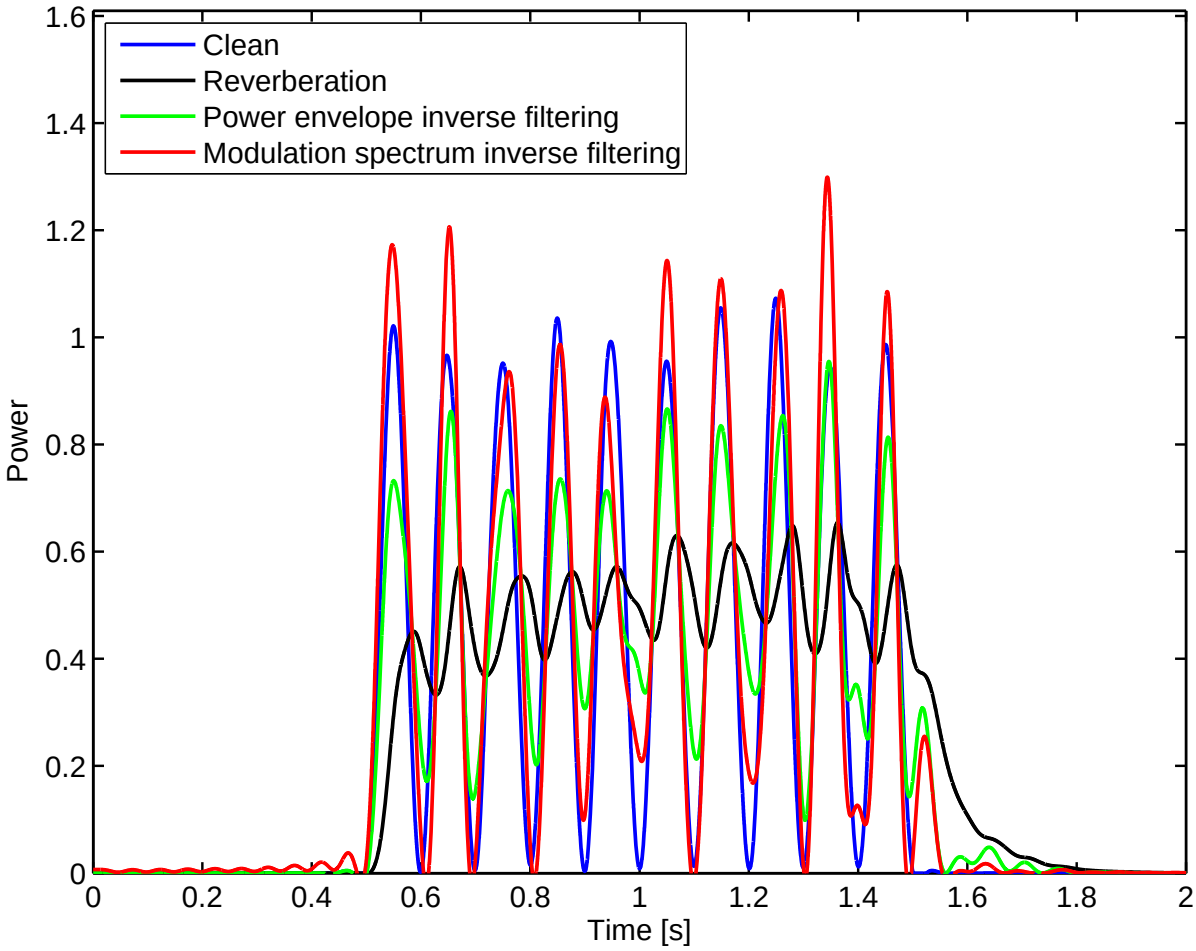

図 6.1: 正弦波信号のパワーエンベロープ

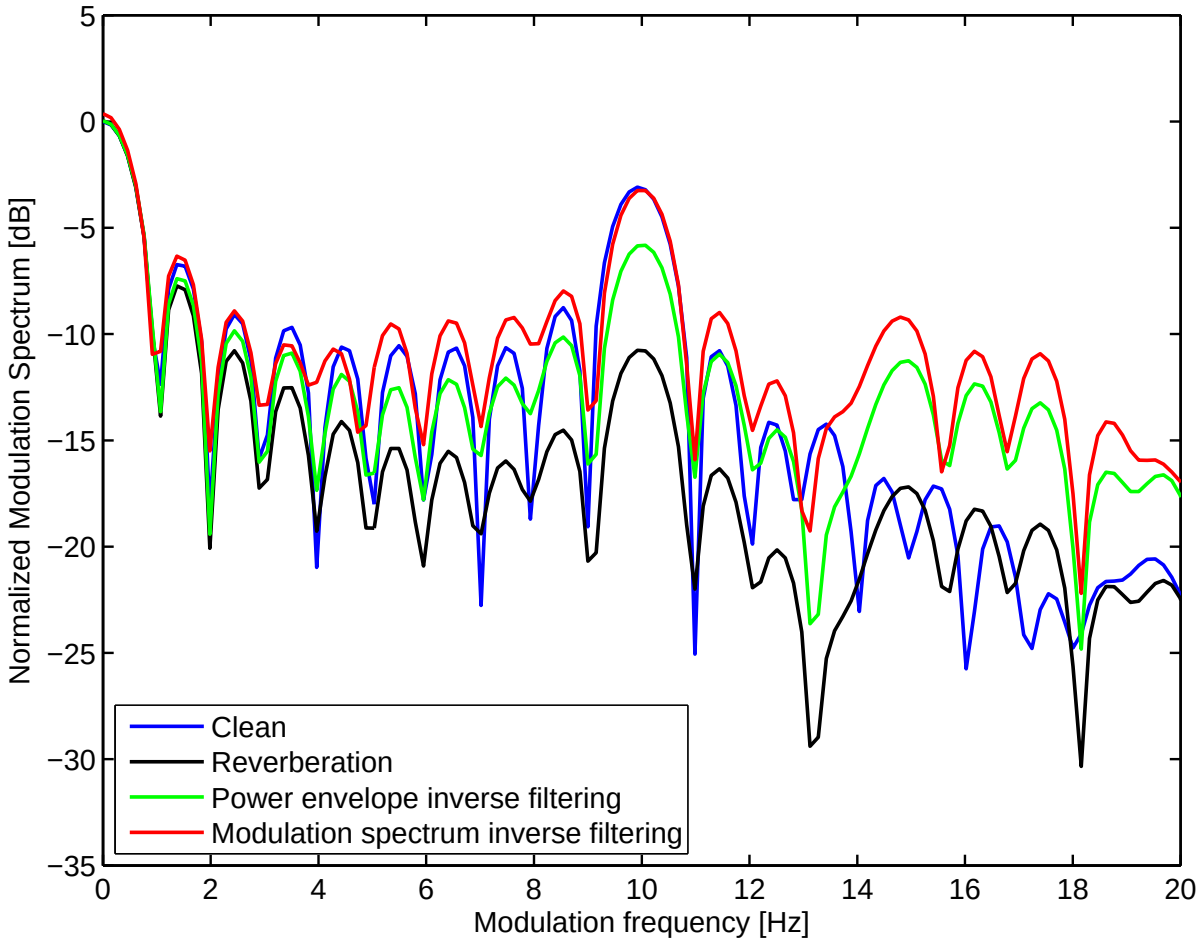

図 6.2: 正弦波信号の変調スペクトル

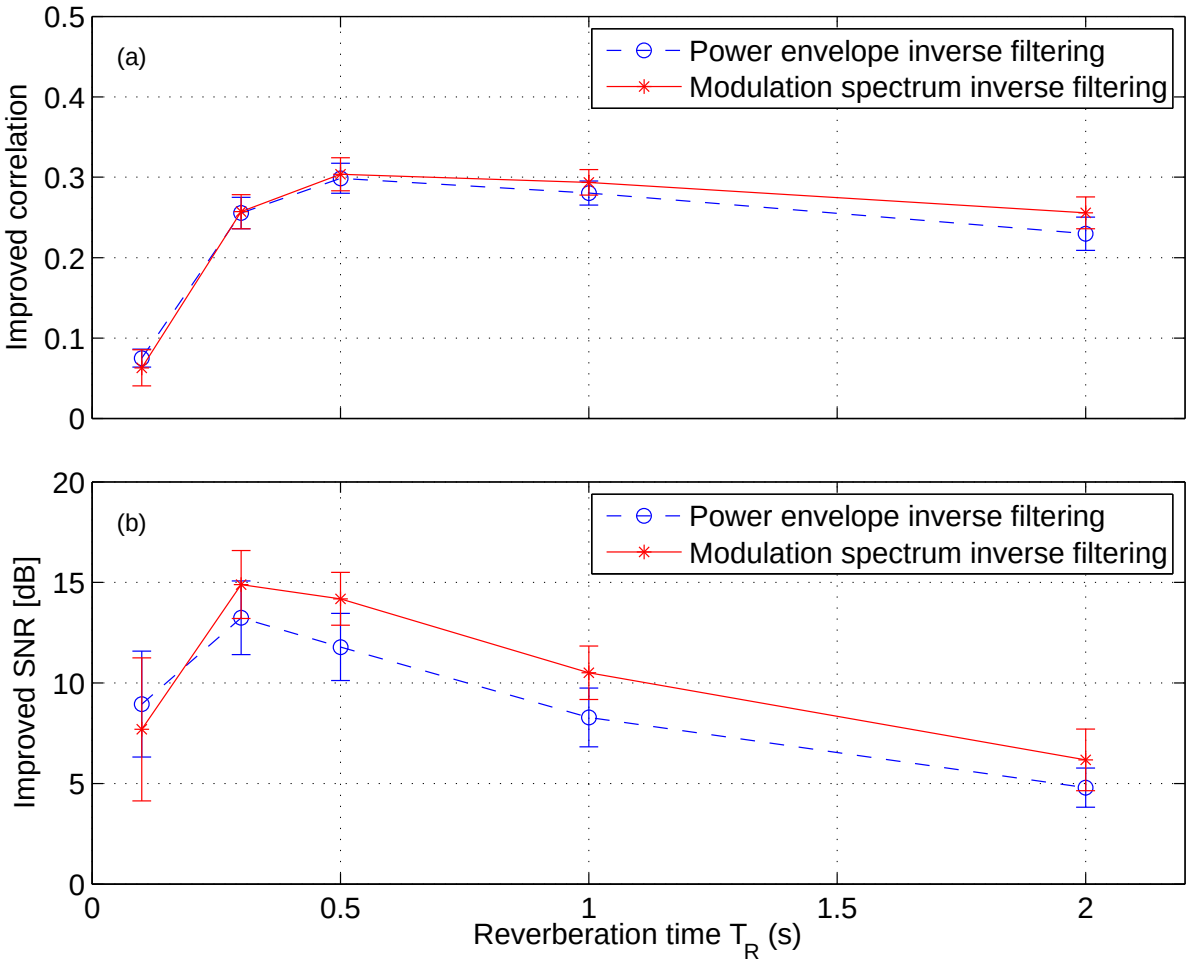

図 6.3: 正弦波で構成されるパワーエンベロープの評価結果：(a) Correlation と (b) SNR の改善度

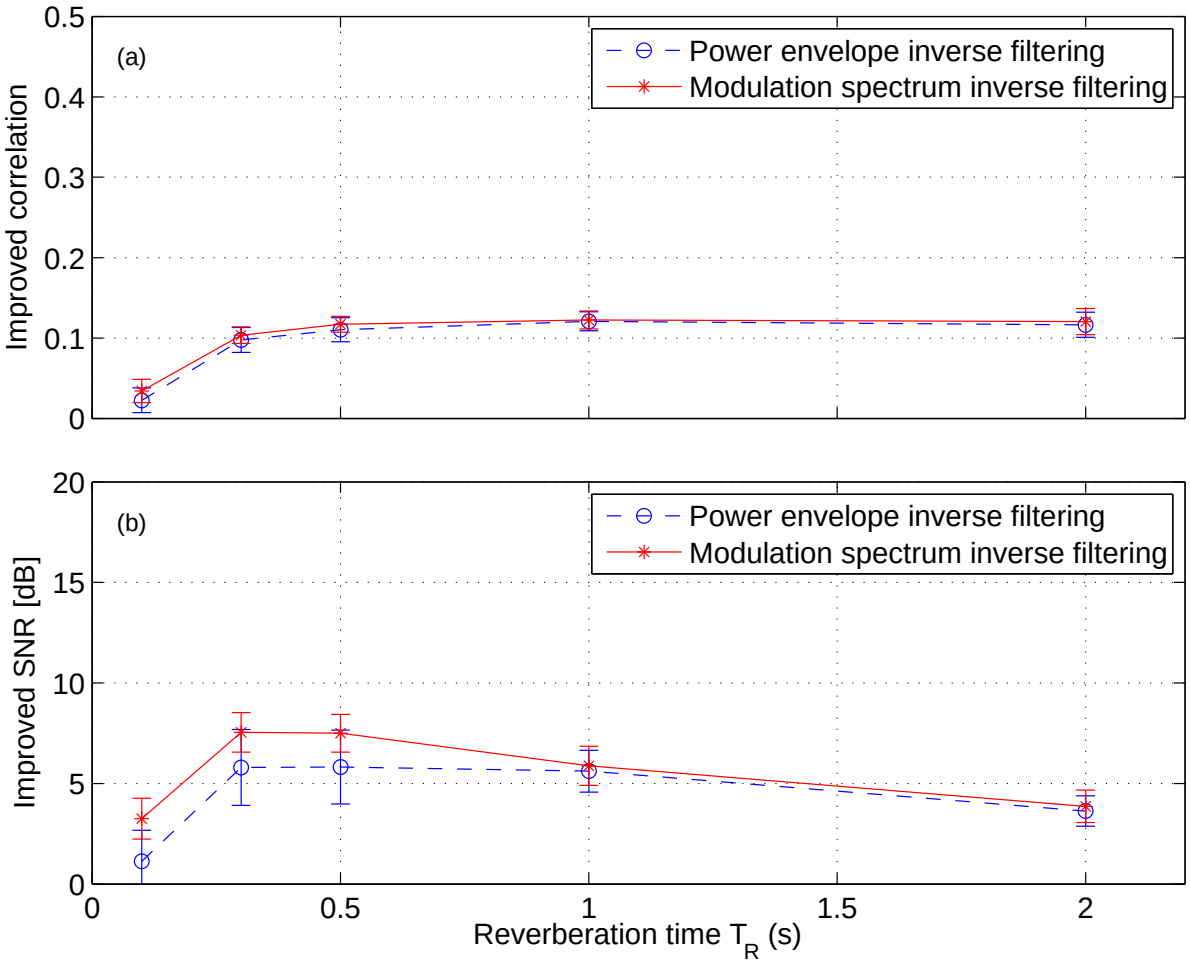

図 6.4: 調波複合音で構成されるパワーエンベロープの評価結果：(a) Correlation と (b) SNR の改善度

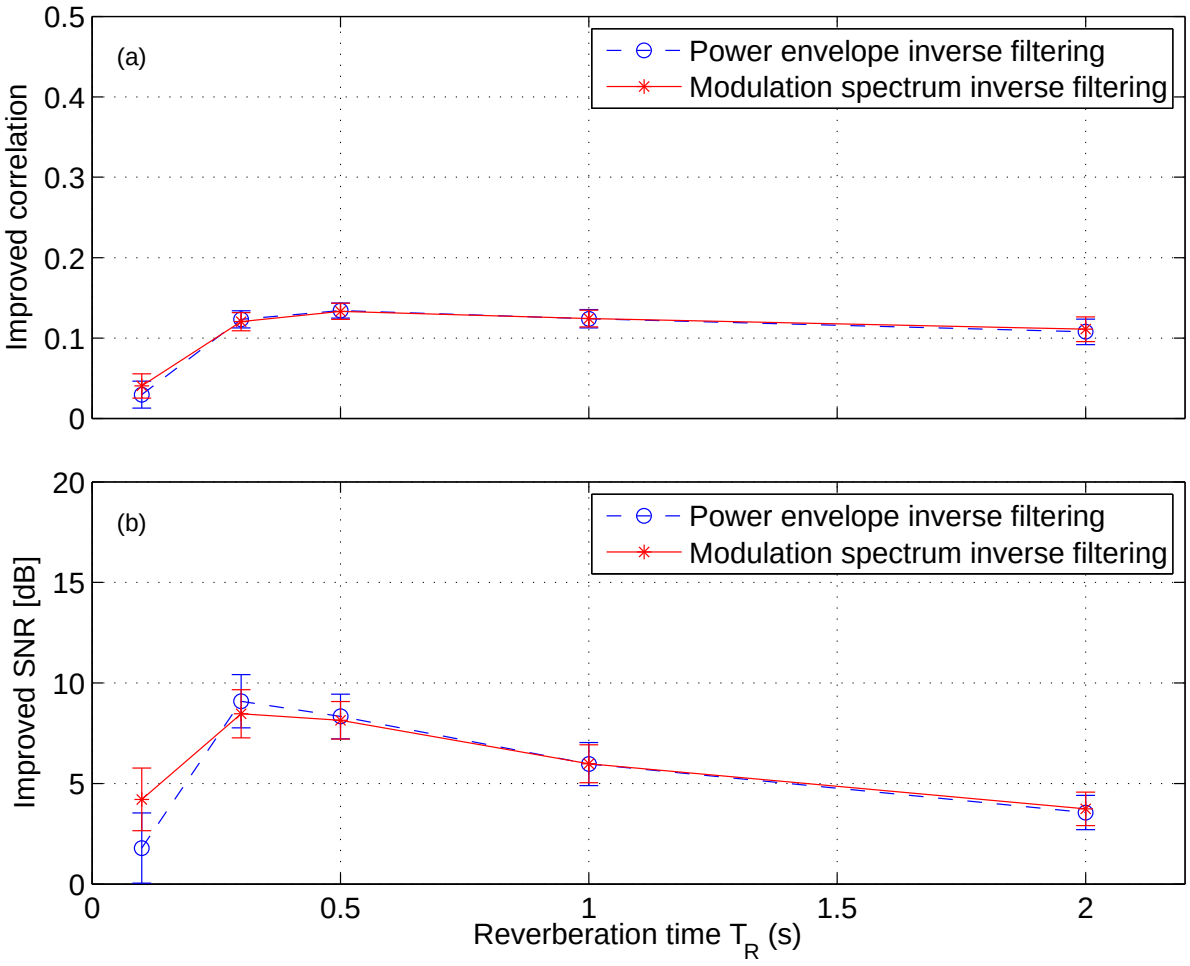

図 6.5: 帯域制限雑音で構成されるパワーエンベロープの評価結果：(a) Correlation と (b) SNR の改善度

参考文献

[1] 高橋 玲，青木 仁志，北脇 信彦，“IP 電話の通話品質評価法に関する標準化動向,” 電子情報通信学会 技術研究報告 CQ，Vol. 103，No. 660，pp. 27–32，Feb. 2004.

[2] 飯田 茂隆，“明領度試験法について,” 日本音響学会誌，Vol. 43，No. 7，pp. 532–536，1987.

[3] 坂本 修一，鈴木 陽一，天野 成昭，小澤 賢司，近藤 公久，曾根 敏夫，“親密度と音韻バランスを考慮した単語了解度試験用リストの構築,” 日本音響学会誌，Vol. 54，No. 12，pp. 842–849，1998.

[4] 坂本 修一，天野 成昭，鈴木 陽一，近藤 公久，小澤 賢司，曾根 敏夫，“単語了解度試験におけるモーラ同定に対する親密度の影響,” 日本音響学会誌，Vol. 60，No. 7，pp. 351–357，2004.

[5] 長谷 芳樹，橘 亮輔，阪口 剛史，細井 裕司，“親密度別単語了解度試験用音声データセット（FW03）単音節音声のラウドネス校正,” 日本音響学会誌，Vol. 64，No. 11，pp. 647–649，2008.

[6] 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 2007（FW07）の作成,” 電子情報通信学会 技術研究報告 SP，Vol. 107，No. 434，pp. 43–48，2008.

[7] 佐藤 洋，佐藤 逸人，吉野 博，鈴木 陽一，天野 成昭，近藤 公久，長友宗重，“単語親密度と加齢による聴力損失が残響及び雑音下における単語了解度に及ぼす影響,” 日本音響学会誌，Vol. 58，No. 6，pp. 346–354，2002.

[8] M. Martin, SPEECH AUDIOMETRY SECOND EDITION, Wwhurr Publishers Ltd London, 1997.

[9] 佐藤 洋，長友 宗重，吉野 博，矢島 吉紀，“残響・騒音の音声聴取に及ぼす影響の評価に関する実験的検討,” 日本建築学会計画系論文集，Vol. 495，pp. 1–8，1996.

[10] M. Morimoto, H. sato, and M. Kobayashi, “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, Sept. 2004.

[11] 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.

[12] 佐藤 洋，“残響騒音下において一度に提示する単語数と話速が単語了解度と「聴き取りにくさ」に及ぼす影響,” 日本音響学会 聴覚研究会，Vol. 38，No. 1，pp. 1–6，2008.

[13] N. R. French and J. C. Steinberg, “Factors governing the intelligibility of speech sounds,” *J. Acoust. Soc. Am.*, Vol. 19, pp. 90–119, 1947.

[14] 山田 武志，北脇 信彦，“PESQ と疑似音声を用いた雑音下音声認識の性能予測の検討,” 情報処理学会研究報告 SLP，Vol. 2003，No. 124，pp. 37–42，Dec. 2003.

[15] 藤田 顕吾，加藤 恒夫，山田 秀昭，河井 恒，中島 康之，“携帯電話音声に対する主観評価の精度及び客観評価尺度 PESQ の有効性の検証,” 電子情報通信学会 技術研究報告 SP，Vol. 104，No. 470，pp. 29–33，Nov. 2004.

[16] 中岡 謙，加藤 正美，“客観的音声品質評価法 PESQ による VoIP 端末の音声品質評価実験,” 電子情報通信学会 技術研究報告 CQ，Vol. 102，No. 190，pp. 35–40，July 2002.

[17] ハインリッヒ・クットルフ，室内音響学 建築の響きとその理論，市ヶ谷出版社，2003.

[18] T. Houtgast and H. J. M. Steeneken, “The Modulation Transfer Function in Room Acoustics as a Predictor of Speech Intelligibility,” *Acustica*, Vol. 28, pp. 66–73, 1973.

[19] T. Houtgast, H. J. M. Steeneken, and R. Plomp, “Predicting Speech Intelligibility in Rooms from the Modulation Transfer Function. I. General Room Acoustics,” *Acustica*, Vol. 46, pp. 60–72, 1980.

[20] T. Houtgast and H. J. M. Steeneken, “A review of the MTF concept in room acoustics and its use for estimating speech intelligibillity in auditoria,” *J. Acoust. Soc. Am.*, Vol. 77, pp. 1069–1077, 1985.

[21] 戸井田 義徳，“小特集-音声の明瞭度と認識率-空間内における音情報伝達,” 日本音響学会誌，Vol. 51，No. 4，pp. 312–316，1995.

[22] 中島 立視，“音声の明瞭度指標 (sti) の測定,” 日本音響学会誌，Vol. 49，No. 2，pp. 103–110，Feb. 1993.

[23] 古井 貞熙，ディジタル音声処理，東海大学出版会，2002.

[24] 大賀 寿郎，山崎 芳男，金田 豊，音響システムとディジタル処理，電子情報通信学会 編，コロナ社，1995.

[25] 金田 豊，“騒音下音声認識のためのマイクロホンアレー技術,” 日本音響学会誌，Vol. 53，No. 11，pp. 872–876，1997.

[26] S. F. Boll, “Suppression of Acoustic Noise in Speech Using Spectral Subtraction,” *IEEE Trans. ASSP*, Vol. 27, No. 2, pp. 113–120, 1979.

[27] Z. Goh, K. Tan, and B. T. G. Tan, “Postprocessing Method for Suppressing Musical noise generated by Spectral Subtraction,” *IEEE Trans. on Speech and Audio Processing*, Vol. 6, No. 3, pp. 287–292, 1998.

[28] 水町 光徳，赤木 正人，“マイクロホン対を用いたスペクトルサブトラクションによる雑音除去法,” 電子情報通信学会論文誌 A，Vol. 82，No. 4，pp. 503–512，1999.

[29] M. R. Sambur, “Adaptive Noise Canceling for Speech Signals,” *IEEE Trans. Acoustics, Speech, and Signal processing*, Vol. 26, No. 5, pp. 419–423, 1978.

[30] W. A. Harrison, J. S. LIM, and E. Singer, “A New Application of Adaptive Noise Cancellation,” *IEEE Trans. ASSP*, Vol. 34, No. 1, pp. 21–27, 1986.

[31] J. E. Greenberg, “Modified LMS Algorithms for Speech Processing with an Adaptive Noise Canceller,” *IEEE Trans. Speech and Audio Processing*, Vol. 6, No. 4, pp. 338–351, 1998.

[32] Y. Ephraim and D. Malah, “Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator,” *IEEE Trans. ASSP*, Vol. ASSP-32, No. 6, pp. 1109–1121, 1984.

[33] 加藤 正徳，杉山 昭彦，芹沢 昌宏，“重み付き雑音推定と MMSE STSA 法に基づく高音質雑音抑圧,” 電子情報通信学会論文誌 A，Vol. J87，No. 7，pp. 851–860，July 2004.

[34] J. S. Lim and A. V. Oppenheim, “All-pole modeling of degraded speech,” *IEEE Trans. ASSP*, Vol. 26, No. 3, pp. 197–210, 1978.

[35] R. J. McAulay and M. L. Malpass, “Speech enhancement using a soft-decision noise suppression filter,” *IEEE Trans. ASSP*, Vol. 28, No. 2, pp. 137–145, 1980.

[36] H. Hermansky and N. Morgan, “RASTA Processing of Speech,” *IEEE Trans. Speech Audio Process.*, Vol. 2, No. 4, pp. 578–586, 1994.

[37] N. Kanedera, T. Arai, H. Hermansky, and M. Pavel, “On the relative importance of various components of the modulation spectrum for automatic speech recognition,” *Speech Commun.*, Vol. 28, pp. 43–55, 1999.

[38] S. T. Neely and J. B. Allen, "Invertibility of a room impulse response," *J. Acoust. Soc. Am.*, Vol. 66, No. 1, pp. 165–169, 1979.

[39] M. Miyoshi and Y. Kaneda, "Inverse filtering of room acoustics," *IEEE Trans. ASSP*, Vol. 36, pp. 145–152, 1988.

[40] 古家 賢一，片岡 章俊，"チャネル間相関行列と音声の白色化フィルタを用いたSemi-blind 残響抑圧," 電子情報通信学会論文誌 A，Vol. J88，No. 10，pp. 1089–1099，Oct. 2005.

[41] K. Furuya and A. Kataoka, "Robust Speech Dereveerberation Using Multichannel Blind Deconvolution With Spectral Subtraction," *IEEE Trans. Audio, Speech, and language processing*, Vol. 15, No. 5, pp. 1579–1591, 2007.

[42] H. Wang and F. Itakura, "Realization of acoustic inverse filtering through multi-microphone sub-band processing," *IEICE Trans. Fundamanetals*, Vol. E75-A, pp. 1474–1483, 1992.

[43] 中谷 智広，三好 正人，木下 慶介，"調波構造に基づくモノラル音声信号のブラインド残響除去," 電子情報通信学会論文誌 D，Vol. 88，No. 3，pp. 509–520，2005.

[44] T. Nakatani, K. Kinoshita, and M. Miyoshi, "Harmonicity-Based Blind Dereverberation for Single-Channel Speech Signals," *IEEE Trans. Audio. Speech, and Lancuage Processing*, Vol. 15, No. 1, pp. 80–95, 2007.

[45] T. Nakatani, Computational auditory scene analysis based on residue-driven architecture and its application to mixed speech recognition, Ph. D. Thesis, Deot. of Applied Analysis and Complex Dynamical Systems, Kyoto Univ., 2002.

[46] 中谷 智広，奥乃 博，"音オントロジーに基づいた音環境理解システムの統合," 人工知能学会誌，Vol. 14，No. 6，p.11，1999.

[47] 鵜木 祐史，赤木 正人，"雑音が付加された波形からの信号波形の一抽出法," 電子情報通信学会論文誌 A，Vol. J80，No. 3，pp. 444–453，1997.

[48] M. Unoki and M. Akagi, "A method of signal extraction from noisy signal based on auditory scene analysis," *Speech Communication*, Vol. 27, pp. 261–279, 1999.

[49] 鵜木 祐史，赤木 正人，"聴覚の情景解析に基づいた雑音下の調波複合音の一抽出法," 電子情報通信学会論文誌 A，Vol. J82，No. 10，pp. 1497–1507，1999.

[50] 浅野 太，"ICA による音響信号の分離," 電子情報通信学会誌，Vol. 87，No. 3，pp. 175–181，2004.

[51] 古屋 武志，金田 圭一，五反田 博，“ブラインド信号分離による雑音除去法の SN 比改善量，” 電子情報通信学会 A，Vol. 87，No. 7，pp. 1054–1058，2004.

[52] 高橋 祐，高谷 智哉，猿渡 洋，鹿野 清宏，“独立成分分析に基づく空間的サブトラクションアレーによる雑音抑圧，” 電子情報通信学会 技術研究報告 EA，Vol. 106，No. 125，pp. 13–18，2006.

[53] 古屋 武志，金田 圭一，五反田 博，“独立成分分析に基づく耐高残響音源分離に関する研究，” 電子情報通信学会 技術研究報告 NC，Vol. 105，No. 131，pp. 7–12，2005.

[54] T. Langhans and H. W. Strube, “Speech enhancement by nonlinear multiband envelope filtering,” *Proc. ICASSP'82*, pp. 156–159, 1982.

[55] C. Avendano and H. Hermansky, “Study on the dereverberation of speech based on temporal envelope filtering,” *Proc. ICSLP'96*, pp. 889–892, 1996.

[56] J. Mourjopoulos and J. K. Hammond, “Modelling and enhancement of reverberant speech using an envelope convolution method,” Proc. ICASSP'83, pp. 1144–1147, 1983.

[57] 広林 茂樹，野村 博昭，小池 恒彦，東山 三樹夫，“パワーエンベロープ伝達関数の逆フィルタ処理による残響音声の回復，” 電子情報通信学会論文誌 A，Vol. J81，No. 10，pp. 1323–1330，Oct. 1998.

[58] 広林 茂樹，寺島 洋行，山淵 龍夫，“帯域分割を用いたパワーエンベロープ逆フィルタ処理，” 電子情報通信学会論文誌 A，Vol. J83，No. 8，pp. 1029–1033，Aug. 2000.

[59] M. Unoki, M. Furukawa, K. Sakata, and M. Akagi, “An improved method based on the MTF concept for restotring the power envelope from reverberant siganal,” *Acoust. Sci. and Tech.*, Vol. 25, No. 4, pp. 232–242, 2004.

[60] M. Unoki, K. Sakata, and M. Akagi, “A speech dereverberation method based on the MTF concept,” *Proc. EUROSPEECH 2003*, pp. 1417–1420, 2003.

[61] M. Unoki, M. Toi, and M. Akagi, “Refinement of an MTF-based speech dereverberation method using an optimal inverse-MTF filter,” *Proc. SPECOM'06*, Vol. 1, pp. 323–326, 2006.

[62] M. Unoki, M. Toi, and M. Akagi, “Development of the MTF-based speech dereverberation method using adaptive time-frequency division,” *Proc. ForumAcusticum 2005*, pp. 51–56, 2005.

[63] R. Petrick, X. Lu, M. Unoki, M. Akagi, and R. Hoffmann, "Robust Front End Processing for Speech Recognition in Reverberant Environments: Utilization of Speech Characteristics," *Proc. Interspeech 2008*, pp. 22–26, 2008.

[64] Y. Yamasaki and M. Unoki, "Study on a method of suppressing noise based on the MTF concept," *J. Signal Processing*, Vol. 13, No. 4, pp. 335–338, 2009.

[65] X. Lu, S. Matsuda, M. Unoki, and S. Nakamura, "Temporal contrast normalization and edge-preserved smoothing of temporal modulation structures of speech for robust recognition," *Speech Communication*, Vol. 52, No. 1, pp. 1–11, 2010.

[66] M. Unoki and Y. Yamasaki, "MTF-BASED POWER ENVELOPE RESTORATION IN NOISY REVERBERANT ENVIRONMENTS," *Proc. EUSPICO 2009*, pp. CD–ROM, 2009.

[67] K. Kinoshita, M. Delcroix, T. Nakatani, and M. Miyoshi, "Multi-step linear prediction based speech enhancement in noisy reverberant environment," *Proc. Interspeech-2007*, pp. 854–857, 2007.

[68] 吉岡 拓也，中谷 智広，三好 正人，"雑音と残響の同時抑圧による音声強調," 日本音響学会講演論文集，pp. 731–732，March 2008.

[69] 吉岡 拓也，中谷 智広，三好 正人，"雑音・残響抑圧を目的とした線形フィルタに非線形フィルタを後置きさせた系の最適化法," 日本音響学会講演論文集，pp. 845–846，Sept. 2008.

[70] M.R. Schroeder, "Modulation transfer functions: definition and measurement," *Acustica*, Vol. 49, pp. 179–182, 1981.

[71] R. Drullman, J. M. Festen, and R. Plomp, "Effect of temporal envelope smearing on speech reception," *J. Acoust. Soc. Am*, Vol. 95, No. 2, pp. 1053–1064, Feb. 1994.

[72] R. Drullman, J. M. Festen, and R. Plomp, "Effect of reducing slow temporal modulations on speech reception," *J. Acoust. Soc. Am.*, Vol. 95, No. 5, pp. 2670–2680, May 1994.

[73] T. Arai, M. Pavel, H. Hermansky, and C. Avendano, "Intelligibility of speech with filtered time trajectories of spectral envelopes," *Proc. ICSLP 1996*, pp. 2490–2493, 1996.

[74] T. Arai, M. Pavel, H. Hermansky, and C. Avendano, "Syllabel intelligibility for temporally filtered LPC cepstral trajectories," *J. Acoust. Soc. Am.*, Vol. 105, No. 5, pp. 2783–2791, 1999.

[75] N. Kanedera, T. Arai, and H. Hermansky, “On properties of modulation spectrum for robust automatic speech recognition,” *Proc. IEEE ICASSP 1998*, pp. 613–616, 1998.

[76] M. Unoki, K. Sakata, M. Furukawa, and M. Akagi, “A speech dereverberation method based on the MTF concept in power envelope restoration,” *Acoust. Sci. and Tech.*, Vol. 25, No. 4, pp. 243–254, 2004.

[77] 鵜木 祐史，“変調伝達関数に基づく音声信号 (1) -パワーエンベロープ逆フィルタ処理の原理とその応用について-,” *J. Signal Processing*，Vol. 12，No. 5，pp. 339–348，Sept. 2008.

[78] 鵜木 祐史，“変調伝達関数に基づく音声信号 (2) -ブラインド残響音声回復法-,” *J. Signal Processing*，Vol. 13，No. 1，pp. 3–12，Jan. 2009.

[79] Pek Kimhuoch，荒井 隆行，金寺 登，吉井 順子，“変調スペクトルによる雑音下における自動音声区間検出：音声周波数帯域及び変調周波数帯域の検討,” 日本音響学会講演論文集，pp. 155–158，2009.

[80] 鵜木 祐史，“変調伝達関数に基づく音声信号 (3) -残響環境下の基本周波数推定法と残響時間のブラインド推定-,” *J. Signal Processing*，Vol. 13，No. 2，pp. 91–101，March 2009.

[81] S. Hiramatsu and M. Unoki, “A study on the blind estimation of reverberation time in room acoustics,” *J. Signal Processing*, Vol. 12, No. 6, pp. 351–361, 2008.

[82] D. Ying, Y. Shi, X. Lu, J. Dang, and F. Soong, “Robust voice activity detection based on noise eigenspace,” *Acoust. Sci. and Tech.*, Vol. 28, No. 6, pp. 413–423, 2007.

[83] D. Ying, M. Unoki, X. Lu, and J. Dang, “Speech Enhancement Based on Noise Eigenspace Projection,” *IEICE TRANS. INF. and SYST.*, Vol. E92-D, No. 5, pp. 1137–1145, 2009.

[84] 石塚 健太郎，藤本 雅清，中谷 智広，“音声区間検出技術の最近の研究動向,” 日本音響学会誌，Vol. 65，No. 10，pp. 537–543，2009.

[85] M. Unoki, T. Hosorogiya, and Y. Ishimoto, “Comparative evaluation of robust and accurate F0 estimates in reverberant environments,” *Proc. ICASSP2008*, pp. 4569–4572, 2008.

[86] R. Petrick, M. Unoki, A. Mittal, Carlos Segura, and R. Hoffmann, “A Comprehensive Study on the Effects of Room Reverberation on Fundamental Frequency Estimation,” *Proc. Interspeech 2008*, pp. 131–134, 2008.

[87] Y. Ishimoto, M. Unoki, and M. Akagi, "A fundamental frequency estimation method for noisy speech based on instantaneous amplitude and frequency," *Proc. EusroSpeech2001*, pp. 2439–2442, 2001.

謝辞

本研究を進めるにあたり，多大な助言と懇切丁寧かつ，熱心な御指導をして頂きました鵜木祐史准教授に心から感謝致します．本研究を進めるにあたり，多大な助言と熱心な御指導をして頂きました赤木正人教授に心から感謝致します．本研究に関して多大な助言をして頂い李軍峰助教，宮内良太助教，博士後期課程の羽二生篤氏，村上泰樹氏，木谷俊介氏に心より感謝致します．有意義な討論，助言を賜った赤木・鵜木研究室，党・徳田研究室の皆様方に心から感謝いたします．

学会発表リスト

1. S. Morita, M. Unoki, amd M. Akagi, “A study on the MTF-based inverse filtering for the modulation spectrum of reverberant speech,” *2010 RISP International Wrokshop on Nonlinear Circuits and Signal Processing*, Mar. 2010

2. 森田 翔太，鵜木 祐史，赤木 正人，“変調伝達関数に基づいた変調スペクトル逆フィルア処理の検討，” 日本音響学会 2010 年 春期研究発表会，講演論文集，2-P-21，2010.3