

Title	変調知覚メカニズムに着目した騒音低減法の検討
Author(s)	磯山, 拓都
Citation	
Issue Date	2018-09
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/15461
Rights	
Description	Supervisor: 鷓木 祐史, 先端科学技術研究科, 修士 (情報科学)

修士論文

変調知覚メカニズムに着目した騒音低減法の検討

1610008 磯山 拓都

主指導教員 鷗木祐史
審査委員主査 鷗木祐史
審査委員 赤木正人
党建武
吉高淳夫

北陸先端科学技術大学院大学
先端科学技術研究科 [情報科学]

平成30年8月

概要

ヒトは日々様々な音を聞いて生活している。生活に必要な音もあれば、騒音などの不必要な音もある。一般に、老化とともに聴力は低下する。さらに、長時間騒音（定常騒音）に暴露すると聴力の低下は加速する。また、短時間であっても非常に音圧レベルの高い騒音（間欠・衝撃騒音）は聴力低下を加速させる。この問題を解決するためには、「アクティブ耳栓」のような補聴技術が必要となる。その第一歩として、本稿は、定常・間欠・衝撃騒音を選択的に低減するためにヒトの変調知覚メカニズムに基づいた騒音低減法を検討した。

従来の研究の多くは、定常騒音の低減と非定常騒音の低減は別に研究されてきた。これらの研究は、個別には十分に騒音を低減できるが、実環境のように様々な騒音が混じった音に対しては、十分に騒音を低減できるとは言えない。また、同時に騒音を低減する手法では、事前学習を用いて騒音低減をするため、決まった環境に対して十分に騒音を低減できるが、新しい環境では、再度、騒音を学習しなければならないため、決定論的に騒音を低減できず、聴力を守ることができない。

音声の時間的包絡線には特有の特徴があることが知られている。さらに、音の包絡線のスペクトルである、変調スペクトルは音声区間検出の研究にも応用されている。もし、騒音の包絡線情報に特有の特徴があれば、その特徴を除去することで騒音を低減できるかもしれない。そこで本稿は、騒音の特徴を明らかにするために、聴覚フィルタバンク出力における特徴分析と変調フィルタバンク出力における特徴分析を行った。さらに、これらの特徴を用いた騒音低減法を提案した。騒音低減法の評価として騒音低減レベルと相対的な低減レベルと音質評価指標の原音と低減後の音の差を調べた。

聴覚フィルタバンク出力における特徴分析の結果、定常騒音のパワー包絡線は時間方向の標準偏差が狭いことが明らかになった。間欠・衝撃騒音と音声のパワー包絡線は時間方向に標準偏差が広いことが明らかになった。変調フィルタバンク出力における特徴分析からは、定常騒音の変調スペクトログラムは極めて低い変調周波数成分だけを持つことが明らかになった。間欠騒音の変調スペクトログラムは調波性を持つことが明らかになった。衝撃騒音の変調スペクトログラムはあらゆる変調周波数成分を持つことが明らかになった。これらの特徴を除去するために以下の3つの処理をした。(1) 定常騒音の除去として、音声区間のパワー包絡線から非音声区間のパワー包絡線の直流成分を減算した。(2) 間欠騒音の除去として、L次のFIR帯域阻止フィルタ(Band-stop filter: BSF)を用いた。(3) 衝撃騒音の除去として、1次のバターワース型IIR低域通過フィルタ(Low-pass filter: LPF_S)を用いた。騒音低減レベルでは定常騒音では8 dB、間欠騒音では6 dB、衝撃騒音では、8 dBの低減ができることが明らかになった。また、相対的な低減レベルの評価ではSNRが-60 ~ -20 dBの間において7 dBの騒音低減ができることが明らかになった。ラウドネスの減少量の評価では、音圧レベルが100 dBのとき、定常騒音で50 sone、間欠騒音で20 sone、衝撃騒音で50 soneの減少ができることが明らかになった。ラフネスの減少量の評価では音圧レベルが100 dBのとき、間欠騒音で0.7 asper、衝撃騒音で0.3

asper の減少ができることが明らかになった．このことから，変調スペクトル上の特徴を利用することで，定常・間欠・衝撃騒音を十分に低減できることが明らかになった．

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	4
1.4	研究の目的	4
1.5	本論文の構成	4
第2章	騒音低減法の原理	6
2.1	ヒトの変調知覚メカニズム	6
2.1.1	変調知覚	6
2.1.2	音質評価指標	7
2.1.3	変調知覚モデル	9
2.2	変調スペクトルの概念	10
2.3	騒音低減の原理	12
第3章	対象音源の特徴抽出	10
3.1	対象音源	10
3.2	聴覚フィルタの実装	10
3.3	聴覚フィルタバンク出力における特徴	14
3.3.1	聴覚フィルタバンク出力の特徴抽出方法	14
3.3.2	聴覚フィルタバンク出力の特徴抽出結果	14
3.4	変調フィルタバンク出力における特徴	17
3.4.1	変調フィルタバンク出力の特徴抽出方法	17
3.4.2	変調フィルタバンク出力の特徴抽出結果	17
第4章	騒音の特徴を利用した騒音低減	21
4.1	騒音低減法	21
4.2	変調スペクトルの直流成分の除去	24
4.3	変調スペクトルの調波性の除去	24
4.4	変調スペクトルの高域成分の除去	24

第5章	提案法の客観評価	25
5.1	騒音を入力としたときの騒音低減レベル	25
5.1.1	評価方法	25
5.1.2	評価結果	25
5.2	音声と騒音を混ぜた混合音における騒音の相対的低減レベル	28
5.2.1	評価方法	28
5.2.2	評価結果	28
5.3	音質評価指標	30
5.3.1	評価方法	30
5.3.2	評価結果	30
5.4	考察	34
5.5	まとめ	34
第6章	結論	34
6.1	本研究で明らかになったこと	34
6.2	残された課題	35
	謝辞	36
	参考文献	37

目次

1.1	騒音環境下における聴力保護技術	2
1.2	本論文の構成	5
2.1	音の包絡線のレベルと興奮レベル	8
2.2	変調知覚モデル	9
2.3	音声の変調スペクトル	10
2.4	変調知覚メカニズムを用いた変調スペクトル分析のブロックダイアグラム	11
2.5	騒音低減法の原理	13
3.1	ガンマトーンフィルタバンクの周波数特性と等価矩形帯域幅	13
3.2	ガンマトーンフィルタバンク出力から求めた特徴：(a) 時間平均と分散，(b) スペクトル傾斜	16
3.3	白色雑音の変調スペクトログラム	18
3.4	ピンク雑音の変調スペクトログラム	19
3.5	バブル雑音の変調スペクトログラム	20
3.6	マシンガン雑音の変調スペクトログラム	21
3.7	インパルス性雑音の変調スペクトログラム	22
3.8	男性音声の変調スペクトログラム	23
3.9	女性音声の変調スペクトログラム	24
4.1	変調知覚メカニズムに基づいた騒音低減法	22
4.2	騒音低減法にマシンガンノイズを入れたときのパワー包絡線：(a) ガンマトーンフィルタバンクの出力，(b) 変調スペクトルに含まれる定常騒音の特徴の除去，(c) 変調スペクトルに含まれる間欠騒音の特徴の除去，(d) 変調スペクトルに含まれる衝撃騒音の特徴の除去	23
4.3	騒音低減のためのフィルタの周波数特性：(a) 帯域阻止フィルタ，(b) 低域通過フィルタ	25
5.1	騒音を入力としたときの騒音低減レベル：(a) 騒音の音圧レベル 100 dB の場合，(b) 80 dB の場合，(c) 60 dB の場合	27
5.2	混合音における騒音の相対的低減レベル：(a) 騒音の音圧レベル 100 dB の場合，(b) 80 dB の場合，(c) 60 dB の場合	29
5.3	ラウドネスの減少量	31

5.4	シャープネスの減少量	32
5.5	ラフネスの減少量	33

表 目 次

3.1	特徴抽出に用いた音源	11
3.2	ガンマトーンフィルタバンクの構成条件	12

第1章 序論

1.1 はじめに

地球に大気が覆うようになって、生物が誕生した。海の生物であったナメクジのようなピカイアがシーラカンスに進化し、単弓類、サヘラントロプス、ヒトへと進化した。ヒトは音を楽しみ、音に悩まされ、音によってコミュニケーションを行ってきた。ヒトと音は切っても切れない重要な関係がある。ヒトは生活する上で様々な音を聞いている。音声や音楽などの生活に必要な音もあれば、エアコンの音、車の音、電車の音など生活に不必要な音である騒音もある。一般に聴力は老化とともに低下する（加齢難聴）が、長時間の騒音（定常騒音）の暴露は聴力低下を加速させる。また、短時間であっても音圧レベルが非常に高い騒音（間欠・衝撃騒音）の暴露はさらに聴力を低下させる [1]。聴力が低下すると話者の話が聞こえず話の内容を理解することが困難になる。さらに、音の選択的聴取も困難となり、騒音下での音声コミュニケーションが困難になる。一日に音圧レベル平均 70 dB の音を 40 年間聴取すると 10% のヒトに無視できない聴力損失が現れることが知られている [2]。例えば、騒音の音圧レベルが 3 dB 上昇すると 20 年後には無視できない聴力損失が起きる。もし、騒音の音圧レベルを 3 dB 下げることができれば無視できない聴力損失が起きにくくなる。世界保健機関 (World Health Organization: WHO) が聴覚的に許容される一日の騒音曝露量を示したガイドラインがある [27]。職業性の騒音暴露では、 $L_{Aeq,8h}$ (A 特性補正をした音の 8 時間の平均エネルギーに等価な白色雑音の音圧レベル) が 75 dB 以下であれば、生涯に渡って暴露したとしても聴力障害は起きないとされている。また、環境騒音及び娯楽による騒音暴露では $L_{Aeq,24h}$ が 70 dB 以下であれば、障害にかけて騒音暴露したとしても聴力障害にはならないとされている。しかし、日本において、工場労働者、トラック運転手、農業、幼稚園教師、小学校教員、技術者において、一日の騒音曝露量が 75 dB を超えている [4]。有職者のうち、一日あたりの騒音曝露量が 75 dB を超えているヒトは 30 % 程度いる。また、主婦において一日あたり騒音曝露量が 75 dB を超えているヒトは 17 % 程度いる。さらに、乗物においても騒音は暴露する。徒歩・自転車、自家用車においては 75 dB 程度の騒音を暴露している。バス、電車においては、77 dB の騒音を暴露している。バスや地下鉄においては 80 dB を超えた騒音を暴露している。このように、ヒトは日々様々な騒音に暴露し、聴力の低下を加速させている。

この問題を解決するためには図 1.1 に示すような、生活に必要な音は通しつつ、生活に不必要な音を選択的に低減する「アクティブ耳栓」のような聴力保護技術が必要となる。この補聴技術は、難聴になるのを防ぐために、普段から装着し騒音を低減する。今まで、

様々な騒音低減法が考えられてきたが，騒音をすべて低減できる方法は実現できていない．その理由は，すべての騒音が定常ではないという点が挙げられる．間欠騒音は，一回の騒音が数秒程度の長さで，間欠的にでる．衝撃騒音は，継続時間は極めて短い [2]．また，非定常騒音はいつ騒音が到来するかわからない．そこで，いつ騒音が来ても聴力を守るために，決定論的に定常・非定常騒音を低減する手法が重要である．



図 1.1: 騒音環境下における聴力保護技術

1.2 関連研究

騒音を低減する方法は、今まで多くの研究がある。マイクロフォンを複数利用して騒音低減を行う手法 [5]、一つのマイクロフォンで騒音低減をする Boll のスペクトル減算法 [6] やスペクトル減算法を改良した手法 [7]- [15]、アクティブノイズコントロール (ANC) [16]、アナログ的に耳栓を用いて騒音を低減 [17] する手法などが考えられる。

複数のマイクロフォンを用いた手法は、2つのタイプがあり、1つ目は、遅延和マイクロホンアレーである [18]。この手法は、複数のマイクロフォンを利用して超指向性マイクロフォンを作り、所要の音だけを取れるようにし、騒音を低減する。2つ目は、適応マイクロホンアレーである [19]。この手法は、騒音が死角になるようにマイクロフォンアレーを構築することで、騒音低減する。

一方、単マイクロフォンで騒音を低減する手法として Boll のスペクトル減算法がある。この手法は、騒音を定常音と仮定し、非音声区間から事前に推定された雑音スペクトルを観測信号のスペクトルから減算することで騒音を低減する。この手法は定常騒音に対しては非常に高い効果を持つ。しかし、衝撃騒音や間欠騒音といった非定常騒音に十分な効果を持たない。このため、エアコンなどの騒音からは聴力を守ることができるが、電車や工場内で起きる非定常騒音には対応することができない。ANC を用いた騒音低減法は、適応フィルタを用いることで騒音を低減している。この手法は、定常騒音に対して非常に有効だが、非定常騒音には十分に低減できない。

非定常騒音の低減法としてゼロ位相信号を利用した手法がある [20]。インパルス性雑音から変換されたゼロ位相信号は原点周りにのみ大きな値を有する。音声信号のゼロ位相信号は、周期的である。そこで、インパルス性雑音のゼロ位相信号の原点周りの値をインパルス性雑音の次の区間の原点周りの値に置き換えることで、騒音を低減する。また、Yoshioka らによる統計量を用いた騒音低減法がある [21]。この手法は、ベイジアン推定法に基づいて非定常雑音を低減している。非定常騒音の低減法として Qu らによるウェーブレット変換を用いた騒音低減法がある [22]。この手法は、ウェーブレット係数を閾値処理し、信号を再構成してノイズを抑圧する。しかし、これらの手法は、非定常騒音は十分に低減できるが、定常騒音を低減することができない。

定常騒音・非定常騒音を同時に低減する手法として、非負スペクトル分解を用いた騒音低減法がある [23]。この方法は、事前学習を用いて雑音特性を訓練することで、騒音を低減する。しかし、この手法は、事前学習が必要なため、学習データによって騒音の低減する能力が変化するという問題がある。

耳栓を用いた騒音の低減法は、耳を守る点においては有効であるが、生活に必要な音も同時に低減してしまうという問題がある。

1.3 問題点

複数のマイクロフォンを利用する手法では、規模が大きくなり、日常生活で使うことができない。また、単マイクロフォンでは、従来の研究の多くは、定常騒音の低減と非定常騒音の低減は別に研究されてきた。これらの研究は、個別には十分に騒音を低減できるが、実環境のように様々な騒音が混じった音に対しては、十分に騒音を低減できるとは言えない。また、同時に騒音を低減する手法では、事前学習を用いて騒音低減をするため、決まった環境に対して十分に騒音を低減できるが、新しい環境では、再度、騒音を学習しなければならないため、様々な環境で騒音を低減できず、聴力を守ることができない。また、耳栓による騒音低減法は、聴力を維持する点では、非常に有効的であるが、音声や音楽など生活に必要な音も同時に低減してしまうという問題がある。そこで、いつ騒音暴露しても聴力を守れるように、定常・非定常騒音を低減する手法が重要である。

1.4 研究の目的

騒音暴露から聴力保護を行うにあたり、学習の必要な低減法は、様々な環境に対応することができない。そこで、本研究は、騒音の暴露から聴力を守るために、定常・間欠・衝撃騒音を決定論的に低減する手法を検討する。

ヒトは音の包絡線を知覚できる。音声の包絡線には、音声知覚において重要な特徴があることが知られている [24]。また、音声の包絡線の周波数情報である変調スペクトルには 4 Hz 付近にピークを持ち、16 Hz 以下に非言語情報があることが知られている [25]。この変調スペクトルを用いた研究に金井らの音声区間検出がある [26]。この研究は、音声とその他の音を変調スペクトルを利用して判断している。もし、定常・間欠・衝撃騒音の変調スペクトル上に特有の特徴があれば、その特徴を除去することで定常・間欠・衝撃騒音を低減できるかもしれない。本論文の目的は、定常・間欠・衝撃騒音を低減するためにヒトの変調知覚メカニズムに基づいて騒音低減法を検討することである。

1.5 本論文の構成

本論文は、6章で構成される。図 1.2 に本論文の構成を図示する。

1章は序論であり、本論文で対象とする研究課題と研究の目的を述べる。

2章は騒音低減法で利用する変調知覚メカニズムと変調スペクトルおよび、騒音低減法の原理について述べる。

3章は定常・間欠・衝撃騒音の特徴を調べるために、各騒音の聴覚フィルタバンク出力における特徴抽出と変調フィルタバンク出力における特徴抽出を述べる。

4章は3章で求めた騒音の特徴を打ち消す処理と騒音低減法について述べる。

5章は騒音低減法の評価方法および、評価結果について述べる。

6章は結論であり、本研究で明らかになったこと、残された課題について述べる。

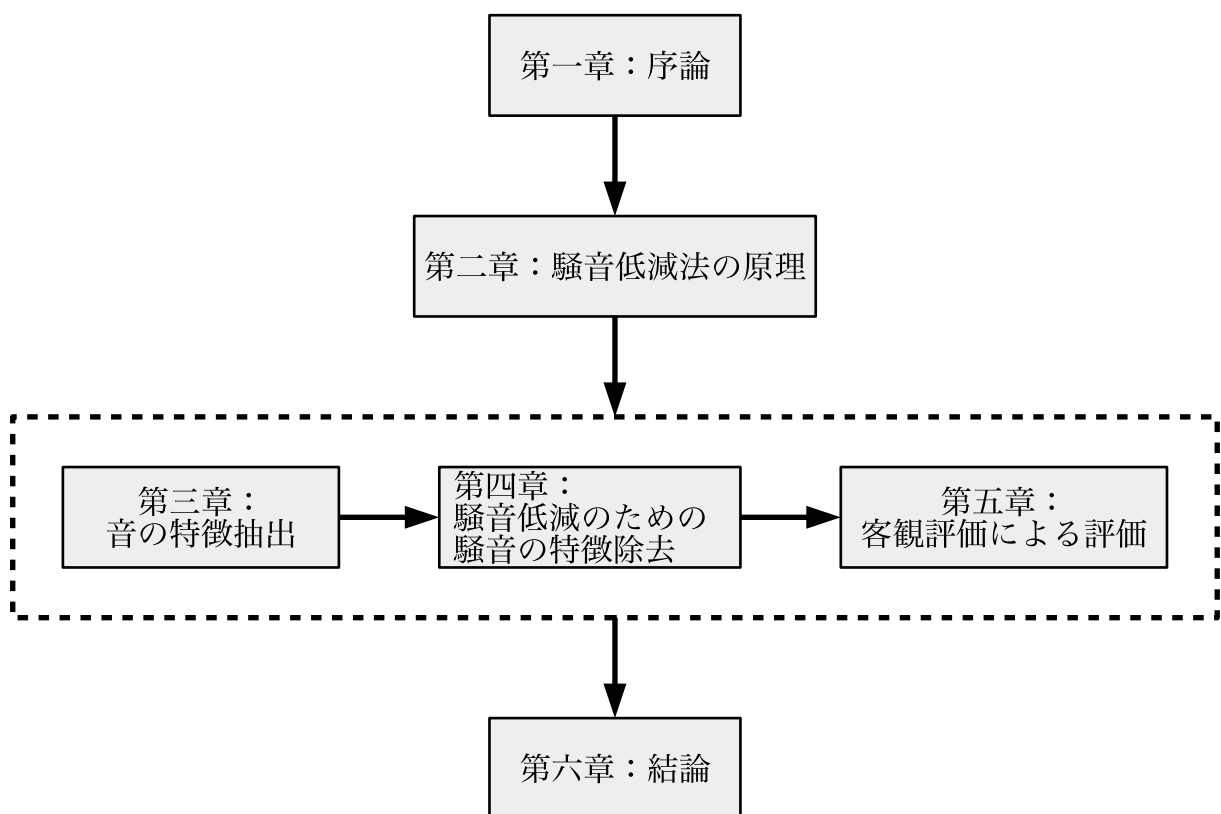


図 1.2: 本論文の構成

第2章 騒音低減法の原理

2.1 ヒトの変調知覚メカニズム

2.1.1 変調知覚

Robert らは音声の時間的包絡線と音声認識の関係を調べた [27]。この研究では、音声の微細構造を取り除くために、まず、音声を帯域分割し、振幅包絡線と微細構造に分け、振幅包絡線を低域通過フィルタ処理する。そして、微細構造と白色雑音を入れ替え、各帯域を足し合わせて雑音駆動音声を作る。このとき、音声の帯域分割数は 1, 2, 3, 4 である。聴取実験では、母音、子音の認識、書き取りを行った。この研究から、音声のスペクトル情報が著しくくずれているにもかかわらず、帯域数が 3 つ以上のときに高い音声認識性能が示された。また、風間らは、狭帯域音声波形包絡線に現れる話者情報について検討した [28]。その結果、狭帯域包絡線の帯域間相互相関行列には話者の個人性が含まれていることを明らかにした。一方で、聴覚メカニズムを利用した Ryosuke らの検討から、4 ~ 8 Hz の変調周波数は言語知覚において重要で有ることが示された [29]。朱らは言語と話者個性情報の知覚に関連する変調周波数帯域について研究を行った [25]。変調知覚モデルの低域通過フィルタのカットオフ周波数を 1 ~ 30 Hz と変化させたときの言語情報と話者の個人性について検討した。その結果、言語情報は、振幅包絡線の 4 Hz の成分が重要であり。また、個人性においては振幅包絡線の 16 Hz 以下の成分が重要であることが示された。このように音声の包絡線には言語知覚や個人性などの重要な情報がある。

2.1.2 音質評価指標

騒音の音質評価をするために、ヒトの知覚に基づいた音質評価指標がある [32]。音質評価とは、ヒトから得た心理量と物理量を結びつけたものである。この指標は、音の大きさを表すラウドネス、音の甲高さを表すシャープネス、音のザラツキ感を表すラフネスなどがある。これらなかで特に、ラフネスは振幅包絡線に関係する指標である。ラフネスは、振幅包絡線と周波数の変動によって変化する。音の変調度が1のときに一番ザラツキ感を感じるということが知られている。さらに、ラフネスは、臨界帯域と関係があることが知られている。音の搬送波周波数（微細構造）が1 kHz 以下の場合、臨界帯域幅の1/2の変調周波数でラフネスが一番高い値になる。一方、搬送波周波数が2 kHz より高い場合、変調周波数が75 Hz のときにラフネスの値が一番高い値になる。ラフネスは音圧レベル (SPL) の影響が少なく、音圧レベルが20 dB 増加すると、ラフネスは2倍になる。

Terhardt は上記の物理量とラフネスの関係から、下記のモデルを提案した。

$$R = A(df_m, f_c)m^2 2^{\frac{SPL-40}{20}} \quad (2.1)$$

ここで、 f_m は変調周波数、 f_c は搬送波周波数、 m は変調度、 SPL は音圧レベルである。

一方、Fastl は時間マスクパターンに基づいたラフネスのモデルを提案した。図 2.1 に音の包絡線のレベルと興奮レベルの模式図を示す。ここで、 ΔL は興奮レベルの谷と山の差である。音の変動に応じで、時間マスクパターンを加味した興奮レベルが変動することが知られている。音の粗さは興奮レベルの変動に依存する。 ΔL は、変調周波数が低いほど大きくなるが、低すぎるとラフネスの値は低くなる。また、変調周波数が高いと、時間マスクの影響で ΔL は小さくなる。このことから、Fastl のモデルでは、臨界帯域ごとに ΔL を求め、すべての帯域の ΔL を足し合わせたものを、分析音の ΔL として、次式で定義されるラフネスを求める。

$$R = 0.3f_m \int_0^{24Bark} \Delta L(z) dz \quad [asper] \quad (2.2)$$

音圧レベルが60 dB、搬送波周波数が1 kHz、変調周波数が70 Hz のときにラフネスが1と定義されている。

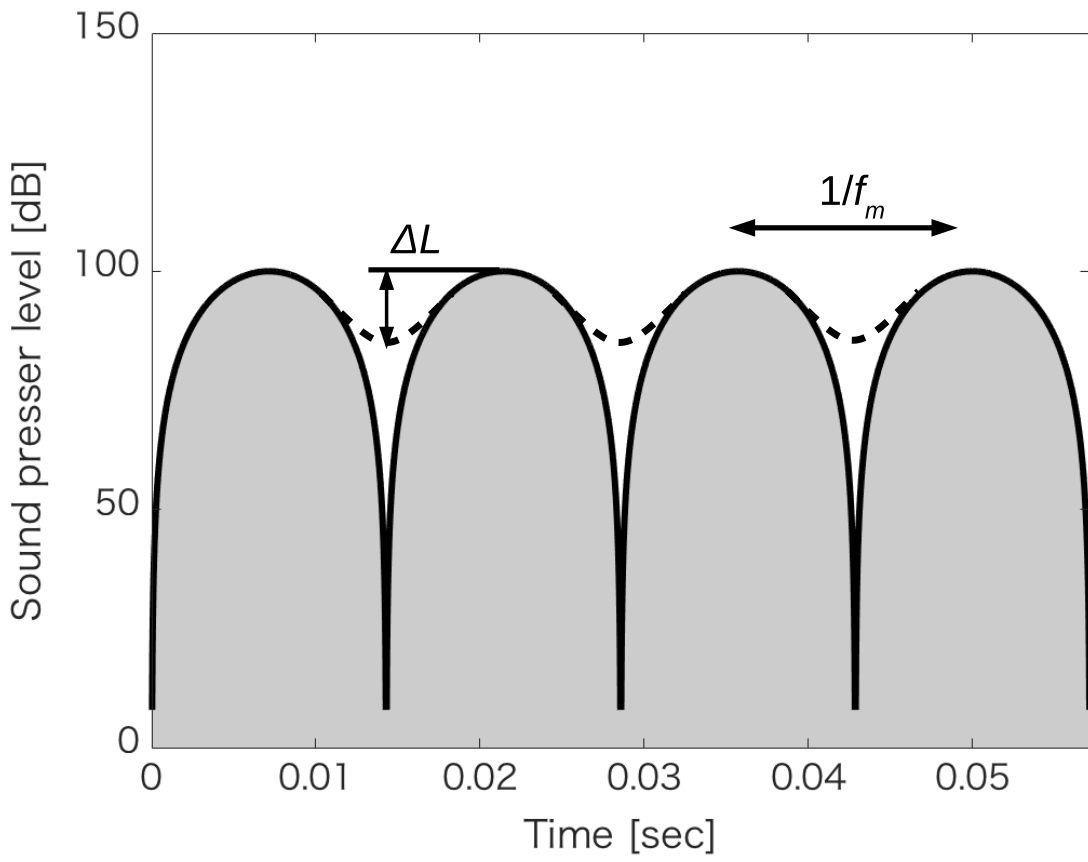


図 2.1: 音の包絡線のレベルと興奮レベル

2.1.3 変調知覚モデル

Dua らは，聴覚系における信号処理を記述するモデルを提案した [33]．図 2.2 に変調知覚に関する実験データを記述するために提案された拡張モデルを示す．信号は前処理として聴覚フィルタバンク（ガンマトーンフィルタバンク）によって帯域分割する．そして，半波整流とカットオフ周波数 1 kHz の低域通過フィルタ処理を行う．次に適応され，線形フィルタバンクがエンベロップの振幅変化をさらに解析すると仮定して，変調フィルタバンクによってフィルタリングする．最後に内部雑音に加えられ，決定処理を行う．この処理は信号を内部表現に変換する．

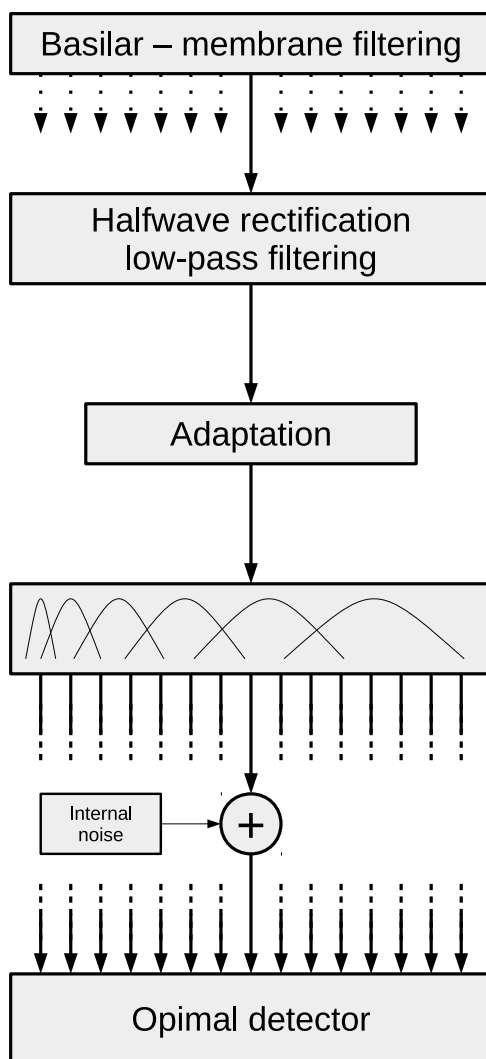


図 2.2: 変調知覚モデル

2.2 変調スペクトルの概念

変調スペクトルとは、音の時間的包絡線のスペクトル情報のことである。図 2.3 に音声の変調スペクトルを示す。変調スペクトルは変調周波数が高ければ高いほど、分析した音は早く変動し、変調周波数が低いと、分析した音は遅く変動することを表す。音声の場合、変調スペクトルの 4 Hz 付近にピークが立つことが知られている [24]。これは、ヒトは一文字 0.25 sec の速度でしゃべることを意味している。

図 2.4 に変調知覚メカニズムを用いた変調スペクトル分析のブロック図を示す。はじめに原信号 $s(t)$ を聴覚フィルタバンクにより、 k 個の周波数帯域 $x_k(t)$ に分解する。次に、各帯域成分のパワー包絡線 $e_k^2(t)$ を次式より求め、次式よて定義する変調スペクトル $E_k(f_m)$ を得る。

$$e_k^2(t) = \text{LPF} (|x_k(t) + j \cdot \text{Hilbert}(x_k(t))|^2) \quad (2.3)$$

$$E_k(f_m) = |\text{DFT}(e_k^2(t))| \quad (2.4)$$

ここで、 k は聴覚フィルタバンクのチャンネル番号、 $\text{Hilbert}(\cdot)$ はヒルベルト変換、 $\text{LPF}(\cdot)$ はカットオフ周波数 64 Hz のローパスフィルタ、 $(\cdot)^2$ は二乗処理、 DFT は離散フーリエ変換、 $|\cdot|$ は絶対値、 f_m は変調周波数である。変調周波数の上限は、 LPF のカットオフ周波数に対応し、64 Hz である。

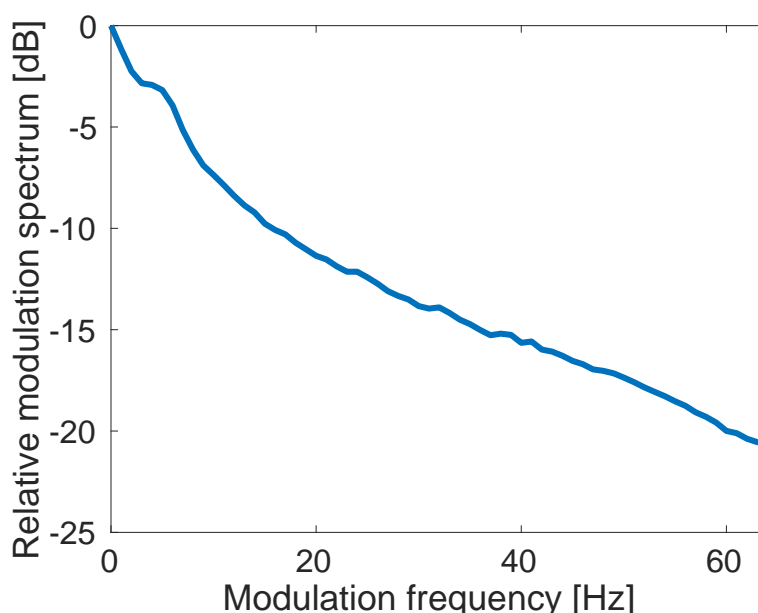


図 2.3: 音声の変調スペクトル

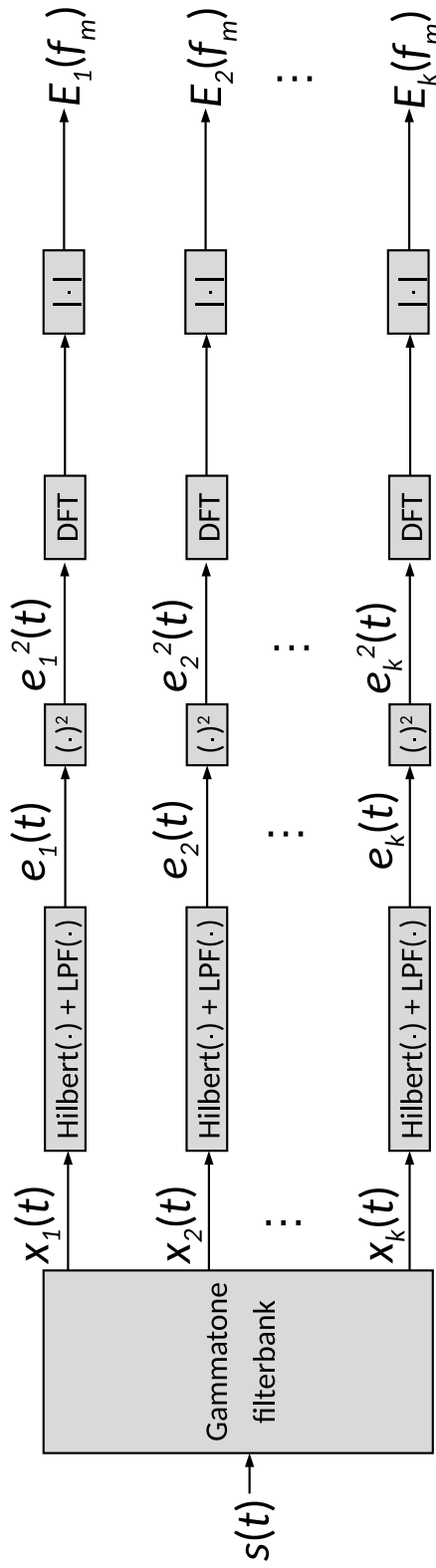


図 2.4: 変調知覚メカニズムを用いた変調スペクトル分析のブロックダイアグラム

2.3 騒音低減の原理

上記で説明したことを基にして、騒音低減法の原理を説明する。この手法は変調知覚モデルに準じている。図 2.5 に騒音低減法の原理の図を示す。はじめに、原信号 $s(t)$ を帯域通過フィルタバンクによって、 k 個の周波数帯域 $x_k(t)$ に分解する。次に、各帯域で変調フィルタバンクを用いてパワー包絡線 $e_k^2(t)$ を求める。各騒音のパワー包絡線上の特徴を除去する。最後に、逆ガンマトーンフィルタバンクを利用して騒音が低減された音信号 $y(t)$ を得る。

3章で帯域通過フィルタバンク出力の特徴と変調フィルタバンク出力に特徴を明らかにする。4章で3章で求めた特徴を利用した騒音低減法を実装する。5章で騒音低減法を評価する。

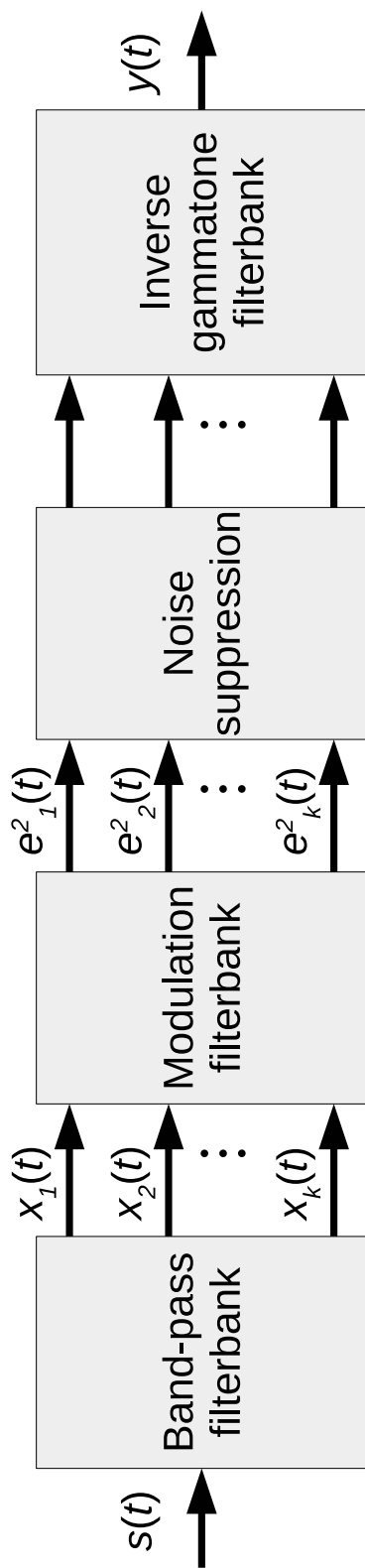


図 2.5: 騒音低減法の原理

第3章 対象音源の特徴抽出

騒音低減法で利用する騒音の特徴を調べるために聴覚フィルタバンク出力における特徴と変調フィルタバンク出力における特徴を調べる，

3.1 対象音源

表 3.1 に分析対象音源を示す． f_s は音源のサンプリング周波数である．分析では，サンプリング周波数をすべて 44.1 kHz にした．騒音として NOISEX-92 [34] に収録された雑音データを利用した．定常騒音として白色雑音，ピンク雑音，バブル雑音を，間欠騒音としてマシンガン騒音を，衝撃騒音としてインパルス性雑音を使用した．音声信号として親密度別単語了解度試験用音声データセット 2007 (FW07) [35] から 4 モーラ単語の男性音声と女性音声を使用した．

3.2 聴覚フィルタの実装

聴覚フィルタバンク（聴覚末梢系の周波数分析処理）の一次近似としてガンマトーンフィルタバンクがある．ここでは，聴覚フィルタの中心周波数を f_k として，次式で定義されるガンマトーンフィルタのインパルス応答を利用する [36] ．

$$g_k(t) = at^{(N-1)} \exp(-2\pi b \text{ERB}_N(f_k)t) \cos(2\pi f_k t + \phi) \quad (3.1)$$

a は振幅， t は時間， N は次数， b は定数， ϕ は位相である．等価矩形帯域幅 $\text{ERB}_N(f_k)$ は，聴覚フィルタの帯域幅であり，次式のように定義される．

$$\text{ERB}_N(f_k) = 24.7(4.37f_k + 1) \quad (3.2)$$

フィルタバンクの構成では， k 個のガンマトーンフィルタを，次式で定義される ERB_N -number に従って配置する．

$$\text{ERB}_N\text{-number} = 21.4 \log_{10}(4.37f_k/1000 + 1) \quad (3.3)$$

ここで， ERB_N -number が 1~35 と並ぶように中心周波数 f_k を求める．

表 3.1: 特徴抽出に用いた音源

音源	f_s [Hz]	データサンプル数	時間 [sec]
白色雑音	19,980	1	235
ピンク雑音	19,980	1	235
バブル雑音	19,980	1	235
マシンガン雑音	19,980	1	235
インパルス性雑音	19,980	4	1
男性音声	48,000	400	1
女性音声	48,000	400	1

表 3.2 に示す構成条件に基づき、ガンマトーンフィルタバンクを構築した。図 3.1 にガンマトーンフィルタバンクの周波数特性と等価矩形帯域幅を示す。横軸は周波数、縦軸はフィルタゲインを示す。青色の点線は等価矩形帯域幅を、実線はガンマトーンフィルタの周波数特性を示す。この図から、周波数が高くなるにつれ聴覚フィルタの帯域幅が広がることがわかる。これは、ガンマトーンフィルタバンクが定 Q フィルタの特性を持っているためである。

次に、聴覚フィルタバンクの出力の導出について説明する。各チャンネル出力 $x_k(t)$ は、次式のように原信号 $s(t)$ と k 番目のガンマトーンフィルタのインパルス応答 $g_k(t)$ の畳み込み演算から得られる。

$$x_k(t) = s(t) * g_k(t) \quad (3.4)$$

ただし、*は畳み込み演算子である。

最後に、構築したガンマトーンフィルタバンクの周波数特性ならびに信号の再合成の精度を次式で定義される SNR を用いて評価する。

$$\text{SNR} = 10 \log_{10} \frac{\int_0^T s^2(t) dt}{\int_0^T (s(t) - \hat{s}(t))^2 dt} \quad (3.5)$$

ただし、 $s(t)$ は原信号、 $\hat{s}(t)$ は再合成音である。

評価には、200 個の音声刺激を利用した。評価の結果、SNR の平均は 19.7 [dB] であった。このことから、本稿で構築したガンマトーンフィルタバンクは分析合成系として十分な周波数分析・信号再合成の能力をもつことが確認できる。

表 3.2: ガンマトーンフィルタバンクの構成条件

チャンネル番号 k	中心周波数 f_k [Hz]	等価矩形帯域幅 $ERB_N(f_k)$
1	26	27
2	55	30
3	87	34
4	123	38
5	163	42
6	207	47
7	257	52
8	312	58
9	374	65
10	442	72
11	518	81
12	603	90
13	698	100
14	803	111
15	921	124
16	1051	138
17	1196	154
18	1358	171
19	1539	191
20	1739	212
21	1963	237
22	2212	263
23	2489	293
24	2798	327
25	3142	363
26	3525	405
27	3951	451
28	4426	502
29	4955	560
30	5544	623
31	6200	694
32	6930	773
33	7743	861
34	8649	958
35	9657	1067

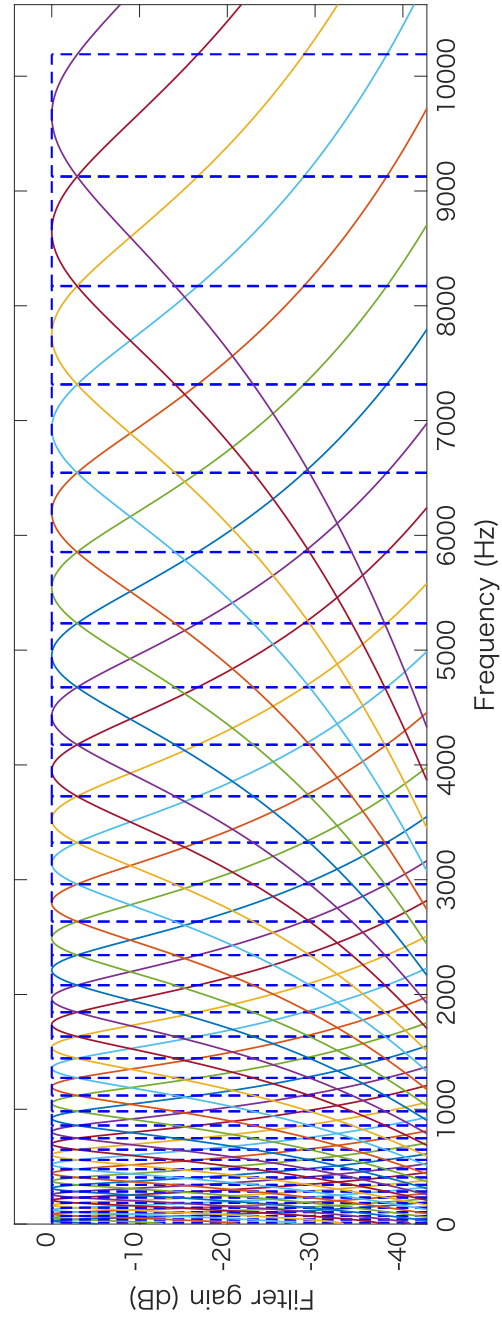


図 3.1: ガンマトーンフィルタバンクの周波数特性と等価矩形帯域幅

3.3 聴覚フィルタバンク出力における特徴

3.3.1 聴覚フィルタバンク出力の特徴抽出方法

ガンマトーンフィルタバンクの出力から，信号の各周波数成分の時間変動の特徴を分析するために，次式で定義されるチャンネル出力のパワー $x_k^2(t)$ の時間平均とその標準偏差を調べる．

$$\mu_k = \frac{1}{T} \int_0^T x_k^2(t) dt \quad (3.6)$$

$$\sigma_k = \sqrt{\frac{1}{T} \int_0^T (x_k^2(t) - \mu_k)^2 dt} \quad (3.7)$$

ただし， T は分析区間長であり，1,000 [ms] とした．音楽や雑音などのデータ長が長い場合は，1,000 [ms] の分析長で 250 [ms] ずつシフトしながら (1/4 オーバーラップ)，平均と標準偏差を求める．

ガンマトーンフィルタバンクの出力から，信号の周波数方向の特徴を分析するために，チャンネル方向のスペクトル傾斜を調べる．ここでは，次式の評価関数が最小となるように μ_k の平均レベルをスペクトルとみなして，一次回帰分析によりスペクトル傾斜を求める．

$$E = \sum_k (10 \log_{10}(\mu_k) - (\alpha k + \beta))^2 \quad (3.8)$$

3.3.2 聴覚フィルタバンク出力の特徴抽出結果

表 3.1 の音源に対して，ガンマトーンフィルタバンクによる周波数分析ならびに変調スペクトル分析を行った．

まず，各音源に対するガンマトーンフィルタバンクによる周波数分析の結果をみる．図 3.2(a1)～図 3.2(a9) に各音源に対する時間方向のパワーの平均とその標準偏差を示す．図 3.2(a) の横軸はガンマトーンフィルタバンクのチャンネル k を，縦軸は音信号のパワーを示す．図 3.2(a) より白色雑音，ピンク雑音，バブル雑音といった定常騒音は標準偏差が小さいことがわかる．一方、マシンガン雑音，インパルス性雑音，男性音声，女性音声は標準偏差が他の音源のものより大きいことがわかる．

図 3.2(b1)～図 3.2(b9) にスペクトル傾斜の分析結果を示す．図 3.2(b) の横軸はガンマトーンフィルタバンクのチャンネル k を，縦軸はパワーレベル [dB] を示す．また，赤色の破線は一次回帰分析により得られたスペクトル傾斜を示す．図 3.2(b) より，音声のスペクトルは帯域性をもつため，スペクトル傾斜を求めたときの誤差は他のものより大きくなっていることがわかる．また，男性音声，女性音声，ピンク雑音では，スペクトル傾斜がほぼ平坦であることがわかる．バブル雑音，マシンガン雑音ではスペクトル傾斜が右下

がり (α が負の値) になっていることがわかる。白色雑音, インパルス性雑音ではスペクトル傾斜が右上がり (α が正の値) になっていることがわかる。なお, ガンマトーンフィルタバンクは定 Q フィルタの特性をもつため, 白色雑音のように平坦なスペクトルをもつ成分は, フィルタバンクの出力上では右上がりの傾斜をもつことに注意されたい。

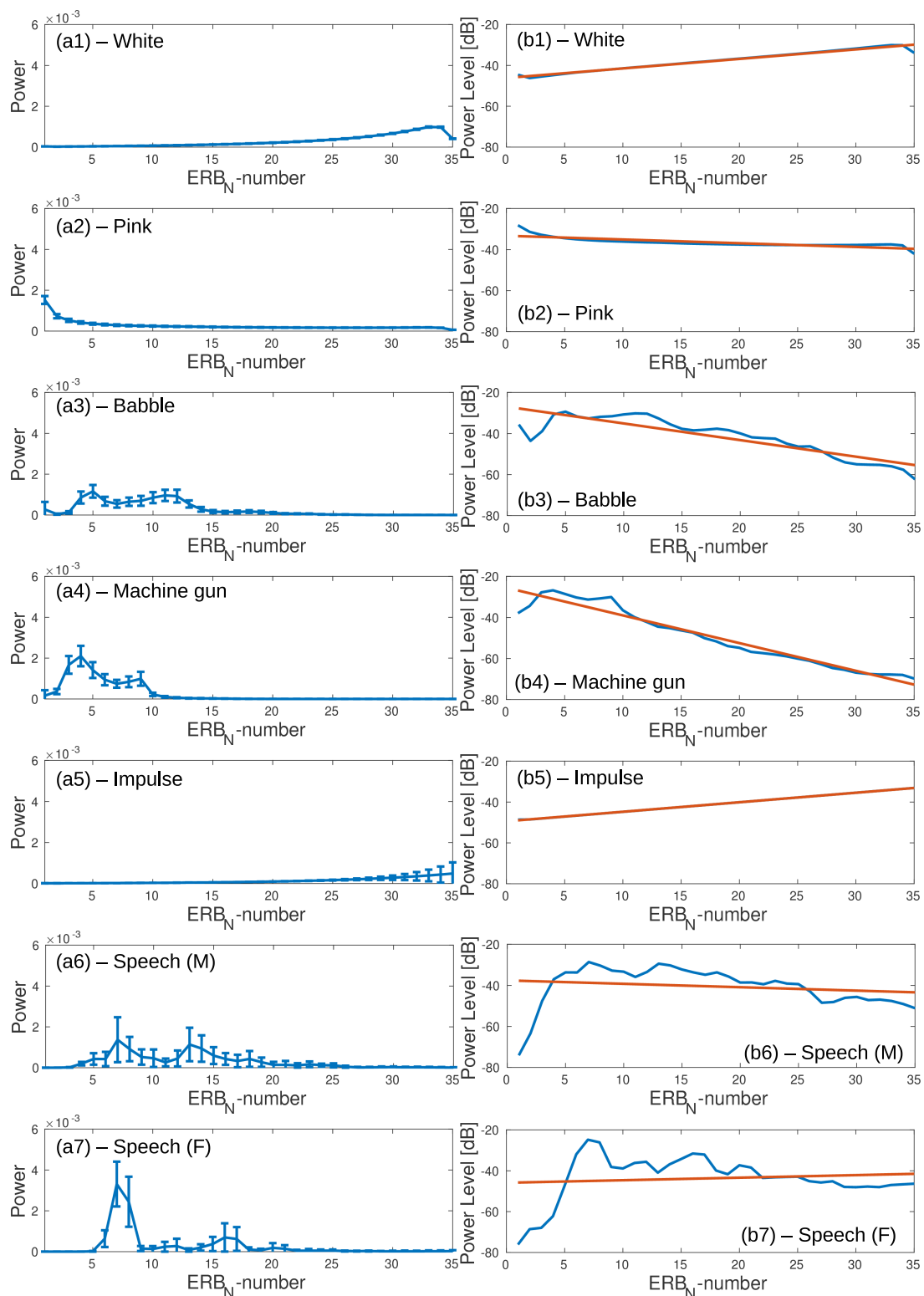


図 3.2: ガンマトーンフィルタバンク出力から求めた特徴：(a) 時間平均と分散，(b) スペクトル傾斜

3.4 変調フィルタバンク出力における特徴

3.4.1 変調フィルタバンク出力の特徴抽出方法

ガンマトーンフィルタバンク出力の時間変動の分析方法（変調スペクトル分析）について説明する．まず，ガンマトーンフィルタバンクの出力から，次式を利用して各周波数帯域のパワー包絡線を求める．

$$e_k^2(t) = \text{LPF} (|x_k(t) + j \cdot \text{Hilbert}(x_k(t))|^2) \quad (3.9)$$

この方法は，信号の瞬時振幅の計算に基づいたものであり，パワー包絡線に含まれる高周波成分を取り除くための後処理として低域通過フィルタを利用している．次に，次式を利用して，パワー包絡線の離散 Fourier 変換から変調スペクトルを求める [37]．

$$E_k(f_m) = |\text{DFT}(e_k^2(t))| \quad (3.10)$$

ここで，DFT は離散フーリエ変換である．

3.4.2 変調フィルタバンク出力の特徴抽出結果

図 3.3～図 3.9 に音源の変調スペクトログラムを示す．図中の横軸は変調周波数 Hz，縦軸はガンマトーンフィルタバンクのチャンネル，色の濃淡は変調周波数 0 Hz でスペクトルレベルを 0 dB としたときの変調スペクトルである．図 3.3 は，白色雑音の変調スペクトログラム，図 3.4 は，ピンク雑音の変調スペクトログラム，図 3.5 は，バブル雑音の変調スペクトログラム，図 3.6 は，マシンガン雑音の変調スペクトログラム，図 3.7 は，インパルス性雑音の変調スペクトログラム，図 3.8 は，男性音声の変調スペクトログラム，図 3.9 は，女性音声の変調スペクトログラムを表す．これらの結果から，白色雑音，ピンク雑音，バブル雑音などの定常騒音は直流成分付近の変調周波数成分だけをもつことがわかる．マシンガン雑音などの間欠騒音は調波性をもつことがわかる．また，今回の音源に関しては 8 Hz を基本変調周波数とする調波性が現れた．インパルス性雑音はあらゆる変調周波数成分をもつことがわかる．

男性音声と女性音声は変調周波数 4 Hz 付近にスペクトルのピークをもつ傾向があることがわかる．

この結果より，定常騒音はパワー包絡線の直流成分を除去することで，間欠騒音は変調スペクトル上の調波性を除去することで，衝撃騒音は，変調スペクトルの高い変調周波数成分を除去することで，定常・間欠・衝撃騒音を低減できると考えられる．

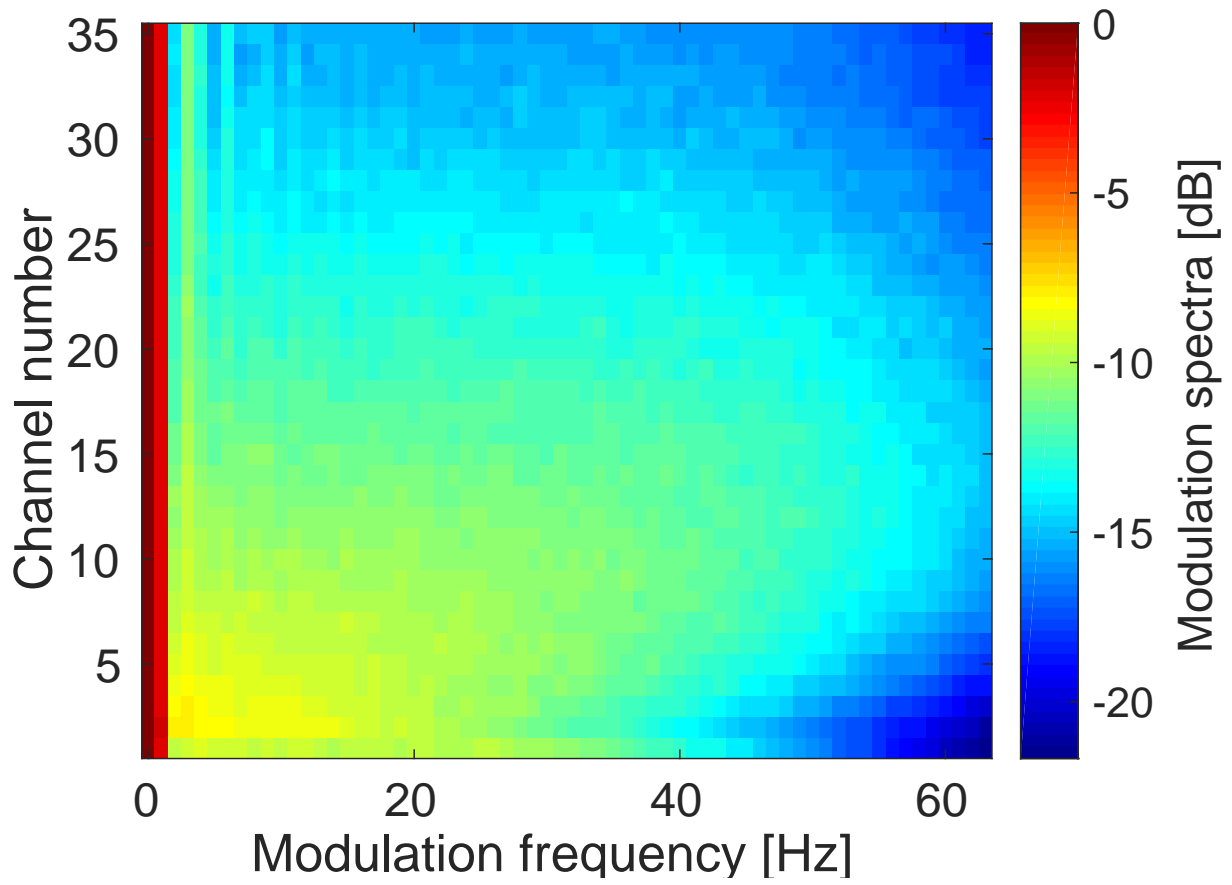


図 3.3: 白色雑音の変調スペクトログラム

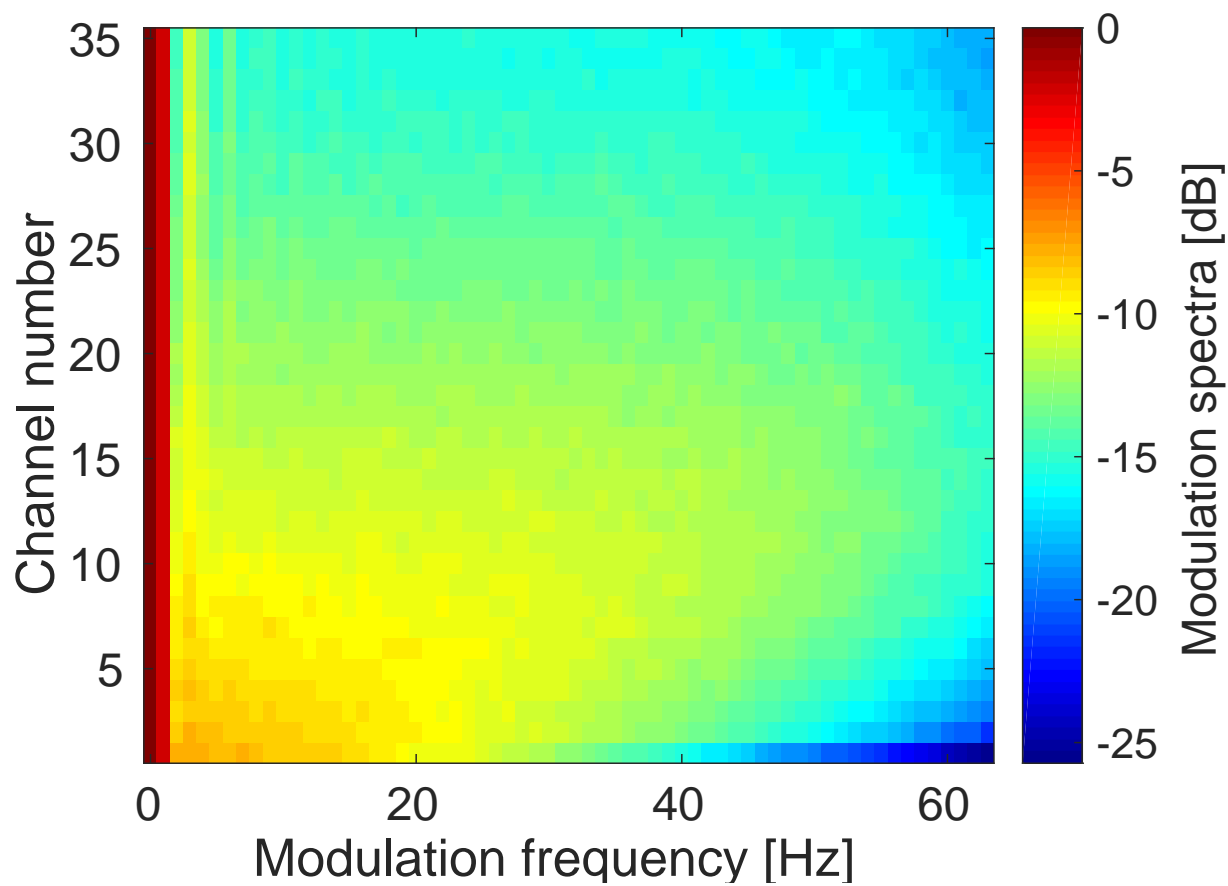


図 3.4: ピンク雑音の変調スペクトログラム

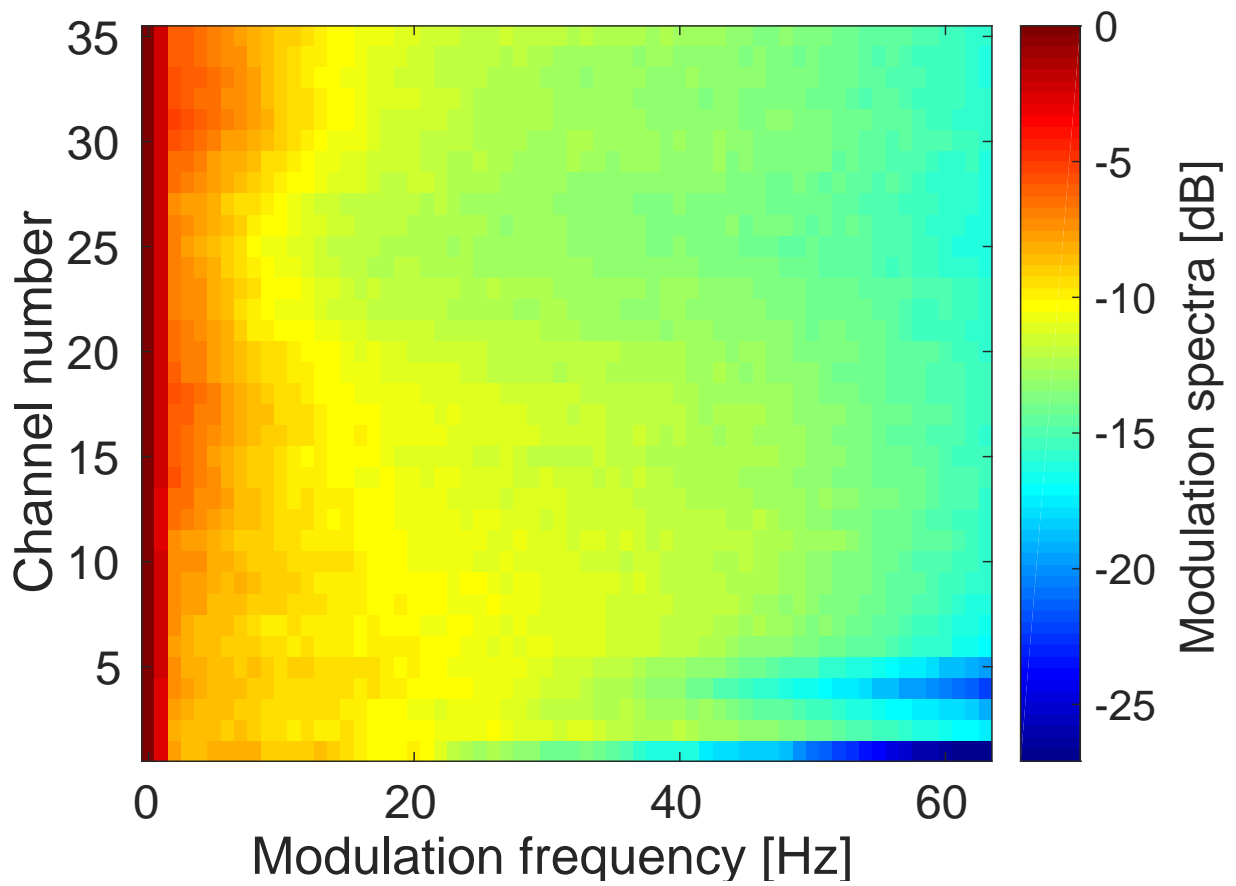


図 3.5: バブル雑音の変調スペクトログラム

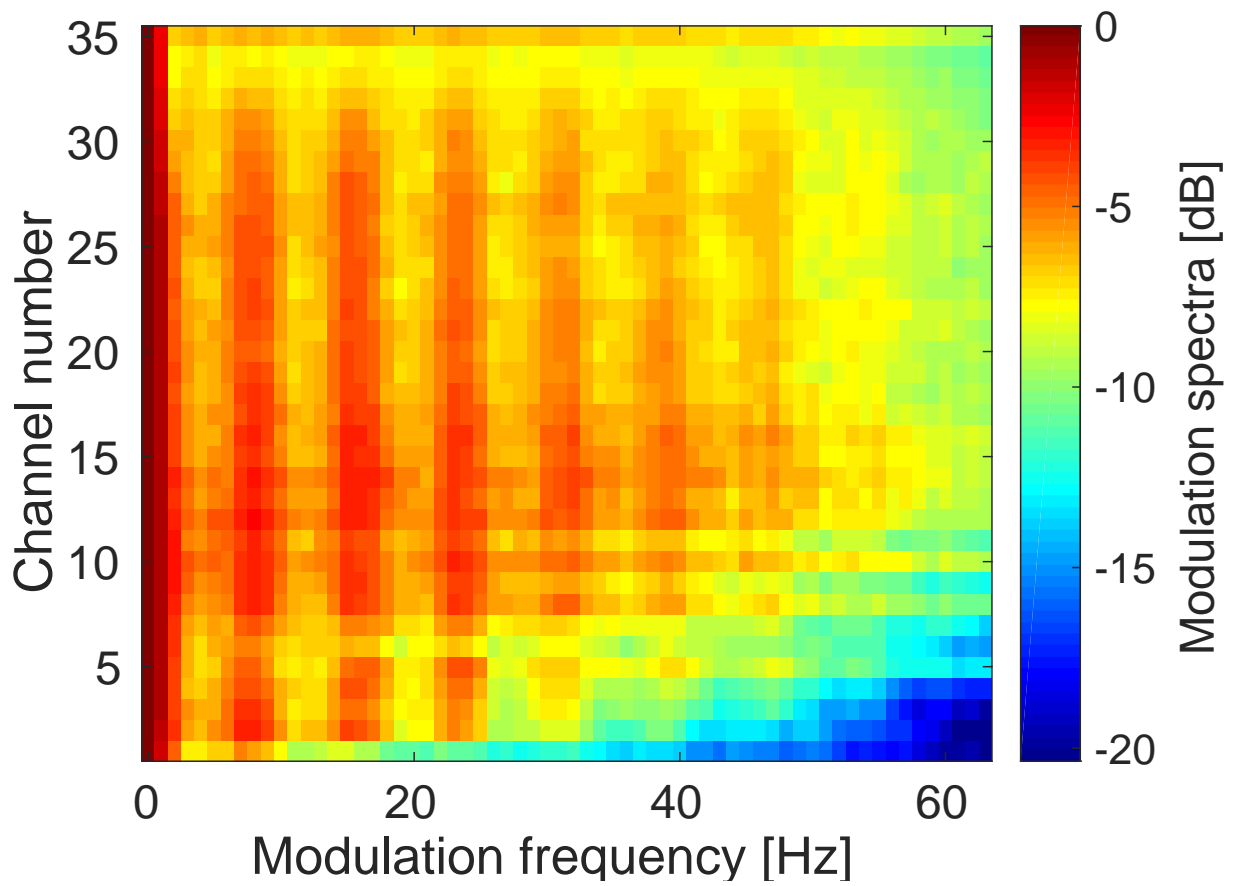


図 3.6: マシンガン雑音の変調スペクトログラム

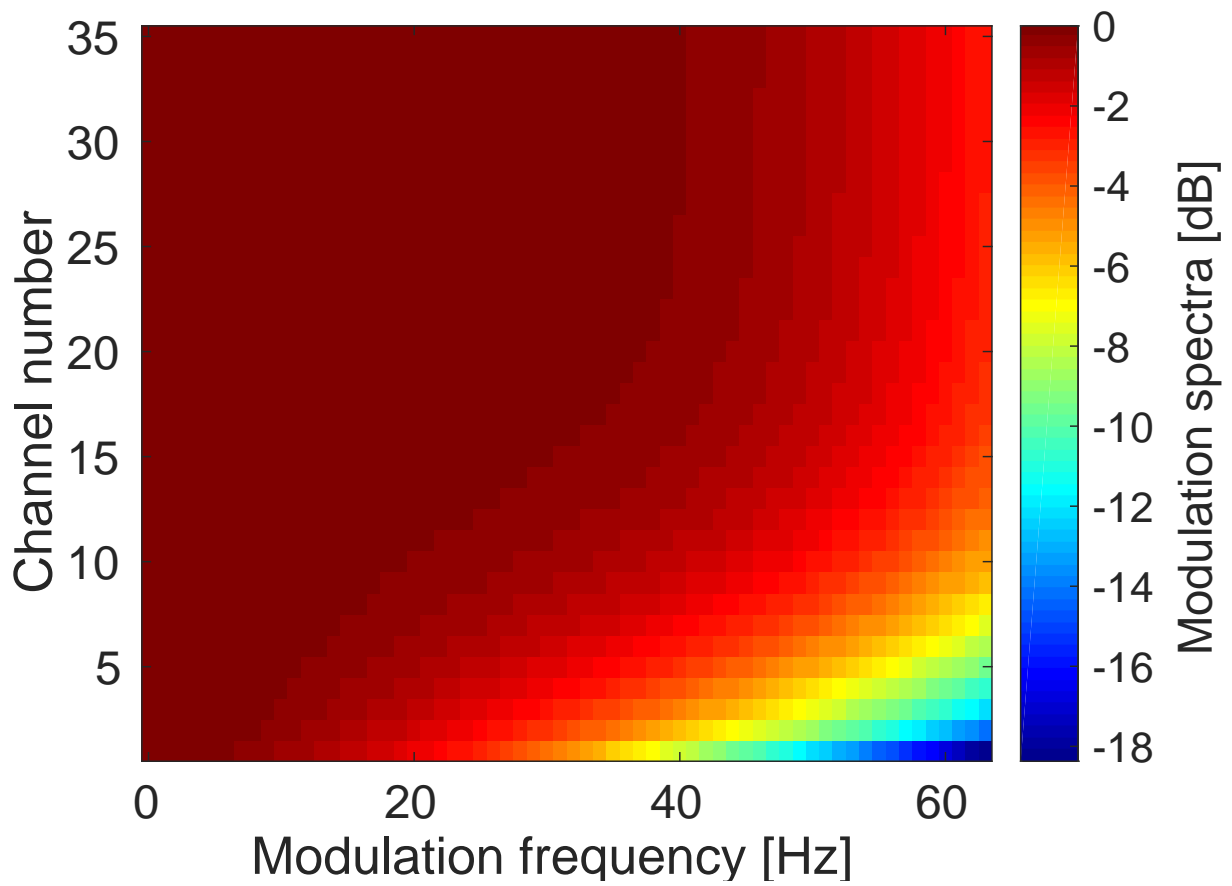


図 3.7: インパルス性雑音の変調スペクトログラム

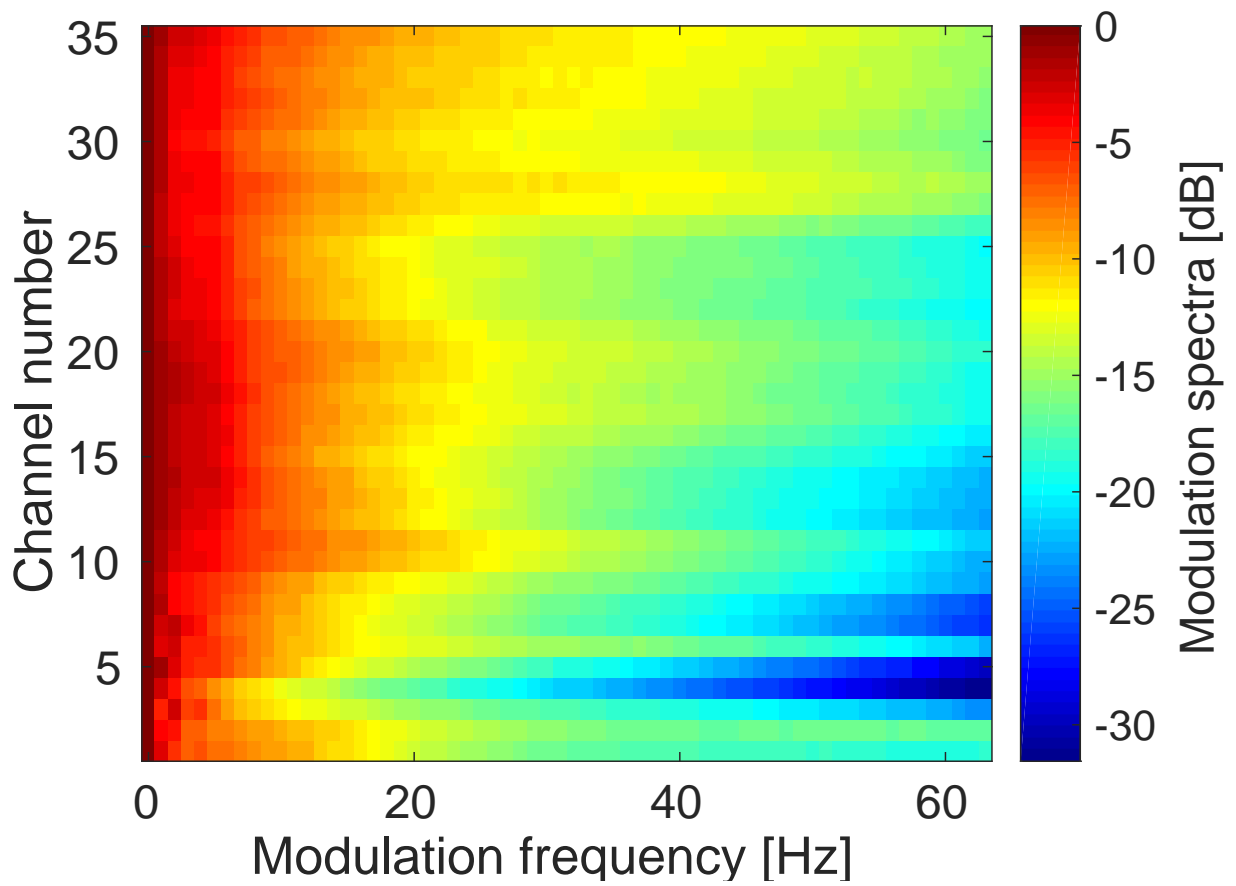


図 3.8: 男性音声の変調スペクトログラム

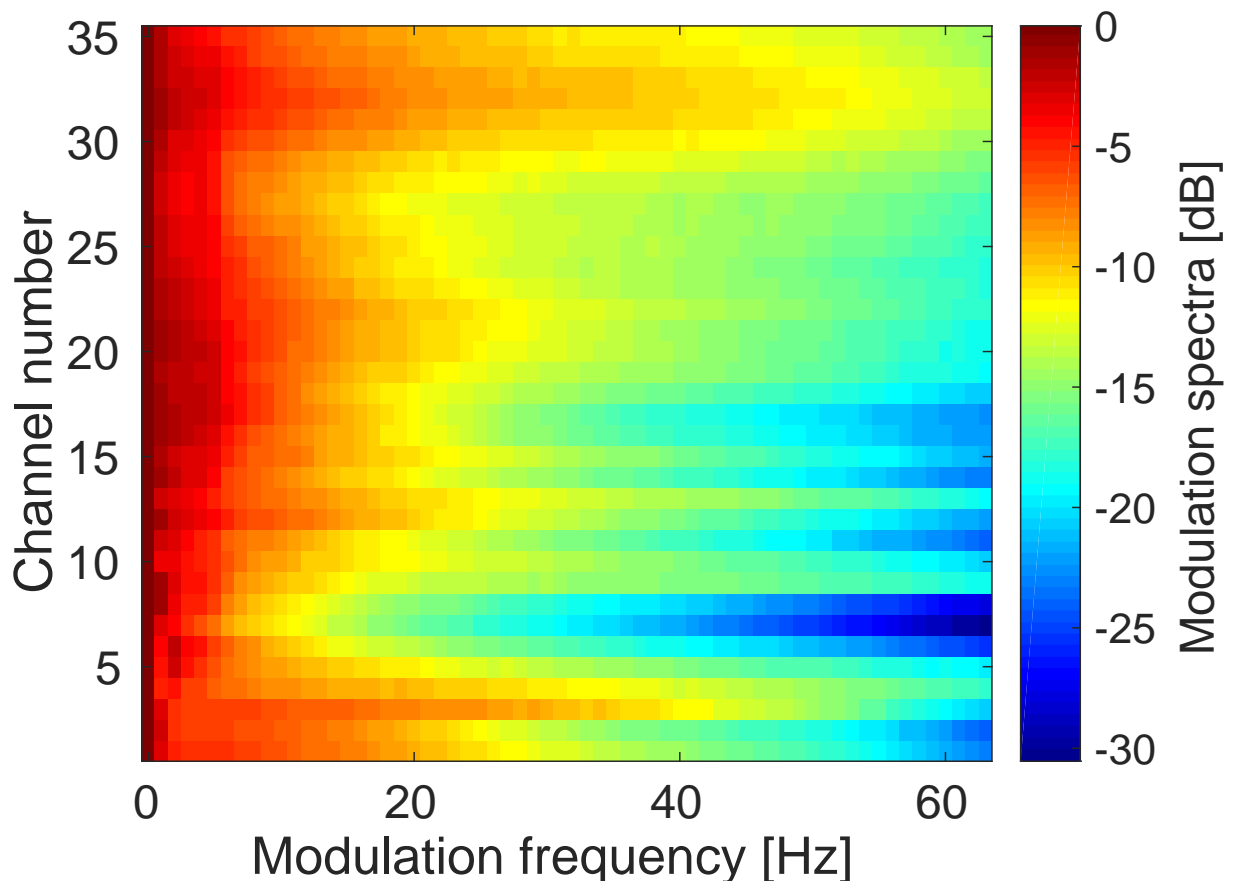


図 3.9: 女性音声の変調スペクトログラム

第4章 騒音の特徴を利用した騒音低減

4.1 騒音低減法

図 4.1 に変調スペクトルに着目した騒音低減法のブロックダイアグラムを示す。はじめに原信号 $s(t)$ をガンマトーンフィルタバンクにより、 k 個の周波数帯域（チャンネル信号） $x_k(t)$ に分解する。次に $x_k(t)$ からパワー包絡線 $e_k^2(t)$ とキャリア $c_k(t)$ を求める。そして、パワー包絡線 $e_k^2(t)$ に対し、次の三つの手順で騒音成分を低減する：(1) 定常騒音成分の除去，(2) 間欠騒音成分の除去，(3) 衝撃騒音成分の除去。これらの騒音の特徴を打ち消したパワー包絡線を振幅包絡線に戻し、キャリアと掛けあわせることでチャンネル信号を復元する。最後に逆ガンマトーンフィルタバンクにより、騒音が低減された信号 $y(t)$ を得る。ここで、図 4.1 の各ブロックにて、 $(\cdot)^2$ は二乗処理， $\text{Mean}(\cdot)$ は時間方向の平均処理， $\text{HWR}(\cdot)$ は半波整流処理， $\text{BPF}(\cdot)$ は帯域阻止フィルタ， $\text{LPF}_S(\cdot)$ は音声の変調スペクトルの概形に合わせた低域通過フィルタを示す。

図 4.2 にマシンガン雑音に対するそれぞれの騒音の特徴を除去する処理を行ったパワー包絡線を示す。図 4.2(a) はガンマトーンフィルタバンクの出力のパワー包絡線である。図 4.2(b) はパワー包絡線の直流成分を除去したパワー包絡線である。図 4.2(c) は変調スペクトルの調波性を除去したパワー包絡線である。図 4.2(d) は変調スペクトル上の高い変調周波数成分を除去したパワー包絡線である。

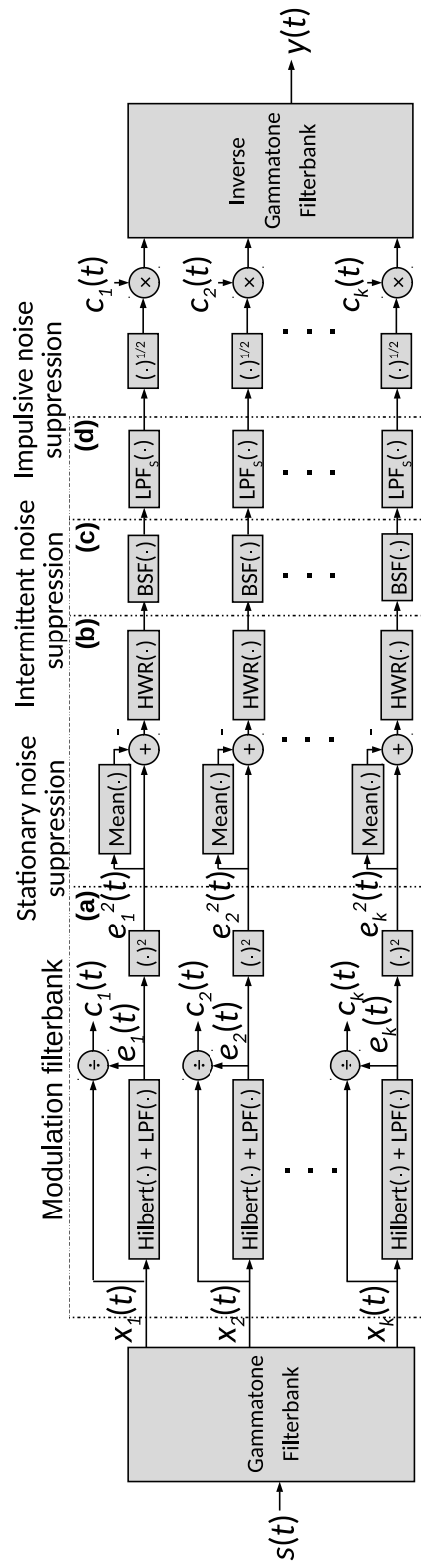


図 4.1: 変調知覚メカニズムに基づいた騒音低減法

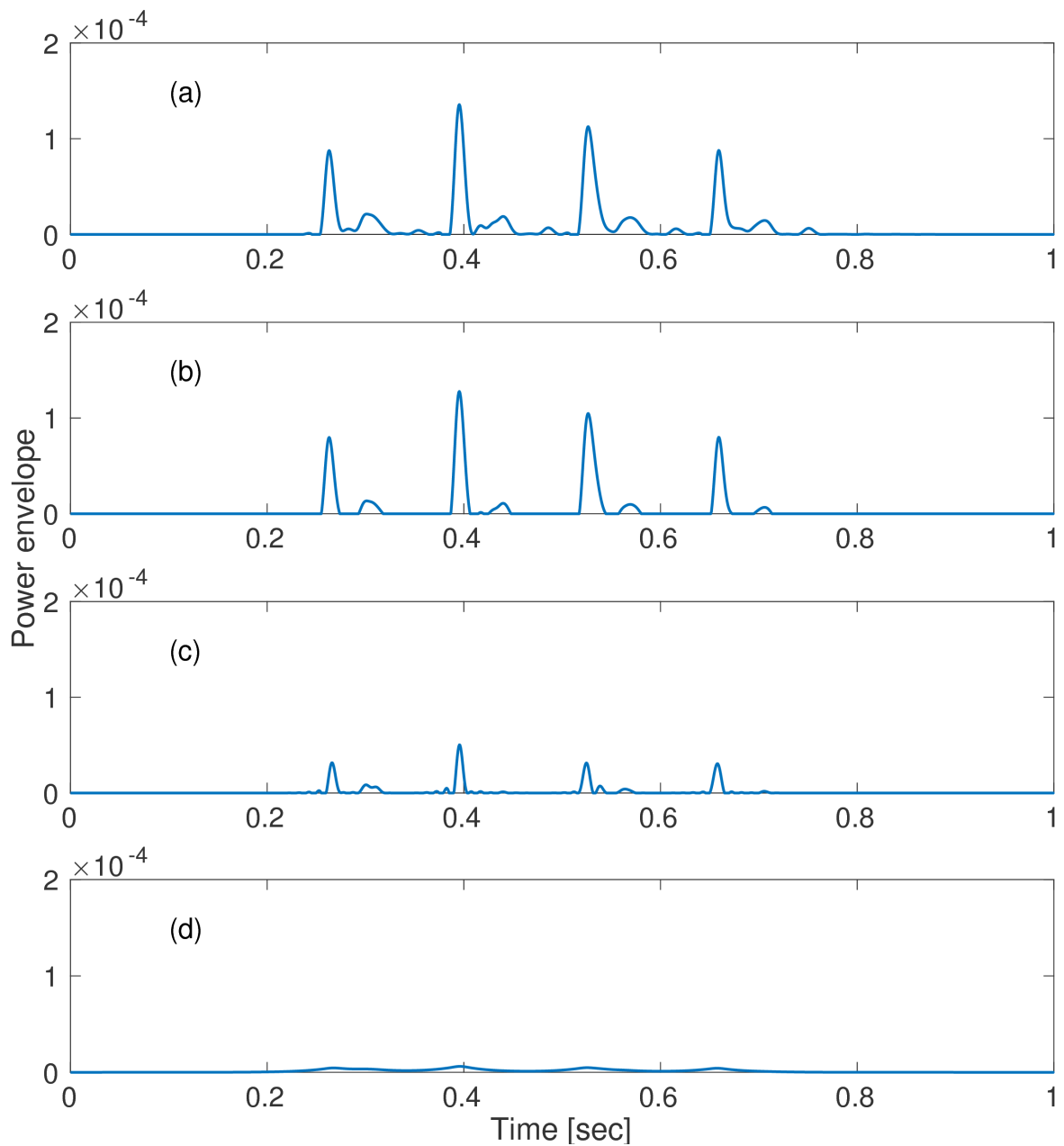


図 4.2: 騒音低減法にマシンガンノイズを入れたときのパワー包絡線 : (a) ガンマトーンフィルタバンクの出力, (b) 変調スペクトルに含まれる定常騒音の特徴の除去, (c) 変調スペクトルに含まれる間欠騒音の特徴の除去, (d) 変調スペクトルに含まれる衝撃騒音の特徴の除去

4.2 変調スペクトルの直流成分の除去

定常騒音の聴覚フィルタバンク出力の特徴は，時間平均の標準偏差が小さいく，変調フィルタバンク出力における特徴は，極めて低い変調周波数成分を持つことである．このことを踏まえて，定常騒音の特徴を打ち消すために変調スペクトルの直流成分を減算する．ここでは，パワー包絡線 $e_k^2(t)$ の非音声区間の平均値を次式から求める．

$$\mu_k = \frac{1}{T_N} \int_0^{T_N} e_k^2(t) dt \quad (4.1)$$

ただし， T_N は非音声区間の時間長である．定常騒音の特徴を打ち消したパワー包絡線を次式のように求める．

$$q_k(t) = \begin{cases} e_k^2(t) - \mu_k & (e_k^2(t) \geq \mu_k) \\ 0 & \text{otherwise} \end{cases} \quad (4.2)$$

4.3 変調スペクトルの調波性の除去

間欠騒音の変調スペクトル上の特徴は，調波性を持つことである．このことから，図 4.1 の半波整流処理されたパワー包絡線の変調スペクトルの基本周波数 f_c を自己相関で求める．そして，次式の L 次の FIR 帯域阻止フィルタ (Band-stop filter: BSF) を用いて，振幅包絡線に畳み込むことで特徴を打ち消す．

$$H(z) = b_0 - r^L z^{-L} \quad (4.3)$$

ただし， $b_0 = 1$ ， $r = 0.995$ ， $L = f_s/f_c$ である．図 4.3(a) に変調スペクトルの基本周波数を 8 Hz として作成した BSF の例を示す．

4.4 変調スペクトルの高域成分の除去

衝撃騒音の変調スペクトル上の特徴は，あらゆる変調周波数成分を持つことである．このことを踏まえて，図 4.1 の帯域阻止フィルタの出力の変調スペクトルを音声の変調スペクトルの概形に似せるように，低域通過フィルタを設計する．そして，衝撃騒音のパワー包絡線に次式に示す低域通過フィルタを畳み込み，音声を残しつつ，衝撃騒音を低減する．

$$H(z) = \frac{b_0 + b_1 z^{-1}}{1 + a_1 z^{-1}} \quad (4.4)$$

ここで，低域通過フィルタは 1 次のバターワース型 IIR フィルタを用いた．カットオフ周波数は音声の変調スペクトルの概形に合わせてチャンネルごとに定めた．例としてカットオフ周波数が 5 Hz のときの係数は， $b_0 = 0.07$ ， $b_1 = 0.07$ ， $a_1 = 0.85$ である．図 4.3(b) にカットオフ周波数を 5 Hz として作成した低域通過フィルタの例を示す．

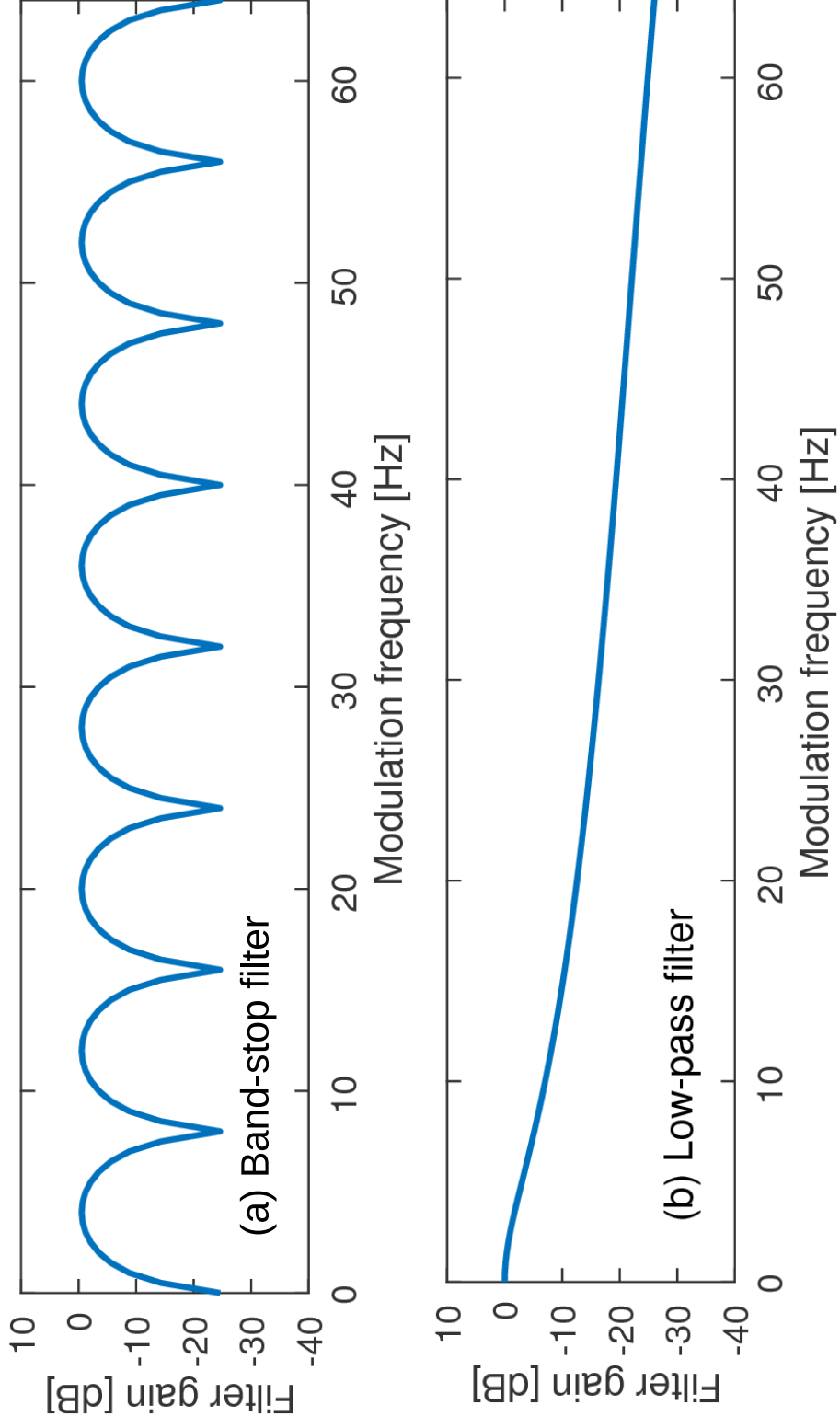


図 4.3: 騒音低減のためのフィルタの周波数特性: (a) 帯域阻止フィルタ, (b) 低域通過フィルタ

第5章 提案法の客観評価

前章では騒音を低減する手法について述べた。本章では、騒音低減法の評価について述べる。

5.1 騒音を入力としたときの騒音低減レベル

5.1.1 評価方法

騒音低減法を評価するために、次式で定義される騒音低減レベル N_R を調べた。

$$N_R = 10 \log_{10} \frac{\int_0^T s^2(t) dt}{\int_0^T y^2(t) dt} \quad (5.1)$$

ただし、 $s(t)$ は原信号、 $y(t)$ は騒音が低減された信号である。

使用音源は表 3.1 から、定常騒音として白色雑音、ピンク雑音、バブル雑音を、間欠騒音としてマシンガン雑音を、衝撃騒音としてインパルス性雑音を、白色・マシンガン・インパルス性雑音を混ぜた騒音 (WMI) を用いる。各音源を 1 秒ずつ 0.5 刻みで 10 回切り出しそれらの平均値を評価値とする。

ここで、騒音の音圧レベルを 60, 80, 100 dB としたときの騒音低減レベルを調べた。

5.1.2 評価結果

各音源に対する騒音低減レベルの評価結果を求めた。図 5.1 に各音源の音圧レベルが 60, 80, 100 dB のときの騒音低減レベルを示す。図 5.1 の横軸は各騒音の音圧レベルを、縦軸は各音源の低減されたレベルを示す。図 5.1(a) は騒音の音圧レベル 100 dB の場合、図 5.1(b) は騒音の音圧レベルが 80 dB の場合、図 5.1(c) は騒音の音圧レベルが 100 dB の場合の騒音低減レベルの結果である。各騒音の紺色の棒グラフは変調スペクトルの直流成分だけを除去したもの、青色の棒グラフは変調スペクトルの調波性だけを除去したもの、緑色の棒グラフは変調スペクトルの高域成分だけを除去したもの、黄色の棒グラフはすべての処理をしたものを示す。図 5.1(a)、図 5.1(b)、図 5.1(c) より、音圧レベルの変化による騒音低減レベルの変化はないことが分かった。図 5.1 より、定常騒音では変調スペクトルの直流成分の除去が、間欠騒音では変調スペクトルの調波性の除去が、衝撃騒音では変調

スペクトルの高域成分の除去が騒音の低減に最も効果的であった。また、騒音からすべての変調スペクトル上の特徴を除去することで、定常騒音で 8 dB、間欠騒音で 6 dB、衝撃騒音で 8 dB、WMI で 6 dB の低減が明らかになった。

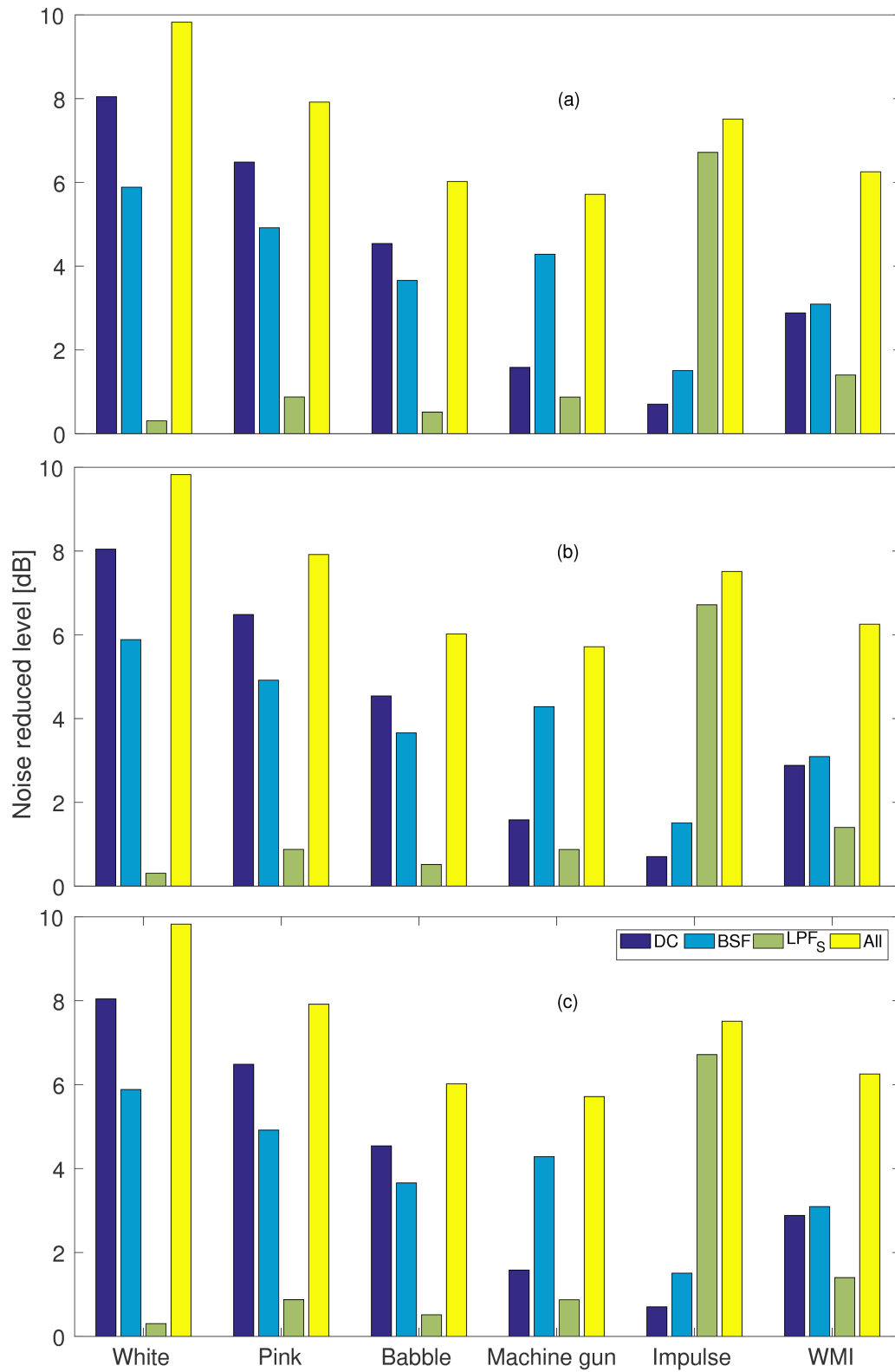


図 5.1: 騒音を入力としたときの騒音低減レベル: (a) 騒音の音圧レベル 100 dB の場合, (b) 80 dB の場合, (c) 60 dB の場合

5.2 音声と騒音を混ぜた混合音における騒音の相対的低減レベル

5.2.1 評価方法

音声と騒音が混ざった混合音から騒音がどれだけ低減できるかを評価するために混合音における騒音の相対的低減レベル N_S を調べた。

$$N_S = 10 \log_{10} \frac{\int_0^T s_s^2(t) dt}{\int_0^T (s_s(t) - y(t))^2 dt} - \text{SNR} \quad (5.2)$$

ただし、 $s_s(t)$ は音声信号、 $y(t)$ は原信号、 T は信号の時間長である。使用音源は表 3.1 から、音声として、1 秒の男性音声と女性音声を 1 つずつ用いた。定常騒音として白色雑音、ピンク雑音、バブル雑音を、間欠騒音としてマシンガン雑音を、衝撃騒音としてインパルス性雑音を、白色・マシンガン・インパルス性雑音を混ぜた騒音を用いる。各音源を 1 秒ずつ 0.5 刻みで 10 回切り出しそれらの平均値を評価値とする。

ここで、騒音の音圧レベルが 60 dB のときに音声の音圧レベルを 40, 50, 60, 70 dB、騒音の音圧レベルが 80 dB のときに音声の音圧レベルを 40, 50, 60, 70 dB、騒音の音圧レベルが 100 dB のときに音声の音圧レベルを 40, 50, 60, 70 dB となるような SNR を設定した。

5.2.2 評価結果

各音源に対し、騒音低減前と騒音低減後の音声の SER から騒音の相対低減レベルを求めた。図 5.2(a) は音圧レベル 100 dB の場合、図 5.2(b) は音圧レベル 80 dB の場合、図 5.2(c) は音圧レベル 60 dB の場合の騒音のときの相対的低減レベルの結果である。図 5.2 の横軸は SNR の値を、縦軸は相対低減レベル N_S を示す。図 5.2 より SNR が $-60 \sim -20$ dB においてどの騒音も 5 dB 程度の相対的低減レベルを示した。しかし、SNR が $-10 \sim 10$ dB において騒音の相対的な低減レベルが下がっている。特に SNR が 10 dB のときの騒音の相対的低減レベルは -5 dB と騒音が増加していることが分かった。

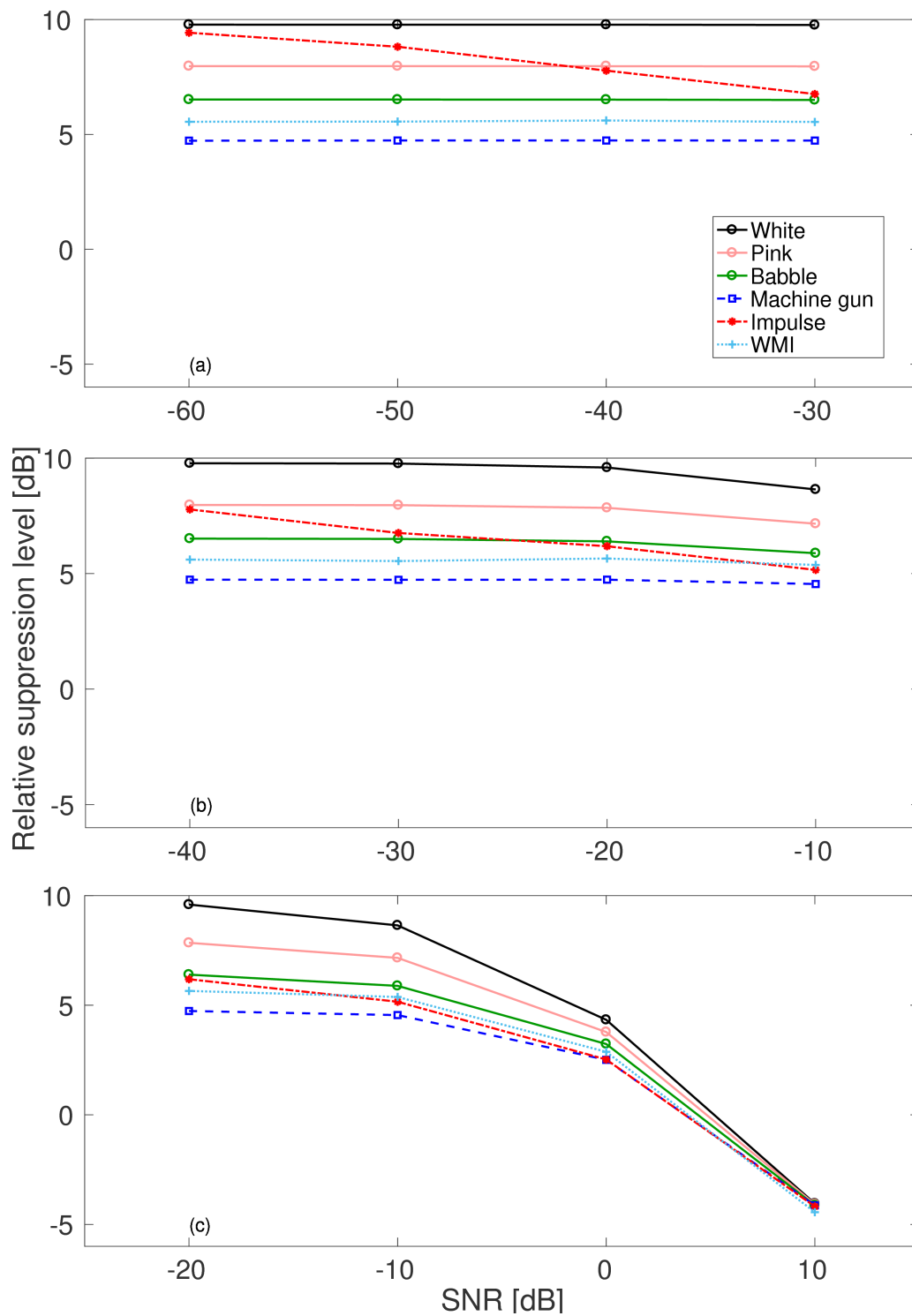


図 5.2: 混合音における騒音の相対的低減レベル: (a) 騒音の音圧レベル 100 dB の場合, (b) 80 dB の場合, (c) 60 dB の場合

5.3 音質評価指標

5.3.1 評価方法

次に、低減前の騒音と低減後の騒音で音の大きさがどれだけ減少するかを評価するために次式で示すラウドネスの減少量 L_R からを調べた。

$$L_R = L_{\text{org}} - L_{\text{sup}} \quad (5.3)$$

ここで、 L_{org} は低減前のラウドネスの値、 L_{sup} は低減後のラウドネスの値である。もし、ラウドネスが減少すれば、騒音が聴覚に与えるダメージが低減できたと言える。

低減前の騒音と低減後の騒音で甲高さがどれだけ減少するかを評価するために次式に示すシャープネスの減少量 K_R を調べた。

$$K_R = K_{\text{org}} - K_{\text{sup}} \quad (5.4)$$

ここで、 K_{org} は低減前のシャープネスの値、 K_{sup} は低減後のシャープネスの値である。もし、シャープネスが減少すれば、低減後の騒音は甲高さが減少したと言える。

最後に、次式に示すラウドネスの時間変動を評価するラフネスの低減前と低減後の減少量 R_R を評価することによって、騒音が低減できたかを調べた。

$$R_R = R_{\text{org}} - R_{\text{sup}} \quad (5.5)$$

ここで、 R_{org} は低減前のラフネスの値、 R_{sup} は低減後のラフネスの値である。もし、ラフネスが減少すれば、騒音の時間変動が小さくなったと言える。

5.3.2 評価結果

図 5.3～図 5.5 に各音質評価指標の低減前から低減後の減少量を示す。図 5.3 はラウドネスの減少量、図 5.4 はシャープネスの減少量、図 5.5 はラフネスの減少量の結果である。横軸は騒音の音圧レベル、縦軸は各指標の減少量である。

図 5.3 より、音圧レベル 100 dB のとき、定常騒音である白色雑音、ピンク雑音、バブル雑音のラウドネスでは 50 sone の減少が、間欠騒音であるマシンガン雑音と衝撃騒音であるインパルス性雑音では 20 sone の減少が確認できた。また、音圧レベルの上昇に伴って、ラウドネスの減少量が増加していることが確認できた。

図 5.4 より、音圧レベルが 100 dB のとき、定常騒音である白色雑音、ピンク雑音、バブル雑音のシャープネスでは 0.1 acum の減少を、間欠騒音であるマシンガン雑音と衝撃騒音であるインパルス性雑音では 0 acum の減少が確認できた。また、音圧レベルの上昇に伴って、ラウドネスの減少量が増加していることが確認できた。また、シャープネスの減少量は騒音の音圧レベルの変化に依存しないことがわかった。

図 5.5 より、音圧レベルが 100 dB のとき、定常騒音である白色雑音、ピンク雑音、バブル雑音のラフネスでは 0.05 asper の減少が、間欠騒音であるマシンガン雑音では 0.73 asper の減少が、衝撃騒音であるインパルス性雑音では 0.25 asper の減少が確認できた。

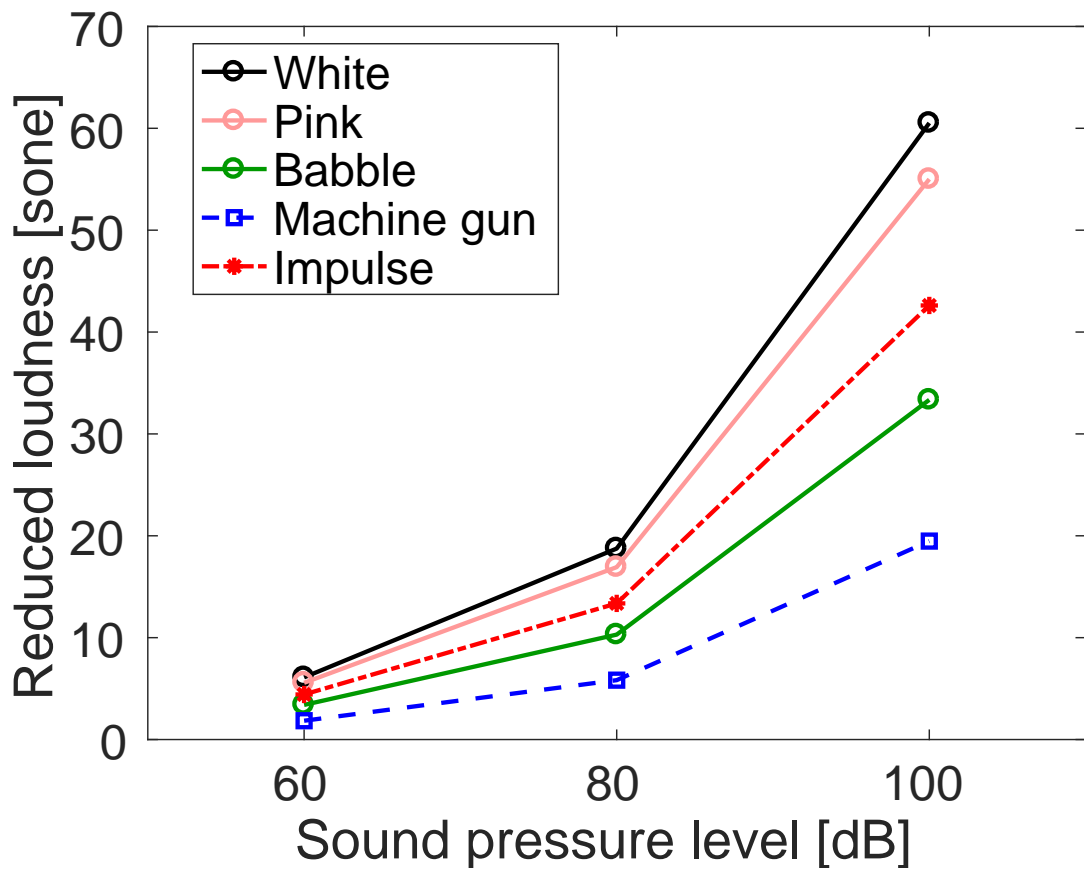


図 5.3: ラウドネスの減少量

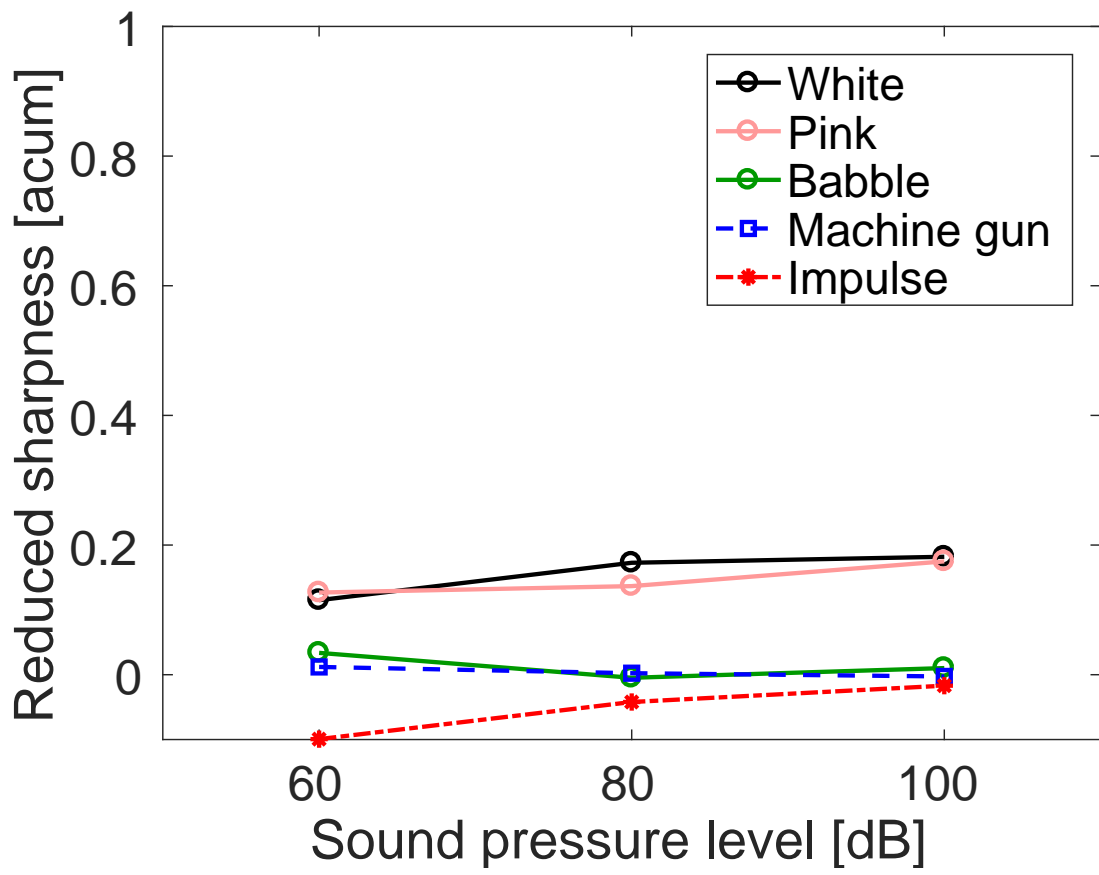


図 5.4: シャープネスの減少量

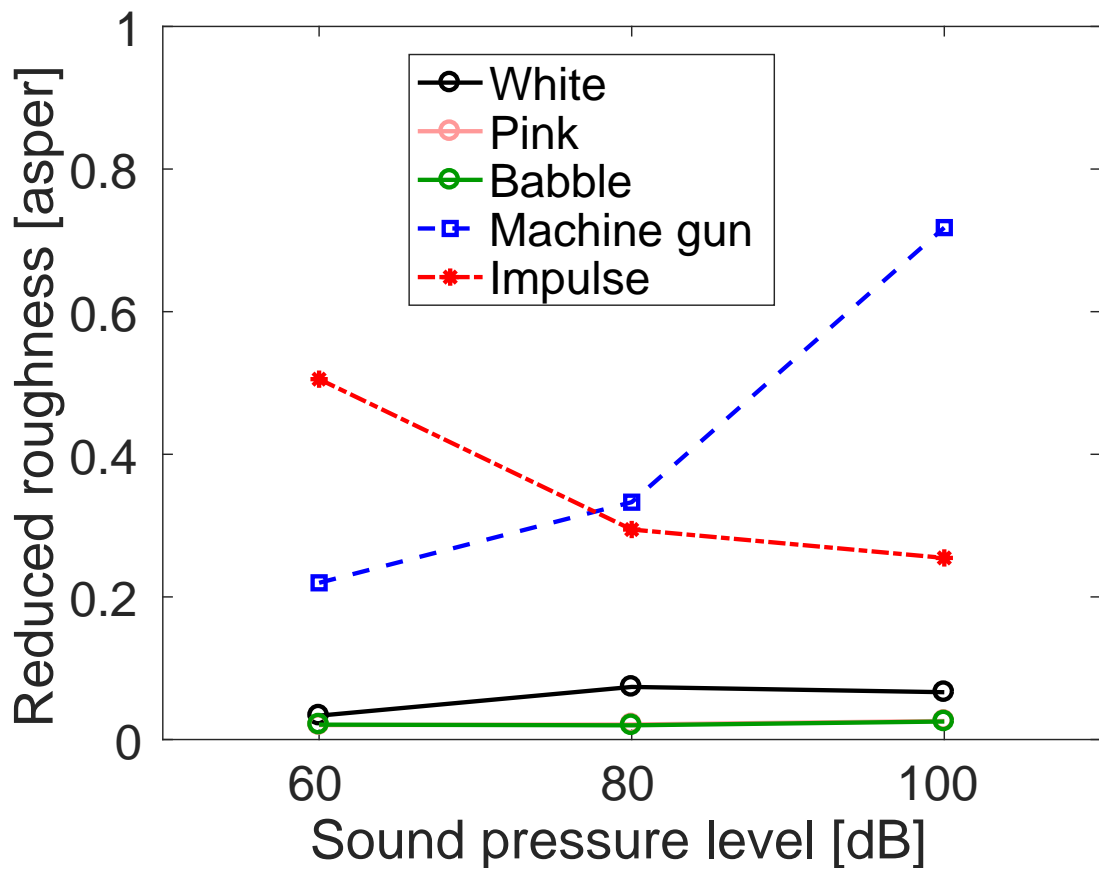


図 5.5: ラフネスの減少量

5.4 考察

騒音の特徴と騒音低減法の評価結果について考察を述べる。

定常騒音の特徴は，時間方向の標準偏差が小さく，変調スペクトル上では極めて低い変調周波数成分を持っていることが明らかになった．このことから，定常騒音のパワーエンベロープは，時間でほとんど変動がなく，直流成分だけを持っていると考えられる．間欠騒音の特徴は，変調スペクトル上に調波性を持っていることが明らかになった．今回の結果に関しては 8 Hz を基本変調周波数とする調波性を持っていた．このことから，今回用いた間欠騒音の音源は，1/8 で sec ごとに突発的な騒音があると考えられる．衝撃騒音の特徴は，変調スペクトル上にあらゆる変調周波数成分を持っていつことが明らかになった．衝撃騒音の特徴はガンマトーンフィルタバンクのガンマ関数の特性であると考えられる．音声の変調スペクトル上の特徴は，4 Hz 付近に変調周波数成分が多いという特徴を持っていた．これは，先行研究の結果と一致した [24]，

騒音低減法の評価結果から，一日に音圧レベル平均 70 dB の音を 40 年間聴取すると 10% のヒトに無視できない聴力損失が現れることが知られている [2]．3 dB 音圧レベルが高くなると，20 年後には無視できない聴力損失が起きる．3 dB 音圧レベルを低くすることができれば，80 年間聴取することができる．今回の騒音低減レベルの結果は，定常騒音では 8 dB，間欠騒音では 6 dB，衝撃騒音では 8 dB 程度の低減ができた．これは，十分な低減量だと考えられる．

騒音の相対的低減レベルの結果から，定常・間欠・衝撃騒音ともに，SNR が 0 dB より SNR が高いときに騒音の相対的低減レベルが下がり，SNR が 10 dB のときにマイナスになってしまった．これは，騒音を低減するために用いた帯域阻止フィルタと低域通過フィルタが音声にダメージを与えてしまったと考えられる．

音質評価指標の結果から，シャープネスはほとんど変化がなかった．これは，ガンマトーンフィルタバンク出力のチャンネル方向に処理をしていないため，変化しなかったと考えられる．

5.5 まとめ

本章では変調知覚メカニズムを利用することにより，定常・非定常騒音を低減できるのか検討した．変調スペクトル分析から，騒音が持つ特徴を確認し，各騒音の特徴を打ち消す処理を行った．騒音低減レベル，相対低減レベル，音質評価指標を利用して，騒音低減法を評価した．その結果，騒音低減レベルの評価にて，定常騒音では 8 dB，間欠騒音では 6 dB，衝撃騒音では 8 dB 程度，騒音を低減できることが分かった．相対的な低減レベルの評価では SNR が $-60 \sim -20$ dB の間において 7 dB の騒音低減が確認できた．ラウドネスの減少量の評価では，音圧レベルが 100 dB のとき，定常騒音で 50 sone，間欠騒音で 20 sone，衝撃騒音で 50 sone の減少が確認できた．ラフネスの減少量の評価では音圧レベルが 100 dB のとき，間欠騒音で 0.7 asper，衝撃騒音で 0.3 asper の減少が確認で

きた．このことから，変調スペクトル上の特徴を利用することで，定常・間欠・衝撃騒音を十分に低減できることが明らかになった．特に，従来の方法では難しかった，間欠騒音で6 dB，衝撃騒音で8 dBの低減が確認できた．

第6章 結論

6.1 本研究で明らかになったこと

本研究は定常・間欠・衝撃騒音を低減するためにヒトの変調知覚メカニズムに基づいた騒音低減法を検討した。そのために、以下の2つのことを行った。

- 騒音と音声の特徴を抽出するために、聴覚フィルタバンク出力における特徴および、変調フィルタバンク出力における特徴を明らかにした。
- 騒音低減法の有効性について明らかにした。

1点目では、騒音と音声の特徴を抽出した。聴覚フィルタバンク出力における特徴の抽出の結果、定常騒音のパワー包絡線はパワーの変化が少ないことが明らかになった。間欠・衝撃騒音と音声のパワー包絡線は変動が大きいことが明らかになった。変調フィルタバンク出力における特徴を抽出した結果、定常騒音の変調スペクトル上の特徴は非常に低い変調周波数成分のみを有していることが明らかになった。間欠騒音の変調スペクトル上の特徴は調波性を有していることが明らかになった。衝撃騒音の変調スペクトル上の特徴は、変調スペクトルが概ね平坦であることが明らかになった。これらの結果から、騒音の聴覚フィルタバンク出力と変調フィルタバンク出力における特徴を明らかにした。

2点目では、求めた特徴を利用して騒音低減法を提案し、その評価を行った。評価には、騒音低減レベル、相対的な低減レベル、ラウドネス、シャープネス、ラフネスを用いた。その結果、騒音低減レベルの評価にて、定常騒音では8 dB、間欠騒音では6 dB、衝撃騒音では8 dB程度、騒音を低減できることが分かった。相対的な低減レベルの評価ではSNRが $-60 \sim -20$ dBの間において7 dBの騒音低減ができることが分かった。ラウドネスの減少量の評価では、音圧レベルが100 dBのとき、定常騒音で50 sone、間欠騒音で20 sone、衝撃騒音で50 soneの減少ができることが分かった。ラフネスの減少量の評価では音圧レベルが100 dBのとき、間欠騒音で0.7 asper、衝撃騒音で0.3 asperの減少できることが分かった。このことから、変調知覚メカニズムに基づくことで、定常・間欠・衝撃騒音を十分に低減できることが分かり、騒音低減法の有効性を明らかにした。

6.2 残された課題

本研究の残された課題は以下の4点である。

- 相対的な騒音低減レベルの改善すること
- 定常騒音を低減するために事前にパワーエンベロープの直流成分を推定すること
- リアルタイムで騒音を低減すること
- 残響を考慮すること

1点目では、音声と騒音のSNRが10 dBより高いときに、騒音が増加してしまう。本研究では、変調スペクトル上の調波性を除去するためにL次のFIR帯域阻止フィルタと変調スペクトルの高域成分を除去するために低域通過フィルタを用いた。この2つのフィルタはSNRに関係なく一定である。そのため、SNRが高いときに音声にダメージを与えてしまう。そこで、騒音低減法のはじめに段階で、SNRを推定し、騒音低減を行うかどうかの判断をすることで、この問題が解決できると考えられる。

2点目では、定常騒音を低減する際、騒音区間から事前に求めたパワーエンベロープの直流成分を利用している。しかし、実環境においては、音声区間と非音声区間を検出する必要がある。

3点目では、聴力を守るためにはリアルタイムで低減する必要があるが、今のプログラムでは処理に時間がかかってしまう。この問題を解決するためには、ガンマトーンフィルタバンクをIIRフィルタで設計する必要がある。

4点目では、残響を考慮していない点である。本研究では、雑音のみに焦点をおいて検討してきた。しかし、実環境において、騒音だけでなく残響も問題になる。そこで、騒音と残響を同時に低減する手法が必要になる。Flanaganらはマイクロフォンアレーを用いて残響除去を行った[38]。しかし、この手法は装置が大きくなるという問題がある。一方、鶴木らは、残響信号のパワーエンベロープから残響除去するためのMTFコンセプトに基づく方法について検討した[39]。この手法はMTFに基づいており、パワーエンベロープから残響を除去する。この手法を提案法に取り入れることで、雑音・残響下においても有効に使うことができると考えられる。

謝辞

研究活動から私生活，社会の一般常識にいたるまで，厳しくも非常に丁寧な指導をしていただいた，主指導教員である鵜木祐史教授には，深謝の意を表する．また，研究室会議やミーティングなどの場において，数多くの助言を頂いた赤木正人教授に深謝の意を表する．研究でわからないときに助けて頂いた，朱治博士に深謝の意を表する．公私共に筆者の悩みや疑問を真剣に考えてくださった，鳥谷輝樹氏に深謝の意を表する．共に切磋琢磨した，関谷伸一氏，高橋響子氏に深く感謝致します．研究で行き詰まったときに相談に乗って頂いた，鵜木・赤木研究室のメンバーの皆様には感謝致します．本学での研究生生活を支え，温かく見守ってくれた両親に心から感謝致します．

参考文献

- [1] 和田哲郎, “騒音性難聴の最近の知見 (疫学, 基礎など),” 専門医通信, 120 巻, 3 号, pp. 252–253, 2017.
- [2] 日本音響学会編, “音響学入門,” コロナ社, 2011.
- [3] Birgita, B., Thomas, L., Dietich, H., S., “GUIDELINES FOR COMMUNITY NOISE,” World Health Organization, 1999.
- [4] 曽根敏夫 著, “くらしと音,” 裳華房, 1991.
- [5] 笹岡直人, 伊藤良生, “騒音抑圧技術—基礎とその応用—,” IEICE Fundamentals Review, Vol. 5, No. 2, pp. 136–145, 2011.
- [6] Boll, S., “Suppression of acoustic noise in speech using spectral subtraction,” IEEE Transactions on Acoustics, Speech and Signal Processing, vol. 27, pp. 113–120, 1979.
- [7] 幸田晃, 宇佐川毅, 江端正直, “雑音スペクトル推定にパワー変動を利用したスペクトルサブトラクション法,” 日本音響学会誌, 58 巻, 8 号, pp. 493–500, 2002.
- [8] Yariv, E., David, M., “Speech Enhancement Using a Minimum Mean Square Error Short-Time Spectral Amplitude Estimator,” IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. 32, No. 6, pp. 1109–1121 1984.
- [9] 野村行弘, 斗澤秀亮, 呂建明, 関屋大雄, 谷萩隆嗣, “モフォロジー処理を用いたスペクトルサブトラクションにおけるミュージカルノイズ除去,” 電子情報通信学会論文誌, vol. J89D, no. 5, pp. 991–1000, 2006.
- [10] Volker, S., Alexander, F., Rolf, B., “Quantile based noise estimation for spectral subtraction and Wiener filtering,” IEEE TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, pp. 1875–1878, 2000.
- [11] Zenton, G., Kah-Chye T., B. T. G. Tan, “Postprocessing Method for Suppressing Musical Noise Generated by Spectral Subtraction,” IEEE TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING, vol. 6, no. 3, 1998.

- [12] 伊藤憲三, 水島昌英, 北脇信彦, “音声と非音声の識別処理に基づく定常雑音抑圧方式,” 日本音響学会誌, 61 巻 8 号, pp. 431–440, 2005.
- [13] 古田訓, 高橋真哉, 中島邦男, “スペクトル減算と振幅抑圧の相互制御に基づく雑音抑圧法の検討,” 電子情報通信学会論文誌, vol. J87DII, no. 2, pp. 464–474, 2004.
- [14] 堀井圭祐, 福森隆寛, 森勢将雅, 中山雅人, 西浦敬信, 山下洋一, 南條浩輝, “雑音下音声受音における Weighted 反復スペクトル減算法を用いたミュージカルノイズの低減,” 電子情報通信学会論文誌, vol. J96D, no. 3, pp. 664–674, 2013.
- [15] 三宅亮太, 福森隆寛, 中山雅人, 西浦敬信, “連の検定に基づく雑音環境適応型重み付き反復スペクトル減算,” 電子情報通信学会論文誌, vol. J98D, no. 6, pp. 1063–1067, 2015.
- [16] 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.
- [17] 松田圭二, 他, “ある騒音職場就業員の聴力の長期的追跡,” *Audiology Japan* 35, pp.187–192, 1992.
- [18] J.L. Flanagan, J.D. Johnston, R. Zahn, G.W. Elko, “Computer-steered microphone arrays for sound transduction in large rooms,” *The Journal of the Acoustical Society of America*, vol.78, no.5, pp. 1508–1518, 1985.
- [19] Y. Kaneda and J. Ohga, “Adaptive microphone-arraysystem for noise reduction,” *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-34, no.6, pp. 1391–1400, 1986.
- [20] Takehara, R., Kawamura, A., Iiguni, Y., “Impulsive Noise Suppression Using Interpolated Zero Phase Signal,” *APSIPA2017*, pp. 1382–1389, 2017.
- [21] Yoshioka, T., Miyoshi, M., “Adaptive suppression of non-stationary noise by using the variational Bayesian method,” *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 4889–4892, 2008.
- [22] Qu, W., Jia, X., Pei, S., “Non-stationary Signal Noise Suppression Based on Wavelet Analysis,” *Congress on Image and Signal Processing*, pp. 303–306, 2008.
- [23] Zhiyao D., Gautham, J., M., Paris S., “Speech enhancement by online non-negative spectrogram decomposition in non-stationary noise environments,” *Proc. Interspeech 2012*, pp. 595–598, 2012.

- [24] Atlas, L., Greenberg, S., and Hermansky, H., “The Modulation Spectrum and Its Application to Speech Science and Technology,” Interspeech Tutorial, Antwerp, Belgium, 2007.
- [25] Zhi, Z., Yasutaka N., Ryota, M., Masashi U., “Study on linguistic information and speaker individuality contained in temporal envelope of speech,” *Acoust. Sci. & tech* 37, pp. 258–261, 2016.
- [26] Kanai, Y., Morita, S., Unoki, M., “Concurrent processing of voice activity detection and noise reduction using empirical mode decomposition and modulation spectrum analysis,” *Proc. INTERSPEECH*, pp. 742–746, 2013.
- [27] Robert, V. S., Fan-Gang, Z., Vivek, K., John, W., Michael, E., “Speech Recognition with primarily Temporal Cues,” *Science*, Vo. 270, pp. 303–304, 1995.
- [28] 風間道子, 東山三樹夫, 山崎芳男, “狭帯域音声波形包絡線の帯域間相関行列に現れる話者情報,” *電子情報通信学会論文誌 A*, vol. J92-A, no. 4, pp. 205–215, 2009.
- [29] Ryosuke, O. T., Sasaki, Y., Riquimaroux, H., “Relative contributions of spectral and temporal resolutions to the perception of syllables, words, and sentences in noise-vocoded speech,” *Acoust. Sci. & Tech.*, vol. 34, pp. 263–270, 2013.
- [30] Moore, B. C. J., “An Introduction to the Psychology of Hearing,” Academic Press Ltd., London, 1989. (大串健吾訳)
- [31] Stephan D. E., and Torsten D., “Characterizing frequency selectivity for envelope fluctuations,” *The Journal of the Acoustical Society of America*, 108, 1181, 2000.
- [32] Zwicker, F., “Psychoacoustics: facts and models,” Berlin, Springer, 2007.
- [33] Torsten, D., Birger, K., “Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers,” *The Journal of the Acoustical Society of America* 101, 1997.
- [34] Varga, A., and J.M. Steeneken H., “Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems,” *Speech Communication*, vol. 12, no. 13, pp. 247–251, 1993.
- [35] Kondo, T., Amano, S., Sakamoto, S., Susuki, Y., “Development of Familiarity-controlled Word-lists (FW07),” *IEICE technical report*, 107(436), pp. 43–48, 2008.

- [36] Patterson, R., Nimmo-Smith, L., Holdsworth, J., Rice, P., “An auditory filter bank based on the gammatone function,” Paper presented at a meeting of the IOC Speech Group on Auditory Modelling at RSRE, pp. 14–15, 1987.
- [37] 鷓木祐史, “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,” Journal of Signal Processing, pp. 339–348, 2008.
- [38] J. L. Flanagan, J. D. Johnston, R. Zahn, and G. W. Elko, “Computer steered microphone arrays for sound transduction in large rooms,” The Journal of the Acoustical Society of America 78, vol. 1508, 1985.
- [39] Unoki, M., Furukawa, M., Sakata, K., Akagi, M., “A METHOD BASED ON THE MTF CONCEPT FOR DEREVERBERATING THE POWER ENVELOPE FROM THE REVEABERANT SIGNAL,” Proc. ICASSP-2003, vol. 1, pp. 840-843, 2003.

研究業績一覧

国際会議

1. ISOYAMA, T., UNOKI, M., “Noise Suppression Method based on Modulation Spectrum analysis,” Proc. SPECOM 2018, Sept. 2018 (accepted).

国内発表

1. 磯山 拓都, 鷗木 祐史, “音声・非音声の変調スペクトルの特徴分析に関する検討,” 信学技報, EA 2017-49, pp. 29–34, 2017.
2. 磯山, 鷗木, “高次統計量を用いた音声・非音声の変調スペクトルの特徴分析に関する検討,” 第32回信号処理シンポジウム, pp. 366–371, 2017.
3. 磯山 拓都, 鷗木 祐史, “変調スペクトルを用いた騒音低減手法の検討,” 日本音響学会 2018 年春季研究発表会, 3-P-1, 2018.
4. 磯山 拓都, 鷗木 祐史, “変調スペクトルに着目した騒音抑圧法の検討,” 信学技報, EA 2017-168, pp. 361–366, 2018.