

Title	聴覚末梢モデルを利用した音質評価指標の計算モデルの構築
Author(s)	磯山, 拓都
Citation	
Issue Date	2024-06
Type	Thesis or Dissertation
Text version	ETD
URL	<a href="http://hdl.handle.net/10119/19332">http://hdl.handle.net/10119/19332</a>
Rights	
Description	Supervisor: 鶴木祐史, 先端科学技術研究科, 博士

博士論文

聴覚末梢モデルを利用した音質評価指標の  
計算モデルの構築

磯山 拓都

主指導教員 鵜木 祐史

北陸先端科学技術大学院大学  
先端科学技術専攻

[情報科学]

令和06年6月

# Abstract

Sound quality metrics (SQMs), such as loudness, sharpness, roughness, and fluctuation strength, have been used for modeling sensory pleasantness and annoyance, for product sound design, and for soundscape analysis and as SQMs for these purposes. Moreover, the ability to convey appropriate SQMs is considered essential in creating more desirable products and environmental sounds. SQMs can be modeled computationally. In particular, those related to temporal properties i.e., roughness, and fluctuation strength, have been modeled on the basis of time-domain auditory filters. The classical models of SQMs use the temporal variation of specific loudness measured using a loudness meter on the basis of ISO 532B:1975.

The international standards ISO 532-1:2017 and ISO 532-2:2017, provide two loudness models called the Zwicker and Moore-Glasberg models, respectively. The Zwicker model calculates the loudness for stationary and time-varying sounds, while the Moore-Glasberg model calculates only the loudness for stationary sounds. Moreover, these models have differences in regard to the auditory filter; the Zwicker model uses the Bark frequency scale based on an auditory filter with a symmetric shape whereas the Moore-Glasberg model uses the equivalent rectangular bandwidth (ERB) scale based on an asymmetric filter.

In regard to the ERB scale, it was derived in psychoacoustics studies on auditory filters, it covers the appropriate bandwidth of the auditory filter by taking into account the asymmetrical shape of the auditory filter and off-frequency listening. On the other hand, the Bark scale does not take these auditory characteristics into account. A physiology study showed that 1 ERB<sub>N</sub>-number corresponds to approximately 0.9 mm of the cochlea length, demonstrating the physiological and psychological importance of ERB scale. Since perceived sound is processed by the brain via the auditory filterbank, the SQMs models should incorporate an elaborate auditory filterbank based on the ERB scale.

The Moore-Glasberg model is a stationary loudness model (based on long-term spectra) and can be used directly only to define sharpness. The model uses a rounded-exponential (roex) auditory filterbank that is defined in the frequency domain as a frequency-specific gain without specifying the exact impulse response. It is, therefore, difficult to use this model as a basis for constructing other SQMs model, e.g., classical roughness and fluctuation-strength models, from the time variability of loudness.

In this study, SQMs models were derived that utilize a time-domain auditory filterbank and involve a two-part calculation of specific loudness and modeling of SQMs. In particular, to calculate specific loudness, a time-domain GTFB or GCFB is employed instead of the roex auditory filterbank used in the Moore-Glasberg model. Minute changes in specific loudness, which are essential for modeling SQMs, are captured by the time-domain filterbank. Furthermore, the impact of asymmetric auditory-filter shapes on SQMs estimations was investigated using the GCFB and the GTFB. The specific loudness is then used to model loudness, sharpness, roughness, and fluctuation strength.

To obtain loudness, the summation of specific loudness was calculated; to obtain sharpness, calculation of the specific-loudness centroid was calculated; and to obtain roughness and fluctuation strength, they were calculated from the difference between the peaks and dips in the time direction of the specific loudness.

The proposed loudness model and the Moore-Glasberg model were compared in terms of the root-mean-squared error (RMSE) of loudness, after which the RMSEs of the three other SQMs models were compared with perceived SQMs. It was found that the proposed loudness model can be regarded as a time-domain model for calculating loudness based on the Moore-Glasberg model, as indicated by its very small RMSEs. In addition, it was found that the proposed sharpness, roughness, and fluctuation-strength models could explain actual SQMs better than the previous SQMs model, as evidenced by their lower RMSEs.

**Keywords:** Sound quality metrics, timbre, specific loudness, gammatone auditory filterbank, gammachirp auditory filterbank

# 目次

<b>1</b>	<b>序論</b>	<b>1</b>
1.1	はじめに . . . . .	2
1.2	本研究の背景 . . . . .	8
1.2.1	音色の定義 . . . . .	8
1.2.2	聴覚末梢系の生理的知見 . . . . .	10
1.2.3	聴覚末梢系の音響心理学的知見 . . . . .	13
1.2.4	音質評価指標に関するヒトの知覚 . . . . .	18
1.2.5	ラウドネスの計算モデル . . . . .	20
1.3	本研究の動機と立ち位置 . . . . .	21
1.4	本研究の目的 . . . . .	24
1.5	本論文の構成 . . . . .	26
<b>2</b>	<b>従来研究</b>	<b>29</b>
2.1	音質評価指標の計算モデル . . . . .	30
2.1.1	ラウドネスの計算モデル . . . . .	30
2.1.2	シャープネスの計算モデル . . . . .	31
2.1.3	ラフネスの計算モデル . . . . .	32
2.1.4	変動強度の計算モデル . . . . .	33
2.2	課題 . . . . .	33
2.3	研究方略 . . . . .	34
<b>3</b>	<b>聴覚フィルタバンクを用いたラウドネス密度の計算</b>	<b>36</b>
3.1	ラウドネス密度の算出 . . . . .	37
3.2	外耳・中耳の特性を模したフィルタによる補正 . . . . .	37
3.3	聴覚フィルタによる帯域分割 . . . . .	39

3.3.1	ガンマトーン聴覚フィルタバンク . . . . .	39
3.3.2	ガンマチャープ聴覚フィルタバンク . . . . .	39
3.4	エキサイテーションの算出 . . . . .	41
3.5	エキサイテーションパターンからラウドネス密度への変換 . . . . .	44
3.6	ラウドネス密度の比較 . . . . .	50
3.7	3章の総括 . . . . .	63
<b>4</b>	<b>音質評価指標の計算</b>	<b>64</b>
4.1	音質評価指標の計算モデル . . . . .	65
4.2	ラウドネスの計算 . . . . .	65
4.3	シャープネスの計算モデル . . . . .	70
4.4	ラフネスの計算モデル . . . . .	74
4.4.1	ラフネスの帯域通過フィルタ . . . . .	76
4.4.2	ラウドネス密度におけるピークとディップの差 . . . . .	76
4.4.3	正規化相互相関の計算 . . . . .	77
4.4.4	ラフネス密度とラフネスの計算 . . . . .	78
4.5	変動強度の計算モデル . . . . .	80
4.5.1	変動強度の帯域通過フィルタ . . . . .	80
4.5.2	ラウドネス密度におけるピークとディップの差 . . . . .	80
4.5.3	変動強度密度と変動強度の計算 . . . . .	82
4.6	音質評価指標の計算モデルの評価と結果 . . . . .	82
4.6.1	ラウドネスの計算モデル . . . . .	82
4.6.2	シャープネスの計算モデル . . . . .	87
4.6.3	ラフネスの計算モデル . . . . .	97
4.6.4	変動強度の計算モデル . . . . .	106
4.7	音質評価指標の計算モデルについての総括 . . . . .	113
<b>5</b>	<b>総合考察</b>	<b>114</b>
5.1	はじめに . . . . .	115
5.2	時間領域で定義された聴覚フィルタバンクを用いたラウドネスの計算モデル . . . . .	115

5.3	精緻化された聴覚末梢系のモデルを用いた音質評価指標の計算モデル	116
5.3.1	ERB 尺度と Bark 尺度の違いが音質評価指標の推定に与える影響	116
5.3.2	聴覚フィルタのフィルタ形状の非対称性が音質評価指標の推定に与える影響	117
<b>6</b>	<b>結論</b>	<b>119</b>
6.1	本論文で明らかにされたことの要約	120
6.2	今後の展望	120
6.2.1	本研究成果の利点と限界	120
6.2.2	聴覚末梢系の時間特性の考慮	121
6.2.3	個人の聴覚特性に合わせた音質評価指標の計算モデルの検討	122
6.2.4	騒音の評価	123
6.2.5	今後への期待	123
	<b>謝辞</b>	<b>124</b>
	<b>参考文献</b>	<b>126</b>
	<b>本研究に関する発表論文</b>	<b>135</b>
	<b>本研究以外の発表論文</b>	<b>138</b>

# 目 次

1.1	Processing system of noise suppression method focusing on timbre . . . . .	6
1.2	Structure of the peripheral auditory system . . . . .	9
1.3	Transfer characteristics of the ear canal . . . . .	11
1.4	Transfer characteristics of the middle ear . . . . .	12
1.5	Relationship between the bandwidth of an auditory filter relative to the center frequency of the auditory filter . . . . .	16
1.6	Hierarchical structure of auditory-sensation models. The 1st block represents the auditory filterbank, the 2nd block represents the specific loudness calculation, and the 3rd block represents the loudness, sharpness, roughness, and fluctuation-strength calculations. . . . .	23
1.7	Research issues . . . . .	25
1.8	Thesis organization . . . . .	28
3.1	Block diagram of specific loudness calculation using gammatone auditory filterbank (GTFB) or gammachirp auditory filterbank (GCFB). . . . .	38
3.2	Excitation pattern . . . . .	42
3.3	Excitation pattern . . . . .	43
3.4	The function relating the excitation level at threshold $E_{\text{THRQ}}$ . . . . .	46
3.5	The function relating low-level gain of cochlear amplification $G$ . . . . .	47
3.6	The function relating power exponent of Stevens' power law $\alpha$ . . . . .	48
3.7	The function relating input/output characteristics of the human auditory system $A$ . . . . .	49
3.8	Specific loudness at 100-Hz sinusoidal signal calculated from GT specific-loudness model and Moore-Glasberg model . . . . .	53

3.9	Specific loudness at 1,000-Hz sinusoidal signal calculated from GT specific-loudness model and Moore-Glasberg model . . . . .	54
3.10	Specific loudness at 2,000-Hz sinusoidal signal calculated from GT specific-loudness model and Moore-Glasberg model . . . . .	55
3.11	Specific loudness at 4,000-Hz sinusoidal signal calculated from GT specific-loudness model and Moore-Glasberg model . . . . .	56
3.12	RMSE with specific loudness calculated from the GT specific loudness model and the Moore-Glasberg model . . . . .	57
3.13	Specific loudness at 100-Hz sinusoidal signal calculated from GC specific-loudness model and Moore-Glasberg model . . . . .	58
3.14	Specific loudness at 1,000-Hz sinusoidal signal calculated from GC specific-loudness model and Moore-Glasberg model . . . . .	59
3.15	Specific loudness at 2,000-Hz sinusoidal signal calculated from GC specific-loudness model and Moore-Glasberg model . . . . .	60
3.16	Specific loudness at 4,000-Hz sinusoidal signal calculated from GC specific-loudness model and Moore-Glasberg model . . . . .	61
3.17	RMSE with specific loudness calculated from the GC specific loudness model and the Moore-Glasberg model . . . . .	62
4.1	Block diagram of proposed computational models of sound quality metrics . . . . .	67
4.2	Block diagram of proposed loudness model . . . . .	68
4.3	Relationship between loudness level and loudness. The input signal is a 1,000 Hz sinusoidal signal. . . . .	69
4.4	Block diagram of proposed computational model of sharpness (proposed sharpness model) . . . . .	72
4.5	Weighting function of sharpness model . . . . .	73
4.6	Block diagram of proposed computational model of roughness (proposed roughness model) . . . . .	75
4.7	Weighting function of roughness model . . . . .	79



4.8	Block diagram of proposed computational model of fluctuation-strength (proposed fluctuation-strength model) . . . . .	81
4.9	RMSEs of both versions of proposed loudness model when the input signal is 100 Hz: GT model and GC model. . . . .	84
4.10	RMSEs of both versions of proposed loudness model when the input signal is 1000 Hz: GT model and GC model. . . . .	85
4.11	RMSEs of both versions of proposed loudness model when the input signal is 3000 Hz: GT model and GC model. . . . .	86
4.12	Sharpness of auditory perception and four sharpness models (FZ, Aures, proposed GT sharpness, and proposed GC sharpness models) for narrow-band noise . . . . .	92
4.13	Sharpness of auditory perception and four sharpness models (FZ, Aures, proposed GT sharpness, and proposed GC sharpness models) for high-pass noise . . . . .	93
4.14	Sharpness of auditory perception and four sharpness models (FZ, Aures, proposed GT sharpness, and proposed GC sharpness models) for low-pass noise . . . . .	94
4.15	RMSEs of four sharpness models (FZ, Aures, proposed GT sharpness, and proposed GC sharpness models) for different types of noise	95
4.16	RMSEs of four sharpness models (FZ, Aures, proposed GT sharpness, and proposed GC sharpness models) for different loudness . . .	96
4.17	Relative roughness was calculated using four roughness models for various modulation frequencies. . . . .	99
4.18	Relative roughness was calculated using four roughness models for various modulation frequencies. . . . .	100
4.19	Relative roughness was calculated using four roughness models for various sound-pressure levels. . . . .	101
4.20	Relative roughness was calculated using four roughness models for various sound-pressure levels. . . . .	102

4.21	Relative roughness was calculated using four roughness models for various center frequencies. . . . .	104
4.22	Roughness was calculated using four roughness models for various modulation index of AM signal. . . . .	105
4.23	Fluctuation strength was calculated using three fluctuation-strength models for various modulation frequencies. . . . .	107
4.24	Fluctuation strength was calculated using three fluctuation-strength models for various modulation frequencies. . . . .	108
4.25	Fluctuation strength was calculated using three fluctuation-strength models for various sound-pressure levels. . . . .	110
4.26	Fluctuation strength was calculated using three fluctuation-strength models for various sound-pressure levels. . . . .	111
4.27	Fluctuation strength was calculated using three fluctuation strength models for various modulation index of AM signal. . . . .	112

# 表 目 次

1.1 Pleasantness/Unpleasantness model . . . . .	5
1.2 Type of auditory filter . . . . .	17
3.1 Sound signal used for evaluating specific loudness . . . . .	50
4.1 Sound signal used for evaluating proposed loudness model . . . . .	83
4.2 Narrow-band noise with loudness of 4 sone used for evaluating pro- posed sharpness model . . . . .	90
4.3 High and low-pass noise with loudness 4 sone used for evaluating proposed sharpness model . . . . .	91

# 記号/略語一覧

略語	正式名称
SD 法	意味微分法 Semantic differential method
CB	臨界帯域 Critical band
ERB	等価矩形帯域幅 Equivalent rectangular bandwidth
Roex	Rounded exponential
Moore-Glasberg モデル	Moore and Glasberg のラウドネスモデル
Zwicker モデル	Zwicker のラウドネスモデル
GTFB	ガンマトーン聴覚フィルタバンク Gammatone auditory filterbank
GCFB	ガンマチャープ聴覚フィルタバンク Gammachirp auditory filterbank
IIR	無限インパルス応答 Infinite impulse response
NB	狭帯域 Narrow band
HP	高域通過 High pass
LP	低域通過 Low pass
RMSE	二乗平均平方根誤差 Root mean square error

変数名	定義
$f_s$	サンプリング周波数
$P$	音圧レベル
$P_{s_k}$	GTFB の各フィルタ出力の音圧レベル
$a$	振幅
$f$	周波数
$\omega$	角周波数
$I$	強度
$t$	時間
$\phi$	位相
$s$	観測信号
$z$	Bark レートを基準とした周波数軸
$x$	基底膜の蝸牛入り口からの位置
$A_g$	基底膜の位置と共振周波数を表すためのパラメータ
$a_g$	基底膜の位置と共振周波数を表すためのパラメータ
$b_g$	基底膜の位置と共振周波数を表すためのパラメータ
$f_k$	聴覚フィルタの中心周波数
$K$	聴覚フィルタのチャンネル総数
$k$	聴覚フィルタのチャンネル
$gt_k$	ガンマトーン聴覚フィルタのインパルス応答
$gc_k$	ガンマチャープ聴覚フィルタのインパルス応答
$M$	フィルタの次数
$b$	聴覚フィルタのパラメータ
$c$	周波数変化 (チャープ) の係数
$H_{\text{LPF}}$	漏洩積分処理の伝達特性
$h$	スティーブンスのパラメータ

変数名	定義
$\alpha$	スティーブンスのべき指数
$E_k$	エキサイテーションパターン
$E_0$	1,000 Hz 純音で音圧レベル 0 dB の時の最大エキサイテーション値
$E_{\text{THRQ}}$	可聴レベル
$G$	蝸牛増幅の低レベルゲイン
$A$	聴覚系の入出力ゲイン
$N$	算定ラウドネス
$N'_k$	ラウドネス密度
$Q_N$	ラウドネス係数
$q_{s,k}$	シャープネスに関するラウドネス依存の重み
$w_{S,k}$	シャープネスに関する重み
$S$	算定シャープネス
$Q_S$	シャープネス係数
$N'_{\text{BP},k}$	帯域制限されたラウドネス密度
$i_k$	相互相関値
$\Delta L_{R,k}$	ラフネスモデルに関するラウドネス密度レベルのピークとディップの差
$C_{F,k}$	ラフネスモデル特有の BPF の中心周波数
$H_0, k$	ラウドネス密度の直流成分
$L'_{R,\text{Upper},k}$	ラフネスモデルに関するラウドネス密度の上側包絡線のレベル
$L'_{R,\text{Lower},k}$	ラフネスモデルに関するラウドネス密度の下側包絡線のレベル
$w_{R,k}$	ラフネスに関する重み
$R$	算定ラフネス
$R'_k$	ラフネス密度
$Q_R$	ラフネス係数

変数名	定義
$\Delta L_{F,k}$	変動強度モデルに関するラウドネス密度レベルのピークとディップの差
$L'_{F,Upper,k}$	変動強度モデルに関するラウドネス密度の上側包絡線のレベル
$L'_{F,Lower,k}$	変動強度モデルに関するラウドネス密度の下側包絡線のレベル
$F$	算定変動強度
$F'_k$	変動強度密度
$Q_F$	変動強度係数

関数名	定義
Hilbert	ヒルベルト変換
LPF	低域通過フィルタ
HPF	高域通過フィルタ
BPF	帯域通過フィルタ
HWR	半波整流処理
LT	対数変換処理
$BPF_{R,k}$	聴覚フィルタごとのラウドネスモデルの帯域通過フィルタ
$BPF_F$	変動強度モデルの帯域通過フィルタ

# 第 1 章

## 序論



## 1.1 はじめに

我々が目を閉じた状態で音を聴くと、その音がどのような物から発せられているかある程度正確に予測できると思う。この能力は、私たちの経験や感覚などの。もともとヒトが持っていたり、日々の生活によって培われてきたものである。例えば、車が走行中の音を聴けば、それが馬の足音だと勘違いすることはないだろう。また、未知の音であっても、何かが回転している音なのか、何かをたたいている音なのか、またはこすっている音なのかといった予測ができるはずである。では、我々は一体どのようにして音を知覚しているのだろうか。音は外耳、中耳、内耳を通り最後に脳によって解釈して音として知覚する。耳の構造や脳の音の処理に関する研究は進んでいるが、未だにその全体像は解き明かされていない。

音の基本的属性に、「音の三要素」がある [1]。これらの属性は、主観的な音の大きさを表すラウドネス、主観的な音の高さを表すピッチ、そしてこれら以外の主観的な要素すべてをまとめた音色で構成されている。いくつかの音を誰かに聞かせ、音が大きい順番に並び替えてと尋ねると、並び変えることが可能であろう。同様の質問を音の高さについても行うと、こちらも並び替えることが可能であろう。しかし、音色についてはどうであろうか。今から聞かせる音を音色の高い-低い、もしくは大きい-小さい順番に並び替えてくださいと尋ねると、ほとんどのヒトが答えられないはずである。では、なぜ答えられないのか。これは、音色と一言で言っても、その実態は、ラウドネス、ピッチ以外の様々な、感覚をひとまとめにしているからである。ラウドネスやピッチが一次元の特徴として定義される一方で、音色は多次元で構成されている [2]。では、多次元な音色を一次元に絞って判断させてみよう。例えば、音を甲高く感じる順番で並べてくださいや、ざらざらしている順番に並べてくださいと言われれば、答えることができるはずである。そのため、音色の研究は音色という多次元的な現象を説明する要素に分けるところから始まる。

音の研究の多くは、ラウドネスやピッチなどに焦点を当てている。しかし、音色も音の要素として重要である。もし、ヒトが音色知覚をないがしろにしていたならば、トラックの音と普通自動車の音の違いも判らなかつたはずである。さらに、誰がしゃべっているのかという判断もつかなくなってしまう。逆に言い換えれば、

音色の研究が進めば、ヒトと同じようなプロセスで話者判断や異常音検知などにも応用できる。また、自分の声の音色を保存しておけば、喉の手術によってしゃべられなくなった際に、自分の声の再現などもできるようになるかもしれない。

音色研究では、意味微分法（SD法：Semantic Differential Method）を用いて音色評価尺度を整理、体系化を目的とした研究と、明らかとなった音色評価尺度を数値化し、音の音色の評価を目的とした研究がある。前者の研究では、音楽や騒音、人工音を用いてSD法と因子分析を用いて、主要な音色因子を調べた [2-4]。その結果、音の快・不快にかかわらず音色因子として迫力因子、金属因子、美的因子が共通の因子として抽出されている。一方で、ヒトが音から感じる音色を数値化しようという立場からの検討もある [5]。迫力因子にかかわる音色の指標としてラウドネス、金属因子にかかわる音色の指標としてシャープネス、美的因子にかかわる指標の指標としてラフネスや変動強度がある。以下にそれぞれの指標をまとめる。

**ラウドネス** 主観的な音の大きさ

**シャープネス** 主観的な音の甲高さ

**ラフネス** 主観的な音のざらつき感

**変動強度** 主観的な音の変動感

これらの音色指標は音質評価指標とも呼ばれている。音色と音質は似た概念であり、音質は騒音などの音を評価するときに使われる言葉である。つまり、音質評価指標は、音色を評価するために指標である。ただし、ラウドネスは、本来音色に含まれないが通例として音質評価指標の一部とされている。

ラウドネスやシャープネス、ラフネス、変動強度などの感覚は、製品のサウンドデザイン [6-8]、サウンドスケープ分析 [9-11]、製品の故障判断 [12-14] などに利用されている。また、ラウドネスやシャープネス、ラフネス、変動強度は、Table 1.1 に挙げるように音の快・不快さのモデル化にとって重要な指標となっている。これらの研究は、物理量から音質評価指標を求め、その音質評価指標を用いている。

上記の研究が進めば騒音低減法への応用も考えることができる。Fig. 1.1 に不快音を快音に変換する騒音低減モデルの処理体系を示す。この処理体系では、ま入

力音の快不快さを音質評価指標を用いて分析する。そして、音の不快さにかかわる音響特徴を明らかにし、その音響特徴を打ち消す処理を時間領域と周波数領域の両方で行う。最後に音を合成することで、音から不快さを取り除くことが可能となる。このような不快音を快音に変換させる研究について考えると、入力と出力だけでなく、その途中にあるブラックボックスつまりは、ヒトの聴知覚メカニズムの解明が鍵となる。

Table 1.1: Pleasantness/Unpleasantness model

Pleasantness/Unpleasantness model	Author	Metrics
Model of Sensory Pleasantness [15]	Aures	Loudness Sharpness Roughness Tonality
Psychoacoustic annoyance model [5]	Fastl & Zwicker	Loudness Sharpness Roughness, Fluctuation strength
BTL-scaling the unpleasantness [16]	Ellermeier, Mader, and Daniel	Roughness Sharpness
Probabilistic choice models [17]	Zimmer, Ellermeier, and Schmid	Roughness Sharpness
Noise annoyance model [18]	Morel, Marquis-Favre, and Gille	Loudness Sharpness Roughness Fluctuation strength

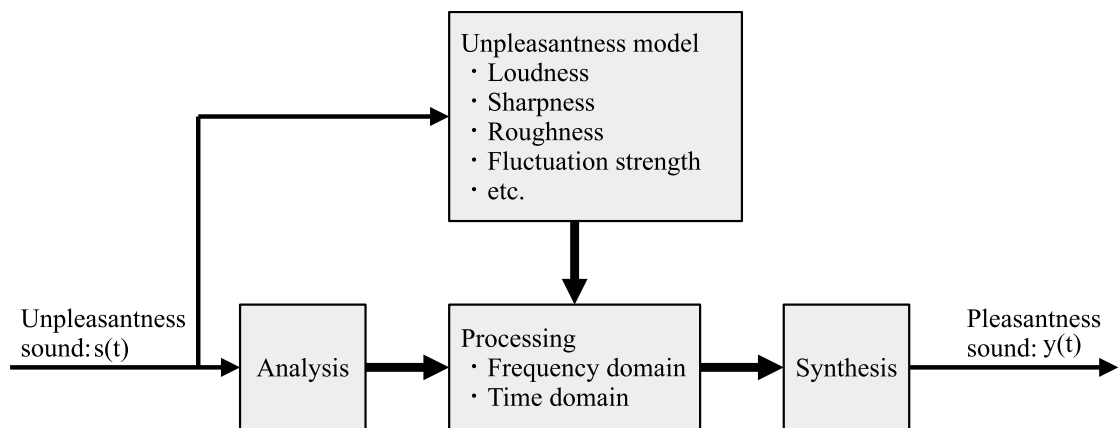


Fig. 1.1: Processing system of noise suppression method focusing on timbre

上記で挙げたシャープネスやラフネス、変動強度のモデル化には、聴覚末梢系の特性を取り入れて検討されてきた [5]。シャープネスやラフネス、変動強度は一般に Zwicker が提案したラウドネスの計算モデル [19] を元に構築されている [20–22]。Zwicker は聴覚末梢系の情報表現（エキサイテーションパターン）からラウドネスは算出できるとし、ラウドネスの計算モデルを提案した [23]。その結果、計算されたラウドネスは主観評価結果とよく一致していることが分かっている。また、ラウドネスの計算モデルを用いて構築された音質評価指標の計算モデルも、音質評価指標に対する主観評価結果とよく一致していることが分かった [20–22]。しかし、Zwicker のラウドネスの計算モデルで使われている聴覚末梢系モデルとヒトの聴覚末梢系の特性は一致していないことがのちの研究で明らかとなっている [24]。

Moore & Glasberg は、ヒトが知覚するラウドネスをモデル化するのだから、ヒトの知覚経路と同じようにラウドネスを計算すべきという立場で、精緻化された聴覚末梢系モデルを用いてラウドネスの計算モデルを提案した [25,26]。その結果、ヒトの聴覚末梢系の機能とよく対応のとれたモデルとなり、聴覚末梢系での情報表現とラウドネスとの関係が明らかとなった。Moore & Glasberg の研究でもあるように、ヒトが知覚する感覚をモデル化するには、精緻化された聴覚末梢系モデルを用いて計算すべきである。人間の聴覚は非常に複雑で、周波数、強度、時間の経過による感度の変化など、多くの因子に依存している。精緻化された聴覚末梢系モデルを用いてこれらの特性を正確にモデル化することで、ヒトの音色知覚により近い音質評価が可能となる。また、精緻化された聴覚末梢系モデルを使うことで、いままではわからなかった、ヒトが音から知覚する様々な音色を理解することが可能となる。この聴覚末梢系の特性を考慮することで、音質評価指標を通じて、より深い音の理解が可能となる。

以上の点から、本研究では、精緻化されたラウドネスの計算モデルを用いて音質評価指標のモデル化を行う。従来のモデルでは明らかにならなかった聴覚機能と音質評価指標との関係の解明をめざす。

## 1.2 本研究の背景

### 1.2.1 音色の定義

音色という言葉聞いてまず初めに思いつくのが，“物理的に異なる二つの音が、たとえ同じラウドネス及びピッチであっても異なった感じに聞こえるとき、その相違に対応する属性”（JIS Z 8106:2000）ではなかろうか。この定義はもともと、Helmholtz の“バイオリン、フルート、クラリネット及び歌声が、同じ音符を同じ高さで演奏されるとき、バイオリンの音をほかのものと区別する要因を音色と呼ぶ”という話からきている。しかし、この定義では、ピッチがない音、例えば民族楽器などでは音色の比較ができなくなるといった問題がある。また、ピッチが明確な楽器の場合、ピッチやラウドネスが違っても、トロンボーンとクラリネットを判別することができるが、JIS Z 8106:2000 の定義では、音色を用いることができない。

難波は、このような問題点を踏まえて、以下のように音色を定義している [27].

1. 音源が何であるか認知（識別）するための手掛かりとなる特性
2. 音を聞いた主体が音から受ける印象の諸側面（多次元的属性）の総称で感情的色彩をおびる。この音色的印象は種々の音色表現語で記述しうる。

また、木村は、音色には狭義の性質と広義の性質があるとした [2]。狭義の音色は、音響スペクトルによって決まるラウドネスやピッチ以外の音の性質を指し、これは、JIS Z 8106:2000 と同じような定義であることがわかる。一方で、広義の音色では、狭義の音色に加えて、ラウドネスやピッチも加えた音の印象すべてを指すと定義した。

本稿では、木村の広義での意味の音色を音色として扱い、この音色の中には、ラウドネスやシャープネス、ラフネス、変動強度を含み、これらを音質評価指標と呼ぶ。

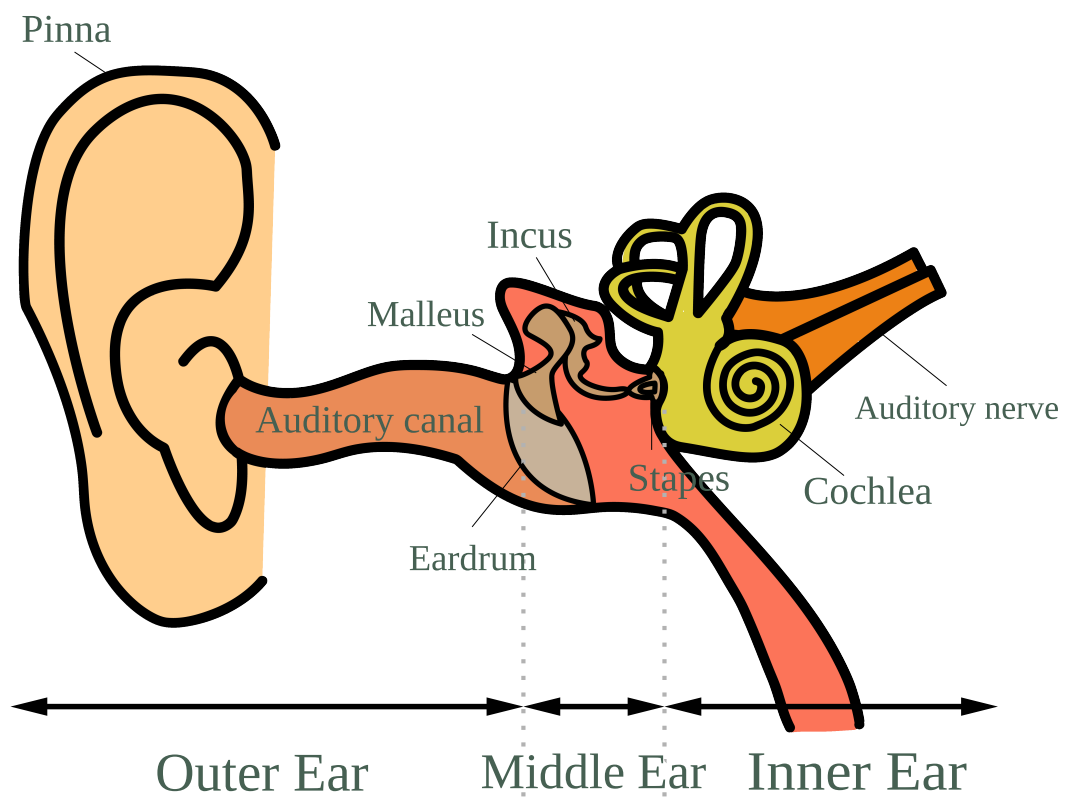


Fig. 1.2: Structure of the peripheral auditory system



## 1.2.2 聴覚末梢系の生理的知見

Fig. 1.2 に聴覚末梢系における音の知覚経路を示す。音は耳介 (Pinna) と外耳道 (Auditory canal) から構成される外耳を通り鼓膜 (Eardrum) を振動させる。鼓膜の振動はツチ骨 (Malleus), キヌタ骨 (Incus), アブミ骨 (Stapes) を伝わり、蝸牛 (Cochlea) と呼ばれる器官に入力される。音は蝸牛によって周波数分析され、帯域ごとに聴神経 (Auditory nerve) を経由し脳に信号が伝達される。この時、蝸牛内では、基底膜と呼ばれる膜の上を音が進行していく。この音が基底膜上を伝わっていく考え方は進行波説として呼ばれている [28]。基底膜には、その位置によって共振周波数が異なっており、ある特定の周波数が大きく振動する部位がいくつも並んでいる。この共振周波数の違いによって、音は周波数分解される。

### 外耳

耳甲介腔では、5,000 Hz あたりで共鳴現象が起き、5,000 Hz 周辺の周波数はそれ以外の周波数と比べて 10 倍ほど大きくなる。また、外耳道は、音響管とみなすことができ、2,500 Hz 付近で共鳴現象が起きる。耳介、耳甲介腔および外耳道を経由した音は鼓膜までに、2,000~7,000 Hz 付近で、低域と高域に比べて音圧が 30~100 倍ほど大きくなる。

Fig. 1.3 に外耳の周波数特性を示す。図中の横軸は周波数、縦軸は音圧レベルを示している。また、拡散音場、自由音場において多少周波数特性が異なっており、それぞれの特性を示している。これらの特性は Killion, Berger & Nuss, Kuhn, および Shaw によって測定された特性の平均値である [29–32]。

### 中耳

空気の比音響インピーダンスはおおよそ 400 Pa·s/m であり、猫における蝸牛の比音響インピーダンスは  $1.5 \times 10^5$  Pa·s/m であり、直接振動を伝達することは非効率である。鼓膜の蝸牛の入り口 (前庭窓) との面積には 35:1 の関係があり、耳小骨は、この面積比によるインピーダンスマッチングを行っている。Fig. 1.4 に中耳の周波数特性を示す。

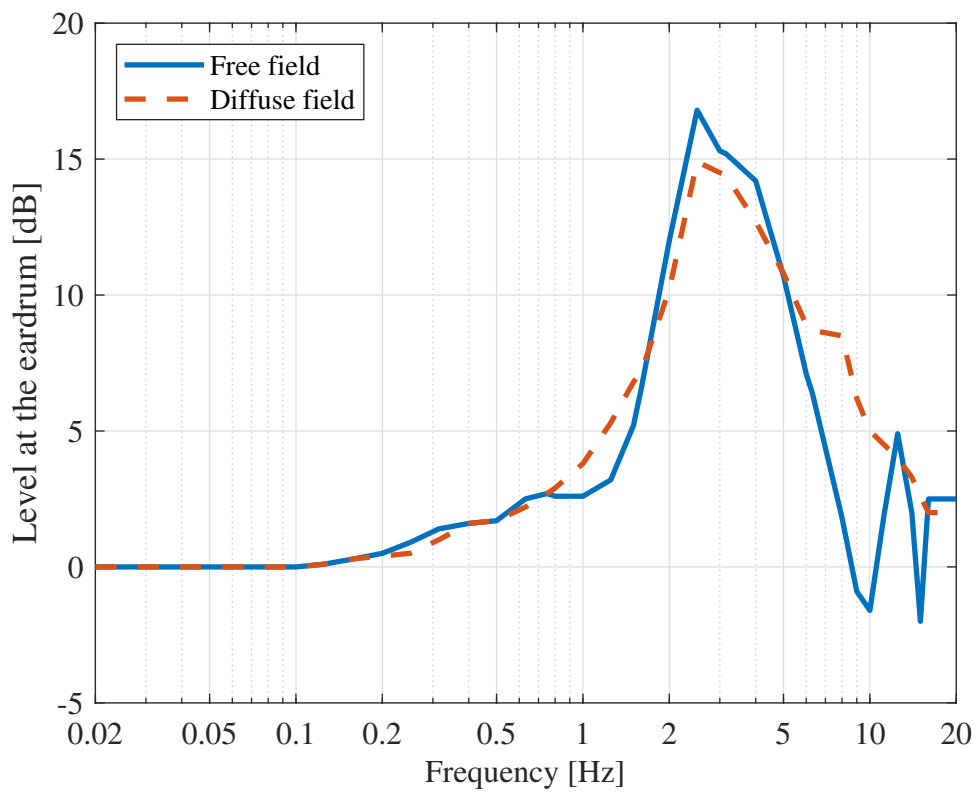


Fig. 1.3: Transfer characteristics of the ear canal (free and diffuse sound fields) [33]

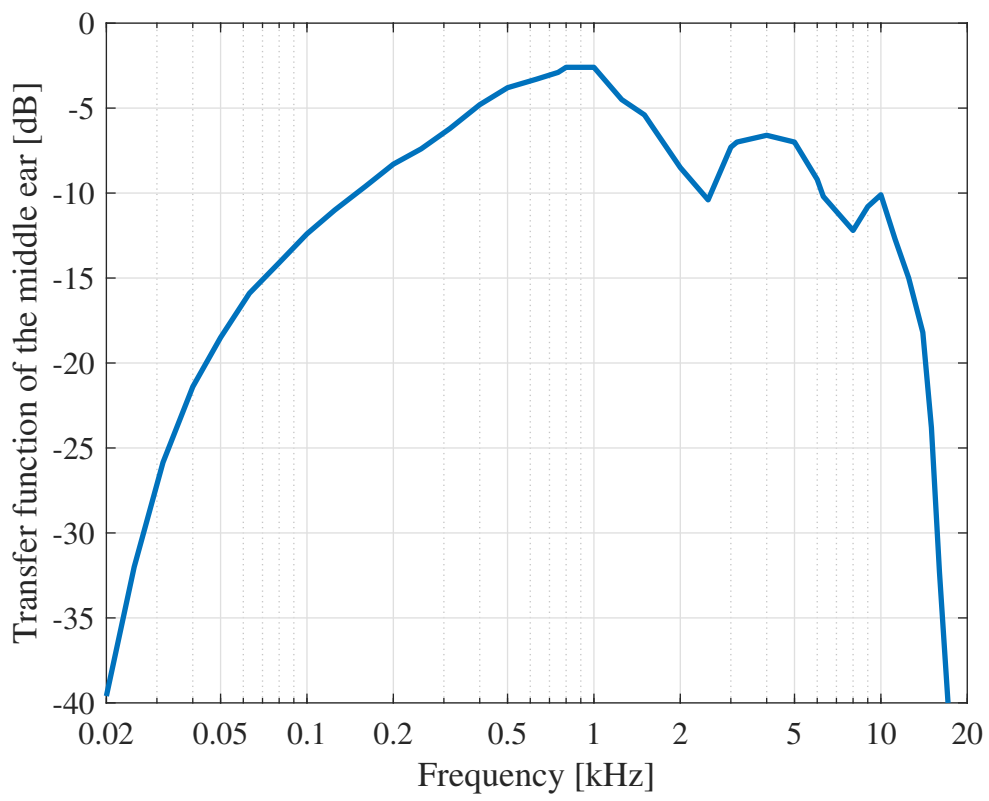


Fig. 1.4: Transfer characteristics of the middle ear [33]

### 1.2.3 聴覚末梢系の音響心理学的知見

#### マスキング

マスキングとは JIS Z 8106:2000 において“(a) 他の (マスクする) 音の存在によって, ある音の聴覚閾値が上昇する現象. (b) (a) の現象による聴覚閾値の上昇量. 単位は, デシベル, 単位記号は, dB.”と定義されている. マスキング実験において, 聞くべき音をマスクと呼び, ほかの音をマスクさせる音のことをマスクと呼ぶ. マスキングは, マスキとマスクが同時に提示されたときに起きる同時マスキングとマスクとマスクが時間的にずれた時に起きる継時マスキングがある [2].

同時マスキングの代表的な検討として, マスキングパターンがある. マスキングパターンは, マスカを固定した状態で, マスキの音圧レベルや周波数を変化させた時の, 聴覚閾値 (dB) をプロットしたものである [34]. また, このグラフはマスキングオーディオグラムとも呼ばれている. Wegel & Lane は, マスカとマスクに純音を用いた. その結果, マスカの周波数とマスクの周波数が近い場合, うなりが生じてしまい, マスキングパターンは複雑なものとなった. この問題を解決するために Egan & Hake は, マスカに純音の代わりに狭帯域雑音を使用した [35]. 狭帯域雑音には, 周波数変調や振幅変調が含まれているため, マスキング実験においてうなりによるマスキングパターンの複雑化を解決することができた. この実験によって, 低域側の傾斜は急峻であり, 高域側の傾斜は緩やかであることが分かった. この結果は, 低域側の音は, 高域側の音をよくマスクすることを意味している. このようなマスキングの低域側から高域側への伸びは upward spread of masking と呼ばれている.

#### 臨界帯域を用いた聴覚フィルタの特徴づけ

Fletcher は純音 (マスク) と狭帯域雑音 (マスク) を用いて聴覚フィルタの臨界帯域を調べた [36]. マスカとその中心周波数を持つマスクを同時にヒトに聞かせる. そうすると, マスキのみを聞かせた時に比べて, マスキングの影響でマスクの検出閾値が上昇する. マスカの帯域幅を広げると, それに伴ってマスキング閾値が帯域幅で変化するため, マスキの検出閾値が上昇する. しかし, 一定の帯域幅を超えるとマスクの検出閾値が上昇しなくなる. これは, 聴覚フィルタの帯域

幅をマスクの帯域幅が超えると、聴覚フィルタによってマスク音が帯域制限され、マスクの検出閾値に影響しなくなることに起因していると考えられる。Fletcherはこの時のマスクの帯域幅を聴覚フィルタの臨界帯域幅（CBW）と名付けた。

一方、Zwickerは臨界帯域幅の概念はラウドネスにも適応できるとし、4つの純音で構成される複合音のラウドネスとその複合音の中心周波数（500, 1,000, 2,000 Hz）を持つ純音のラウドネスをマッチングさせ臨界帯域幅を測定した [37]。この時、臨界帯域幅は複合音の最高周波数と最低周波数の差の関数として、評価された。その結果、最高周波数と最低周波数の差が小さいときは、ラウドネスは一定であったが、ある一定の最高周波数と最低周波数の差を超えるとラウドネスは増加した。また、複合音でなく狭帯域雑音を使用した実験においても同様の結果が得られた。Zwickerは、ヒトの可聴域を24個の臨界帯域が隣接するように並んだ聴覚モデルを提案した [38]。この時Zwickerは、臨界帯域に1から24の番号を付け、この時の帯域の周波数幅をBarkと名付けた。のちに、Zwicker & Terhardtによって、臨界帯域の番号を基準とした周波数軸  $z$  上の値を臨界帯域レートと呼び、以下のような数式で表された [39]。

$$\text{CBW} = 25 + 75(1 + 1.4f_k^2)^{0.69} \quad (1.1)$$

$$z = 13 \arctan(0.76f_k) + 3.5 \arctan(f_k/7.5)^2 \quad (1.2)$$

$f_k$  は聴覚フィルタの中心周波数である。この尺度はBark尺度と呼ばれ、この尺度を用いた聴覚フィルタバンクとして、1/3オクターブ帯域を用いたものがある。

### Greenwoodによる臨界帯域の推定

Greenwoodは、Békésyの基底膜上の共振周波数の位置は対数になっているという考えのもと、マスクングパターンを調べ、基底膜の位置とそこでの共振周波数の関係を示した [40,41]。その結果、基底膜上の位置  $x$  と周波数  $f$  の以下は以下のようなになった。

$$f = A_g(10^{a_g x} - b_g) \quad (1.3)$$

パラメータ  $A_g$ ,  $a_g$ ,  $b_g$  は動物の種類ごとに調整することで、この式のまま使うことができる。ヒトのパラメータは  $A_g = 165.4$ ,  $a_g = 0.06$  であり、 $b_g = 1$  である。

この関係は、死体や、様々な動物から得られた生理学データをよく説明できた。また、この検討から、臨界帯域は基底膜上の 0.9 mm に相当することが明らかとなり、臨界帯域は以下のようにあらわすことができる。

$$\text{CBW} = 0.9a_g A_g \ln(10) \left( \frac{f_k}{A_g} + 1 \right) \quad (1.4)$$

### 等価矩形帯域幅を用いた聴覚フィルタの特徴づけ

Moore & Glasberg は、等価矩形帯域幅 (ERB) を用いた聴覚フィルタのフィルタ形状の推定を行った [42]。その結果、Bark 尺度とは異なり、中心周波数が 500 Hz 以下であっても、聴覚フィルタの帯域幅が狭くなることが明らかとなった。また、Bark 尺度と同様に ERB レートと周波数を関連づける式を導き出した。さらに Moore & Glasberg は、rounded exponential (roex) フィルタを用いて、音圧レベルを考慮した聴覚フィルタの形状推定を行った [43]。その結果、roex フィルタを用いた聴覚フィルタの音圧依存性についても数式で表すことができた。その後、Moore & Glasberg は、実験に使用するイヤフォンや、外耳、中耳の伝達特性を補正した新たな聴覚フィルタのフィルタ形状の推定法を提案した。この推定法には、ノッチノイズデータから聴覚フィルタのフィルタ形状をより正確に導き出すことができることが示された。この時、ERB レートは  $\text{ERB}_N$ -number と呼称が変わり、以下の式のように定義される [44]。

$$\text{ERB} = 24.7(4.37f_k/1,000 + 1) \quad (1.5)$$

$$\text{ERB}_N\text{-number} = 21.4 \log_{10}(4.37f_k/1,000 + 1) \quad (1.6)$$

$\text{ERB}_N$ -number は、Hartmann によって、Moore らの研究所の所在地である Cambridge にちなんで “Cam” という単位がつけられている [45]。この関係は中心周波数が高周波数の聴覚フィルタの推定からも同じであることが実証されている [46,47]。

Fig. 1.5 に Greenwood の検討により明らかとなった臨界帯域と Bark 尺度と ERB 尺度の比較を示す。図中の横軸は周波数であり、縦軸は聴覚フィルタの帯域幅を表している。この結果より、Greenwood の臨界帯域と ERB 尺度が Bark 尺度と比べて、よく対応が取れていることがわかる。この結果より、ERB 尺度は心理学及び生理学の両方から有効性が示されている [48]。

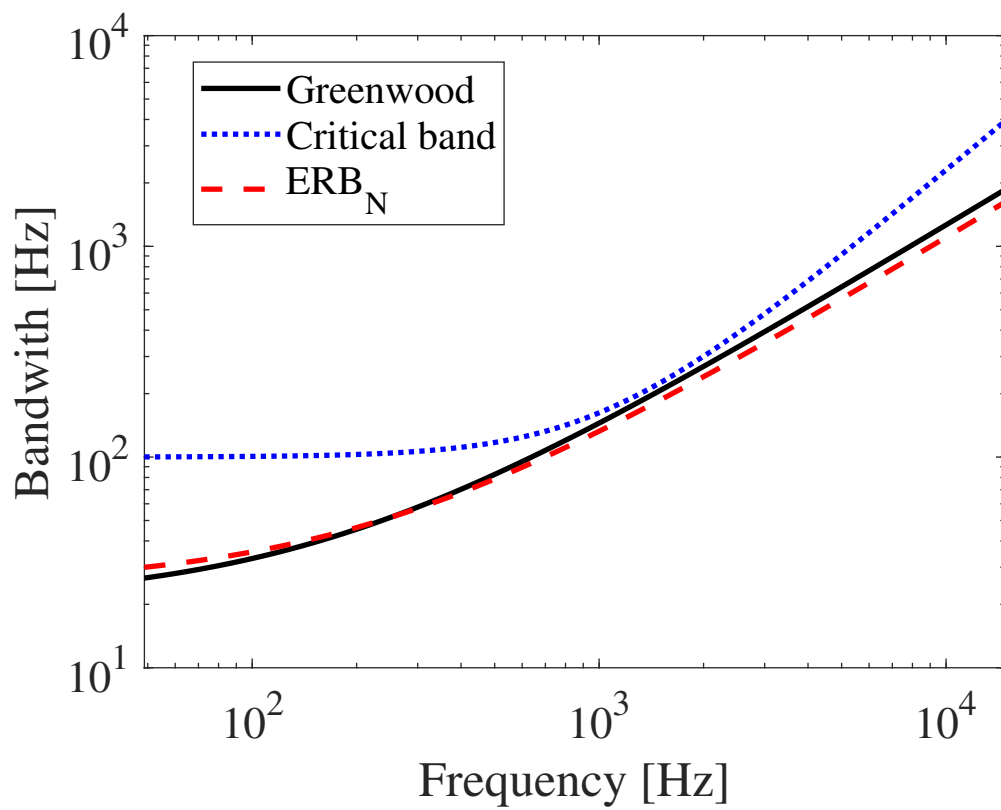


Fig. 1.5: Relationship between the bandwidth of an auditory filter relative to the center frequency of the auditory filter [48]

Table 1.2: Type of auditory filter

Filter name	Defied	Filter shape
Roex filter	Frequency domain	asymmetry
Gammatone filter	Time domain	symmetry
Gammachirp filter	Time domain	asymmetry

## 聴覚フィルタのモデル化

roex フィルタは、聴覚フィルタのフィルタ形状の推定に使用された。このフィルタは、聴覚フィルタのフィルタ形状の非対称性が考慮されている。しかし、周波数領域でフィルタのピーク周波数を中心に低域・高域側で独立に定義されているため、フィルタのインパルス応答を持たない。そのため、時間方向の特性についての検討が難しい。そこで、Patterson らはインパルス応答を持つ聴覚フィルタとしてガンマトーン聴覚フィルタを提案した [49]。現在においても、ガンマトーン聴覚フィルタは、聴覚研究において様々なところで使用されている。このガンマトーン聴覚フィルタに聴覚フィルタのフィルタ形状の非対称性を考慮したガンマチャープ聴覚フィルタが Irino & Patterson によって提案された [50]。Irino & Unoki は分析合成できる聴覚フィルタとして解析的ガンマチャープフィルタを提案した [51]。さらに、解析的ガンマチャープフィルタを無限インパルス応答で実装した聴覚フィルタを提案した [52]。Irino & Patterson は解析的ガンマチャープフィルタでは聴覚の時間特性が表現できないとして、圧縮型ガンマチャープフィルタやダイナミック圧縮型ガンマチャープフィルタバンクを提案した [53, 54]。Table 1.2 に現在聴覚フィルタのモデルとして使われている聴覚フィルタを示す。

## エキサイテーションパターン

Moore-Glasberg は、roex 聴覚フィルタバンクを使用することで、さまざまな周波数や音圧レベルに応じた聴覚系の神経活動の分布を求められることを示した [43]。この神経活動の分布もしくは基底膜振動のことは、エキサイテーションパターンと呼ばれている。マスクが起こすエキサイテーションパターンをそのままマスクングパターンだと解釈して、マスクングパターンはエキサイテーションパターン



のレベルをを全体的に高くしたり低くしたりと平行移動させたものだと考えることができる。

#### 1.2.4 音質評価指標に関するヒトの知覚

各指標について説明する。

##### ラウドネス

ラウドネスは，“小さい”から“大きい”までの尺度上で表される聴覚の属性の一つである。ラウドネスは，Stevensによって検討されてきた心理尺度法の一つであり，比例尺度である。比例尺度は，ラウドネスの順序，差に加えてその比までに意味を持たせている。ラウドネスの単位は“sone”であり，1,000 Hzの純音で，音圧レベル 40 dBの時のラウドネスが1 soneと定義されている。

Stevensは，物理刺激の強さと様々な感覚的主観量がべき乗の関係で結ばれることを示した（Stevensのべき乗則）。次式にラウドネス  $N$  と音の強さ  $I$  もしくは音圧レベル  $P$  としたときのべき乗の関係を示す。

$$N = hI^\alpha = h'P^{2\alpha} \quad (1.7)$$

ここで， $h$ と $h'$ は，条件によって決まる定数であり， $\alpha$ はべき指数である。1,000 Hzの純音におけるべき指数の平均値は0.296である [55]。

しかし，最小可聴値付近では，ヒトが知覚する主観的なラウドネスと比較して，Stevensのべき乗則を用いて計算されたラウドネスは，大きいことが知られている。これは，最小可聴値付近では血流などの生体雑音によって音がマスクされるためにおこる。また，生体雑音以外にも，ピンク雑音や1/3オクターブ雑音などをマスクとして純音におけるラウドネスを測定すると，最小可聴値付近でのラウドネスは，それぞれ異なったふるまいをする。このような純音のラウドネスが別の音のラウドネスによって減少してしまうことをパーシャルラウドネスと呼ぶ。

## ラウドネスレベル

ラウドネスレベルは、ある音のラウドネスと等しいラウドネスに聞こえる 1,000 Hz の純音の強さをレベルで表したものである。ラウドネスレベルの単位は“phon”であり、心理量であるラウドネスとは異なりラウドネスレベルは物理量である。例えば、音圧レベル 60 dB の 1,000 Hz の純音のラウドネスと等しいラウドネスの音は 60 phon である。

Fletcher & Munson は 1,000 Hz の純音を基準として様々な純音のラウドネスと比較した [56]。この時、Fletcher & Munson は基準音の音圧レベルごとに同じラウドネスになる様々な純音の音圧レベルをプロットし、線をつないだ。この線は等ラウドネスレベル曲線と呼ばれている。Robinson & Dadson も同様の実験から等ラウドネス曲線を求めた [57]。さらに、この曲線は ISO226:1987 として規格化された。しかし、この特性には大きな誤差が含まれるとして、Suzuki らは、実験参加者数を増やすため、世界各国の研究室と協力し聴取実験を行い、2003 年に ISO226:1987 は全面改訂され、ISO226:2003 として規格化された [55]。

## シャープネス

シャープネスは、ヒトが感じる音の甲高さを表す指標である [5]。シャープネスの単位は“acum”である。Bismarck は、定常音を対象とした音色の言語属性を調査し、定常音の音色を言語属性によって特徴付けるいくつかの尺度を使って一意に説明できるか判断した [58]。分散が最も大きい因子がシャープネスであり、定常音の音色はシャープネスが最も重要であることを明らかにした。また、Bismarck は、スペクトルの形状とシャープネスの関係を調査し、シャープネスとスペクトル重心の関係を明らかにした [20]。このことから、シャープネスは周波数方向の特徴が重要であることがわかる。

## ラフネス

ラフネスは、ヒトが感じる音のざらつきを表す指標である。ラフネスの単位は“asper”である。Terhardt は、振幅変調音及び周波数変調音に対して、ラフネスと変調度、変調周波数、音圧レベル、キャリア周波数との関係を調査した [59]。その

結果、ラフネスは変調周波数 70 Hz にピークを持つ帯域制限された形状を持つことが明らかとなった。また、ラフネスは、周波数依存性を持っており、聴覚フィルタバンクの帯域幅が 140 Hz 以上では、ラフネスは 70 Hz でピークを持つ。一方、聴覚フィルタの帯域幅が 140 Hz 以下では、ラフネスは、帯域幅の半分の周波数でピークを持つことが明らかとなった。このことから、ラフネスは、時間方向の特徴が重要であることがわかる。

## 変動強度

変動強度は、ヒトが音から感じる変動感を表す指標である。変動強度の単位は“vacil”である。Terhardt は、変動強度についても、ラフネスと同様に、変調度、変調周波数、音圧レベル、搬送波周波数の関係を検討した。その結果、変動強度は変調周波数 4 Hz にピークを持つことが明らかとなった [59]。このことから、変動強度はラフネスと同様に時間方向の特徴が重要であることがわかる。

### 1.2.5 ラウドネスの計算モデル

ラウドネスやシャープネス、ラフネス、変動強度などの音質評価指標はラウドネス密度から推定することができる [5]。ここでは、これら音質評価指標の計算モデルの大前提となる、ラウドネス密度を計算することのできる、ラウドネスの計算モデルについて説明する。

#### Zwicker のラウドネスの計算モデル

ラウドネスは音の強さから求まるものではなく、エキサイテーションパターンから求まるものとして、ラウドネスの計算モデルを Zwicker は提案した [23]。このラウドネスの計算モデルは ISO 532B:1975 として規格化されている [19]。このラウドネスの計算モデルを元として臨界帯域フィルタバンクとして時間領域の 1/3 オクターブフィルタバンクを用いてラウドネスを算出する方法もある。このモデルは定常音に対してのみ使えるラウドネスの計算モデルであるが、時変動音に適應できるラウドネスの計算モデルも検討された [60]。このラウドネスの計算モデルは、

ISO532-1:2017として規格化されている [60]. この方法で実装されたラウドネスの計算モデルは Zwicker モデルと呼ばれる. 計算の基本的な流れは ISO 532B:1975 と同じであるが, 聴覚非線形時間減衰処理を追加することで, 時変動音に対応した.

### Moore-Glasberg のラウドネスの計算モデル

Moore & Glasberg は精緻化された聴覚末梢系モデルを用いたラウドネスの計算モデルを検討した [25,61]. Moore & Glasberg のラウドネスの計算モデルは ISO532-2:2017として規格化された [33]. このラウドネス計算法は周波数領域の聴覚フィルタバンクを用いてモノラル音とバイノーラル音のラウドネスを算出する. この方法で実装されたラウドネスの計算モデルは Moore-Glasberg モデルと呼ばれる. 基本的な計算の流れは ISO 532B:1975 と同じである. Zwicker モデルと Moore-Glasberg モデルの違いは, 周波数スケール (Bark 尺度または ERB 尺度 [62]) や聴覚フィルタの形状 (対称または非対称) などの違いがある. また, Zwicker モデルは古い規格の等ラウドネスレベル曲線 (ISO226:1987) に対応しているが, Moore-Glasberg モデルは新しい等ラウドネスレベル曲線 (ISO226:2003) に対応しているという点も異なっている.

## 1.3 本研究の動機と立ち位置

音の質感を客観的に評価する指標として, ラウドネスやシャープネス, ラフネス, 変動強度などの音質評価指標がある. これらの音質評価指標は, Bark 尺度に基づいたフィルタバンクを利用した Zwicker モデルを用いて計算されたラウドネス密度をもとに推定されている [5]. しかし, Zwicker モデルで使用されている Bark 尺度に基づいたフィルタバンクは, 生理学的に得られた蝸牛内の周波数を分解するための聴覚フィルタバンクと完全に対応していないことが知られている. また, これまでの心理物理学の実験から, ヒトの聴覚フィルタの形状は音圧レベルに依存して非対称になることが分かっているが, この Zwicker モデルではこの特性が考慮されていない. そのため, このラウドネスの計算法では聴覚末梢系での音の表現が正しく行われているか疑問が残る. このラウドネス計算法を用いて算出された音質評価指標では, ヒトの音色知覚を真に理解することができない.

一方、Moore-Glasberg モデルは、生理学と心理学の両方から重要性が示された ERB 尺度を用いている。また、Moore-Glasberg モデルに使われている roex 聴覚フィルタは聴覚フィルタのフィルタ形状の非対称性を考慮している。Moore-Glasberg モデルを用いた音質評価指標の計算モデルが構築できれば、音質評価指標の計算における ERB 尺度の有効性や聴覚フィルタのフィルタ形状の非対称性の重要性を明らかにすることができる。しかしながら、Moore-Glasberg モデルで使われている roex 聴覚フィルタは、周波数領域でフィルタのピーク周波数を中心に低域・高域側で独立に定義されているため、フィルタのインパルス応答を持たない。そのため、ラフネスや変動強度といったラウドネスの時間変動から計算される指標に適應することが難しい。時間シフトしながらラウドネスを計算することは可能だが、ラウドネスの微小な時間変化を捉えることが難しい。そのため、時間領域で計算するラウドネスの計算モデルを提案することが重要となる。音質評価指標をモデル化する手法として、機械学習もあげられる。機械学習を用いて音質評価指標をモデル化することで、高い推定精度を実現可能であるが、ヒトがどのように音質を知覚しているかは不明のままとなってしまう。そのため、本研究の趣旨である“我々は一体どのようにして音を知覚しているのだろうか。”とはずれてしまう。

本研究の新規性は音質評価指標の計算において、従来の Bark 尺度フィルタバンクを用いた Zwicker モデルの代わりに、ERB 尺度を使用し、聴覚フィルタの非対称性を考慮したラウドネスの計算モデルを中心としたアプローチにある。また、この研究の独創性は、時間領域で定義された聴覚モデルを使用して人間が音の質感をどのように知覚しているかを解明しようとする点にある。

本研究では、Fig. 1.6 に示すように、精緻化された聴覚末梢系モデルを用いて算出されたラウドネス密度を中心に、入力信号の物理特徴と音質評価指標との関係について取り扱う。そのために、(1) Moore-Glasberg モデルと同程度に時間領域の聴覚フィルタを用いたラウドネスの計算モデルでラウドネスを計算できるか、(2) ERB 尺度や、聴覚フィルタのフィルタ形状の非対称性が音質評価指標の計算に与える影響は何かを焦点を当てて検討する。

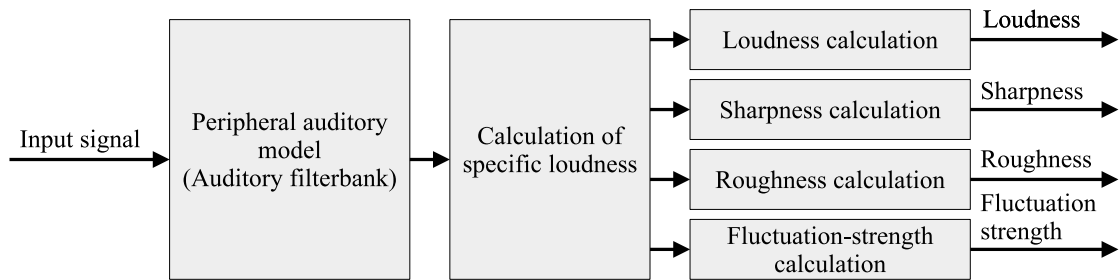


Fig. 1.6: Hierarchical structure of auditory-sensation models. The 1st block represents the auditory filterbank, the 2nd block represents the specific loudness calculation, and the 3rd block represents the loudness, sharpness, roughness, and fluctuation-strength calculations.

## 1.4 本研究の目的

本研究では、「ヒトが知覚する感覚をモデル化するには、精緻化された聴覚末梢系モデルを用いて計算すべきである」という立場をとり、精緻化された聴覚末梢系モデルを用いて音質評価指標の計算モデルを構築する。音質評価指標を構築するにあたり、ラウドネス密度の時間変化が重要となり Moore-Glasberg モデルでは、表現が難しい。本稿では、この問題を解決するために、時間領域で定義される聴覚フィルタとして、GTFB と GCFB を使用してラウドネス密度を計算するモデルを構築する。GTFB と GCFB のどちらの聴覚フィルタも聴覚フィルタのモデルとして様々な聴覚研究に用いられている。特に GCFB は roex フィルタと同様に音圧レベルに依存した聴覚フィルタのフィルタ形状の非対称性を考慮している。

構築したラウドネス密度の計算モデルを用いて音質評価指標の計算モデルを構築する。これによって、Bark 尺度と ERB 尺度の違いによる音質評価指標の推定に与える影響や、聴覚フィルタ形状の非対称性が、音質評価指標の推定に与える影響を明らかにすることができる。

本研究の目的は、ヒトが音の質感を知覚している様相を時間領域で定義される聴覚モデルを用いて解明することである。本研究では (1) 定常音の Moore-Glasberg モデルを元に、時間領域で定義される聴覚フィルタバンクを用いたラウドネスの計算モデルを構築し、このラウドネスの計算モデルと Moore-Glasberg モデルが同程度にラウドネスを算出できるか評価する。(2) 音質評価指標に関する主観評価結果を構築した音質評価指標のモデルが説明できるかを評価する。Fig. 1.7 に本研究で明らかにすることを図示する。

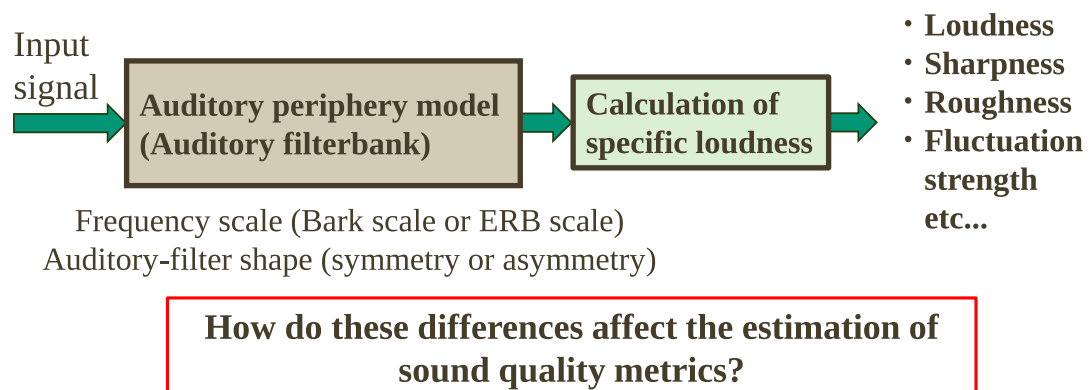


Fig. 1.7: Research issues



## 1.5 本論文の構成

本論文は、7章で構成される。Fig. 1.8 に本論文の構成を図示する。

### 第1章:

本研究の背景について述べ、その問題点を指摘した。また、本研究の立ち位置を明確にし、目的を述べた。

### 第2章:

本研究に関連する研究を述べる。音質評価指標についての知見について説明する。最後に本研究の方針について説明する。

### 第3章:

定常音の Moore-Glasberg モデルに基づいた時間領域で定義された聴覚フィルタバンクを用いたラウドネス密度の計算モデルを構築する

### 第4章:

第3章で構築したラウドネス密度計算モデルの出力であるラウドネス密度を用いて音質評価指標の計算モデルを構築する。先行研究によって明らかにされた音質評価指標の主観評価結果とモデル出力を比較することで、提案した音質評価指標の有効性を確かめる。

### 第5章:

音圧を変化させたときのラウドネス密度と各音質評価指標の密度を比較することで、聴覚フィルタのフィルタ形状の非対称性が音質評価指標の計算モデルに与える影響を調べる。

## 第6章:

本論文によって明らかになったことを要約し，今後の展望について述べる．

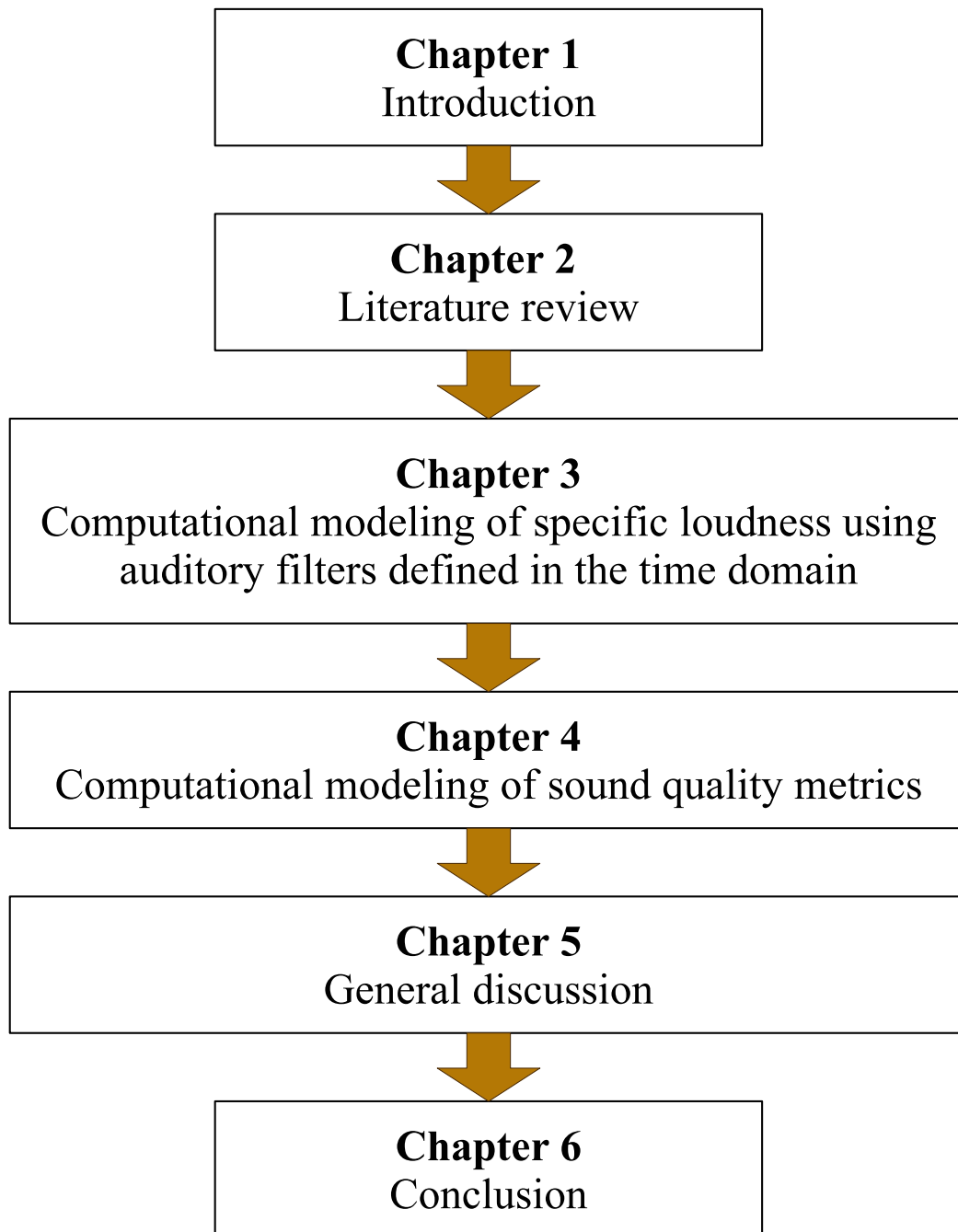


Fig. 1.8: Thesis organization

## 第 2 章

### 従来研究

## 2.1 音質評価指標の計算モデル

### 2.1.1 ラウドネスの計算モデル

ISO532A:1975 規格では、広帯域音のラウドネスの計算モデルが記述されている [19]. Stevens のべき乗則を用いて、3種類のオクターブ帯域 (1 オクターブ, 1/2 オクターブ, 1/3 オクターブ帯域) のラウドネスレベルを計算する. ただし, 低い音圧レベルにおけるラウドネスレベルは計算することができない.

Zwicker は Bark 尺度に従った聴覚フィルタを用いたラウドネスに計算モデルを提案した [19]. また, このモデルは ISO 532B:1975 にラウドネスの計算法として規定された. このモデルは, 時間領域の 1/3 オクターブフィルタバンクを臨界帯域フィルタバンクとして使用してラウドネスを計算している. このモデルは次のようにラウドネスを計算する. (1) 音場を補正し, 外耳と中耳の特性を伝達する. (2) 各臨界帯域のエキサイテーションパターンを計算する. (3) 各聴覚フィルタに対するラウドネス密度を計算する. (4) ラウドネス密度を合計する. 各臨界帯域の励振パターンは, Bark 尺度に従って構築された 1/3 オクターブのフィルタバンクを使用して帯域幅を分割することによって得られる. このモデルは, パーシャルラウドネスの考え方を取り入れることにより, 低い音圧レベルにおけるラウドネスの計算にも適応することができる.

Zwicker は ISO 532B:1975 のラウドネスの計算モデルを改良し, 時間変化音の計算にも適応できるようにした. このラウドネスの計算モデルを Zwicker モデルと呼ぶ. 計算方法は基本的に ISO 532B:1975 と同様だが, 以下の 5つの改良がおこなわれた [60]. (i) 1/3 オクターブフィルタを用いた聴覚フィルタバンクの改良, (ii) エキサイテーションと内部雑音の比からラウドネス密度を計算する改良, (iii) ラウドネスとラウドネスレベルの関係の改良, (iv) 聴覚における非線形時間減衰処理の追加, (v) ラウドネス密度を合計する際の 2 次漏れ積分処理の組み込み.

Moore & Glasberg は精緻化された聴覚末梢系モデルを用いてラウドネスの計算モデルを提案した [61]. このラウドネスの計算モデルを Moore-Glasberg モデルと呼ぶ. このラウドネスの計算モデルは,  $ERB_N$ -number 尺度に従った聴覚フィルタを使用することで, 蝸牛における周波数分解能との対応がよく取れたモデルとなっている. また, ノッチノイズ実験より明らかになった聴覚フィルタの非対称性も

考慮されている。このモデルは、ISO 532-2:2017 にラウドネスの計算法として規定された [33]。このモデルはモノラルおよびバイノーラル音を対象にラウドネスを計算することができる。モノラル音の基本的な計算は ISO 532B:1975 と同様である。Moore-Glasberg モデルは、周波数尺度（Bark 尺度：Zwicker モデル，ERB 尺度：Moore-Glasberg モデル）と聴覚フィルタの形状（対称：Zwicker モデル，非対称：Moore-Glasberg モデル）が Zwicker モデルと異なる。

これらのラウドネスの計算モデルから算出されたラウドネスはあくまでも推定値である。本稿では、主観的なラウドネスと分けるため、ラウドネスの計算モデルによって計算されたラウドネスを算定ラウドネスと呼ぶ。

### 2.1.2 シャープネスの計算モデル

Bismarck は、ラウドネス密度とシャープネスの関係を明らかにし、その後、Zwicker のラウドネスの計算モデルを用いて算出されたラウドネス密度  $N'(z)$  を用いたシャープネスの計算モデルを提案した [20]。Bismarck のシャープネスの計算モデルを以下に示す。

$$S_B = 0.11 \frac{\int_{z=0}^{z=24 \text{ Bark}} N'(z) g_B z / \text{Bark} dz}{\int_{z=0}^{z=24 \text{ Bark}} N'(z) dz} \quad (2.1)$$

$$g_B = \begin{cases} 1 & \text{if } z \leq 15 \text{ Bark} \\ 0.2 \exp(0.308(z/\text{Bark} dz - 15)) + 0.8 & \text{otherwise} \end{cases} \quad (2.2)$$

Fastl & Zwicker は Bismarck のシャープネスの計算モデルを一般化した [5]。Aures は Bismarck のシャープネスの計算モデルの重み  $g_B$  をラウドネス依存モデルに修正した [63]。Aures のシャープネスの重み関数  $g_A$  を次式に示す。

$$g_A = 0.078 \frac{\exp(0.171z/\text{Bark})}{z/\text{Bark}} \frac{N/\text{sone}}{\ln(0.05N/\text{sone} + 1)} \quad (2.3)$$

Swift と Gee は、Moore-Glasberg モデルを用いて計算されたラウドネス密度からシャープネスの計算モデルを提案した [64]。ISO532B:1975 を用いて算出されたラウドネスの計算モデルを用いた Bismarck と Aures のシャープネスの計算モデルは DIN45692:2009 で標準化されている [65]。

これらのシャープネスの計算モデルから算出されたシャープネスはあくまでも推定値である。本稿では、主観的なシャープネスと分けるため、シャープネスの計算モデルを用いて算出されたシャープネス値を算定シャープネスと呼ぶ。

### 2.1.3 ラフネスの計算モデル

Terhardt は結果から、物理量とラフネスをつなげるラフネスの計算モデルを提案した [66]。Terhardt のラフネスの計算モデルを次式に示す。

$$R_T = A(f_{\text{mod}}, f_c) m^2 2^{\frac{P-40}{20}} \quad (2.4)$$

ここで、 $f_{\text{mod}}$  は変調周波数、 $f_c$  は搬送波周波数、 $m$  は変調度、 $P$  は音圧レベルである。また、 $A$  は搬送波周波数ごとに変調周波数とラフネスの関係を表した関数である。Aures は、異なる臨界帯域に対する変調度とラフネスの関係に基づいてラフネスの計算モデルを提案した [67]。

$$R_A = c \sum_{i=1}^n r'_i \Delta z \frac{\rho_{(i-1),i} + \rho_{i,(i+1)}}{2} \quad (2.5)$$

各臨界帯域のラフネス  $r'_i$  は変調度 ( $m_i$ ) のべき乗の関数として表現される。 $\rho_{(i-1),i}$  は臨界帯域間のエンベロップの相関係数であり、無変調の狭帯域雑音や広帯域雑音などの音に対しても適切なラフネスの推定ができるように導入されたものである。このモデルは Daniel & Weber によってラフネスに対する主観評価結果に合うように最適化された [68]。Duisters は、Daniel & Weber のラフネスの計算モデルを元にして、ERB 尺度に基づくラフネスの計算モデルを提案した [69]。このモデルは Daniel & Weber のモデルよりもラフネスに関する主観評価結果と一致した。一方で Widmann & Fastl は、時間的マスキングパターンの山と谷の差  $\Delta L(z)$  がラフネスの知覚に関係することを発見し [70]、ISO 532B:1975 を改良したラウドネスメータを用いて算出したラウドネス密度の時間変化を用いたラフネスの計算モデルを提案した。次式に Widmann & Fastl のラフネスの計算モデルを示す。

$$R_W = 0.3 f_{\text{mod}} \int_{z=0}^{z=24 \text{ Bark}} \Delta L(z) dz \quad (2.6)$$

これらのラフネスの計算モデルから算出されたラフネスはあくまでも推定値である。本稿では、主観的なラフネスと分けるため、ラフネスの計算モデルによって計算されたラフネスを算定ラフネスと呼ぶ。

## 2.1.4 変動強度の計算モデル

Terhardt は変動強度の計算モデルを提案した [59].

$$F_T = \frac{m^2 2^{\frac{F-40}{20}}}{1 + \left(\frac{f_{\text{mod}}}{13}\right)^2} \quad (2.7)$$

Fastl は Terhardt の研究で使われた変調周波数より低い周波数まで変動強度について調べた [22]. その結果, 変動強度は変調周波数 4 Hz にピークを持つ帯域通過フィルタのような特性を持つことが明らかとなった. この結果をもとに, Fastl は時間マスクパターンに基づいた変動強度の計算モデルを提案した. 次に Fastl の変動強度の計算モデルを示す.

$$F_F = \frac{0.008 \int_{z=0}^{z=24 \text{ Bark}} \Delta L(z) dz}{\frac{f_{\text{mod}}}{4} + \frac{4}{f_{\text{mod}}}} \quad (2.8)$$

Vecchi は変動強度と変調周波数の関係から Daniel & Weber のラフネスの計算モデルを元として, 変動強度の計算モデルを提案した [71]. Fastl は, 時間的マスクパターンと変動強度の関係を明らかにし, 変動強度の計算モデルを提案した [22]

これらの変動強度の計算モデルから算出された変動強度はあくまでも推定値である. 本稿では, 主観的な変動強度と分けるため, 変動強度の計算モデルによって計算された変動強度を算定変動強度と呼ぶ.

## 2.2 課題

Fastl の研究は, ラウドネス密度から音質評価指標が計算できるという一貫した視点を持っていた. このことは, 音質評価指標の計算モデル化においてラウドネス密度が重要であることを意味している. また, Duisters の研究は  $ERB_N$ -number 尺度を用いることで聴覚感覚モデルの精度が向上することを示唆している.

ここで, ラウドネスやシャープネス, ラフネス及び変動強度の計算モデルは, ラウドネス密度に基づくことが望ましいと考える. また, 音は脳で一貫して処理されるため, その前段階である聴覚末梢系モデルには精巧な聴覚フィルタバンクが必要であると考え.

このような音質評価指標のモデルを実現するためには以下の問題があげられる. Moore-Glasberg モデルを用いた音質評価指標の計算モデルの研究はシャープネス



に限定されており、ラウドネスの時間変化から得られるラフネスや変動強度のモデル化は困難である。フレーム単位のラウドネスの計算も可能であるが、具体的なラウドネスの微細な変化を捉えることが困難であるため、具体的なラウドネスを算出するためのインパルス応答を持つ時間領域の聴覚フィルタとして用いる必要がある。

Moore-Glasberg モデルをベースに音質評価指標の計算モデルが構築できれば、次の疑問を明らかにすることができる。(A) 聴覚末梢系の情報表現と音質評価指標との対応付けが可能になる。(B) 音圧レベルの変化に対する聴覚フィルタの形状の非対称性が音質評価指標に及ぼす影響を明らかにできる。

(A) については、心理学と生理学の検討によって示された、ERB 尺度の周波数分解能と蝸牛の機能と対応関係から、聴覚末梢系の聴覚フィルタバンクの各中心周波数における音質を決定することができる。これにより、聴覚末梢系における情報表現と聴覚感覚を対応付けることができる。

(B) については、音圧レベルに応じて非対称になる聴覚フィルタの形状は、蝸牛内で低周波音が高周波数へ広がることを upward spread of masking を変化させる。この特性は、音圧レベルを変化させたときの音質評価指標に影響を与えているはずである。

## 2.3 研究方略

これらの検討を行うために、以下に示す順序で精緻化された聴覚末梢系モデルを用いて計算された算定ラウドネス密度を中心に、入力信号の物理特徴と音質評価指標との関係について調査する。

### (1) ラウドネス密度の計算モデルの構築

はじめに、時間領域で定義された聴覚フィルタバンクとして、GTFB 及び GCFB を用いてラウドネス密度の計算モデルの構築を構築する。GTFB と GCFB は roex 聴覚フィルタと同様に聴覚フィルタ形状の推定に使われたものであり、音圧レベルが低い場合は、GTFB と GCFB 及び roex 聴覚フィルタは同程度に聴覚フィルタを推定することができる。一方、音圧レベルが高いときは、音圧依存の聴覚フィルタ形状の非対称性を考慮した GCFB と音圧依存の聴覚フィルタ形状の非対称性を考

慮した roex 聴覚フィルタは同程度にヒトの主観評価結果を説明できる。このことから、Moore-Glasberg モデルで使われている roex 聴覚フィルタの代わりに GTFB や GCFB を使用して、Moore-Glasberg モデルと同程度にラウドネス密度を計算することができると考えられる。聴覚フィルタのフィルタ形状の非対称性を考慮していない GTFB と聴覚フィルタのフィルタ形状が非対称性を考慮した GCFB を用いてラウドネス密度を計算することには、音質評価指標の推定に聴覚フィルタのフィルタ形状の非対称性がどのように影響しているか調べるという狙いがある。

### (2) 音質評価指標の構築及び評価

次に、(1) で構築したラウドネス密度の計算モデルを用いて、音質評価指標の計算モデルを構築する。ここでは、従来のモデルを参考に、構築したラウドネス密度の計算モデルより算出したラウドネス密度に合わせたパラメータの設定を行う。この時のパラメータ設定には、Fastl & Zwicker の検討により明らかになった音質評価指標に関する主観評価結果 [5] をもとにフィッティングを行う。構築した音質評価指標の計算モデルの出力及び従来の音質評価指標の計算モデルの出力と音質評価指標に関する主観評価結果を比較することで、構築した音質評価指標の計算モデルの有効性を評価する。

### (3) 総合考察

最後に、GTFB 及び GCFB を用いた音質評価指標を比較することで、聴覚フィルタのフィルタ形状の非対称性が音質評価指標の推定にどのような影響を与えるか調査し考察する。この検討によって、聴覚フィルタのフィルタ形状の非対称性が音質評価指標の推定にどのような影響を与えるか明らかにすることができる。

## 第 3 章

# 聴覚フィルタバンクを用いたラウドネス密度の計算

### 3.1 ラウドネス密度の算出

Fig. 3.1 に時間領域で定義された聴覚フィルタ（ガンマトーン/ガンマチャープ聴覚フィルタ）を用いたラウドネス密度の計算ブロックダイアグラムを示す。まず、観測信号  $s(t)$  を外耳・中耳の特性を模したフィルタで補正し、聴覚フィルタによって  $K$  個の帯域に分割される。分割された信号に半波整流と漏洩積分処理を行うことで時間領域のエキサイテーションパターン  $E_k(t)$  を得る。次に、エキサイテーションパターンからラウドネス密度  $N'_k(t)$  を求める。最後に、ラウドネス  $N(t)$  はラウドネス密度  $N'_k(t)$  の総和計算から求められる。なお、図中の HWR は半波整流処理、 $(\cdot)^2$  は二乗処理、LPF( $\cdot$ ) は低域通過フィルタ処理である。

### 3.2 外耳・中耳の特性を模したフィルタによる補正

Fig. 1.3 に外耳のフィルタ応答を示す。横軸は周波数、縦軸はフィルタの利得である。青の実線は自由音場における外耳道フィルタであり、赤の破線は拡散音場における外耳道のフィルタである。自由音場、拡散音場のフィルタはラウドネスの計算モデルの使用者が自由に選択することができるが、本稿では自由音場の外耳道フィルタを用いて検討する。Fig. 1.4 に中耳フィルタの周波数特性を示す。横軸は周波数、縦軸はフィルタの利得である。これらのフィルタは Moore-Glasberg モデルで使われているものと同じである。

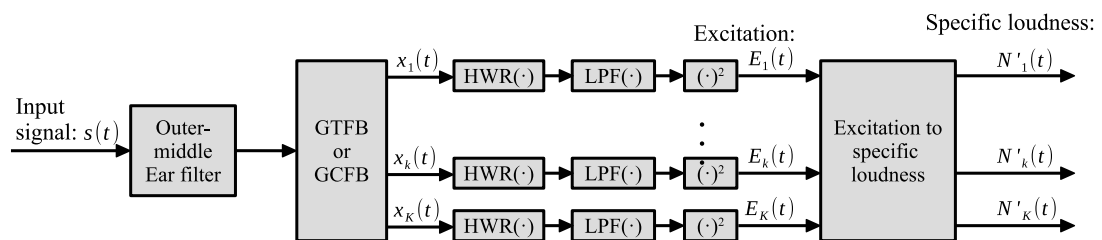


Fig. 3.1: Block diagram of specific loudness calculation using gammatone auditory filterbank (GTFB) or gammachirp auditory filterbank (GCFB).

## 3.3 聴覚フィルタによる帯域分割

### 3.3.1 ガンマトーン聴覚フィルタバンク

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 [49] を利用して、ガンマトーン聴覚フィルタバンク (GTFB) を構築する.

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

ただし、 $a$  は振幅、 $t$  は時間、 $M = 4$  は次数、 $b = 1.019$  は定数、 $f_k$  はフィルタの中心周波数、 $\phi$  は位相である. 等価矩形帯域幅  $\text{ERB}_N(f_k)$  ならびにフィルタの中心周波数  $f_k$  は次式のように定義される.

$$\text{ERB}_N(f_k) = 24.7(4.37f_k/1,000 + 1) \quad (3.2)$$

$$f_k = \left(10^{\frac{\text{ERB}_N\text{-number}}{21.4}} - 1\right) \frac{1,000}{4.37} \quad (3.3)$$

ここで、添え字  $N$  は健聴者の実験から導かれたことを示す. GTFB は Slaney によって提案された 4 つのカスケードの 1 ゼロ 2 極型ガンマトーンフィルタを用いている [72]. GTFB は  $\text{ERB}_N\text{-number}$  が 1.8 Cam ( $k = 1$ ) から 38.9 Cam ( $k = K = 372$ ) まで 0.1 Cam 刻みになるように配置されている. この聴覚フィルタの配置は Moore-Glasberg モデルと同様である.

### 3.3.2 ガンマチャープ聴覚フィルタバンク

解析的ガンマチャープフィルタは、GTFB における聴覚フィルタ形状の非対称性を説明するもので、Irino & Patterson によって提案された [50]. また、この聴覚フィルタバンクには、圧縮特性を考慮した圧縮ガンマチャープ聴覚フィルタ (compressive gammachirp auditory filter)、動的圧縮ガンマチャープ聴覚フィルタ (dynamic-compressive gammachirp auditory filter) も提案されている [54, 73]. Moore-Glasberg モデルで規定された roex 聴覚フィルタバンクは、聴覚フィルタ形状の非対称性を考慮しているが、圧縮特性は考慮していない. Moore-Glasberg モデルと同じ処理を行うために、ラウドネス密度の計算には、圧縮特性を考慮していない解析的な gammachirp フィルタを使用する.

ガンマチャープ聴覚フィルタバンク (GCFB) は、次のように定義される実係数を持つ解析的ガンマチャープ聴覚フィルタのインパルス応答を用いて構築される。

$$gc_k(t) = at^{(M-1)} \exp(-2\pi b \text{ERB}_N(f_k)t) \cos(2\pi f_k t + c \ln(t) + \phi) \quad (3.4)$$

ここで、 $c$ は周波数変化(チャープ)の係数を表し、 $\ln(\cdot)$ は自然対数演算子を表す。ガンマチャープ聴覚フィルタとガンマトーン聴覚フィルタのインパルス応答(式(3.1))との違いは、チャープ項  $c \ln(t)$  だけである。 $c = 0$ のとき、チャープ項 ( $c \ln(t)$ ) は消滅し、ガンマトーン聴覚フィルタと同じ周波数特性になる。式(3.4)を複素インパルス応答に変換し、フーリエ変換すると、ガンマチャープ聴覚フィルタの周波数特性は次のように表される。

$$|G_c(f)| = \frac{|\Gamma(M - jc)| \exp(c\theta(f))}{|2\pi \sqrt{(b \text{ERB}_N(f_k))^2 + (f - f_k)^2}|} \quad (3.5)$$

$$= a_\Gamma |G_T(f)| \exp(c\theta(f)) \quad (3.6)$$

$$\theta(f) = \arctan\left(\frac{f - f_k}{b \text{ERB}_N(f_k)}\right) \quad (3.7)$$

ここで、 $f$ は周波数、 $a_\Gamma$ は振幅を表す。 $\theta(f)$ は  $f_k$  付近で非対称性があるので、 $\exp(c\theta(f))$ は非対称関数である。 $c$ が負のとき、 $\exp(c\theta(f))$ は低域通過フィルタ(LPF)になり、 $c$ が正のとき、 $\exp(c\theta(f))$ は高域通過フィルタ(HPF)になる。これにより、ガンマチャープ聴覚フィルタの低域/高域の非対称性を制御することができる。また、次式のように  $c$ を音圧レベルの関数とすることで、聴覚フィルタのレベル依存性と非対称性を付加することができる。

$$c = 3.38 - 0.107 P_{S_k} \quad (3.8)$$

ここで  $P_{S_k}$ はGTFBの各フィルタの出力による音圧レベルを表す。 $c$ は加重移動平均を用いて  $\text{ERB}_N$ -number 方向に平滑化される。

本稿で使用するガンマチャープ聴覚フィルタのチャープ項  $\exp(c\theta(f))$ は、無限インパルス応答(IIR)フィルタの最小位相フィルタをカスケードして設計されている[74]。GCFBは  $\text{ERB}_N$ -number が 2.6 Cam ( $k = 1$ ) から 36.9 Cam ( $k = K = 344$ ) まで 0.1 Cam 刻みで変化するように配置される。

### 3.4 エキサイテーションの算出

内有毛細胞と聴神経の応答を模擬するために、GTFB及びGCFBの出力に半端整流処理、LPFによる漏洩積分処理、二乗処理を施すことで、そのエキサイテーションパターン  $E_k(t)$  を得る。漏洩積分器の伝達特性  $H_{\text{LPF}}(\omega)$  は次式で示される2つのLPFをカスケード接続した2次LPFとして定義される。

$$H_{\text{LPF}}(\omega) = \frac{a_{\text{LPF}}}{1 - \exp(-2\pi f_c/f_s) \exp(j\omega)} \quad (3.9)$$

$$a_{\text{LPF}} = \frac{1}{1 - \exp(-2\pi f_c/f_s)} \quad (3.10)$$

ここで、 $f_s$  はサンプリング周波数で 44,100 Hz、 $\omega$  は角周波数 ( $2\pi f$ )、 $f_c$  は漏洩積分器のカットオフ周波数で 1,200 Hz である。 $a_{\text{LPF}}$  は漏洩積分器の振幅である。

Fig. 3.2 に roex 聴覚フィルタバンクと GTFB を用いて計算したエキサイテーションパターンを示す。また、Fig. 3.3 に roex 聴覚フィルタバンクと GCFB を用いて計算したエキサイテーションパターンを示す。このエキサイテーションパターンは周波数 1.000 Hz、音圧レベルを 40, 60, 80 dB と変化させたときの結果である。図中の横軸は  $\text{ERB}_N$ -number であり、縦軸はエキサイテーションレベルである。

Fig. 3.2 より、入力信号の音圧レベルが 40 dB の時は、roex 聴覚フィルタを用いたエキサイテーションパターンとガンマトーン聴覚フィルタを用いたエキサイテーションパターンはおおよそ同じであることがわかる。音圧レベルが 80 dB の時のエキサイテーションパターンの高域側のスロープをみると、roex 聴覚フィルタを用いたエキサイテーションパターンに比べて、ガンマトーン聴覚フィルタを用いたエキサイテーションパターンのほうが急峻であることがわかる。これは、ガンマトーン聴覚フィルタが音圧レベルに依存した聴覚フィルタのフィルタ形状の非対称性を考慮していないためだと考えられる。

Fig. 3.3 より、入力信号の音圧レベルが 40 dB の時は、roex 聴覚フィルタを用いたエキサイテーションパターンとガンマチャープ聴覚フィルタを用いたエキサイテーションパターンはおおよそ同じであることがわかる。音圧レベルが高くなると、ガンマチャープ聴覚フィルタを用いたエキサイテーションパターンが高域側に移動していることがわかる。これは、ガンマチャープ聴覚フィルタのフィルタ形状が非対称になるにつれガンマチャープ聴覚フィルタの中心周波数が低域にシフトすることに起因すると考えられる。



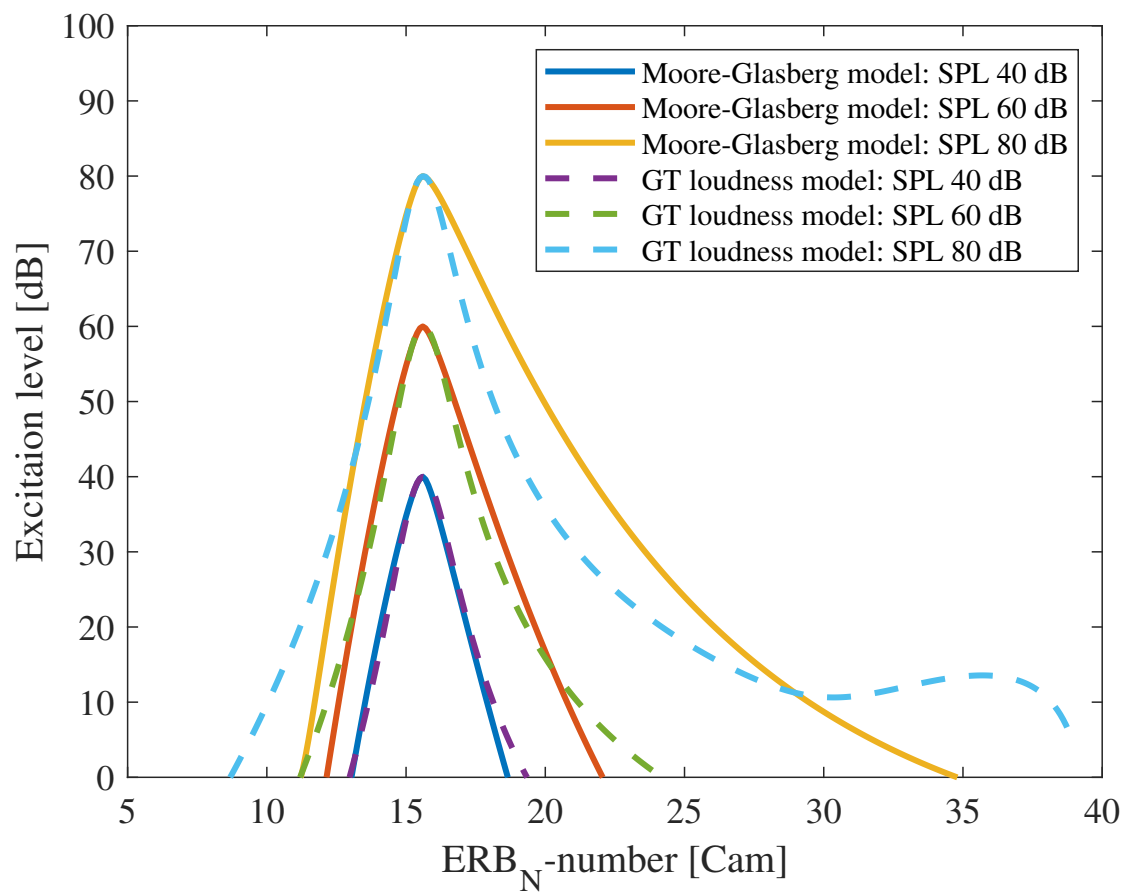


Fig. 3.2: Excitation pattern

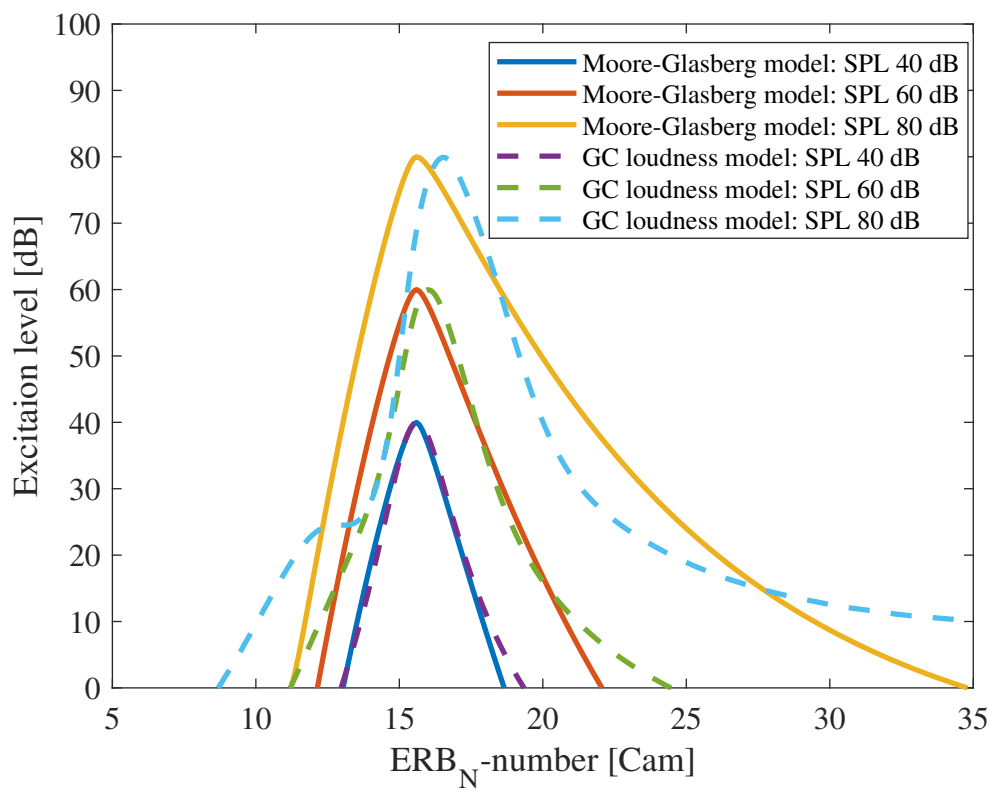


Fig. 3.3: Excitation pattern

### 3.5 エキサイテーションパターンからラウドネス密度への変換

Moore-Glasberg モデルと同じ式を用いて、エキサイテーションパターン  $E_k(t)$  からラウドネス密度  $N'_k(t)$  に変換する。  $N'_k(t)$  は  $E_k(t)$  から次の3つの条件にしたがって計算される。

- Condition of  $E_k(t)/E_0 < E_{\text{THRQ},k}$

$$N'_k(t) = Q_N \left( \frac{2E_k(t)/E_0}{(E_k(t)/E_0 + E_{\text{THRQ},k})} \right)^{1.5} \left( (GE_k(t)/E_0 + A)^\alpha - A^\alpha \right) \quad (3.11)$$

- Condition of  $E_{\text{THRQ},k} \leq E_k(t)/E_0 < 10^{10}$

$$N'_k(t) = Q_N \left( (GE_k(t)/E_0 + A)^\alpha - A^\alpha \right) \quad (3.12)$$

- Condition of  $E_k(t)/E_0 > 10^{10}$

$$N'_k(t) = Q_N \left( \frac{E_k(t)/E_0}{0.99 \cdot 10^{-3}} \right)^{0.2} \quad (3.13)$$

ここで、第一条件 (3.11) は、エキサイテーションが可聴最小値付近にあるときのラウドネス密度を表す式であり、第二条件 (3.12) は、スティーブンスのべき乗則で説明できるエキサイテーションパターンからラウドネス密度への変換式である。第三の条件 (3.13) は、エキサイテーションが非常に大きい ( $E_k(t)/E_0 > 10^{10}$ ) ときのラウドネス密度を表す。  $Q_N$  はラウドネス係数を表し、GTFBを使用する時は  $54.6 \times 10^{-3}$ 、GCFBを使用する時は  $54.8 \times 10^{-3}$ 、 $E_{\text{THRQ},k}$  は可聴レベル、 $G$  は蝸牛増幅の低レベルゲイン、 $\alpha$  は Stevens のべき乗則のべき指数、 $A$  は人間の聴覚系の入出力特性を示す。  $Q_N$  は、次章で説明する構築されたラウドネスの計算モデルを用いて計算したラウドネスが基準音に対して 1 sone となるように設定した。 Fig. 3.4 に  $E_{\text{THRQ},k}$ 、Fig. 3.5 に  $G$  を示す。 図中の横軸は  $\text{ERB}_N$ -number であり、縦軸は各パラメータの値である。 また、Fig. 3.6 に  $\alpha$ 、Fig. 3.7 に  $A$  を示す。  $\alpha$  と  $A$  は、 $G$  の関数で定義される。 Moore-Glasberg モデルで指定した roex 聴覚フィルタバンクと GTFB や GCFB で指定したフィルタバンクの形状は異なるため、これらのフィルタバンクで計算した比ラウドネスの面積も異なる。 そこで、 $\alpha$  に GT ラ

ウドネスモデルで 0.049, GC ラウドネスモデルで 0.047 を加え, Moore-Glasberg モデルによって計算される算定ラウドネス密度と構築されたラウドネス密度の計算モデルによって計算される算定ラウドネス密度の面積を一致させた.

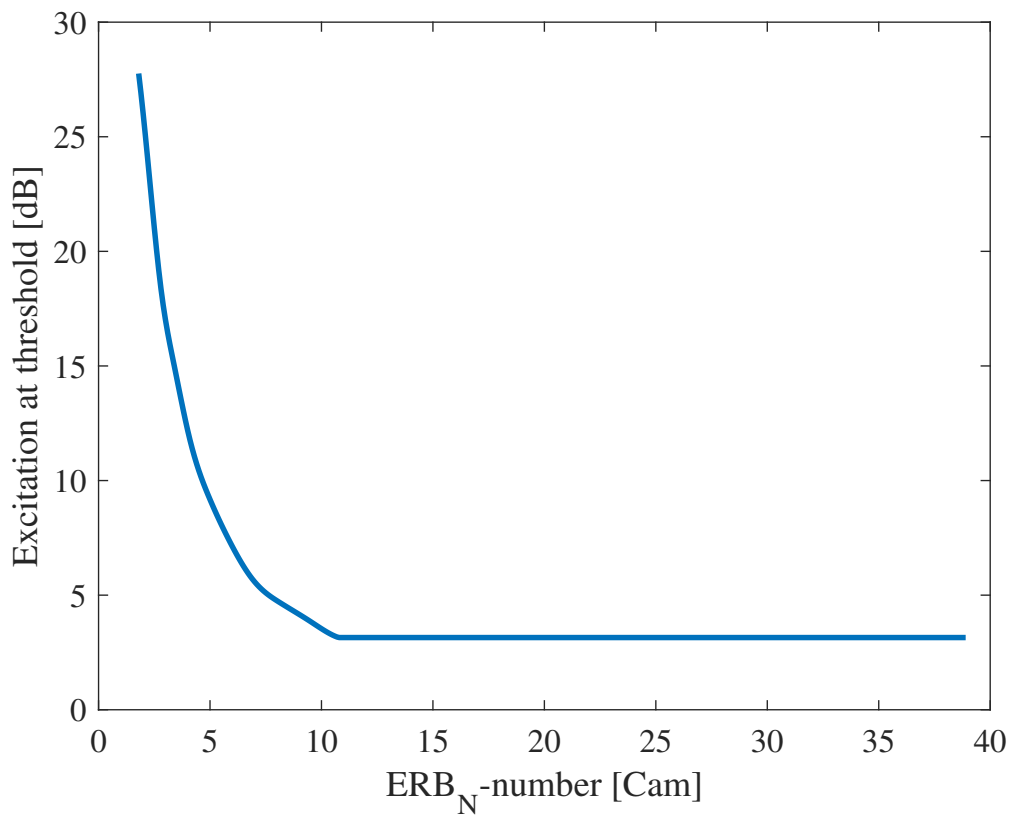


Fig. 3.4: The function relating the excitation level at threshold  $E_{\text{THRQ}}$

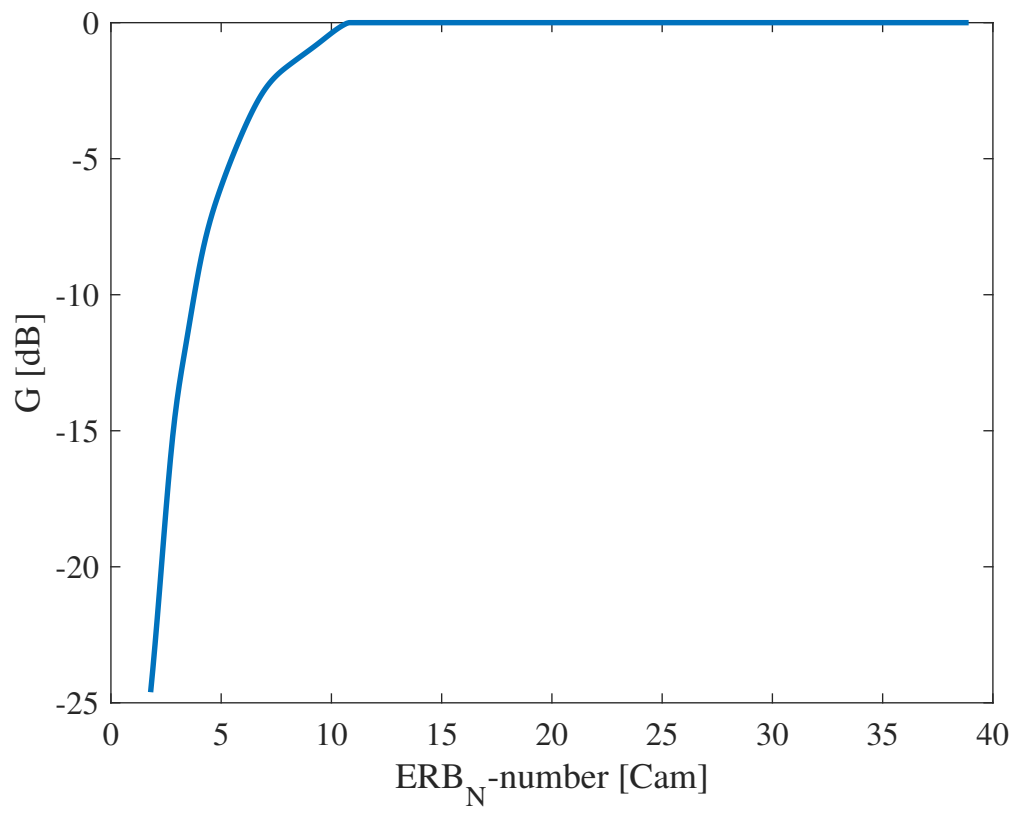


Fig. 3.5: The function relating low-level gain of cochlear amplification  $G$

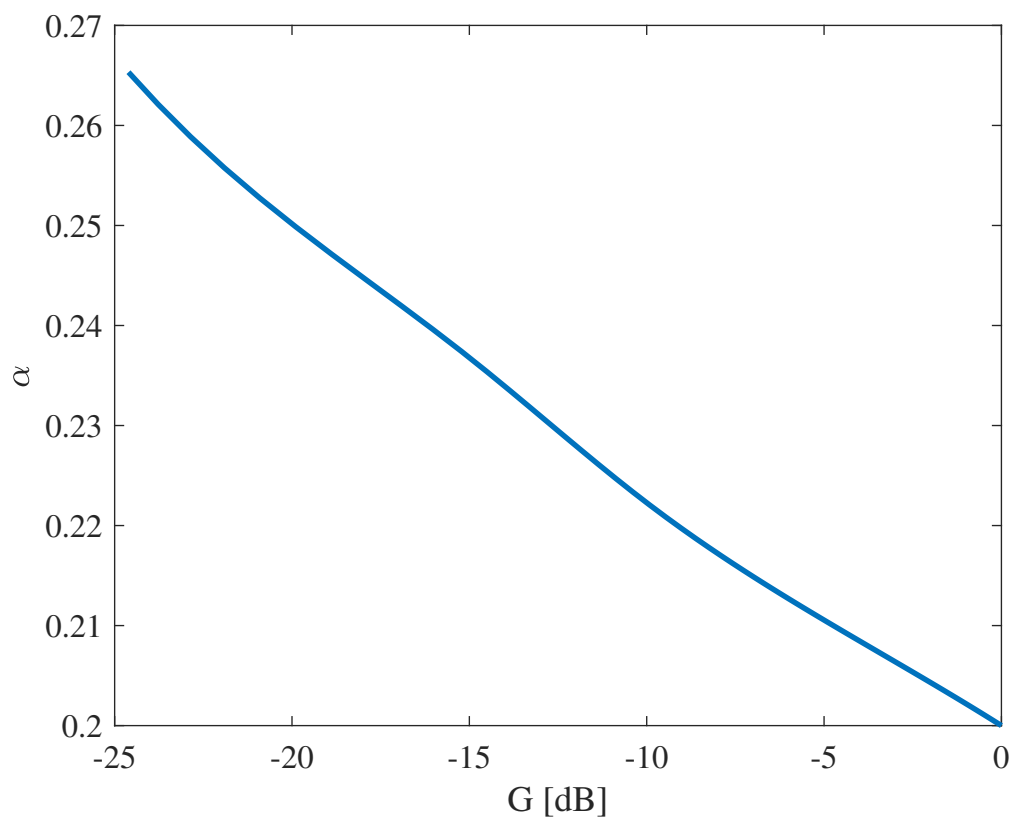


Fig. 3.6: The function relating power exponent of Stevens' power law  $\alpha$

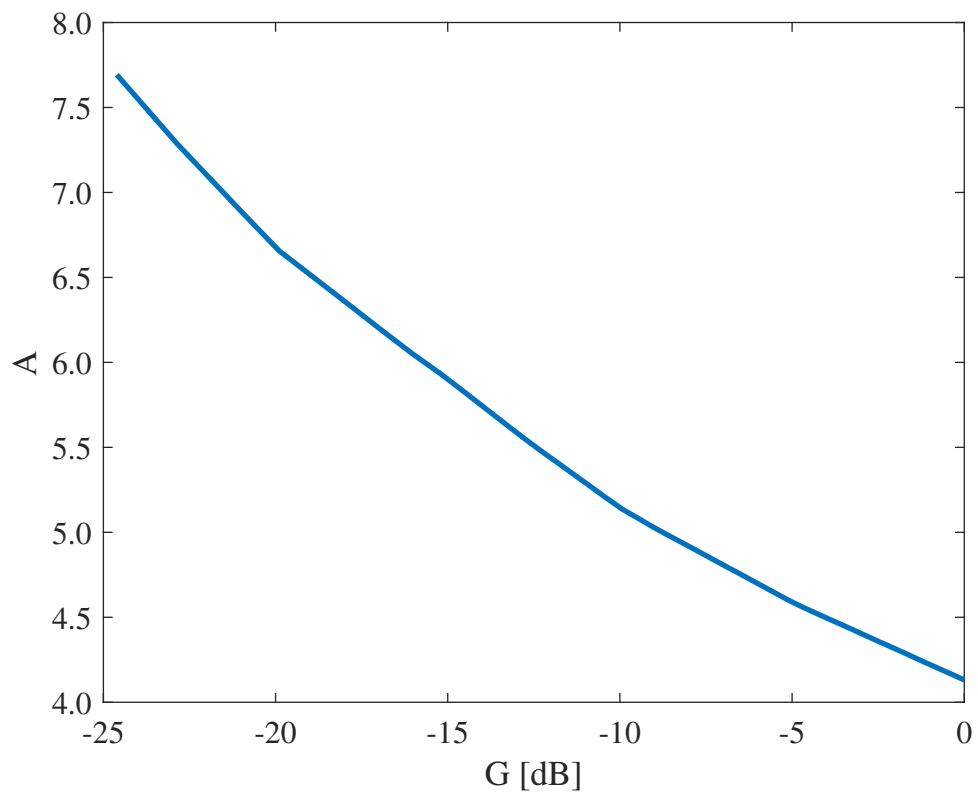


Fig. 3.7: The function relating input/output characteristics of the human auditory system  $A$



Table 3.1: Sound signal used for evaluating specific loudness

Frequency [Hz]	Sound-pressure level(s) [dB]
100	40, 60, 80
1,000	40, 60, 80
2,000	40, 60, 80
4,000	40, 60, 80

### 3.6 ラウドネス密度の比較

Moore-Glabseg モデルを用いて計算した算定ラウドネス密度と GTFB 及び GCFB を用いて計算した算定ラウドネス密度が同程度であることを確かめるために、それぞれのモデルから算出した算定ラウドネス密度の比較を行う。評価方法として、Moore-Glabseg モデルを用いて計算したラウドネス密度に対しての二乗平均平方根誤差 (RMSE) を求めた。

評価には純音の音圧レベルを変化させたときの算定ラウドネス密度を比較する。評価刺激のパラメータを Table 3.1 に示す。ここで、評価刺激のサンプリング周波数は 44,100 Hz とし、時間長は 1 秒とした。Moore-Glasberg モデルを用いて算出された算定ラウドネス密度は長時間スペクトルから求まるものである。一方、GTFB 及び GCFB を用いて算出された算定ラウドネス密度は時間の関数として求めている。そのため、GTFB 及び GCFB を用いた算定ラウドネス密度は時間平均することで、Moore-Glasberg モデルのラウドネス密度と比較した。また、定常部で時間平均をするため、求めた算定ラウドネス密度の前後 0.1 秒を除いた 0.8 秒で時間平均した。

Fig. 3.8, Fig. 3.9, Fig. 3.10, Fig. 3.11 に GTFB を用いて算出した算定ラウドネス密度と Moore-Glasberg モデルによって算出された算定ラウドネス密度を示す。図中の横軸は  $ERB_N$ -number であり、縦軸は算定ラウドネス密度を表している。また、Fig. 3.8, Fig. 3.9, Fig. 3.10, Fig. 3.11 はそれぞれ、評価刺激が 100, 1,000, 2,000, 4,000 Hz の時の算定ラウドネス密度である。実線は Moore-Glasberg モデルを用いて算出された算定ラウドネス密度であり、破線は GTFB を用いて計算された算定ラウドネス密度である。

Fig. 3.8, Fig. 3.9, Fig. 3.10, Fig. 3.11 より, 音圧レベルが低いときは, Moore-Glabseg モデルで計算した算定ラウドネス密度と GTFB を用いて計算された算定ラウドネス密度はおおよそ同じである. 一方で, 音圧レベルが高くなるに付随して, その差も大きくなるように見える. また, 評価刺激の周波数に対応した  $ERB_N$ -number より低い  $ERB_N$ -number での算定ラウドネス密度は, Moore-Glabseg モデルで計算した算定ラウドネス密度と GTFB を用いて計算された算定ラウドネス密度には大きな差がないことがわかる.

Fig. 3.12 に, 音圧レベル及び周波数を変化させたときの Moore-Glabseg モデルを用いて計算した算定ラウドネス密度に対しての GTFB を用いて計算された算定ラウドネス密度との RMSE を示す. 図中の横軸は, 周波数であり, 縦軸は RMSE である. この結果を見ると, 音圧レベルが高くなるについて RMSE も付随して高くなっていることがわかる. また, 周波数が高くなるにつれ RMSE が高くなっている.

これらの結果は, GTFB のフィルタ形状は左右対称であり, 音圧レベルにフィルタ形状は依存しないためだと考えられる. Moore-Glasberg モデルで使われている roex 聴覚フィルタは音圧レベルに依存した聴覚フィルタのフィルタ形状の非対称性を考慮している. この違いが算定ラウドネス密度の計算結果の違いとなったと考えられる.

Fig. 3.13, Fig. 3.14, Fig. 3.15, Fig. 3.16 に GCFB を用いて算出した算定ラウドネス密度と Moore-Glasberg モデルによって算出された算定ラウドネス密度を示す. 図中の横軸は  $ERB_N$ -number であり, 縦軸は算定ラウドネス密度を表している. また, Fig. 3.13, Fig. 3.14, Fig. 3.15, Fig. 3.16 はそれぞれ, 評価刺激が 100, 1,000, 2,000, 4,000 Hz の時の算定ラウドネス密度である. また, 実線は Moore-Glasberg モデルを用いて算出された算定ラウドネス密度であり, 破線は GCFB を用いて計算された算定ラウドネス密度である.

Fig. 3.13, Fig. 3.14, Fig. 3.15, Fig. 3.16 より, 音圧レベルが低いときは, Moore-Glabseg モデルで計算した算定ラウドネス密度と GCFB を用いて計算された算定ラウドネス密度には大きな差がないように見える. 一方で, 音圧レベルが高くなると, GCFB を用いて計算された算定ラウドネス密度のピークは,  $ERB_N$ -number の高いほうに遷移していることがわかる. また, 評価刺激の周波数に対応した

ERB<sub>N</sub>-number より低い ERB<sub>N</sub>-number での算定ラウドネス密度は、Moore-Glabseg モデルで計算した算定ラウドネス密度と GCFB を用いて計算された算定ラウドネス密度で異なっていることがわかる。

Fig. 3.17 に、周波数及び音圧レベルを変化させたときの Moore-Glabseg モデルを用いて計算した算定ラウドネス密度に対しての GCFB を用いて計算された算定ラウドネス密度との RMSE を示す。図中の横軸は、周波数であり、縦軸は RMSE である。この結果を見ると、音圧レベルが高くなるにつれて RMSE も付随して高くなっていることがわかる。また、周波数が高くなるにつれ RMSE が高くなっている。

これらの結果は、GCFB のフィルタ形状の非対称性は音圧レベルに依存しているが、その非対称性は roex 聴覚フィルタとは異なっているためであると考えられる。GCFB のフィルタ形状の非対称性は、音圧レベルが高くなると、ガンマチャープ聴覚フィルタの最大振幅の周波数が低域に遷移する。また、ガンマチャープ聴覚フィルタの左右の裾は音圧レベルが高くなると、低域側に移動する。しかし、roex 聴覚フィルタは、フィルタの中心周波数を中心に、低域側の裾のみが低域に遷移するように定義されている。このようなフィルタ形状の違いによって、算定ラウドネス密度の算出結果が異なると考えられる。

ラウドネスの知覚は、対数であり、ラウドネス密度も対数として考えることができる。Fig. 3.9, Fig. 3.10, Fig. 3.11 の縦軸を見ると、音圧レベル 80 dB の時の Moore-Glasberg モデルの算定ラウドネス密度の最大値はおおよそ 4 sone/ERB であり、この音が 2 倍に聞こえるときは 8 sone/ERB となり、1/2 に聞こえるときは 2 sone/ERB となる。このことを踏まえると、GTFB や GCFB を用いた算定ラウドネス密度の計算モデルから算出された算定ラウドネス密度と Moore-Glasberg モデルから算出された算定ラウドネス密度の違いは、大きくないと考えられる。

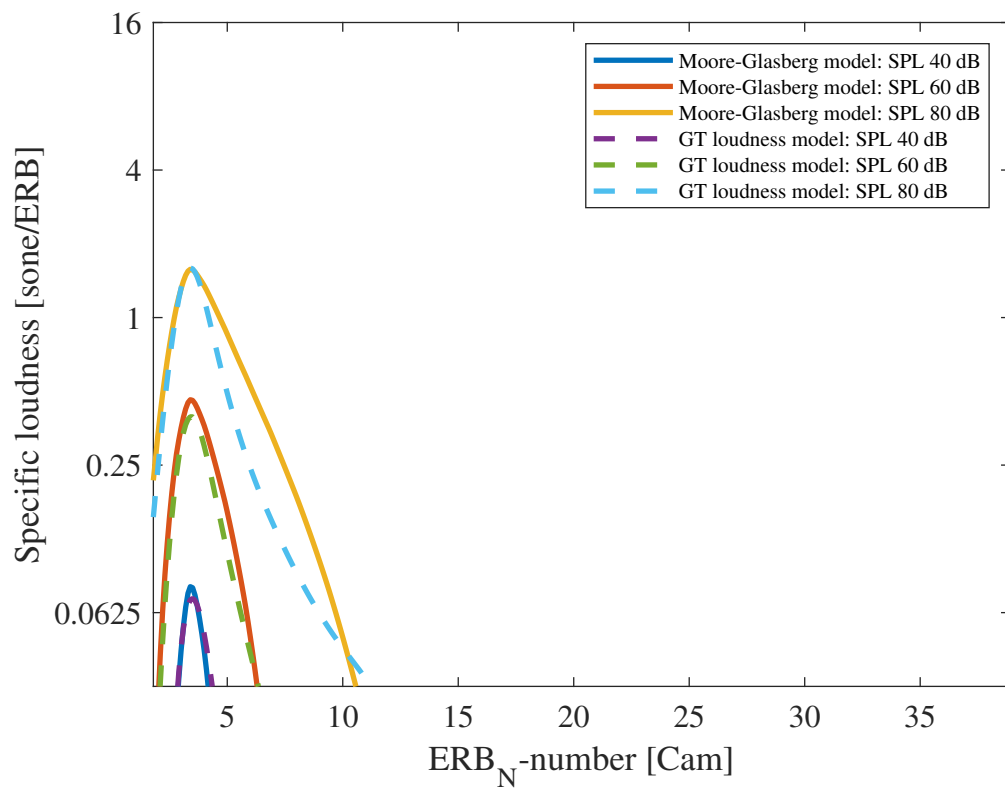


Fig. 3.8: Specific loudness at 100-Hz sinusoidal signal calculated from GT specific-loudness model and Moore-Glasberg model

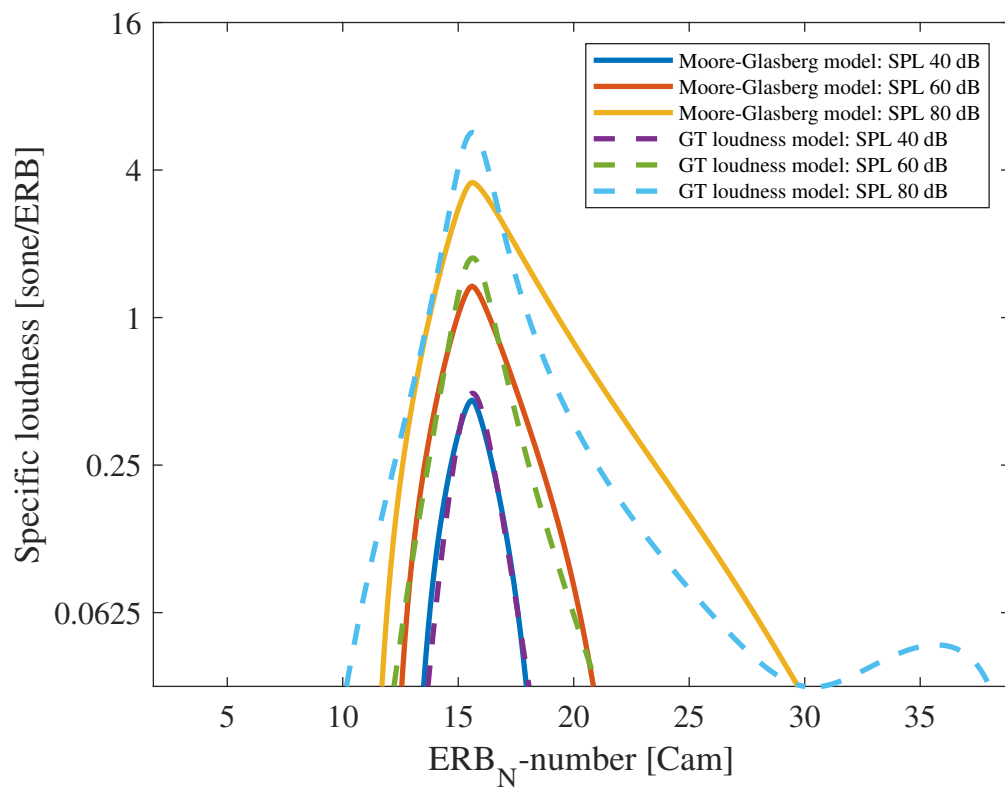


Fig. 3.9: Specific loudness at 1,000-Hz sinusoidal signal calculated from GT specific-loudness model and Moore-Glasberg model

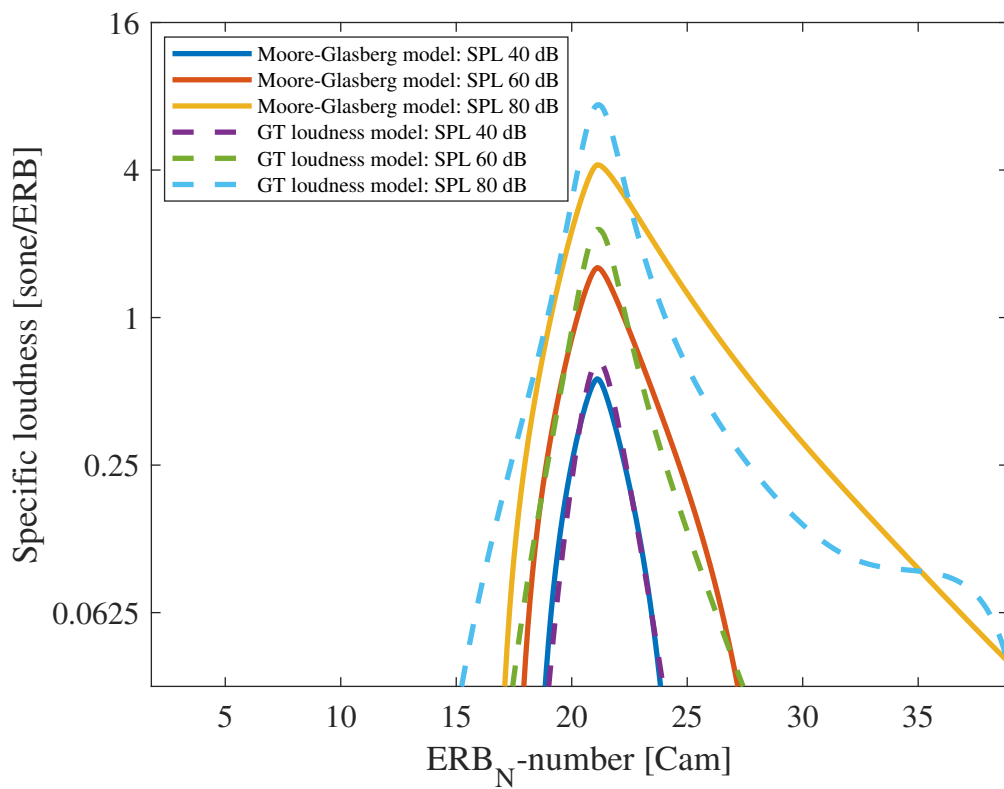


Fig. 3.10: Specific loudness at 2,000-Hz sinusoidal signal calculated from GT specific-loudness model and Moore-Glasberg model

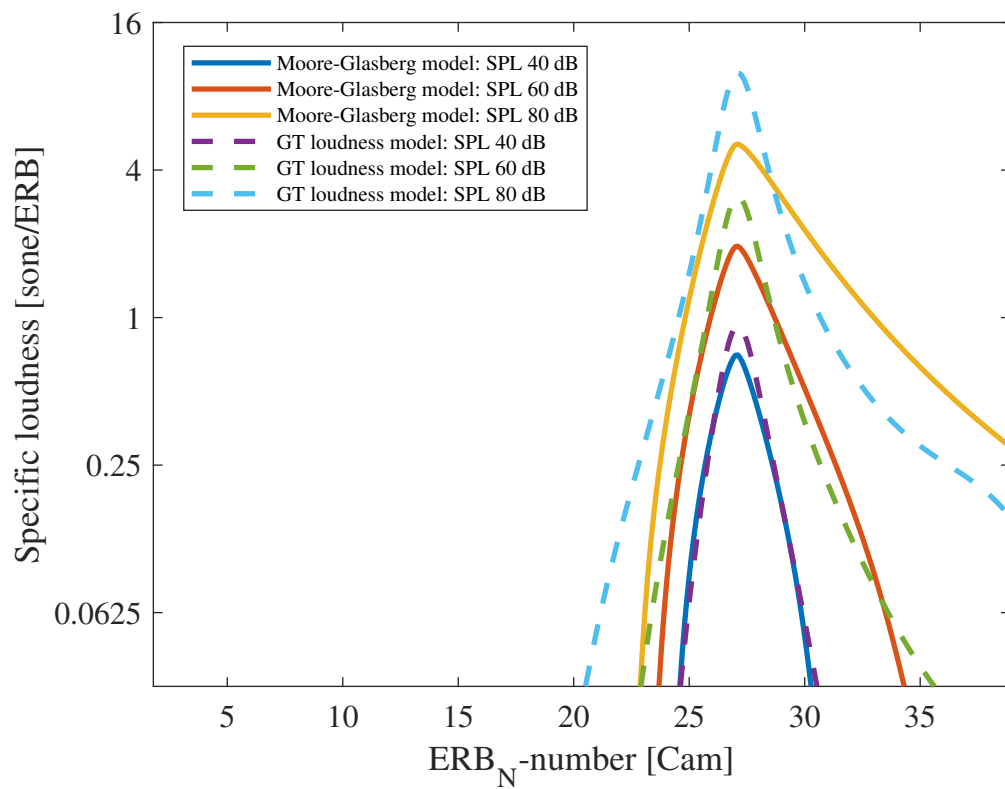


Fig. 3.11: Specific loudness at 4,000-Hz sinusoidal signal calculated from GT specific-loudness model and Moore-Glasberg model

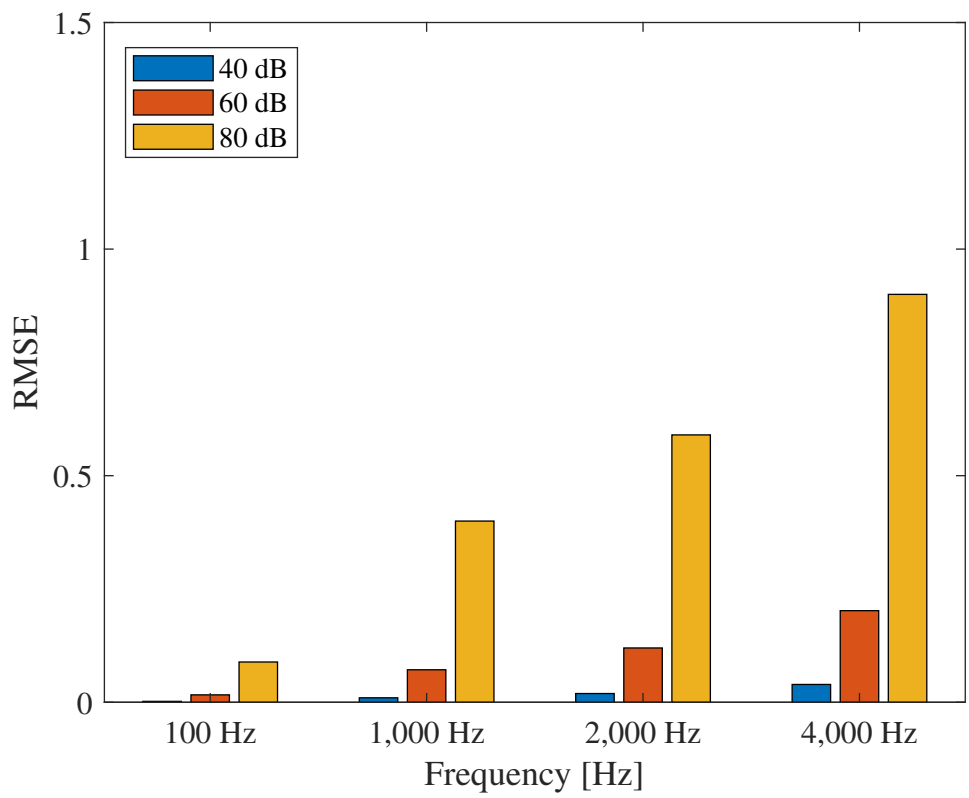


Fig. 3.12: RMSE with specific loudness calculated from the GT specific loudness model and the Moore-Glasberg model



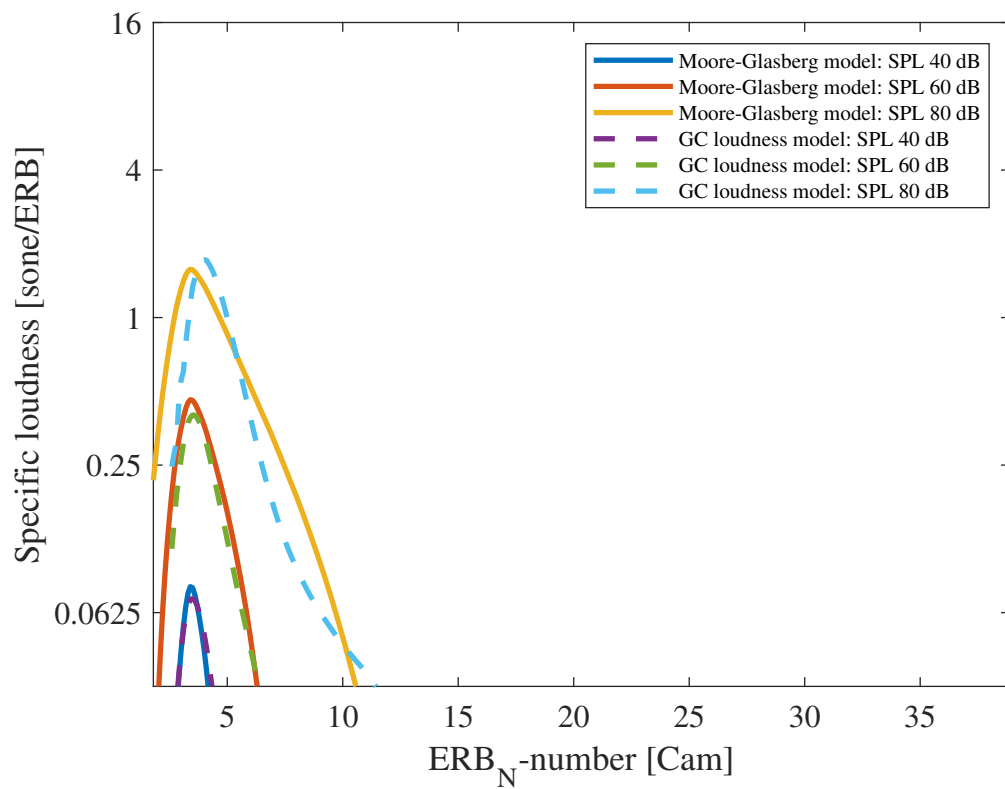


Fig. 3.13: Specific loudness at 100-Hz sinusoidal signal calculated from GC specific-loudness model and Moore-Glasberg model

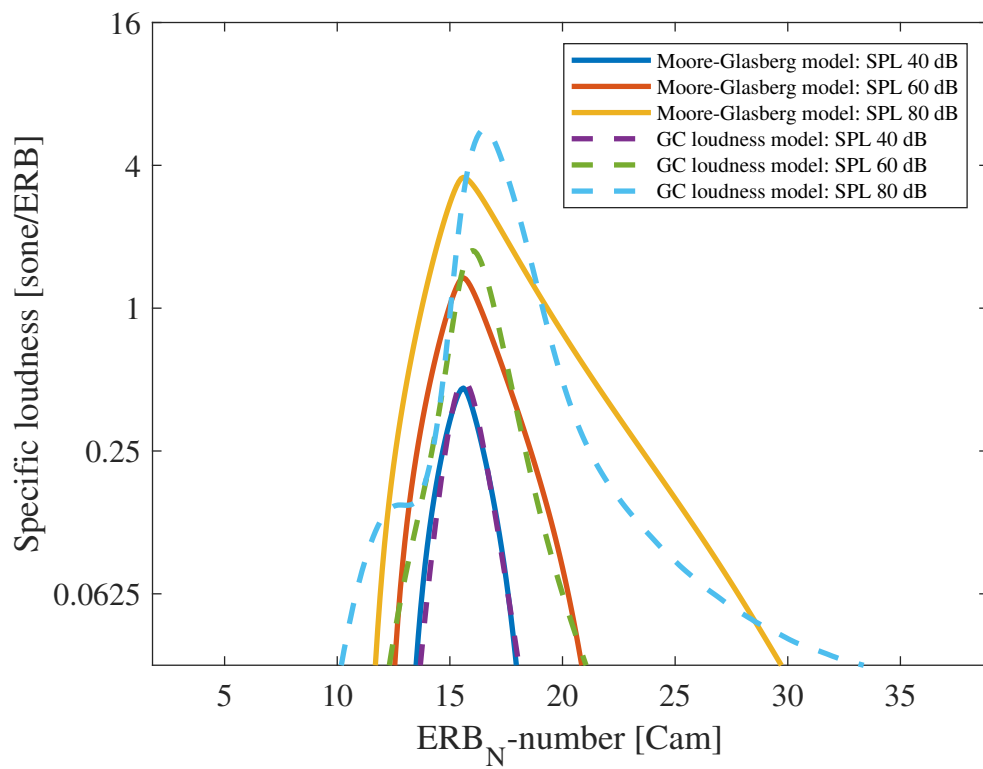


Fig. 3.14: Specific loudness at 1,000-Hz sinusoidal signal calculated from GC specific-loudness model and Moore-Glasberg model

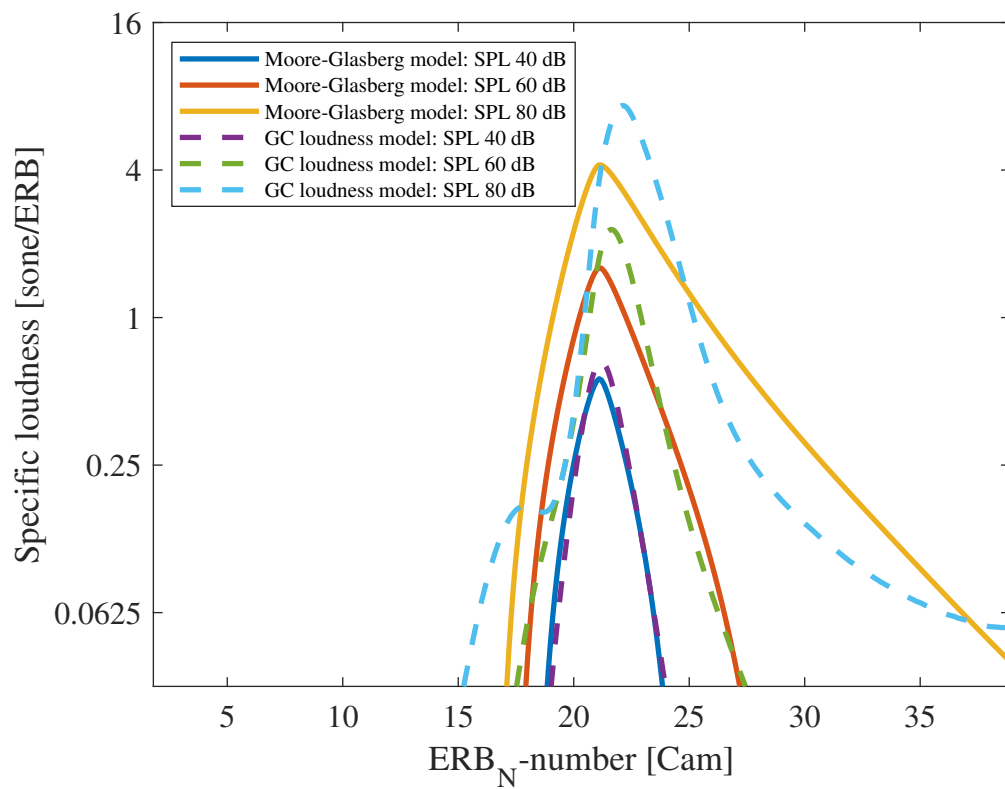


Fig. 3.15: Specific loudness at 2,000-Hz sinusoidal signal calculated from GC specific-loudness model and Moore-Glasberg model

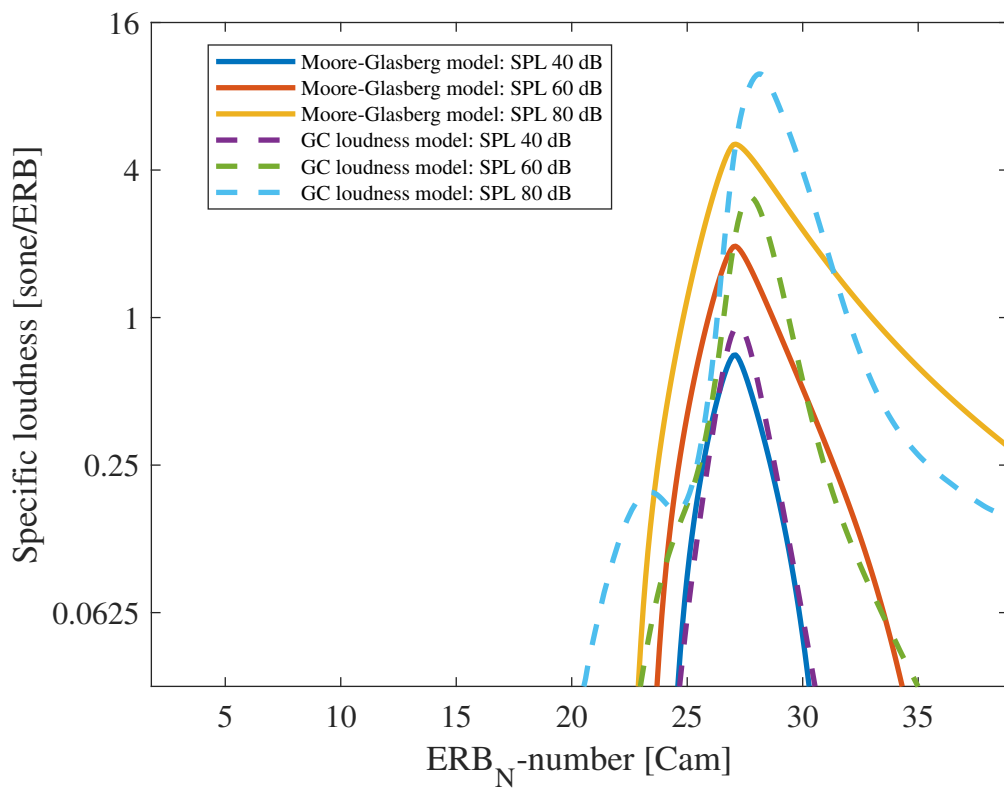


Fig. 3.16: Specific loudness at 4,000-Hz sinusoidal signal calculated from GC specific-loudness model and Moore-Glasberg model

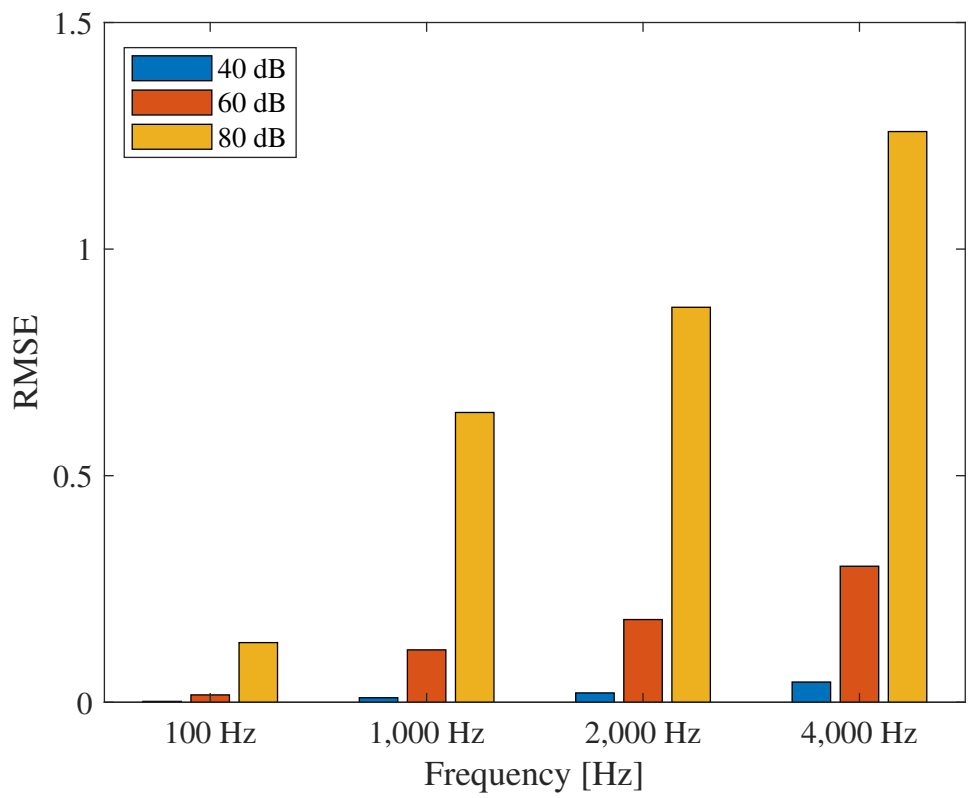


Fig. 3.17: RMSE with specific loudness calculated from the GC specific loudness model and the Moore-Glasberg model

### 3.7 3章の総括

本章では、音質評価指標の計算モデルの前提となるラウドネス密度の計算モデルの構築と評価を行った。構築したラウドネス密度の計算モデルは Moore-Glasberg モデルに基づいて時間領域で定義される聴覚フィルタバンクを用いた。評価の結果、構築したラウドネス密度の計算モデルから算出された算定ラウドネス密度は Moore-Glasberg モデルから算出された算定ラウドネス密度を基準として RMSE が 1 程度であった。ラウドネスが対数知覚であることを考えれば、この誤差は小さいと考えられる。

次章では、本章で構築したラウドネス密度の計算モデルを前提としたシャープネス、ラフネス、変動強度の計算モデルの構築を行う。算定シャープネスは、ラウドネス密度の重心の計算によって、算定ラフネスと算定変動強度は、ラウドネス密度の時間変動から計算される。

## 第 4 章

### 音質評価指標の計算

## 4.1 音質評価指標の計算モデル

Fig. 4.1 に構築する音質評価指標の計算モデルのブロックダイアグラムを示す。3章において構築したラウドネス密度の計算モデルを用いて、観測信号  $s(t)$  からラウドネス密度  $N'(t)$  が計算される。そして、得られたラウドネス密度  $N'(t)$  をラウドネス積分すると算定ラウドネスが計算され、 $N'(t)$  の重み付き重心を計算することで、算定シャープネスを計算する。また、算定ラフネスと算定変動強度は  $N'(t)$  を時間方向に分析することで計算される。

## 4.2 ラウドネスの計算

Fig. 4.2 にラウドネスの計算モデルのブロックダイアグラムを示す。GTFB を使ったものを GT ラウドネスモデル、GCFB を使ったものを GC ラウドネスモデルと呼ぶ。算定ラウドネス  $N(t)$  は次式で定義される算定ラウドネス密度の総和から得られる。

$$N(t) = 0.1 \sum_{k=1}^K N'_k(t) \quad (4.1)$$

このラウドネスの計算モデルから算出された算定ラウドネスはあくまでも推定値である。

Fig. 4.3 にラウドネスレベルを 0 から 120 dB まで変化させたときの、周波数 1,000 Hz の純音に対する Moore-Glasberg, GT ラウドネス, GC ラウドネスモデルを用いて計算した算定ラウドネスを示す。横軸はラウドネスレベル phon であり、縦軸は算定ラウドネス sone である。黒の破線は Moore-Glasberg モデルから算出された算定ラウドネスであり、青の点線は GT ラウドネスモデルから算出された算定ラウドネスであり、赤の実線は GC ラウドネスモデルから算出された算定ラウドネスである。

ラウドネスレベルが 80 phon 以上の場合、GT ラウドネスモデルの算定ラウドネスは Moore-Glasberg モデルの算定ラウドネスよりも低く計算されていることがわかる。しかし、80 phon 以下では、GT ラウドネスモデルの算定ラウドネスは Moore-Glasberg モデルの算定ラウドネスと一致する。GC ラウドネスモデルの算定ラウドネスは、全ラウドネスレベルにわたって Moore-Glasberg モデルの算定ラ



ウドネスとほぼ同じである。これらの結果は、Moore-Glasberg モデルで使われている roex 聴覚フィルタのフィルタ形状が左右非対称であるのに対して、ガンマトーンフィルタのフィルタ形状は対称であることが起因していると考えられる。一方、ガンマチャープフィルタのフィルタ形状は Moore-Glasberg モデルで使われている聴覚フィルタと同様に音圧に依存して非対称になることから、Moore-Glasberg モデルと GC ラウドネスモデルの算定ラウドネスは同程度になったと考えられる。

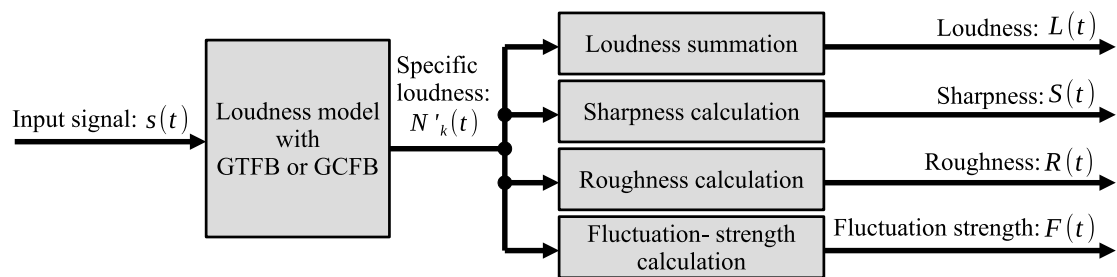


Fig. 4.1: Block diagram of proposed computational models of sound quality metrics

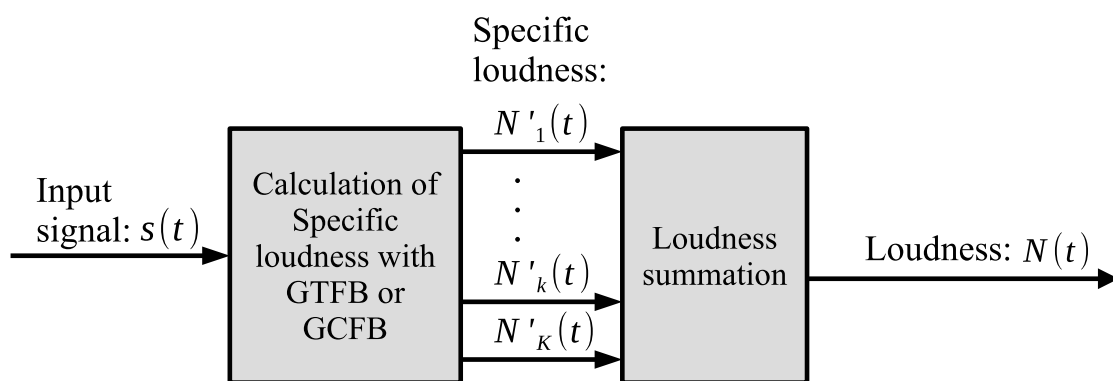


Fig. 4.2: Block diagram of proposed loudness model

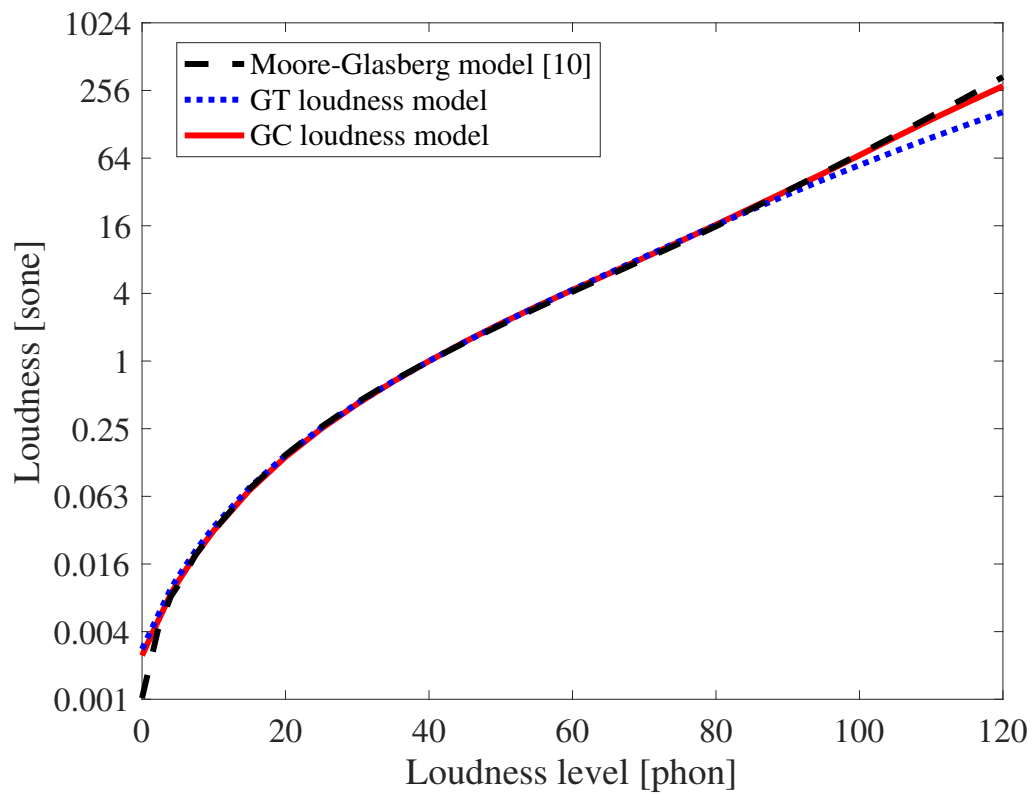


Fig. 4.3: Relationship between loudness level and loudness. The input signal is a 1,000 Hz sinusoidal signal.

### 4.3 シャープネスの計算モデル

構築するシャープネスの計算モデルは、Bismarck のシャープネスの計算モデル [20] を改良した、ラウドネス依存性を考慮した Aures のシャープネスの計算モデル [63] に基づいて設計されている。

Fig. 4.4 に GTFB または GCFB を用いたシャープネスの計算モデルのブロックダイアグラムを示す。GTFB を用いたものを GT シャープネスモデル、GCFB を用いたものを GC シャープネスモデルと呼ぶ。シャープネス  $S(t)$  は次式で定義される  $N'_k(t)$  から重み付き重心を計算することによって得られる。

$$S(t) = Q_s \frac{\sum_{k=1}^K q_{s,k}(t) N'_k(t) \text{ERB}_{\text{N-number}}}{\sum_{k=1}^K N'_k(t)}, \quad (4.2)$$

$$q_{s,k}(t) = \frac{w_{s,k} \sum_{k=1}^K N'_k(t)}{\text{ERB}_{\text{N-number}} \times \ln \left( \frac{N(t)+20}{20} \right)}, \quad (4.3)$$

ここで、 $Q_s$  はシャープネス係数を表し、GTFB を用いた場合は  $2.29 \times 10^{-3}$ 、GCFB を用いた場合は  $2.23 \times 10^{-3}$ 、 $q_{s,k}(t)$  はラウドネスに応じて変化する重み、 $w_{s,k}$  は重み関数を表す。 $Q_s$  は、構築するシャープネスの計算モデルを用いて計算した算定シャープネスが基準音に対して 1 acum となるように設定した。また、 $w_{s,k}$  は、Fastl & Zwicker の研究から引用されたシャープネスに関する主観評価と本モデルの出力との間の二乗平均平方根誤差 (RMSE) を最小化するようにフィッティングした。 $w_{s,k}$  は次式によって定義される。

$$\begin{aligned} w_{s,k} = & 1.19 \times 10^{-3} \text{ERB}_{\text{N-number}}^3 \\ & - 4.90 \times 10^{-2} \text{ERB}_{\text{N-number}}^2 \\ & + 7.17 \times 10^{-1} \text{ERB}_{\text{N-number}} \\ & - 2.01. \end{aligned} \quad (4.4)$$

このシャープネスの計算モデルから算出された算定シャープネスはあくまでも推定値である。

Fig. 4.5 にシャープネスの重みを示す。図の横軸は周波数であり、縦軸はシャープネスの重みである。シャープネスの重みを比較するために、Bark 尺度に基づいたシャープネスの計算モデルの重みは Bark 尺度から周波数に変換してプロットし

た. ERB 尺度に基づいたシャープネスの計算モデルの重みも同様に, ERB 尺度から周波数に変換してプロットした. また, それぞれのシャープネスの計算モデルの重みが 1,000 Hz で 1 となるように正規化した.

構築したシャープネスの計算モデルの重みと Fastl & Zwicker のシャープネスの計算モデル (FZ モデル) の重みを比較する. 構築したシャープネスの計算モデルの重みは周波数が 500 Hz 以下においてシャープネスをより低く, 周波数が 2,000 Hz 以上ではシャープネスをより高くする形状となっている. 一方, FZ モデルの重みは周波数が 4,000 Hz 以上のみシャープネスを高くする形状となっている.

構築したシャープネスの計算モデルの重みと Aures モデルの重みを比べると, Aures モデルの重みは 800 Hz 以上で急激に増加しており, 低域側は一定であるところに違いがある. こちらも FZ モデルと同様で, 高域側でよりシャープネスが高くなるように重み付けされている.

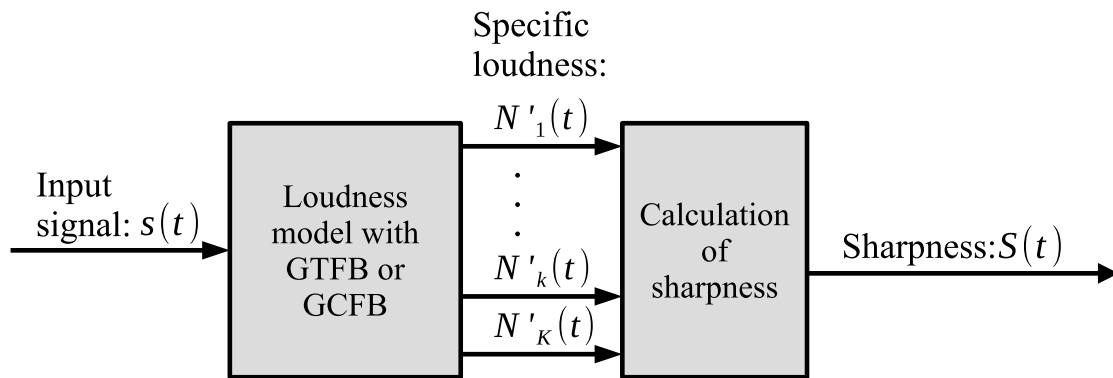


Fig. 4.4: Block diagram of proposed computational model of sharpness (proposed sharpness model)

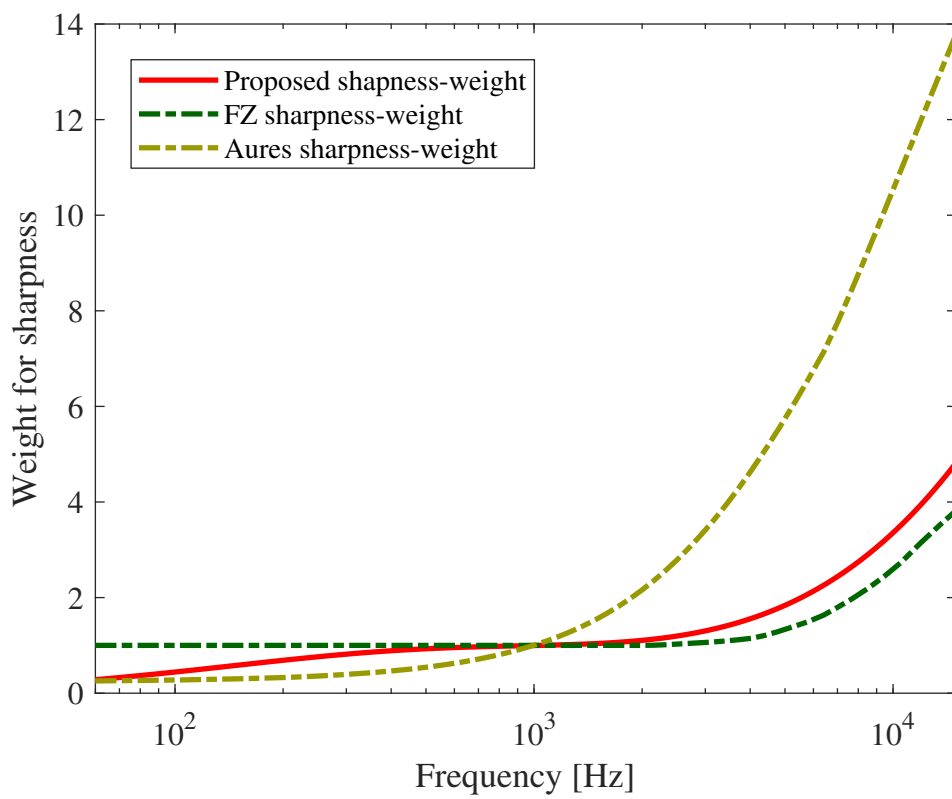


Fig. 4.5: Weighting function of sharpness model



## 4.4 ラフネスの計算モデル

ラウドネスの計算モデルの違いにより、ラフネスを計算するための Fastl のモデル [21] は直接使用できない。そのため、本研究で構築するラフネスの計算モデルは、Daniel & Weber のラフネスの計算モデル [68] をベースに、対数変換した時間マスクパターンピークとディップの差  $\Delta L_{R,k}(t)$  がラフネスに関係するという Fastl の知見を取り入れたものである。

Fig. 4.6 に構築するラフネスの計算モデルのブロックダイアグラムを示す。GTFB を用いたものを GT ラフネスモデル、GCFB を用いたものを GC ラフネスモデルと呼ぶ。構築するラフネスの計算モデルは、はじめに、構築したラウドネスの計算モデルを用いて、観測された  $s(t)$  からラウドネス密度  $N'_k(t)$  を求める。次に、ラフネス固有の帯域通過フィルタを用いてラウドネス密度の帯域を制限し、ヒルベルト変換と LPF を用いて帯域制限されたラウドネス密度の上下の包絡線を求める。これらの包絡線にレベル変換を適用して、ピークとディップの差  $\Delta L_{R,k}(t)$  を求める。正規化相互相関  $i_k$  と  $i_{k-10}$  は、それぞれ  $k$  番目の帯域制限されたラウドネスレベルと  $k \pm 10$  番目 (1 Cam) の帯域制限ラウドネスレベルから計算される。最後に、ラフネス密度  $R'_k(t)$  を  $\Delta L_{R,k}$  と  $i_k$  と  $i_{k-10}$  を用いて計算し、その面積を計算することによりラフネス  $R(t)$  を求める。

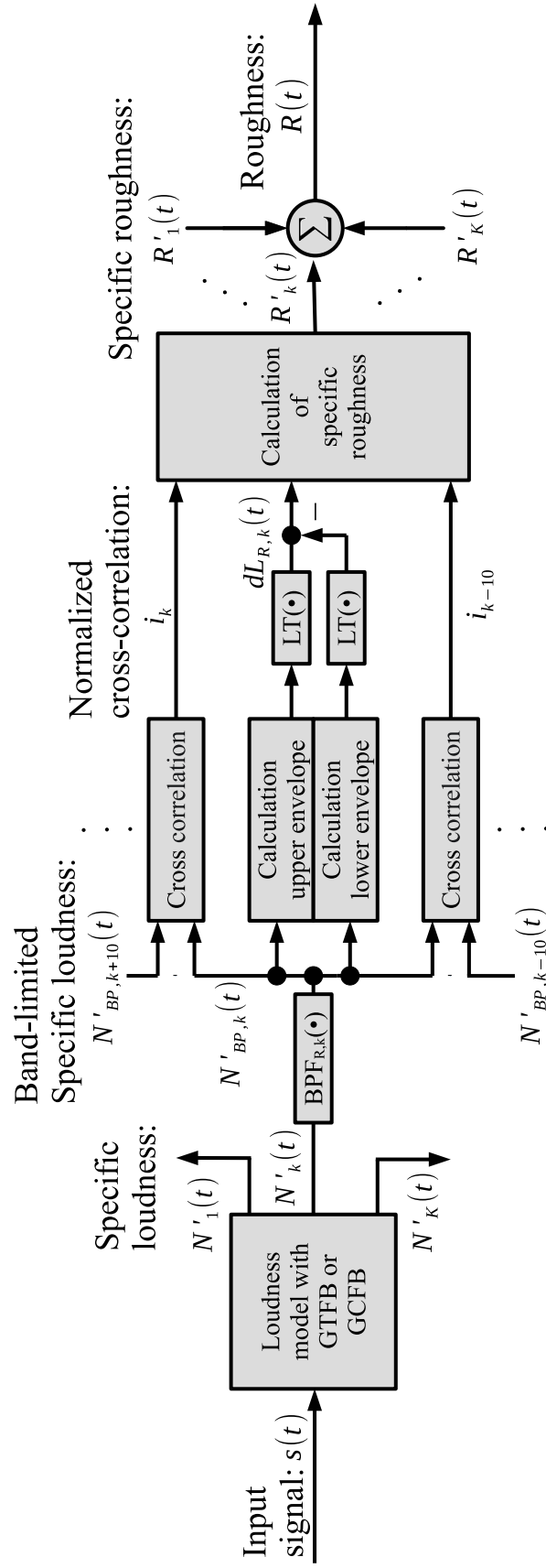


Fig. 4.6: Block diagram of proposed computational model of roughness (proposed roughness model)

#### 4.4.1 ラフネスの帯域通過フィルタ

ラフネス固有の帯域通過フィルタ  $\text{BPF}_{\text{R,itk}}$  は, Fastl & Zwiker の研究から引用されたラフネスの主観評価結果にしたがってシグモイド関数を用いて最適化される. 帯域通過フィルタの中心周波数  $C_{\text{F},k}$  と帯域幅  $W_{\text{B},k}$  はそれぞれ次式で定義される.

$$C_{\text{F},k} = \frac{69.2}{1 + \exp\left(-(\text{ERBN-number} - \alpha)/\beta\right)}, \quad (4.5)$$

$$W_{\text{B},k} = 1.58C_{\text{F},k}, \quad (4.6)$$

ここで,  $\alpha = 4.58$  と  $\beta = 1.48$  は定数である. これらのパラメータは, 搬送波周波数とエンベロープ周波数が変化する音に対して, Fastl & Zwiker の研究におけるラフネスの主観評価結果と, 構築するラフネスの計算モデルを用いて計算されたラフネスとの間の RMSE を最小化するように決定された. 帯域通過フィルタは, 次のように定義する.

$$\text{BPF}_{\text{R},k}(t) = at^{(M-1)} \exp(-2\pi W_{\text{B},k}(C_{\text{F},k})t) \cos(2\pi C_{\text{F},k}t + \phi), \quad (4.7)$$

ここで, フィルタの次数は3である.

帯域制限されたラウドネス密度  $N'_{\text{BP},k}(t)$  は次式で定義される.

$$N'_{\text{BP},k}(t) = \text{BPF}_{\text{R},k}\left(N'_k(t) - H_{0,k}\right), \quad (4.8)$$

ここで  $H_{0,k}$  は  $N'_k(t)$  の直流成分である.

#### 4.4.2 ラウドネス密度におけるピークとディップの差

帯域制限されたラウドネス密度の上側包絡線  $L'_{\text{R,Upper},k}(t)$  と下側包絡線  $L'_{\text{R,Lower},k}(t)$  はそれぞれ次式で定義される.

$$L'_{\text{R,Upper},k}(t) = \text{LT}\left(\text{LPF}\left(|N'_{\text{BP},k}(t) + j\text{Hilbert}(N'_{\text{BP},k}(t))|\right)\right), \quad (4.9)$$

$$L'_{\text{R,Lower},k}(t) = -\text{LT}\left(\text{LPF}\left(|-N'_{\text{BP},k}(t) + j\text{Hilbert}(-N'_{\text{BP},k}(t))|\right)\right), \quad (4.10)$$

ここで、LPF はカットオフ周波数 7 Hz の 9 次 IIR バターワース LPF である。カットオフ周波数は、ラフネスが最大になる変調周波数である 70 Hz の 1/10 に設定した。Duister のラフネスの計算モデルでは、ラフネスの計算に変調度を用いているが、構築するラフネスの計算モデルは Fastl & Zwicker の知見を取り入れ、変調度を使う代わりに、ラウドネス密度を対数変換し、次式で定義される  $\Delta L_{R,k}(t)$  を使う、

$$\Delta L_{R,k}(t) = \left( L'_{R,Upper,k}(t) - L'_{R,Lower,k}(t) \right) W_{Calib,k}(t), \quad (4.11)$$

$$W_{Calib,k}(t) = \frac{LT(N'_k(t))}{\max(LT(N'_k(t)))}, \quad (4.12)$$

ここで、 $LT(\cdot)$  は Fig. 4.3 のラウドネスとラウドネスレベルの関係から作成したレベル変換関数を表す。データ欠損点は線形補間を用いて補間し、関数を完成させる。 $W_{Calib,k}(t)$  は、各時刻におけるラウドネスレベルの最大値が 1 になるようにする重み付け関数である。この補正項は、ラウドネス密度における上包絡線と下包絡線のレベル変換により、信号レベルが小さい場合に  $\Delta L_{R,k}(t)$  が高く計算されすぎるのを防ぐために導入した。

#### 4.4.3 正規化相互相関の計算

ピンク雑音やホワイト雑音のような音では、ざらつき感がほとんど感じられない。しかし、このような雑音をラフネスの計算モデル入力とするとラフネスは大きく計算される。このような問題を解決するために、離れた聴覚フィルタ間の正規化相互相関  $i$  を計算し、構築するラフネスの計算モデルに組み込む。 $i$  は、次式で定義される離れた聴覚フィルタ間で計算される  $N'_{BP,k}(t)$  の時間変動から計算される。

$$i = \max \left( \frac{\int_{\tau=-0.01}^{0.01} x(t)y(t+\tau)}{\sqrt{V_x V_y}} \right), \quad (4.13)$$

ここで、 $\tau$  と  $V$  はそれぞれラグと分散を表す。GTFB または GCFB から計算される  $N'_{BP,k}(t)$  の時間変化には、フィルタ間で時間遅れがある。そこで、 $i$  を 1 サンプル単位で 10 ms までシフトさせ、最大  $i$  をチャンネル  $k$  における相互相関の値とす

る。ここで、 $i_{k-10}$  は  $x$  を  $N'_{BP,k-10}(t)$ 、 $y$  を  $N'_{BP,k}(t)$  としたときの  $i$  の結果である。一方、 $i_k$  は、 $x$  が  $N'_{BP,k}(t)$  で、 $y$  が  $N'_{BP,k+10}(t)$  のときの  $i$  の結果である。

#### 4.4.4 ラフネス密度とラフネスの計算

ラフネス密度  $R'_k(t)$  は  $w_{R,k}$ 、 $\Delta L_{R,k}(t)$ 、 $i_k$  を用いて以下の3つの条件にしたがって計算される。

$$R'_k(t) = \begin{cases} (w_{R,k} \Delta L_{R,k}(t) i_k)^2 & k \in [1, 10], \\ (w_{R,k} \Delta L_{R,k}(t) (i_{k-10} i_k))^2 & k \in [11, K-11], \\ (w_{R,k} \Delta L_{R,k}(t) i_{k-10})^2 & k \in [K-10, K], \end{cases} \quad (4.14)$$

ここで、 $w_{R,k}$  は Fastl & Zwicker の研究から引用したラフネスの主観評価結果と一致するように最適化した重み付け関数である。 $R(t)$  は次式で定義されるラフネス密度の面積計算から得られる。

$$R(t) = Q_R \sum_{k=1}^K R'_k(t), \quad (4.15)$$

ここで  $Q_R$  はラフネスの係数を表し、GTFB を用いた場合は  $3.15 \times 10^{-3}$ 、GCFB を用いた場合は  $3.20 \times 10^{-3}$  である。これらの値は、構築するラフネスの計算モデルを用いて計算した算定ラフネスが基準音に対して 1 asper になるように設定した。ラフネスの計算モデルから算出された算定ラフネスはあくまでも推定値である。

Fig. 4.7 に構築したラフネスの計算モデルと基となった Daniel & Weber のラフネスの計算モデル (DW モデル) の重みを示す。二つのラフネスの計算モデルの重みを比較するために、横軸は周波数で表している。また、縦軸はラフネスの計算モデルの重みであり、DW モデルと今回構築したラフネスの計算モデルの重みの最大値が 1 となるように正規化したものである。二つのモデル重みを比べると、今回構築したラフネス重みのピークは 200 Hz 付近にある一方で、DW モデルのラフネスの重みは 1,000 Hz 付近にある。この二つの重みの違いは、ERB 尺度を使用するのか Bark 尺度を使用するのか、変調度を使用するのかラウドネス密度の山と谷の差を使用するのかによる違いであると考えられる。どちらのモデルも主観評価に合わせるため重みのフィッティングを行っているため、この違いは大きな問題にならないと考えられる。

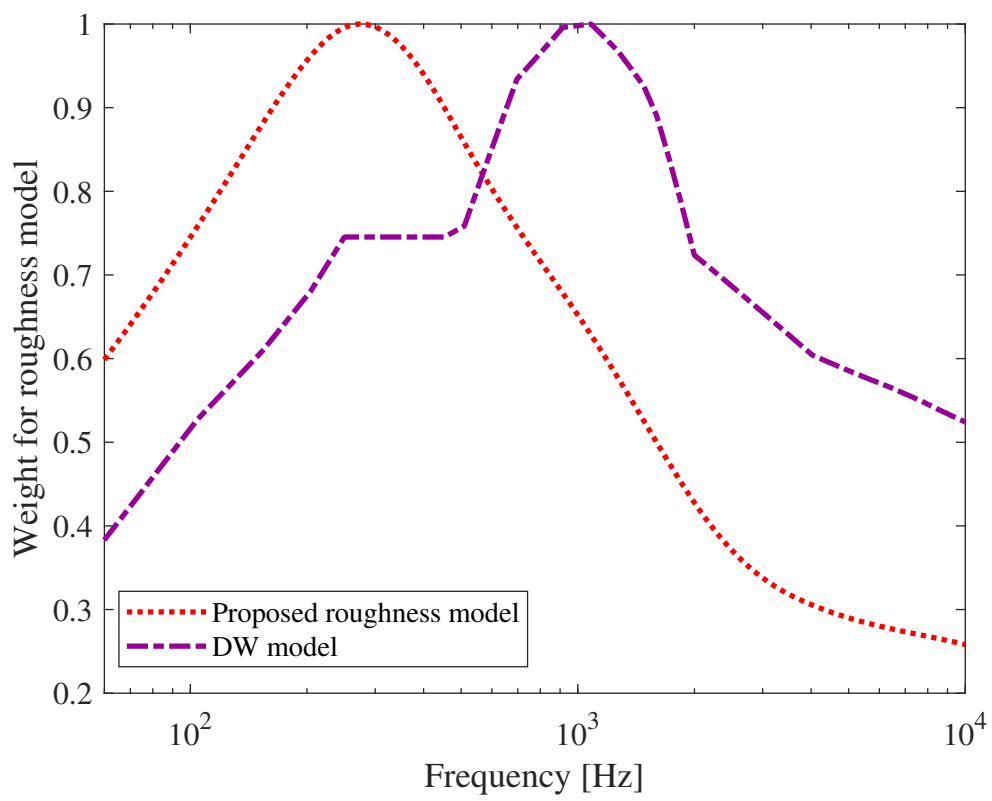


Fig. 4.7: Weighting function of roughness model

## 4.5 変動強度の計算モデル

Fig. 4.8 に構築するラフネスの計算モデルのブロックダイアグラムを示す。このモデルは構築されたラフネスの計算モデルと同様のアプローチで構築したものである。パラメータは変動強度に応じて変更された。主な違いは、変動強度が高くなる変調周波数の範囲である。構築されたラフネスの計算モデルは構築したラウドネスの計算モデルを用いて構築された。GTFB を用いたものを GT 変動強度モデル、GCFB を用いたものを GC 変動強度モデルと呼ぶ。

### 4.5.1 変動強度の帯域通過フィルタ

帯域制限されたラウドネス密度  $N'_{rmBP,k}(t)$  は次式で定義される。

$$N'_{BP,k}(t) = \text{BPF}_F \left( N'_k(t) - H_{0,k} \right) \quad (4.16)$$

帯域通過フィルタは、2 次の IIR バターワース LPF と 2 次 IIR のバターワース HPF をカスケード接続したものである。カットオフ周波数はそれぞれ 5 Hz と 2 Hz である。

### 4.5.2 ラウドネス密度におけるピークとディップの差

帯域制限されたラウドネス密度の上側包絡線  $L'_{F,Upper,k}(t)$  と下側包絡線  $L'_{F,Lower,k}(t)$  はそれぞれ次式で定義される。

$$L'_{F,Upper,k}(t) = \text{LT} \left( \text{LPF} \left( |N'_{BP,k}(t) + j\text{Hilbert}(N'_{BP,k}(t))| \right) \right), \quad (4.17)$$

$$L'_{F,Lower,k}(t) = - \text{LT} \left( \text{LPF} \left( | - N'_{BP,k}(t) + j\text{Hilbert}(-N'_{BP,k}(t))| \right) \right), \quad (4.18)$$

ここで、LPF はカットオフ周波数 0.4 Hz の 9 次の IIR バターワース LPF である。ピークとディップの差  $\Delta L_{F,k}(t)$  は次式で求められる。

$$\Delta L_{F,k}(t) = \left( L'_{F,Upper,k}(t) - L'_{F,Lower,k}(t) \right) W_{\text{Calib},k}(t), \quad (4.19)$$

$$W_{\text{Calib},k}(t) = \frac{\text{LT} \left( N'_k(t) \right)}{\max \left( \text{LT} \left( N'_k(t) \right) \right)}, \quad (4.20)$$

ここで、 $\text{LT}(\cdot)$  は 4.4.2 節で説明した関数と同じである。

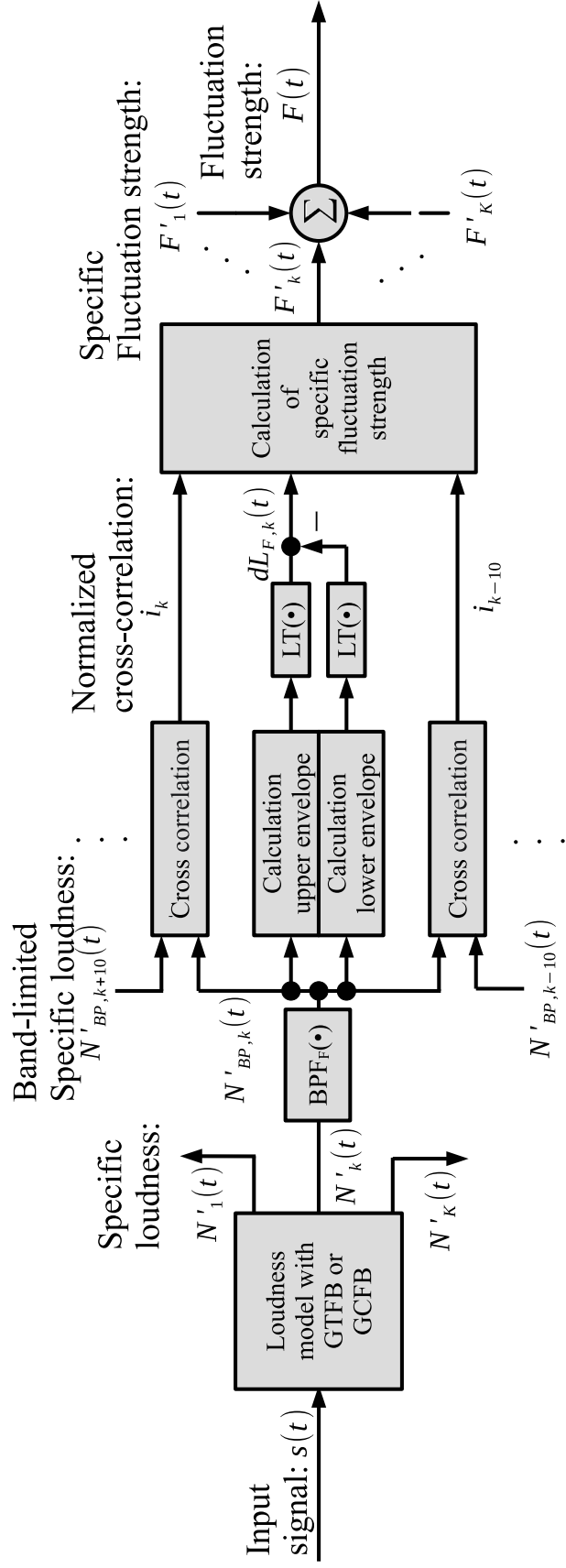


Fig. 4.8: Block diagram of proposed computational model of fluctuation-strength (proposed fluctuation-strength model)



### 4.5.3 変動強度密度と変動強度の計算

変動強度密度  $F'_k(t)$  は  $\Delta L_{F,k}(t)$ ,  $i_k$  を用いて以下の3つの条件にしたがって計算される.

$$F'_k(t) = \begin{cases} \Delta L_{F,k}^{0.6}(t) i_k^2 & k \in [1, 10], \\ \Delta L_{F,k}^{0.6}(t) (i_{k-10} i_k)^2 & k \in [11, K-11], \\ \Delta L_{F,k}^{0.6}(t) i_{k-10}^2 & k \in [K-10, K], \end{cases} \quad (4.21)$$

ここで,  $i_k$  と  $i_{k-10}$  は構築するラフネスの計算モデルと同様に式(4.13)から求める.  $F(t)$  は次式で定義される変動強度密度の面積計算から得られる.

$$F(t) = Q_F \sum_{k=1}^K F'_k, \quad (4.22)$$

ここで  $Q_F$  は変動強度の係数を表し, GTFB を用いた場合は  $3.02 \times 10^{-2}$ , GCFB を用いた場合は  $3.00 \times 10^{-2}$  である. これらの値は, 構築する変動強度の計算モデルを用いて計算した算定変動強度が基準音に対して 1 vacil となるように設定された. この変動強度の計算モデルから算出された算定変動強度はあくまでも推定値である.

## 4.6 音質評価指標の計算モデルの評価と結果

構築する音質評価指標の計算モデルの有効性を判断するため, 従来の聴覚計算モデルについても同様の評価を行い, その結果を構築する聴覚計算モデルの結果と比較した. 構築する音質評価指標計算モデルの入力は定常音であると仮定しているため, 対象とする感覚の大きさは時間とともに変化しない. そこで, 構築する聴覚計算モデルの出力の平均値を評価に用いた.

### 4.6.1 ラウドネスの計算モデル

構築したラウドネスの計算モデルによって推定された算定ラウドネスが Moore-Glasberg モデルから算出される算定ラウドネスと同程度かどうかを 1,000 Hz 純音以外の様々な周波数の純音を用いて評価する. 構築したモデルから計算される算

Table 4.1: Sound signal used for evaluating proposed loudness model

Frequency [Hz]	Sound-pressure level(s) [dB]
100	10, 20, 30, 40, 50, 60, 70, 80
1,000	10, 20, 30, 40, 50, 60, 70, 80
3,000	10, 20, 30, 40, 50, 60, 70, 80

定ラウドネスを、既存の ISO 532-2 で規定されるラウドネスと比較した。比較評価の項目は音圧レベルに対するラウドネスとした。

評価では構築したラウドネスの計算モデルで計算された算定ラウドネスと Moore-Glasberg モデルで計算された算定ラウドネスを比較した。異なる周波数に対する算定ラウドネスを評価した。ラウドネスの評価に用いた評価信号（純音）を Table 4.1 に示す。

これらの評価刺激は、ISO 532-2:2017 の付録に記載されている評価刺激と同一である。構築したラウドネスの計算モデルによって推定された算定ラウドネスの計算には 0.5 秒の評価信号を使用し、推定された算定ラウドネスの前後 0.1 秒を除いた 0.3 秒の定常区間における時間平均ラウドネスを用いて計算した。また、評価刺激のサンプリング周波数は 44,100 Hz とした。

### 音圧レベルを変化させた純音に対するラウドネス

Fig. 4.9, Fig. 4.10, Fig. 4.11 に評価刺激の周波数が 100, 1000, 3000 Hz の時のラウドネスの計算結果を示す。図中の横軸は音圧レベル、縦軸は算定ラウドネスである。評価刺激の周波数が 3,000 Hz では、ほかの周波数の時の結果と比べて Moore-Glasberg モデルの結果との誤差が大きくなった。ただし、RMSE の値は 1 程度であり、評価刺激の周波数が 3,000 Hz の算定ラウドネスの最大値が 30 sone であることから、この誤差は十分に小さいと考えられる。その他の周波数では、構築したモデルの結果は、Moore-Glasberg モデルの結果とおおよそ一致した。これらの結果より、構築したラウドネスの計算モデルは Moore-Glasberg モデルの時間領域版であるといえる。

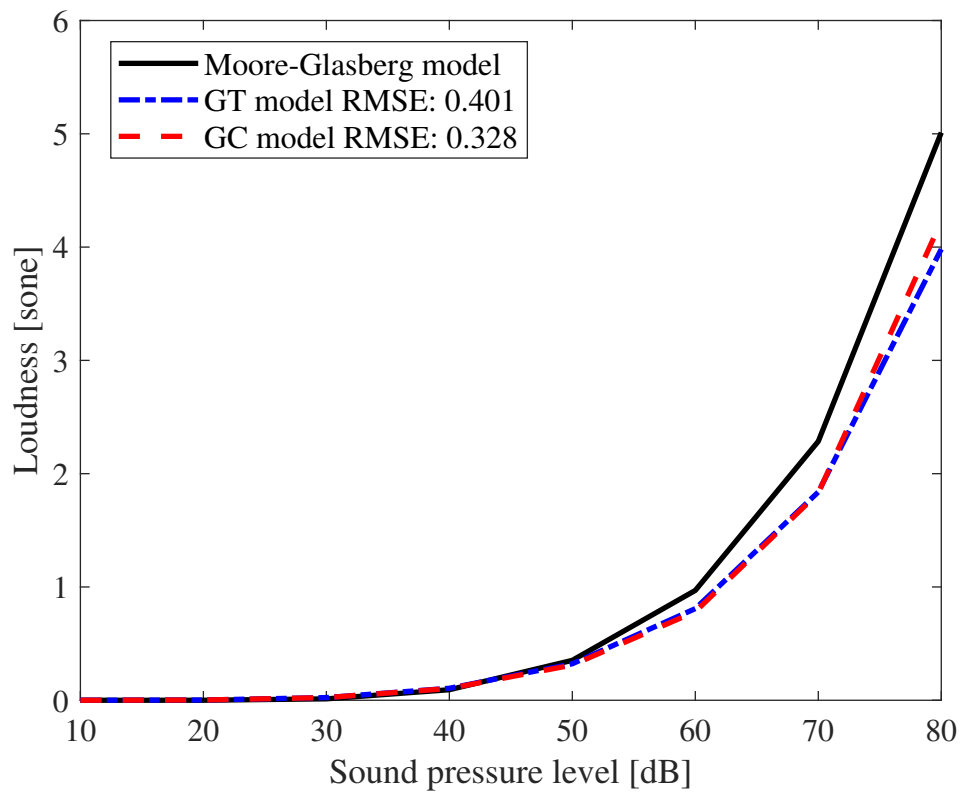


Fig. 4.9: RMSEs of both versions of proposed loudness model when the input signal is 100 Hz: GT model and GC model.

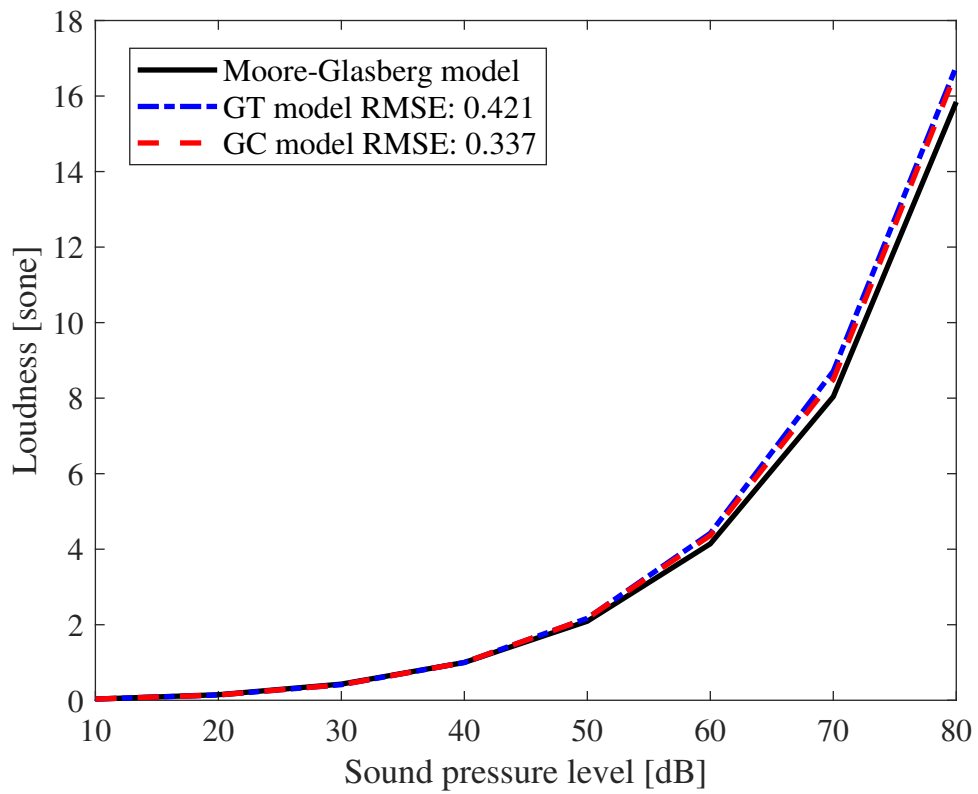


Fig. 4.10: RMSEs of both versions of proposed loudness model when the input signal is 1000 Hz: GT model and GC model.

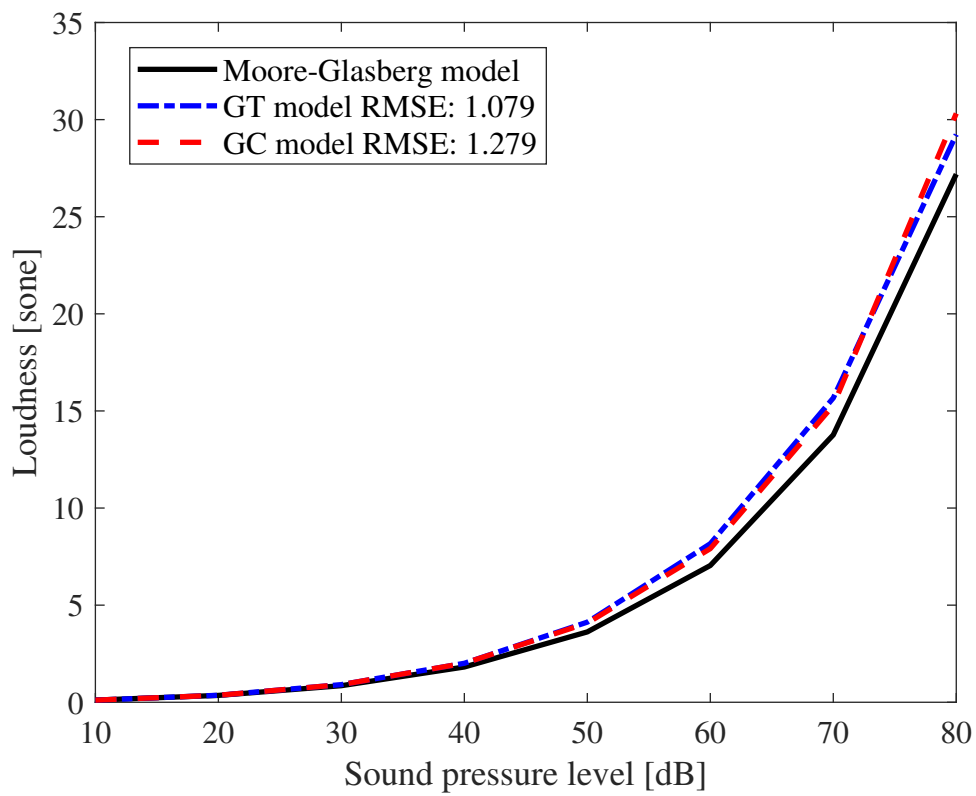


Fig. 4.11: RMSEs of both versions of proposed loudness model when the input signal is 3000 Hz: GT model and GC model.

## 4.6.2 シャープネスの計算モデル

構築したシャープネスの計算モデルの両バージョン（GT シャープネスモデルと GC シャープネスモデル）を，従来のシャープネスの計算モデルである Fastl & Zwicker のモデル（FZ モデル）とラウドネスに依存する Aures のモデル（Aures モデル）と比較した．従来のモデルの入力には ISO532B:1975 に規定されたラウドネスの計算モデルを用いて算出した算定ラウドネスを使用した．シャープネスは，3 つの異なる雑音とラウドネスについて評価した．シャープネス値は，構築されたシャープネスの計算モデルの 2 つのバージョンから，0.3 秒の定常区間における算定シャープネスの時間平均をとることによって計算された．

### 帯域騒音に対するシャープネス

刺激の初期セットは，狭帯域（NB）雑音，高域通過（HP）雑音，低域通過（LP）雑音からなる．NB 雑音の一覧は Table 4.2 に，HP 雑音と LP 雑音の一覧は Table 4.3 に示す．これらの音刺激の振幅は，構築したラウドネスの計算モデルを用いて計算した 0.3 秒間の定常区間における算定ラウドネスが 4 sone になるように調整した．

Fig. 4.12 に NB 雑音を入力信号としたときの算定シャープネスの計算結果を示す．Fig. 4.12 の横軸は NB 雑音の中心周波数であり，縦軸は算定シャープネスである．図中の黒い実線はシャープネスに対する主観評価結果である．Fig. 4.12 より，FZ モデルは，2,000 Hz より高い中心周波数を持つ NB 雑音で主観評価結果より高くシャープネスを推定していることがわかった．Aures モデルは，1,000 Hz より高い中心周波数を持つ NB 雑音で主観評価結果より高くシャープネスを推定していることがわかった．構築したシャープネスの計算モデルは，4,000 Hz 付近までは FZ モデルとおおよそ同じ推定値を算出していることがわかった．一方 4,000 Hz 以降では，構築したシャープネスの計算モデルの推定値は，FZ モデルの推定値と比べて主観評価結果と近いことがわかった．GT シャープネスモデルと GC シャープネスモデルのシャープネスの推定には違いがないことがわかった．

Fig. 4.13 に HP 雑音を入力信号としたときの算定シャープネスの計算結果を示す．Fig. 4.13 の横軸は HP 雑音の低域側の周波数であり，縦軸は算定シャープネスである．図中の黒い実線はシャープネスに対する主観評価結果である．Fig. 4.13

より、FZモデルより算出された算定シャープネスは、2,000 Hz以下では主観評価結果より高く推定されていることがわかった。Auresモデルより算出された算定シャープネスは、主観評価結果より、全周波数において高く推定されていることがわかった。構築したシャープネスの計算モデルより算出された算定シャープネスは、1,000 Hz以下では主観評価結果とよく一致していることがわかった。一方、構築したシャープネスの計算モデルより算出された算定シャープネスは、1,000 Hz以上の周波数では、主観評価結果より低く推定されていることがわかった。

Fig. 4.14にLP雑音を入力信号としたときの算定シャープネスの計算結果を示す。Fig. 4.14の横軸はLP雑音の高域側の周波数であり、縦軸は算定シャープネスである。図中の黒い実線はシャープネスに対する主観評価結果である。Fig. 4.14より、FZモデルとAuresモデルより算出された算定シャープネスは2,000 Hz以上では主観評価結果より高く推定されていることがわかった。構築したシャープネスの計算モデルより算出された算定シャープネスは2,000 Hz以下では主観評価結果とよく一致していることがわかった。一方、構築したシャープネスの計算モデルより算出された算定シャープネスは、2,000 Hz以上の周波数では、主観評価結果より高く推定されていることがわかった。

Fig. 4.15は、雑音の種類を変えた場合の算定シャープネスの結果を示しており、縦軸は、シャープネスの主観評価結果の結果と4つのシャープネスの計算モデルから得られた値とのRMSEを表している。構築したシャープネスの計算モデルのRMSEは、GTシャープネスモデルとGCシャープネスモデルのいずれのもFZモデルやAuresモデルよりも低いことがわかった。さらに、構築されたシャープネスの計算モデルから得られた算定シャープネスは、シャープネスの主観評価結果とほぼ一致することがわかった。

## ラウドネスを変化させた純音に対するシャープネス

2つ目の刺激セットは、500, 1,000, 2,000, 4,000, 8,000 Hzの0.5秒間の正弦波信号で、ラウドネスは2, 7, 14, 28 soneであった。RMSEを用いて、4つのシャープネスの計算モデルから得られたシャープネス値と、Fastl & Zwickerの研究から引用したシャープネスの主観評価結果との差を定量化した。

Fig. 4.16は、ラウドネスレベルを変えたときのシャープネスの結果を示してお

り、縦軸はシャープネスの主観評価結果と4つのシャープネスの計算モデルから得られた値とのRMSEを表している。構築するシャープネスの計算モデルのRMSEは、FZモデルやAuresモデルのRMSEよりも低いことがわかった。さらに、構築されたシャープネスの計算モデルから得られたシャープネスの値は、シャープネスの主観評価結果とほぼ一致することがわかった。GCシャープネスモデルのRMSEはGTシャープネスモデルのRMSEと比較して低いことがわかった。これは、GCFBのフィルタ形状の非対称性がシャープネスの推定により働きを与えたためだと考えられる。



Table 4.2: Narrow-band noise with loudness of 4 sone used for evaluating proposed sharpness model

Center frequency [Hz]	Bandwidth [Hz]
250	104
350	109
450	114
570	122
700	133
840	145
1000	162
1170	182
1370	207
1600	239
1850	277
2150	325
2500	386
2900	460
3400	559
4000	685
4800	867
5800	1111
7000	1426
8500	1851
10500	2463

Table 4.3: High and low-pass noise with loudness 4 sone used for evaluating proposed sharpness model

High-pass noise		Low-pass noise	
Low freq. [Hz]	High freq. [Hz]	Low freq. [Hz]	High freq. [Hz]
250	10,000	200	350
350			450
450			570
570			700
700			840
840			1,000
1,000			1,170
1,170			1,370
1,370			1,600
1,600			1,850
1,850			2,150
2,150			2,500
2,500			2,900
2,900			3,400
3,400			4,000
4,000			4,800
4,800			5,800
5,800			7,000
7,000			8,500
8,500			10,500

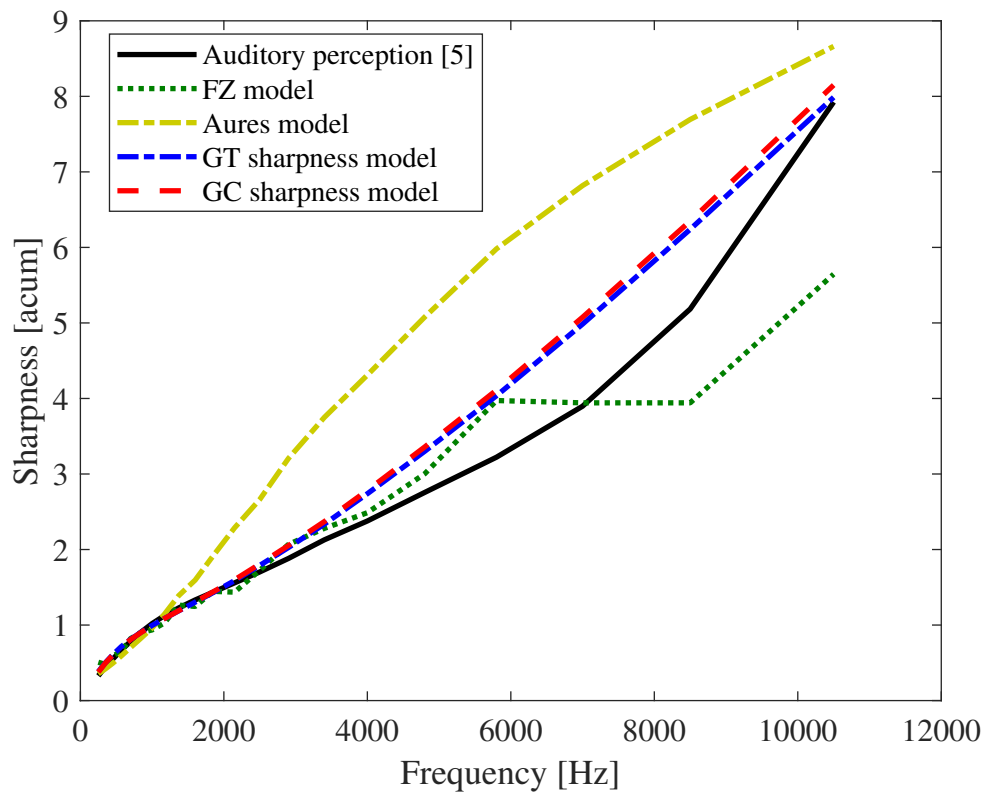


Fig. 4.12: Sharpness of auditory perception and four sharpness models (FZ, Aures, proposed GT sharpness, and proposed GC sharpness models) for narrow-band noise

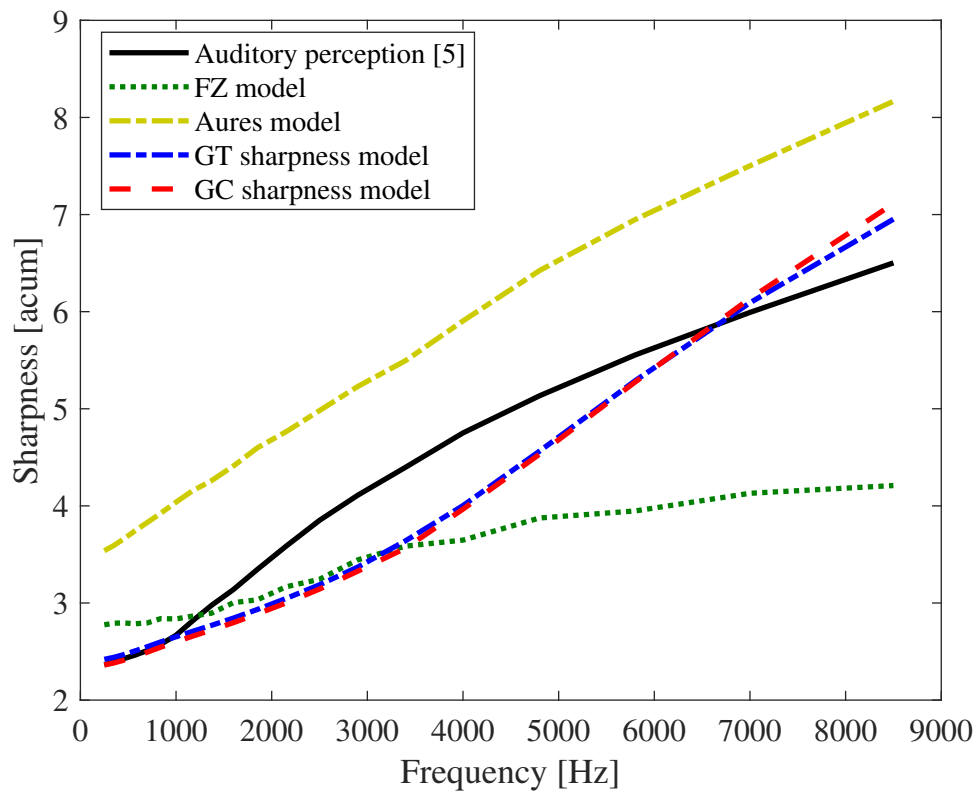


Fig. 4.13: Sharpness of auditory perception and four sharpness models (FZ, Aures, proposed GT sharpness, and proposed GC sharpness models) for high-pass noise

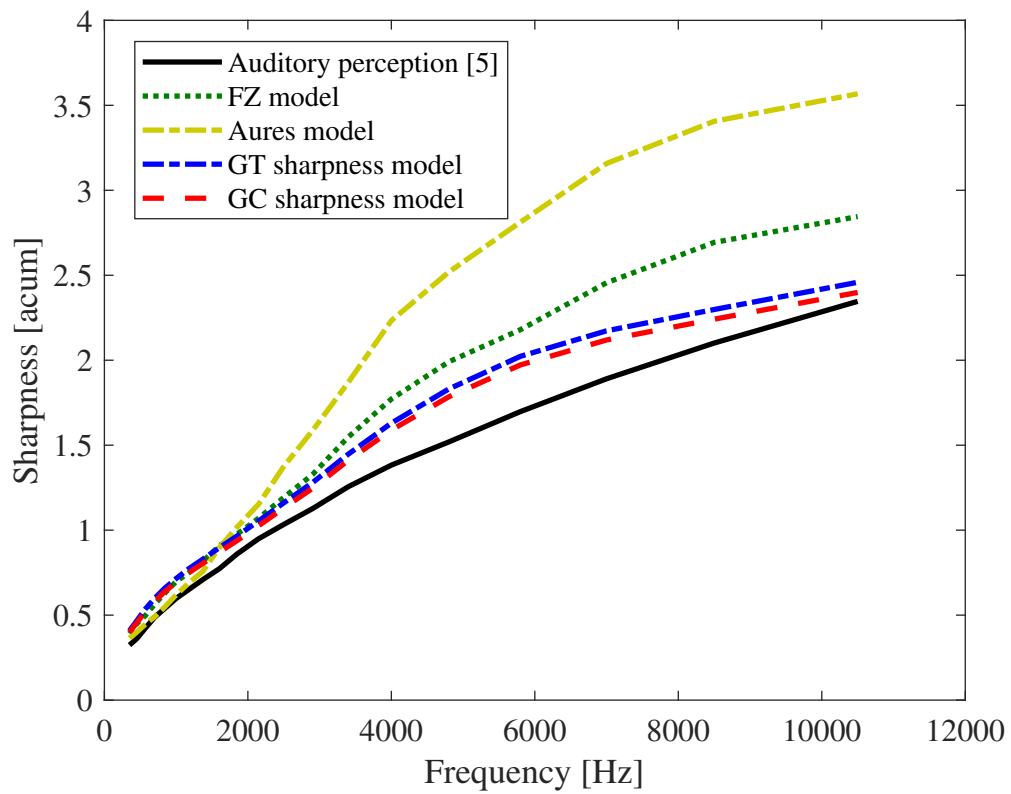


Fig. 4.14: Sharpness of auditory perception and four sharpness models (FZ, Aures, proposed GT sharpness, and proposed GC sharpness models) for low-pass noise

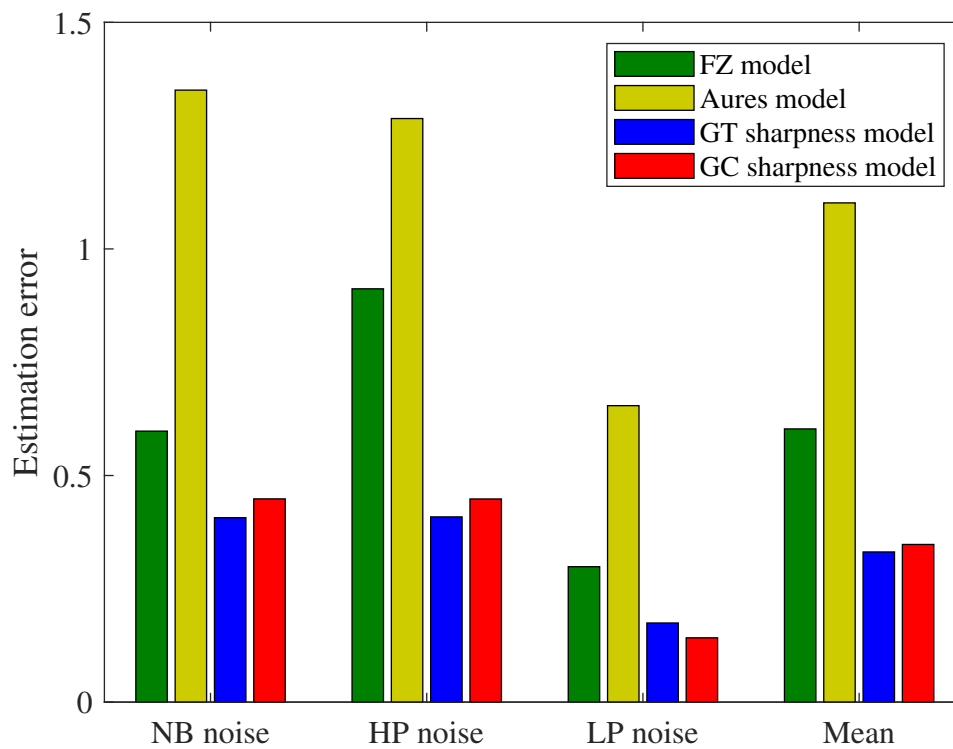


Fig. 4.15: RMSEs of four sharpness models (FZ, Aures, proposed GT sharpness, and proposed GC sharpness models) for different types of noise

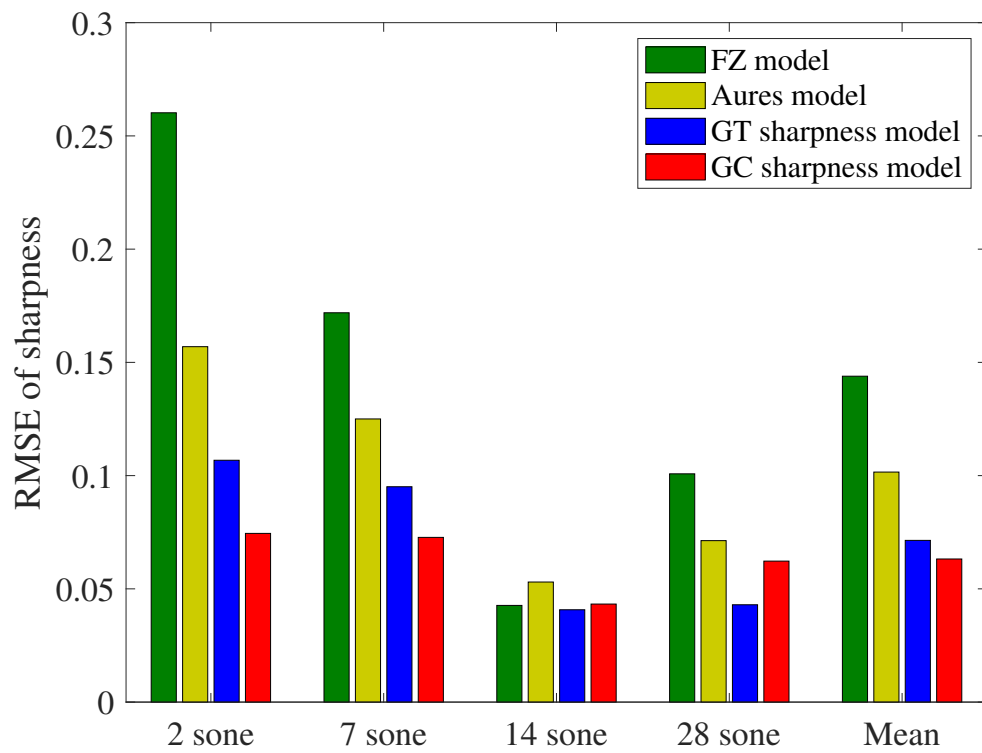


Fig. 4.16: RMSEs of four sharpness models (FZ, Aures, proposed GT sharpness, and proposed GC sharpness models) for different loudness

### 4.6.3 ラフネスの計算モデル

構築したラフネスの計算モデルを Widmann のラフネスの計算モデル (WF モデル), Daniel のラフネスの計算モデル (DW モデル) と比較した。ラフネスの知覚は, 変調周波数, 音圧レベル, 中心周波数, 変調度によって変化する。評価では, これらのパラメータを組み込んだ信号を用いた。これらの評価刺激は, 従来モデルの評価に用いたものと同じである。

ラフネス値は, 構築されたラフネスの計算モデルから時間平均ラフネスをとって計算した。RMSE は, ラフネスの主観評価結果と 4 つのラフネスの計算モデルの予測値との誤差を評価するために用いた。

#### 変調周波数を変化させた変調音に対するラフネス

振幅変調 (AM) 信号と周波数変調 (FM) 信号を用いて変調周波数を変化させたときのラフネスを評価した。AM 信号は, 1,000 Hz の正弦波搬送波を 0.2 秒持続させ, 変調周波数 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200 Hz で変調度 1.0 の振幅変調を行った。FM 信号は, 1,500 Hz で 0.2 秒の正弦波搬送波を用い, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200 Hz で 700 Hz の周波数偏差で周波数変調された。これらの信号の音圧レベルは 70 dB とした。

Fig. 4.17 と Fig. 4.18 は, AM 信号と FM 信号について, 変調周波数を変えて 4 つのモデルから得られたラフネスの結果を示している。横軸は変調周波数, 縦軸は正規化したラフネスである。黒い実線は, Fastl & Zwicker の研究 [5] から引用したラフネスの主観評価結果を示している。

AM 変調音の結果 (Fig. 4.17) を見ると, WF モデルは 70 Hz 以下の変調周波数での算定ラフネスをラフネスの主観評価結果より低く推定しており, 他のモデルはラフネスの主観評価結果に近い値を推定している。本研究で構築した GT ラフネスモデルと GC ラフネスモデルは, 70 Hz 以上の変調周波数において, ラフネスの主観評価結果と比較して低い結果を推定した。RMSE が最も小さい計算モデルは DW モデルで, GC ラフネス, GT ラフネス, WF モデルの順であった。

FM 変調音の結果 (Fig. 4.18) を見ると, WF モデルの推定結果が 70 Hz にピークがあることがわかった。DW モデルの推定結果が 60 Hz にピークがあることが



わかった。GT ラフネスモデルと GC ラフネスモデルの推定結果が 50 Hz にピークがあることがわかった。

主観評価結果と各モデルによって推定された算定ラフネスを比較する。WF モデルの推定結果は、70 Hz 以下において主観評価結果より低いことがわかった。DW モデルの推定結果は、20 Hz から 60 Hz の間で主観評価結果より低いことがわかった。GT ラフネスモデルの推定結果は、20 Hz から 60 Hz の間で主観評価結果より高いことがわかった。また、GT ラフネスモデルの推定結果は、80 Hz から 200 Hz において主観評価結果より低いことがわかった。GC ラフネスモデルの推定結果は、80 Hz から 200 Hz において主観評価結果より低いことがわかった。

RMSE が最も小さいモデルは GC ラフネスモデルで、GT ラフネスモデル、DW モデル、WF モデルの順であった。

### 音圧レベルを変化させた変調音に対するラフネス

AM 信号と FM 信号を用いて音圧レベルを変化させたときのラフネスを評価した。AM 信号は、1,000 Hz の正弦波搬送波を 0.2 秒持続させ、変調周波数 70 Hz で変調度 1.0 の振幅変調を行った。FM 信号は、1,500 Hz で 0.2 秒の正弦波搬送波を用い、70 Hz で 700 Hz の周波数偏差で周波数変調した。これらの信号の音圧レベルは 40, 50, 60, 70, 80 dB とした。

Fig. 4.19 と Fig. 4.20 は、AM 信号と FM 信号について、音圧レベルを変えて 4 つのモデルからラフネスを推定した結果である。横軸は音圧レベル、縦軸は正規化ラフネス。黒い実線は、Fastl & Zwicker の研究 [5] から引用したラフネスの主観評価結果を示す。

AM 変調音の結果 (Fig. 4.19) は、WF モデルの結果がラフネスの主観評価結果よりも低く見積もられていることを示している。RMSE が最も小さいモデルは GC ラフネスモデルで、GT ラフネスモデル、DW モデル、WF モデルの順であった。

FM 変調音の結果 (Fig. 4.20) より、WF、GC ラフネス、GT ラフネスモデルの推定結果は、ラフネスに関するリスニング実験の推定結果よりも低く、DW モデルの推定結果は、ラフネスに関する主観評価結果の推定結果よりも高いことを示している。RMSE が最も小さいモデルは WF モデルで、GT ラフネス、GC ラフネス、DW モデルの順であった。

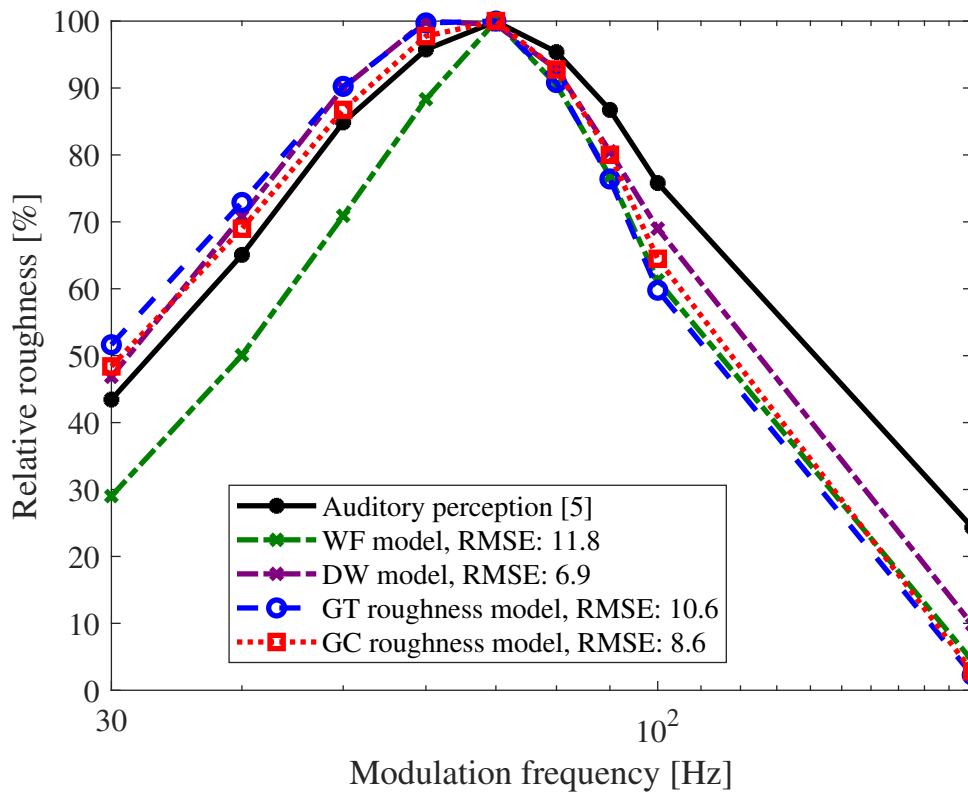


Fig. 4.17: Relative roughness was calculated using four roughness models (WF, DW, proposed GT roughness, and proposed GC roughness models) for various modulation frequencies. AM signal was 1,000-Hz sinusoidal carrier with 0.2-sec duration with 1.0 modulation index amplitude modulation and sound-pressure level of 70 dB.

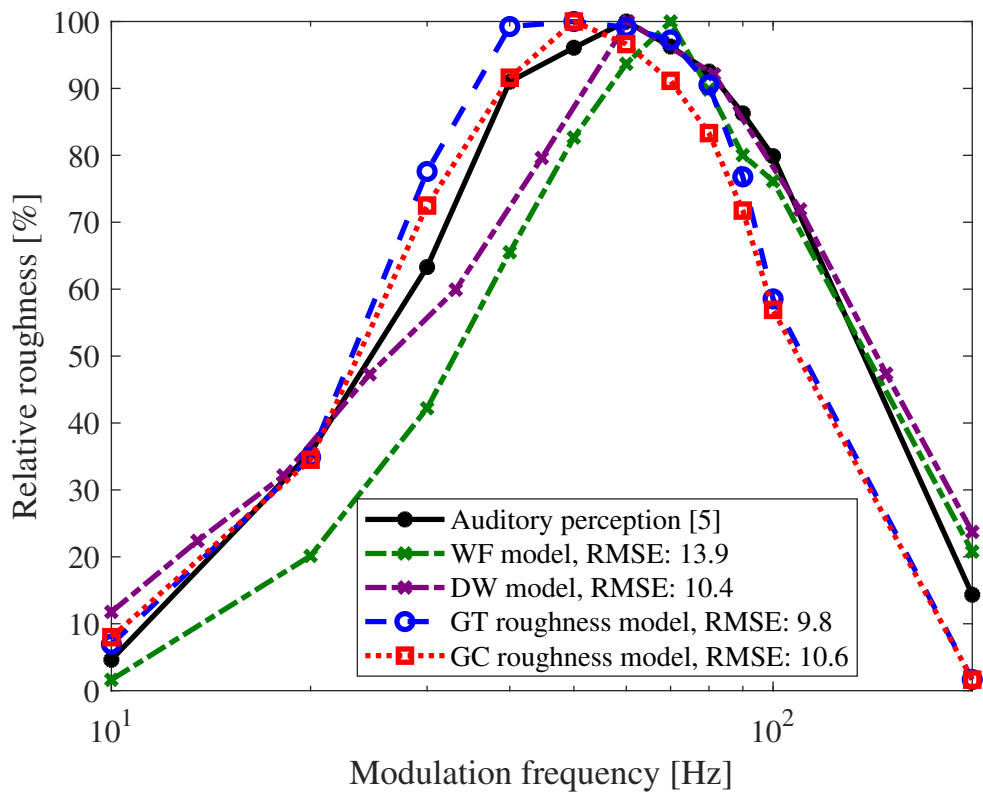


Fig. 4.18: Relative roughness was calculated using four roughness models (WF, DW, proposed GT roughness, and proposed GC roughness models) for various modulation frequencies. FM signal was 1500-Hz sinusoidal carrier with 0.2-sec duration with frequency deviation of 700 Hz and sound-pressure level of 70 dB.

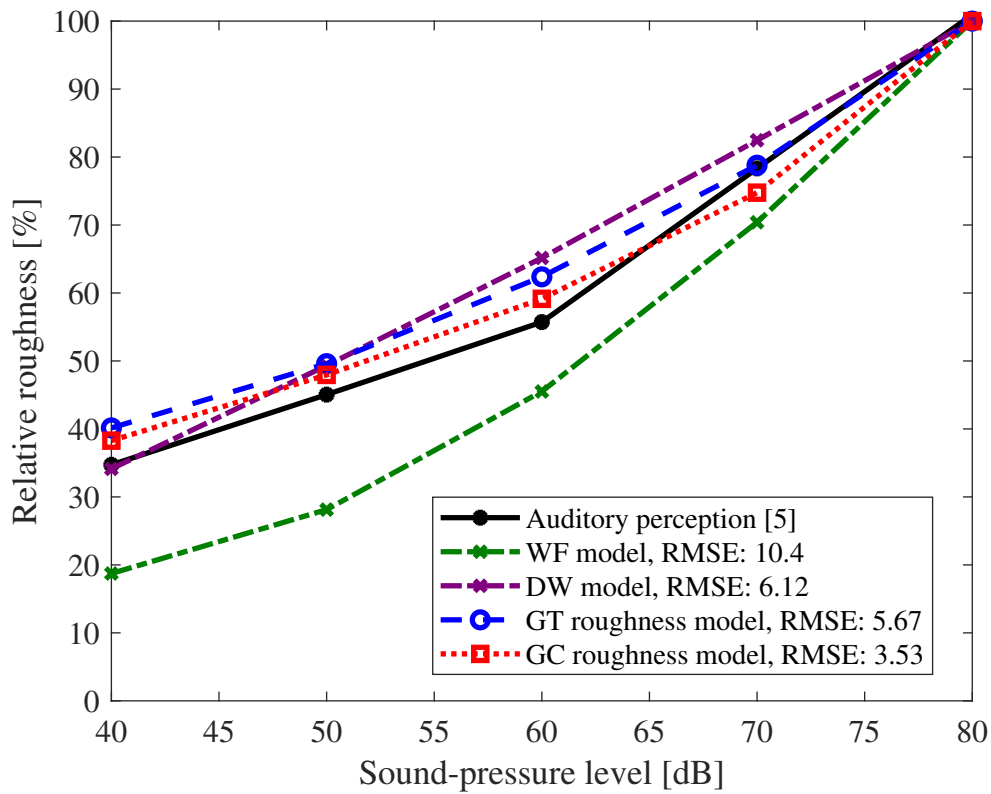


Fig. 4.19: Relative roughness was calculated using four roughness models (WF, DW, proposed GT roughness, and proposed GC roughness models) for various sound-pressure levels. AM signal was 1,000-Hz sinusoidal carrier with 0.2-sec duration with 1.0 modulation index amplitude modulation at modulation frequencies of 70 Hz.

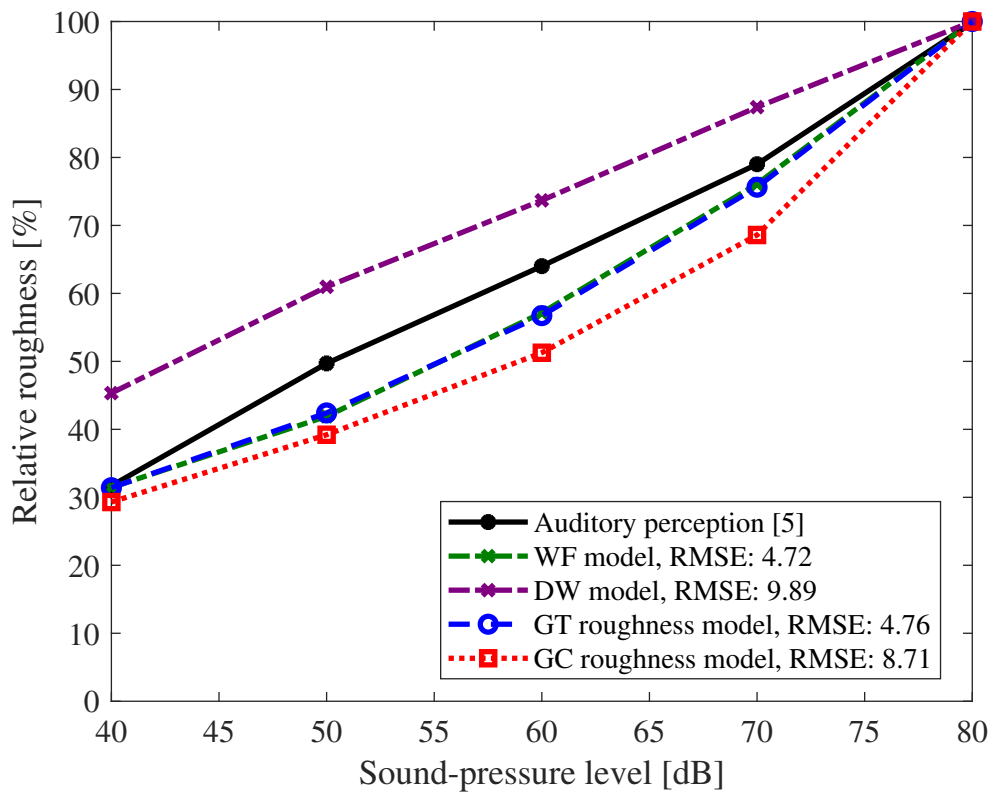


Fig. 4.20: Relative roughness was calculated using four roughness models (WF, DW, proposed GT roughness, and proposed GC roughness models) for various sound-pressure levels. FM signal was 1,500-Hz sinusoidal carrier with 0.2-sec duration with frequency modulated at 70 Hz with frequency deviation of 700 Hz.

## 搬送周波数と変調周波数を変化させた変調音に対するラフネス

変調周波数と搬送波周波数を変えた AM 信号のラフネスを評価した。刺激は 1,000, 2,000, 4,000, 8,000 Hz の正弦波搬送波で、変調周波数 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 200 Hz で変調度 1.0 の振幅変調を 0.2 秒持続させた。音圧レベルは 60 dB とした。

Fig. 4.21 に Fastl & Zwicker の研究 [5] から引用したラフネスの主観評価結果に対する、中心周波数に対する 4 つのモデルの RMSE を示す。横軸は中心周波数、縦軸は RMSE である。すべての中心周波数の RMSE の平均が最も小さかったモデルは DW モデルで、GC, GT, WF ラフネスモデルがそれに続いた。

## 変調度を変化させた変調音に対するラフネス

変調の度合いを変えた AM 信号のラフネスを評価した。刺激は、0.2 秒持続の 1,000 Hz の正弦波搬送波と、変調度が 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90, 1.0 の振幅変調、変調周波数 70 Hz で構成された。音圧レベルは 60 dB とした。

Fig. 4.22 は、変調度の関数としてのラフネスの推定結果を示している。横軸は変調度、縦軸はラフネス。黒い実線は、Fastl & Zwicker の研究から引用したラフネスの主観評価結果を示す。4 つのモデルは、ラフネスの主観評価結果よりも高いラフネスを推定した。RMSE が最も小さかったモデルは GC ラフネスモデルで、DW, WF, GT ラフネスモデルがそれに続いた。

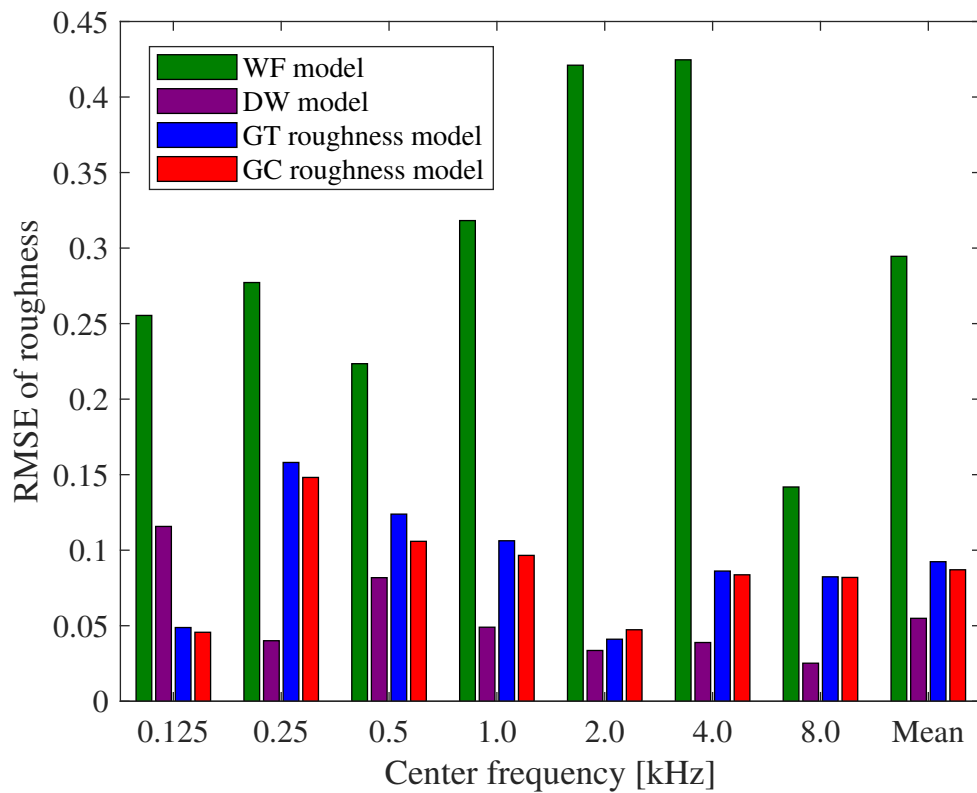


Fig. 4.21: Relative roughness was calculated using four roughness models (WF, DW, proposed GT roughness, and proposed GC roughness models) for various center frequencies. Stimuli had 0.2-sec duration with 100% amplitude modulation with sound-pressure level of 60 dB.

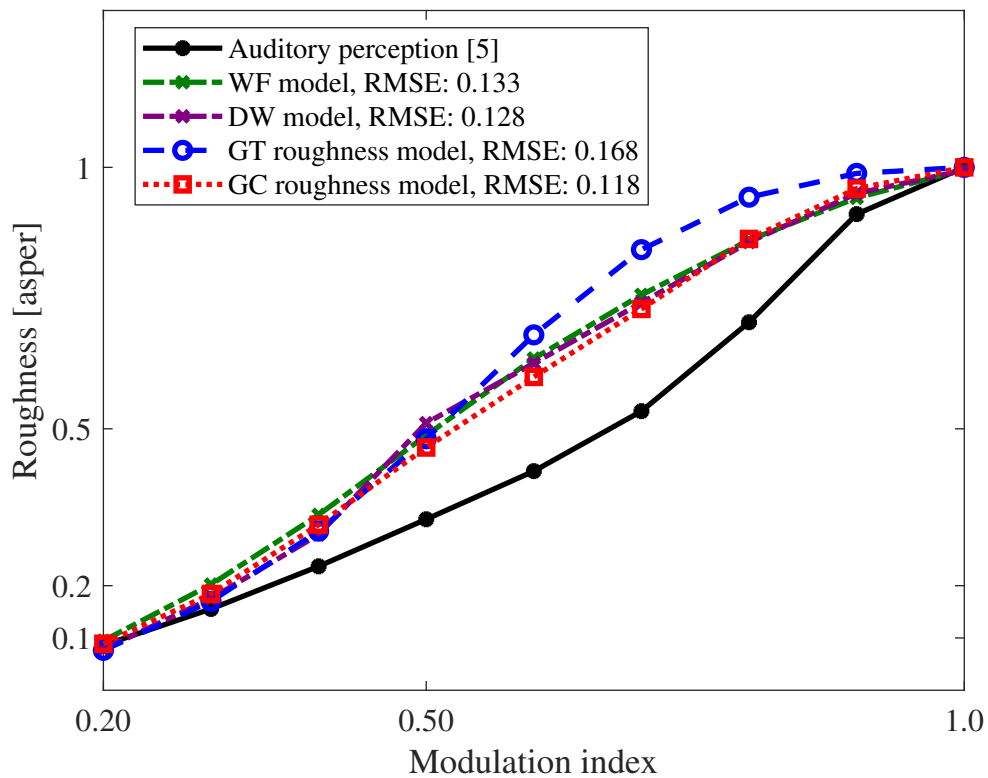


Fig. 4.22: Roughness was calculated using four roughness models (WF, DW, GT roughness, and GC roughness models) for various modulation index of AM signal. Stimuli were sinusoidal carrier of 1,000-Hz with 0.2-sec duration with modulation frequency of 70 Hz and sound-pressure level of 60 dB.



#### 4.6.4 変動強度の計算モデル

構築した変動強度の計算モデルの両バージョン（GT 変動強度モデル，GC 変動強度モデル）を，従来の変動強度の計算モデル（Fastl モデル）と比較した．変動強度の知覚は，変調周波数，音圧レベル，変調度によって変化する．そこで，評価刺激にはこれらのパラメータを取り入れた．これらの評価刺激は，従来モデルの評価に用いたものと同様の信号である．

算定変動強度の代表値は，算定変動強度の定常部を切り出し時間平均した値を用いた．RMSE は，変動強度の主観評価結果と 3 つのモデルの予測値との誤差を評価するために用いた．

##### 変調周波数を変化させた変調音に対する変動強度

変調周波数を変化させた AM 信号と FM 信号の変動強度を評価した．AM 信号は，1,000 Hz の正弦波搬送波を 4 秒間持続させ，変調周波数 0.25, 0.5, 1, 2, 4, 8, 16, 32 Hz で変調度が 1.0 の振幅変調を行った．FM 信号は，1,500 Hz で 4 秒間の正弦波搬送波を用いて，700 Hz の周波数偏差で 0.25, 0.5, 1, 2, 4, 8, 16, 32 Hz で周波数変調された．これらの信号の音圧レベルは 70 dB に設定された．

Fig. 4.23 と Fig. 4.24 は，変調周波数を変えた AM 信号と FM 信号の変動強度の推定結果である．横軸は変調周波数，縦軸は変動強度である．黒い実線は Fastl & Zwicker の研究から引用した変動強度の主観評価結果を示す．

AM 変調音の結果（Fig. 4.23）から，4 Hz 以下の変調周波数において，構築した変動強度の計算モデルの算定変動強度は，変動強度の主観評価結果よりも高く推定していることがわかった．一方，4 Hz 以上の周波数では，構築した変動強度の計算モデルの算定変動強度は，主観評価結果よりも低く推定していることがわかった．RMSE が最も小さいモデルは Fastl モデルであり，次いで GC 変動強度モデル，GT 変動強度モデルであった．

FM 変調音の結果（Fig. 4.24）から，変調周波数が 4 Hz 以上では，Fastl モデルの推定結果が変動強度の主観評価結果よりも高いことを示している．それに対し，本研究で構築した変動強度の計算モデルは，主観評価結果とよく一致していることがわかった．RMSE が最も小さいモデルは GC 変動強度モデルで，GT 変動強度モデル，Fastl モデルの順であった．

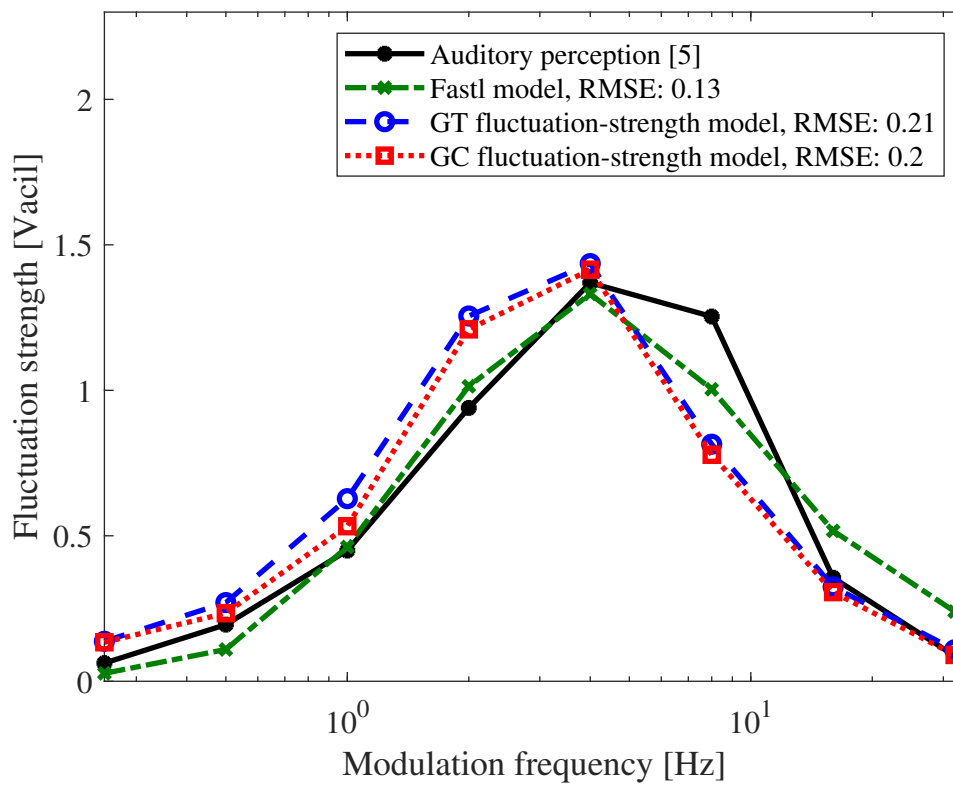


Fig. 4.23: Fluctuation strength was calculated using three fluctuation-strength models (Fastl, GT fluctuation-strength, and GC fluctuation-strength models) for various modulation frequencies. AM signal was 1,000-Hz sinusoidal carrier with 4-sec duration with 1.0 modulation index amplitude modulation and sound-pressure level of 70 dB.

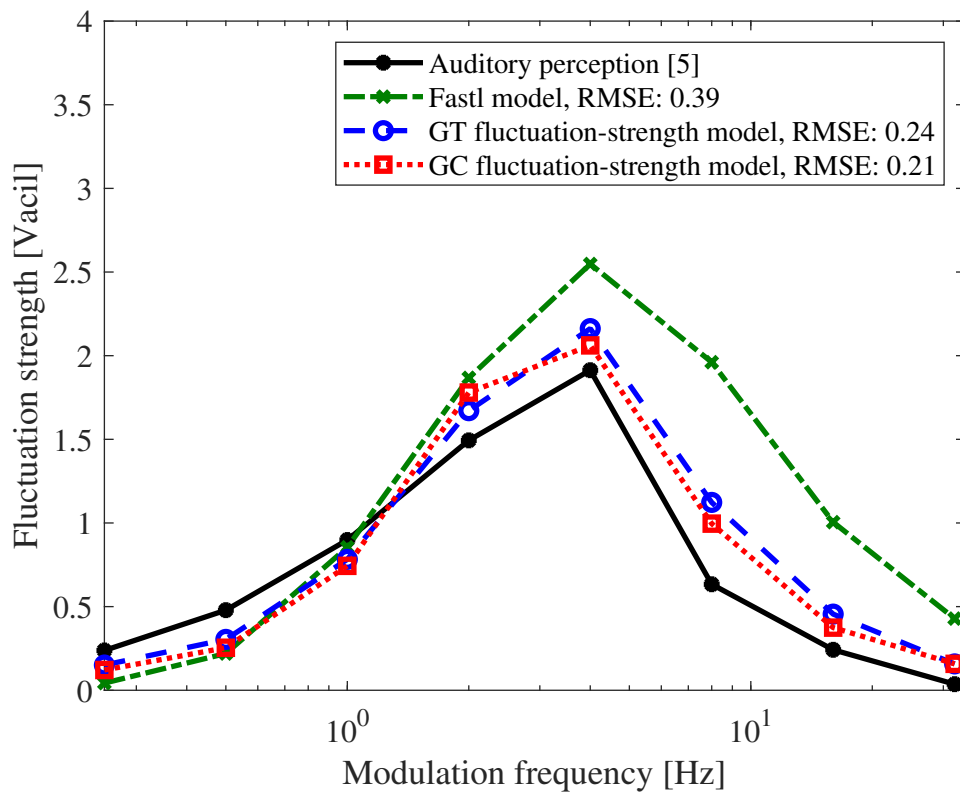


Fig. 4.24: Fluctuation strength was calculated using three fluctuation-strength models (Fastl, GT fluctuation-strength, and GC fluctuation-strength models) for various modulation frequencies. FM signal was 1500-Hz sinusoidal carrier with 4-sec duration with frequency deviation of 700 Hz and sound-pressure level of 70 dB.

## 音圧レベルを変化させた変調音に対する変動強度

次に、音圧レベルを変化させた AM 信号の変動強度を評価した。刺激は、1,000 Hz の正弦波搬送波を 4 秒間持続させ、変調周波数 4 Hz で変調度 1.0 の振幅変調を加えたものである。音圧レベルは 50, 60, 70, 80 dB とした。

Fig. 4.25 と Fig. 4.26 は、AM 信号と FM 信号について、音圧レベルを変化させたときの 3 つのモデルから変動強度を推定した結果を示している。横軸は音圧レベル、縦軸は変動強度である。黒い実線は、Fastl & Zwicker の研究から引用した変動強度の主観評価結果を示す。

AM 変調音の結果 (Fig. 4.25) を見ると、音圧レベル 50 dB から 60 dB では、3 つのモデルの推定結果には違いがみられなかった。一方、60 dB 以上の音圧レベルでは、Fastl モデルと比べて、本研究で構築した変動強度の計算モデルはより主観評価結果と近い値を算出した。最も RMSE が小さかったモデルは GC 変動強度モデルであり、次いで GT 変動強度モデル、Fastl モデルであった。

FM 変調音の結果 (Fig. 4.26) を見ると、WF モデルの推定結果が全音圧レベルにおいて変動強度の主観評価結果よりも高いことがわかった。本研究で構築したラフネスの計算モデルの推定結果は、音圧レベルが高くなるにつれ変動強度も高くなったが、その変化は主観評価結果よりも急峻であった。RMSE が最も小さいモデルは GC 変動強度モデルで、GT 変動強度モデル、Fastl モデルの順であった。

## 変調度を変化させた変調音に対する変動強度

次に、変調度を変化させた AM 信号の変動強度を評価した。刺激は、1,000 Hz の正弦波搬送波を 0.2 秒持続させ、変調度が 0.20, 0.30, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90, 1.0 で振幅変調し、変調周波数は 4 Hz とした。音圧レベルは 70 dB とした。

Fig. 4.27 は変調度に対する変動強度の推定結果を示している。横軸が変調度、縦軸が変動強度。横軸は変調度、縦軸はゆらぎの強さ。黒実線は Fastl & Zwicker の研究 [5] から引用した変動強度の主観評価結果を示す。RMSE が最も小さいモデルは Fastl モデルであり、次いで GC 変動強度モデル、GT 変動強度モデルであった。

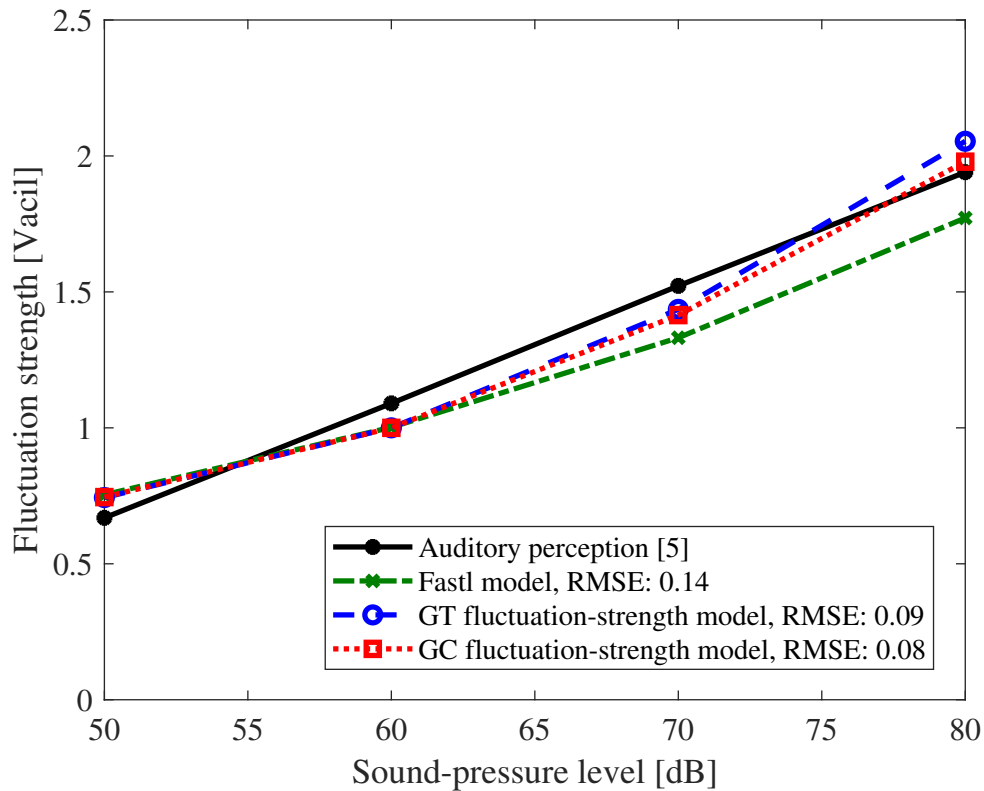


Fig. 4.25: Fluctuation strength was calculated using three fluctuation-strength models (Fastl, GT fluctuation-strength, and GC fluctuation-strength models) for various sound-pressure levels. AM signal was 1,000-Hz sinusoidal carrier with 4-sec duration with 1.0 modulation index amplitude modulation at modulation frequencies of 4 Hz.

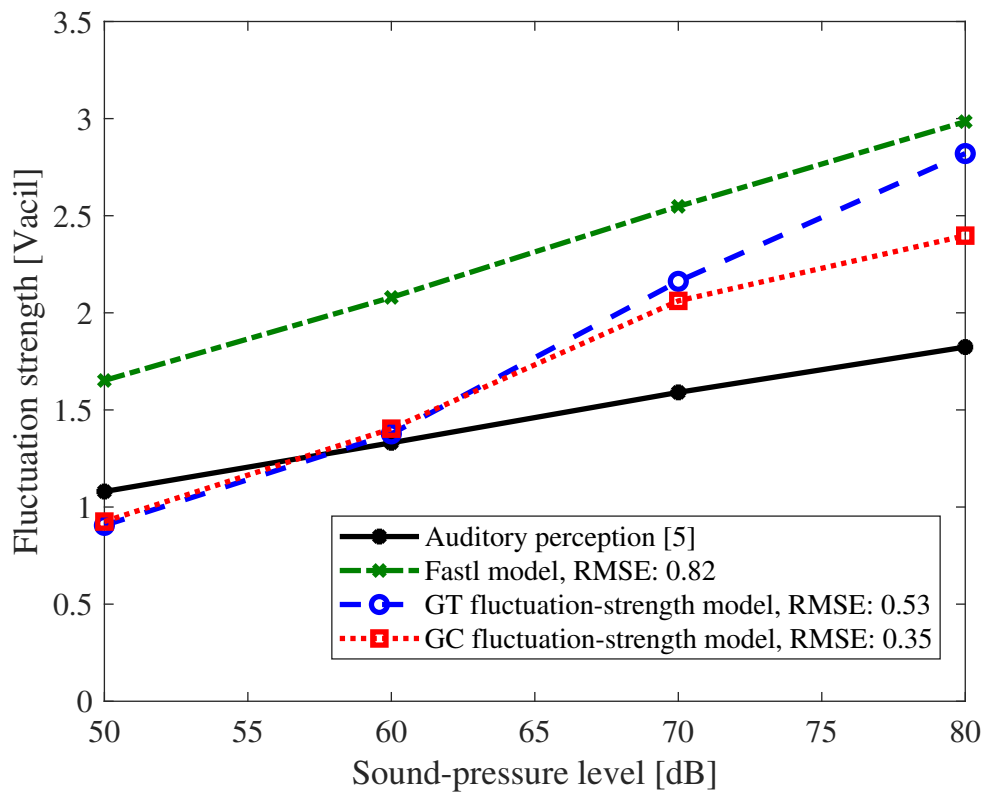


Fig. 4.26: Fluctuation strength was calculated using three fluctuation-strength models (Fastl, GT fluctuation-strength, and GC fluctuation-strength models) for various sound-pressure levels. FM signal was 1,500-Hz sinusoidal carrier with 4-sec duration with frequency modulated at 4 Hz with frequency deviation of 700 Hz.

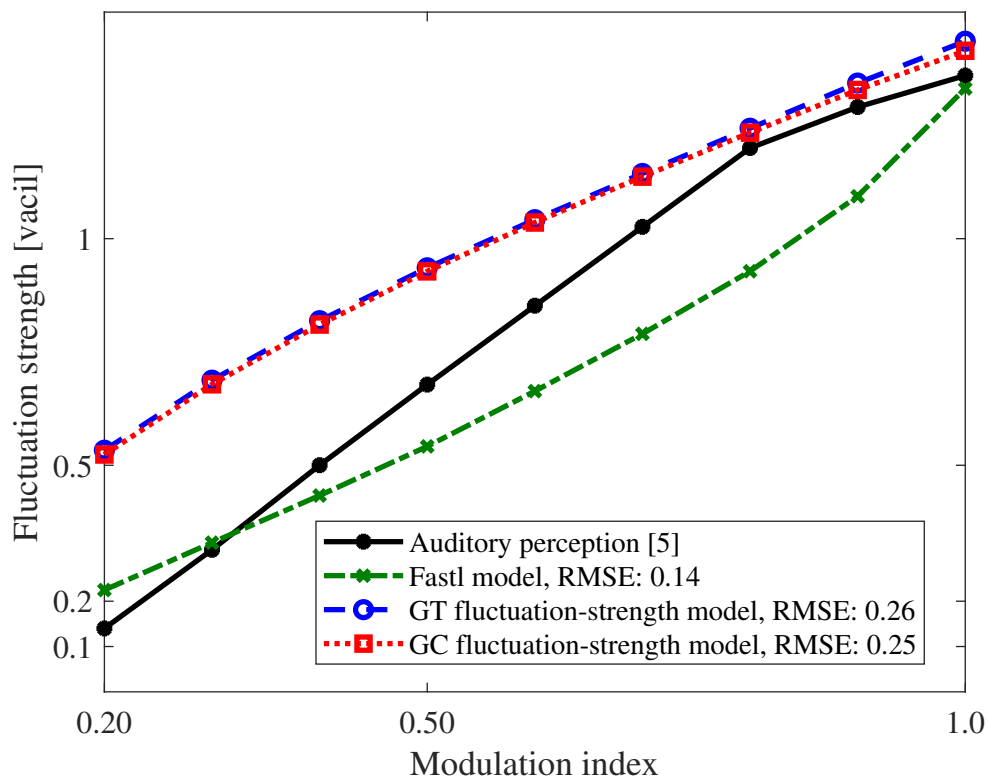


Fig. 4.27: Fluctuation strength was calculated using three fluctuation strength models (Fastl, GT fluctuation-strength, and GC fluctuation-strength models) for various modulation index of AM signal. Stimuli were sinusoidal carrier of 1,000-Hz with 4-sec duration with modulation frequency of 4 Hz and sound-pressure level of 60 dB.

## 4.7 音質評価指標の計算モデルについての総括

本稿では、GTFBやGCFBを使用したラウドネス密度の計算モデルによって計算された算定ラウドネス密度を用いた音質評価指標の構築及び評価を行った。シャープネスの計算モデルは、Auresのラウドネス依存のモデルに基づき構築された。ラフネスの計算モデルは、Daniel & Weberのラフネスの計算モデルに基づきFastlのラフネスに関する知見を取り入れて構築された。変動強度の計算モデルは、構築したラフネスの計算モデルと同様のアプローチで構築された。

音質評価指標の計算モデルは、従来モデルの評価でも使用された評価刺激を入力信号として、音質評価指標に関する主観評価結果と比較することで評価された。また、本研究で構築した音質評価指標の計算モデルは従来モデルと比較することで、本研究で構築した音質評価指標の計算モデルの有効性を評価した。評価の結果、構築したラウドネスの計算モデルは、Moore-Glasbergモデルと同程度のラウドネスを算出した。構築したシャープネスの計算モデルは、従来モデルと比較してシャープネスの推定精度が高いことがわかった。構築したラフネスの計算モデルは、従来モデルと比較して、ラフネスの推定精度が同程度以上であることがわかった。構築した変動強度の計算モデルは、従来モデルと比較して、変動強度の推定精度が同程度以上であることがわかった。

GCFBを用いた音質評価指標の計算モデルは、GTFBを用いた音質評価指標の計算モデルと比較して、音圧レベルを変化させた場合の評価においてRMSEが小さくなる傾向があった。これは、音圧レベルの変化に依存した聴覚フィルタのフィルタ形状の非対称性が音質評価指標の推定に影響を与えたためだと考えられる。

本研究で構築した音質評価指標の計算モデルは、従来モデルと同程度以上の推定精度をもつことから、構築した音質評価指標の計算モデルの有効性が確認された。また、ERB尺度を用いることや聴覚フィルタのフィルタ形状の非対称性を考慮することの重要性が示された。



## 第 5 章

### 総合考察

## 5.1 はじめに

本章では，Bark 尺度と ERB 尺度の違いや，聴覚フィルタのフィルタ形状の違いが音質評価指標の推定へ与える影響について考察する．

## 5.2 時間領域で定義された聴覚フィルタバンクを用いたラウドネスの計算モデル

本節では，時間領域で定義された聴覚フィルタバンクを用いたラウドネスの計算モデルについて考察する．

GTFBを使用したラウドネス密度計算モデルのラウドネス密度の計算結果 (Fig. 3.8, Fig. 3.9, Fig. 3.10, Fig. 3.11) と GCFBを使用したラウドネス密度モデルのラウドネス密度の計算結果 (Fig. 3.13, 3.14, 3.15, 3.16) を見ると，Moore-Glasberg モデルから算出した算定ラウドネス密度と構築したラウドネス密度の計算モデルから算出した算定ラウドネス密度で音圧レベルが低いときは，おおよそ同じ算定ラウドネス密度が算出された．一方で，音圧レベルが高いときは，Moore-Glasberg モデルから算出したラウドネス密度と構築したラウドネス密度の計算モデルから算出したラウドネス密度の形状は異なっていた．これは，roex 聴覚フィルタのフィルタ形状と GTFB 及び GCFB のフィルタ形状の主に中心周波数から離れた裾のフィルタゲインの違いが影響していると考えられる．

本研究で構築したラウドネスの計算モデルから算出された算定ラウドネスと Moore-Glasberg モデルから算出された算定ラウドネスを合わせるために，エキサイテーションパターンからラウドネス密度を計算する際の，スティーブンスのべき指数  $\alpha$  を修正した．この修正によって，音圧レベルが高いときの構築したラウドネス密度の計算モデルから算出した最大ラウドネス密度と，Moore-Glasberg モデルから算出した最大ラウドネス密度は異なると考える．

Fig. 4.9 と Fig. 4.10 と Fig. 4.11 の結果を見ると，Moore-Glasberg モデルとの RMSE は小さいことから，今回構築した時間領域で定義される聴覚フィルタバンクを用いたラウドネスの計算モデルは Moore-Glasberg モデルの時間領域版とみなせるものと考えられる．

## 5.3 精緻化された聴覚末梢系のモデルを用いた音質評価指標の計算モデル

### 5.3.1 ERB 尺度と Bark 尺度の違いが音質評価指標の推定に与える影響

本節では、ERB 尺度と Bark 尺度の違いが音質評価指標の推定に与える影響について考察する。

Fig. 1.5 に示したように、ERB 尺度と Bark 尺度の違いは主に、低域周波数における聴覚フィルタの帯域幅である。このことから、ERB 尺度と Bark 尺度の違いが音質評価指標の推定に与える影響は低域周波数に大きく出ると考えられる。

シャープネスについて考えると、ERB 尺度と Bark 尺度の違いはシャープネスの重みの違いとしても現れるはずである。Fig. 4.5 を見ると、FZ モデルも Aures モデルも低域での重みは一定であるが、構築したシャープネスの計算モデルの重みは、500 Hz 以下で重みが小さくなっている。この重みは、低い周波数を持つ音の算定シャープネスをより低くするように働くことを意味している。この違いは、ERB 尺度と Bark 尺度の低域での違いに起因していると考えられる。また、シャープネスの計算モデルの評価では、異なる雑音に対するシャープネスの結果 (Fig. 4.15) を見ると、GTFB や GCFB を用いたシャープネスの結果は、従来のモデルの算定シャープネスより RMSE が小さい。これは、人間の蝸牛の機能に対応する ERB 尺度の周波数分解が、聴覚感覚の推定を改善することを示唆している。

異なる雑音に対するシャープネスの結果のうち LP 雑音を用いた評価結果 (Fig. 4.14) を見ると、従来のシャープネスの計算モデルの結果と比べて、本研究で構築したシャープネスの計算モデルの結果は主観評価結果と近い。この評価で使用した LP 雑音は、下限周波数を 200 Hz と固定し、上限周波数と徐々に高く変化させた音である。つまり、聴覚フィルタの低域の周波数分解能が重要となる評価となっている。そのためこの評価の結果から、ERB 尺度の低域での周波数分解能の高さが、算定シャープネスの推定精度の向上に貢献したと考えられる。

ラフネスについて考えると、ERB 尺度と Bark 尺度の違いはラフネスの重みの違いとしても現れるはずである。Fig. 4.7 を見ると、DW モデルの重みは 1 kHz を

中心とした帯域通過フィルタのような形状をしている。一方で、今回構築したラフネスの重みを見ると、300 Hz 付近を中心とした帯域通過フィルタのような形状をしている。こちらもシャープネスの重みと同様に、ERB 尺度と Bark 尺度の低域での違いに起因していると考えられる。

Fig. 4.17 や Fig. 4.18 に示した AM 音や FM 音を用いた変調周波数を変化させたときのラフネスの評価では、Bark 尺度と用いたラフネスの計算モデルの結果も ERB 尺度を用いた今回構築したラフネスの計算モデルの結果の RMSE はおおよそ同じであった。ラフネスの知覚は主に、ラウドネスの時間方向の変動によって決まるものであり、Bark 尺度と用いたラフネスの計算モデルも ERB 尺度を用いた今回構築したラフネスの計算モデルも、ラウドネスの時間方向の変動に大きな差がなかったためだと考える。

Fig. 4.23 や Fig. 4.24 に示した AM 音や FM 音を用いた変調周波数を変化させたときの変動強度の評価結果からもラフネスと同様に、Bark 尺度を用いた変動強度の計算モデルの結果と ERB 尺度を用いた変動強度の計算モデルの結果には、大きな差がみられなかった。

これらの結果は、音質評価指標における Bark 尺度と ERB 尺度の違いが、特に周波数方向に関連する音色の推定精度に影響を与えることを示唆している。

### 5.3.2 聴覚フィルタのフィルタ形状の非対称性が音質評価指標の推定に与える影響

本節では、聴覚フィルタのフィルタ形状の非対称性が音質評価指標の推定に与える影響について考察する。

聴覚フィルタのフィルタ形状の非対称性は音圧レベルに依存することが知られている [44]。そのため、聴覚フィルタのフィルタ形状の非対称性が音質評価指標の推定に与える影響は音圧レベルを変化させたときの音質評価指標の推定精度に影響すると考える。

Fig. 4.16 の GTFB と GCFB を使用したシャープネスの計算モデルの結果を比較すると、全体的に聴覚フィルタのフィルタ形状の非対称性を表現できる GC シャープネスモデルの結果のほうが RMSE が小さくなった。聴覚フィルタのフィルタ形

状の非対称性は、エキサイテーションパターンにおける高域への広がりに影響を与える。さらにラウドネス密度は、このエキサイテーションパターンから計算される。このことから、音圧レベルが高くなるに付随して、ラウドネス密度も高域へ広がる。シャープネスはラウドネス密度の重心から計算されることから、ラウドネス密度も高域へ広がりにはラウドネス密度の重心を高域側へ変化させる。この効果によって、音圧レベルを変化させたときのシャープネスの RMSE が対称なフィルタ形状を持つ GT シャープネスモデルに比べて聴覚フィルタのフィルタ形状の非対称性を表現できる GC シャープネスモデルのほうが小さくなったと考えられる。

Fig. 4.19 の GTFB と GCFB を用いたラフネスの計算モデルの結果を見ると、GC ラフネスモデルの結果のほうが RMSE が小さくなった。この結果もシャープネスの結果同様に、GCFB の聴覚フィルタのフィルタ形状の非対称性が音圧レベルを変化させたときのラフネスの推定精度の向上につながったと考えられる。一方で、Fig. 4.20 に示す FM 音を用いた結果では、GT ラフネスモデルの結果のほうが RMSE が小さくなった。これは、聴覚フィルタのフィルタ形状の非対称性がラフネスの推定結果を逆に低下させてしまったと考えられる。しかし、この結果は最大値を 100 となるように正規化したことによってラフネスの推定結果が見かけ上、低下している可能性がある。

Fig. 4.25 及び Fig. 4.26 の GTFB と GCFB を用いた変動強度の計算モデルの結果を見ると、この二つのモデルの結果には大きな差がないことがわかる。変動強度が最大となる変調周波数は 4 Hz であり、聴覚フィルタの帯域幅と比べてその周波数はとても低い。聴覚フィルタのフィルタ形状の非対称性が変動強度の推定に大きな影響を与えなかったと考えられる。

聴覚フィルタのフィルタ形状の非対称性が音質評価指標の推定に与える影響についてまとめると、聴覚フィルタのフィルタ形状の非対称性は、音圧レベルを変化させたときの音色の知覚に影響していると考えられる。

本研究で構築した音質評価指標の計算モデルのすべての評価結果をまとめて考えると、従来モデルの評価結果と比較して、構築した音質評価指標の計算モデルの結果は、主観結果に対する RMSE が小さい。以上のことから、本研究で構築した音質評価指標の計算モデルは、従来の音質評価指標の計算モデルよりも、ヒトが感じる音質を説明できることと考えられる。

## 第 6 章

### 結論

## 6.1 本論文で明らかにされたことの要約

本論文では、「ヒトが知覚する音色を解明するためには精緻化された聴覚末梢系のモデルを使うべき」という立場のもと、様々な音質を推定するために、時間領域で定義される聴覚フィルタを用いた音質評価指標の計算モデルの構築を行った。そして、定常音の Moore-Glasberg モデルを元に、時間領域で定義される聴覚フィルタバンクを用いたラウドネスの計算モデルを構築し、このラウドネスの計算モデルの出力と定常音用の Moore-Glasberg モデルの出力が同様であるかを評価した。また、音質評価指標に関する主観評価結果を構築した音質評価指標の計算モデルが説明できるかを評価した。結論として以下に示すことが明らかとなった。

(1) 時間領域で定義される聴覚フィルタバンクを使用しても、Moore-Glasberg モデルと同程度にラウドネスを計算可能であることから、今回構築したラウドネスの計算モデルは Moore-Glasberg モデルの時間領域版といえる。

(2) 精緻化された聴覚末梢系モデルを使うことは、シャープネスやラフネス、変動強度などの音質評価指標の推定誤差を小さくすることに寄与する。

以上のことから、時間領域での聴覚フィルタバンクの利用が Moore-Glasberg モデルと同等の性能を発揮することが可能であり、聴覚末梢系モデルの精緻化が音質評価の精度を高める重要な要素であることを明らかにした。これにより、音質評価指標の計算モデルの開発において、これらの要素が重要な役割を果たすことが明らかとなった。

## 6.2 今後の展望

以下に本研究の今後の展望を列挙する。

### 6.2.1 本研究成果の利点と限界

本研究では、精緻化された聴覚末梢系のモデルを用いて音質評価指標の構築及び評価を行った。音質評価の精度を高めるためには聴覚末梢系のより詳細なモデリングが重要であることを示している。

本研究の利点は、精緻化されたモデルを使用することによって、実際の聴覚の挙動により近い形で音質評価を行うことができ、その結果として推定誤差を減少させることが可能になったことである。これにより、より正確な音質評価指標の計算が可能となり、音色知覚の科学的な理解や、音響機器の開発、音質改善への応用などに貢献することが期待できる。

本研究の限界は、構築した音質評価指標の計算モデルは定常音の評価にしか適応できない点である。聴覚末梢系では、入力音の音圧レベルに依存した圧縮特性があることが知られている [62]。また、今回のモデルでは、同時マスクングしか考慮していない。しかし、フォワードマスクングとバックワードマスクングなど継時マスクングの影響が音質評価指標の推定精度に影響することが考えられる。これらの時間方向の聴覚特性を本音質評価指標の計算モデルは考慮していないため、本モデルは時変動音の評価に適応することができない。例えば、シャープネスは、Ueda & Akagi らの検討から、振幅包絡線の影響を受けることが知られている [75]。また、Daniel & Stephen による、音の立ち上がりと立下りの形状が異なるダンプ音とランプ音に対するラフネスの評価から、ダンプ音とランプ音は明確にラフネスの主観評価結果が違うことが知られている [76]。こういった音に対する評価をするためには、時変動音に対応した音質評価指標の計算モデルの開発が重要である。

Moore-Glasberg モデルはバイノーラルのモデル [61] であるが、今回研究したラウドネスの計算モデルは、モノラルで構築した。そのため、両耳での音色知覚について検討することができない。例えば、音質評価師匠は、音圧レベルに依存するが、知覚レベルでは、ラウドネスの変化に対応する。両耳のラウドネスは、左右のラウドネスを単純に足し合わせたものでないことが知られている。よって、バイノーラルラウドネスの計算モデルを使用することで、左右で音圧レベルの違う音に対して、音質評価指標の推定精度が向上すると考えられる。

### 6.2.2 聴覚末梢系の時間特性の考慮

聴覚末梢系では、入力音の音圧レベルに依存した圧縮特性があることが知られている [53]。しかし、今回のラウドネスの計算モデルに用いた GCFB は聴覚の圧縮特性が表現できないことを Irino & Patterson は指摘している。時変動音に適応し



たラウドネスの計算モデルを構築するためには、Iriano & Patterson が提案した圧縮型ガンマチャープフィルタ [53] やダイナミック圧縮型ガンマチャープフィルタバンク [54] を用いるほうが良いと考えられる。また、Dau らは聴覚末梢の圧縮特性を考慮した聴覚末梢系のモデルを提案している [77,78]。このような圧縮特性を聴覚フィルタとは別のモデルとして定義して用いることも良いかもしれない。

Moore らは時変動音に対応したラウドネスの計算法を提案した (Moore-Glasberg-Schlittenlacher 法) [26,79]。この計算法は、Moore-Glasberg 法にヒトの時間に関する聴覚特性を考慮した自動利得制御回路を組み込むことで実現された。また、この計算法は ISO 532-3 として規格化された [80]。このような聴覚特性を考慮した自動利得制御回路を今回提案したラウドネスの計算モデルに組み込むことで時変動音に対応できると考えられる。このようなラウドネスの計算モデルを検討することで、ランプ音やダンプ音などの時間包絡線が立ち上がりと立下りで非対称な音に対するラフネスや、振幅包絡線を変化させたときのシャープネスなどを説明できるかもしれない。

### 6.2.3 個人の聴覚特性に合わせた音質評価指標の計算モデルの検討

ラウドネスやシャープネス、ラフネス、変動強度などの聴覚感覚は、個人差がある。今回構築した音質評価指標はすべて平均的な特性を用いて構築している。例えば、ヒトは歳を取れば、聴力が低下し、ラウドネスの知覚もそれに伴って変化する。ほかの聴感に関しても聴力低下に伴って変化するはずである。また、聴力低下に伴って聴覚末梢系においては、聴覚フィルタの帯域幅が広がることが知られている [81]。また、聴覚フィルタの圧縮特性が変化することも知られている。聴力低下を模擬した聴覚末梢モデルを用いたラウドネスの計算モデルを Moore-Glasberg は検討している [82]。このようなラウドネスの計算モデルを参考に聴力低下を考慮したラウドネスの計算モデルを検討することで、個人の聴力に基づいた音色知覚を明らかにできると考えられる。

## 6.2.4 騒音の評価

難波ら (2011) は, さまざまな環境音を対象としたラウドネスレベルの計算法を提案した [83]. この方法は, 125 ms ごとに Zwicker モデルや Moore-Glasberg モデルを用いてラウドネスレベルを計算し, 計算されたラウドネスレベルを強さの次元に変換し, 時間平均し, ラウドネスレベルに戻すことで, 時間変動音のラウドネスレベルを求める. この方法によって計算されたラウドネスレベルは, Zwicker モデルや Moore-Glasberg モデルから計算されたラウドネスレベルや等価騒音レベル,  $N_5$  と比較して主観値との相関値が高いことが示された. このことからエネルギー的に算出した全体のラウドネスレベルと主観的評価の間に良い対応関係があることがわかった. しかし, この方法では, ラウドネスレベルを一回強さの次元へ変換してから平均を求めてラウドネスレベルに戻すという計算を行っていることから, なぜ主観値と一致するのかという点が疑問として残る.

ラウドネス  $N$  とラウドネスレベル  $L$  にはラウドネスレベルが 40 phon から 120 phon の間であれば以下のような関係が成り立つことが知られている.

$$L = 40 + 10 \log_2 N \quad (6.1)$$

この関係に基づけば, モデルによって計算されたラウドネスの平均値をラウドネスレベルに変換することでも, 難波らと似た計算を行うことができると考えられる. このような計算をすることで, ある長時間の時変動音の全体のラウドネスレベルを求めることができるようになる.

## 6.2.5 今後への期待

本研究では, 聴覚末梢の特に周波数スケールに着目した. このような, ヒトの聴知覚にそくした研究は近年徐々に増えてきており, 例えば, 異常音検知であったり, ディープフェイクの検知などである. 音のことを考えれば, 必ずそこにはヒトがいて最後はヒトの聴知覚に行き着くことになる. これらの音研究に対して本研究の成果が一助となればと考えている.

# 謝辞

この研究を行うにあたり、8年間もの長きにわたり、指導と励ましを賜りました本学の鵜木祐史教授に深謝の意を表します。研究者として未熟な私に対して、鵜木教授は「磯山くんとは何年でも一緒に研究できるよ」という温かいが、同時に恐ろしいほどの言葉で励ましてくださいました。そのおかげで、研究を最後まで続けることができました。日々の研究活動では、私が間違った方向に進みそうになった時、常に厳しいご指導を賜り、最終的には誤った道を選ばずに済むよう導いていただきました。日常生活においても、生活習慣が乱れがちな時は、すぐに正しい方向へと導いてくださり、感謝しています。この論文を書き上げることができたのは、鵜木先生の熱心な指導の賜物です。このような環境を提供して下さった鵜木教授に対し、私の研究及び個人的な成長に対する深い感謝の意を表します。鵜木教授のご恩は、今後の研究活動を通じて、さらなる成果を上げることで恩返ししたいと思います。

私の予備審査をしてくださいました、本学の岡田将吾准教授、長谷川忍教授、九州工業大学工学部電気電子工学科水町光徳教授、九州大学大学院芸術工学研究院音響設計部門山内勝也准教授、東北文化学園大学工学部知能情報システム学科鈴木陽一教授に感謝の意を表します。予備審査の際には献身的な質問をしていただき、そのおかげで私の博士論文はより良いものとなりました。

この研究を進めるにあたり、熱心な御指導御鞭撻賜りました本学名誉教授の赤木正人教授に心より感謝申し上げます。研究室会議や合宿での赤木教授からの鋭い質問は、私の研究にとって大変貴重でした。専門性を追求する中で視野が狭くなりがちな時も、赤木教授の広い視野からのコメントが度々、研究の大きな助けとなりました。

本研究を遂行するにあたり、たくさんの相談に乗っていただきました本学の木谷俊介講師に心より感謝申し上げます。木谷講師からは、研究への取り組み姿勢や、

データの分析方法などたくさんのお話を教えてくださいました。木谷講師から、私が行き詰まりを感じた時にいただいた助言は、この研究を続ける上で非常に役立ちました。

上江洲安史博士、鳥谷輝樹博士をはじめ、研究室のすべてのメンバーに対して、たくさんのお話や貴重なコメントをいただいたことに深く感謝します。彼らからの質問は、私自身が気づけなかった研究の隙間を明らかにし、私の研究をより強固なものにするのに非常に役立ちました。

副テーマの研究を行うにあたり、御指導御鞭撻を賜りました東京工業高等専門学校情報工学科小嶋徹也教授に感謝の意を表します。副テーマ研究に関する会議で、小嶋教授からいただいた的確な質問は、私の研究にとって非常に価値があるものでした。さらに、自分では思いつかないような多くのアイデアを提供していただき、これが大きな助けとなりました。

また、これまでの長い道のりを支え、面倒を見てくださった家族への感謝の気持ちを忘れることはありません。彼らの絶え間ない支援と愛情があったからこそ、困難な時期も乗り越え、研究を続けることができました。

家族の理解とサポート、そして研究室のメンバーからの刺激的な交流が、私の航海という名の研究活動を豊かなものにし、多大な成長を促してくれました。これらすべての方々への深い感謝の意を表し、今後の研究と人生において彼らの期待に応えるよう努力していきます。

## 参考文献

- [1] 鈴木陽一, 赤木正人, 伊藤彰則, 佐藤洋, 菖木禎史, 中村健太郎. 音響学入門. コロナ社, 東京, 日本, 2011.
- [2] 大串健吾. 音響聴覚心理学. 誠信書房, 東京, 日本, 2019.
- [3] Seiichiro Namba, Sonoko Kuwano, and Masaru Koyasu. The measurement of temporal stream of hearing by continuous judgments—in the case of the evaluation of helicopter noise. *Journal of the Acoustical Society of Japan (E)*, Vol. 14, No. 5, pp. 341–352, 1993.
- [4] 曾根敏夫, 城戸健一, 二村忠元. 音の評価に使われることばの分析. 日本音響学会誌, Vol. 18, No. 6, pp. 320–326, 1962.
- [5] Hugo Fastl and Eberhard Zwicker. *Psycho-acoustics: Facts and Models*. Springer, La Vergne, TN USA, 2010.
- [6] Bodden Markus. Instrumentation for sound quality evaluation. *Acta Acustica united with Acustica*, Vol. 83, pp. 775–783, 1997.
- [7] Nykänen Arne and Sirkka Anna. Specification of component sound quality applied to automobile power windows. *Applied Acoustics*, Vol. 70, No. 6, pp. 813–820, 2009.
- [8] Kwon Gahee, Jo Hyeonho, and June Kang Yeon. Model of psychoacoustic sportiness for vehicle interior sound: Excluding loudness. *Applied Acoustics*, Vol. 136, pp. 16–25, 2018.
- [9] Matteo Lionello, Francesco Aletta, and Jian Kang. A systematic review of prediction models for the experience of urban soundscapes. *Applied Acoustics*,

Vol. 170, p. 107479, 2020.

- [10] Margret Sibylle Engel, André Fiebig, Carmella Pfaffenbach, and Janina Fels. A review of the use of psychoacoustic indicators on soundscape studies. *Current Pollution Reports*, Vol. 7, No. 3, pp. 359–378, 2021.
- [11] Mercede Erfanian, Andrew J. Mitchell, Jian Kang, and Francesco Aletta. The psychophysiological implications of soundscape: A systematic review of empirical literature and a research agenda. *International Journal of Environmental Research and Public Health*, Vol. 16, No. 19, 2019.
- [12] Mian Tauheed, Choudhary Anurag, and Fatima Shahab. An efficient diagnosis approach for bearing faults using sound quality metrics. *Applied Acoustics*, Vol. 195, p. 108839, 2022.
- [13] Roberto Merino-Martinez, Ana Vieira, Mirjam Snellen, and Dick G. Simons. Sound quality metrics applied to aircraft components under operational conditions using a microphone array. In *25th AIAA/CEAS Aeroacoustics Conference*, 2019.
- [14] Yasuji Ota and Masashi Unoki. Anomalous sound detection for industrial machines using acoustical features related to timbral metrics. *IEEE Access*, Vol. 11, pp. 70884–70897, 2023.
- [15] Von W. Aures. Berechnungsverfahren für den sensorischen wohlklang beliebiger schallsignale. *Acustica*, Vol. 58, pp. 130–141, 1985.
- [16] Scaling the unpleasantness of sounds according to the btl model: Ratio-scale representation and psychoacoustical analysis. *Acta Acustica*, Vol. 90, No. 1, 2004.
- [17] Using probabilistic choice models to investigate auditory unpleasantness. *Acta Acustica*, Vol. 90, No. 6, 2004.

- [18] J. Morel, C. Marquis-Favre, and L.-A. Gille. Noise annoyance assessment of various urban road vehicle pass-by noises in isolation and combined with industrial noise: A laboratory study. *Applied Acoustics*, Vol. 101, pp. 47–57, 2016.
- [19] ISO 532: 1975, acoustics – method for calculating loudness level, 1975.
- [20] G von Bismarck. Sharpness as an attribute of the timbre of steady sounds. *Acustica*, Vol. 30, No. 3, pp. 159–172, 1974.
- [21] Hugo Fastl. The hearing sensation roughness and neuronal responses to am-tones. *Hearing Research*, Vol. 46, No. 3, pp. 293–295, July 1990.
- [22] Hugo Fastl. Fluctuation strength and temporal masking patterns of amplitude-modulated broadband noise. *Hearing Research*, Vol. 8, pp. 59–69, 1982.
- [23] Eberhard Zwicker and B. Scharf. A model of loudness summation. *Psychological Review*, Vol. 72, No. 1, pp. 3–26, 1965.
- [24] Roy D. Patterson and G. Bruce Henning. Stimulus variability and auditory filter shape. *The Journal of the Acoustical Society of America*, Vol. 62, No. 3, pp. 649–664, 1977.
- [25] Brian C. J. Moore and Brian R. Glasberg. A revision of zwicker’s loudness model. *Acta Acustica united with Acustica*, Vol. 82, pp. 335–345, 03 1996.
- [26] Brian C. J. Moore, Brian R. Glasberg, Ajanth Varathanathan, and Josef Schlittenlacher. A loudness model for time-varying sounds incorporating bin-aural inhibition. *Trends in hearing*, Vol. 20, pp. 1–16, 2016.
- [27] 難波精一郎. 精神物理学と音色. 日本學士院紀要, Vol. 71, No. 2, p. 53, 2017.
- [28] Georg von Békésy. The variation of phase along the basilar membrane with sinusoidal vibrations. *Journal of the Acoustical Society of America*, Vol. 19, pp. 452–460, 1947.

- [29] Edgar A. G. Shaw. Transformation of sound pressure level from the free field to the eardrum in the horizontal plane. *The Journal of the Acoustical Society of America*, Vol. 56, No. 6, pp. 1848–1861, 08 1974.
- [30] Mead C. Killion, Elliott H. Berger, and Richard A. Nuss. Diffuse field response of the ear. *The Journal of the Acoustical Society of America*, Vol. 81, No. S1, pp. S75–S75, 08 1987.
- [31] George F. Kuhn. The pressure transformation from a diffuse sound field to the external ear and to the body and head surface. *The Journal of the Acoustical Society of America*, Vol. 65, No. 4, pp. 991–1000, 04 1979.
- [32] Edgar A. G. SHAW. The acoustics of the external ear. *Acoustical Factors Affecting Hearing Aid Performance*, 1980.
- [33] ISO 532-2: 2017, acoustics – methods for calculating loudness – part 2: Moore-glasberg method, 2017.
- [34] R. L. Wegel and C. E. Lane. The Auditory Masking of One Pure Tone by Another and its Probable Relation to the Dynamics of the Inner Ear. *Phys. Rev.*, Vol. 23, pp. 266–285, Feb 1924.
- [35] James P. Egan and Harold W. Hake. On the Masking Pattern of a Simple Auditory Stimulus. *The Journal of the Acoustical Society of America*, Vol. 22, No. 5, pp. 622–630, 09 1950.
- [36] Harvey Fletcher. Auditory patterns. *Rev. Mod. Phys.*, Vol. 12, pp. 47–65, Jan 1940.
- [37] Eberhard Zwicker, Gordon Flottorp, and Stanley Smith Stevens. Critical Band Width in Loudness Summation. *The Journal of the Acoustical Society of America*, Vol. 29, No. 5, pp. 548–557, 05 1957.
- [38] Eberhard Zwicker. Subdivision of the audible frequency range into critical bands (frequenzgruppen). *The Journal of the Acoustical Society of America*, Vol. 33, No. 2, pp. 248–248, 07 1981.



- [39] Eberhard Zwicker and Ernst Terhardt. Analytical expressions for critical - band rate and critical bandwidth as a function of frequency. *The Journal of the Acoustical Society of America*, Vol. 68, No. 5, pp. 1523–1525, 11 1980.
- [40] Donald D. Greenwood. Critical Bandwidth and the Frequency Coordinates of the Basilar Membrane. *The Journal of the Acoustical Society of America*, Vol. 33, No. 10, pp. 1344–1356, 10 1961.
- [41] Donald D. Greenwood. A cochlear frequency - position function for several species—29 years later. *The Journal of the Acoustical Society of America*, Vol. 87, No. 6, pp. 2592–2605, 06 1990.
- [42] Brian C. J. Moore and Brian R. Glasberg. Suggested formulae for calculating auditory - filter bandwidths and excitation patterns. *The Journal of the Acoustical Society of America*, Vol. 74, No. 3, pp. 750–753, 09 1983.
- [43] Brian C.J. Moore and Brian R. Glasberg. Formulae describing frequency selectivity as a function of frequency and level, and their use in calculating excitation patterns. *Hearing Research*, Vol. 28, No. 2, pp. 209–225, 1987.
- [44] Derivation of auditory filter shapes from notched-noise data. *Hearing Research*, Vol. 47, No. 1, pp. 103–138, 1990.
- [45] Hartmann William M. *Signals, Sound, and Sensation*. Springer, American Institute of Physics Melville, NY, 1997.
- [46] M. Unoki, K. Ito, Y. Ishimoto, and C. T. Tan. Estimate of auditory filter shape using notched-noise masking for various signal frequencies. *Acoustical Science and Technology*, Vol. 27, No. 1, pp. 1–11, 2006.
- [47] R. J. Baker and S. Rosen. Auditory filter nonlinearity across frequency using simultaneous notched-noise masking. *Journal of the Acoustical Society of America*, Vol. 119, No. 1, pp. 454–462, 2006.

- [48] 赤木正人. 聴覚フィルタとそのモデル. *The journal of the Institute of Electronics, Information and Communication Engineers*, Vol. 77, No. 9, pp. 948–956, 1994.
- [49] Roy D. Patterson, Ian Nimmo-Smith, J. Holdsworth, and P. Rice. An efficient auditory filterbank based on the gammatone function. In *A Meeting of the IOC Speech Group on Auditory Modelling at RSRE*, 1987.
- [50] Toshio Irino and Roy D. Patterson. A time-domain, level-dependent auditory filter: the gammachirp. *Journal of the Acoustical Society of America*, Vol. 101, No. 1, pp. 412–419, 1997.
- [51] Toshio Irino and Masashi Unoki. A time-varying, analysis/synthesis auditory filterbank using the gammachirp. In *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP '98 (Cat. No.98CH36181)*, Vol. 6, pp. 3653–3656, 1998.
- [52] Toshio Irino and Masashi Unoki. An analysis/synthesis auditory filterbank based on an iir implementation of the gammachirp. *Journal of the Acoustical Society of Japan (E)*, Vol. 20, No. 6, 12 1999.
- [53] Toshio Irino and Roy D. Patterson. A compressive gammachirp auditory filter for both physiological and psychophysical data. *Journal of the Acoustical Society of America*, Vol. 109, No. 5 Pt 1, pp. 2008–2022, May 2001.
- [54] Toshio Irino and Roy D. Patterson. A dynamic compressive gammachirp auditory filterbank. *IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING*, Vol. 14, No. 6, pp. 2222–2232, 2006.
- [55] Yôiti Suzuki and Hisashi Takeshima. Equal-loudness-level contours for pure tones. *The Journal of the Acoustical Society of America*, Vol. 116, No. 2, pp. 918–933, 08 2004.
- [56] H. Fletcher and W. A. Munson. Loudness, its definition, measurement and calculation. *Journal of the Acoustical Society of America*, Vol. 5, pp. 82–108,

1933.

- [57] D W Robinson and R S Dadson. A re-determination of the equal-loudness relations for pure tones. *British Journal of Applied Physics*, Vol. 7, No. 5, pp. 166–181, may 1956.
- [58] G von Bismarck. Timbre of steady sounds: A factorial investigation of its verbal attributes. *Acustica*, Vol. 30, No. 3, pp. 146–159, 1974.
- [59] von E. Terhardt. Über akustische rauigkeit und schwankungsstärke. *Acustica*, Vol. 20, pp. 215–224, 1968.
- [60] ISO 532-1: 2017, acoustics – methods for calculating loudness – part 1: Zwicker method, 2017.
- [61] Brian C. J. Moore and Brian R. Glasberg. Modeling binaural loudness. *The Journal of the Acoustical Society of America*, Vol. 121, No. 3, pp. 1604–1612, 2007.
- [62] Brian C. J. Moore. *An Introduction to the Psychology of Hearing*. Brill, Leiden, Boston, 2013.
- [63] Von W. Aures. Der sensorische wohlklang als funktion psychoakustischer empfindungsgrößen. *Acustica*, Vol. 58, pp. 282–290, 1985.
- [64] S Hales Swift and Kent L Gee. Extending sharpness calculation for an alternative loudness metric input. *Journal of the Acoustical Society of America*, Vol. 142, , 2017.
- [65] DIN 45692: 2009. measurement technique for the simulation of the auditory sensation of sharpness. German, 2009.
- [66] von E. Terhardt. Über die amplitudenmodulierte sinustöne hervorgerufene hörempfindung. *Acustica*, Vol. 20, pp. 210–214, 1968.
- [67] von W. Aures. Ein berechnungsverfahren der rauigkeit. *Acustica*, Vol. 58, pp. 268–281, 1985.

- [68] Peter Daniel and R. Weber. Psychoacoustical roughness: Implementation of an optimized model. *Acta Acustica united with Acustica*, Vol. 83, No. 1, pp. 113–123, 1997.
- [69] R. P. N. Duisters. *The modeling of auditory roughness for signals with temporally asymmetric envelopes*. PhD thesis, Technische Universiteit Eindhoven, 2005.
- [70] Ulrich Widmann and Hugo Fastl. Calculating roughness using time-varying specific loudness spectra. In *Proc. Sound quality symposium 98*, pp. 55–60, 1998.
- [71] Vecchi Vecchi, Osses, Rodrigo Garcia León, and Armin Kohlrausch. Modelling the sensation of fluctuation strength. In *Proceedings of the 22nd International congress on acoustics*, Vol. 28, p. 050005, 2016.
- [72] Slaney Malcolm. An efficient implementation of the patterson-holdsworth auditory filter bank. Technical report, Apple Computer Tech. Rep. #35, 1993.
- [73] Roy D. Patterson, Masashi Unoki, and Toshio Irino. Extending the domain of center frequencies for the compressive gammachirp auditory filter. *The Journal of the Acoustical Society of America*, Vol. 114, No. 3, pp. 1529–1542, 08 2003.
- [74] Masashi Unoki, Toshio Irino, B. Glasberg, Brian C. J. Moore, and Roy D. Patterson. Comparison of the roex and gammachirp filters as representations of the auditory filter. *Journal of the Acoustical Society of America*, Vol. 120, No. 3, pp. 1474–1492, 2006.
- [75] Kazuo Ueda and Masato Akagi. Sharpness and amplitude envelopes of broadband noise. *The Journal of the Acoustical Society of America*, Vol. 87, No. 2, pp. 814–819, 02 1990.

- [76] Daniel Pressnitzer and Stephen McAdams. Two phase effects in roughness perception. *The Journal of the Acoustical Society of America*, Vol. 105, No. 5, pp. 2773–2782, 05 1999.
- [77] Torsten Dau, Birger Kollmeier, and Armin Kohlrausch. Modeling auditory processing of amplitude modulation. I. Detection and masking with narrow-band carriers. *The Journal of the Acoustical Society of America*, Vol. 102, No. 5, pp. 2892–2905, 11 1997.
- [78] Torsten Dau, Birger Kollmeier, and Armin Kohlrausch. Modeling auditory processing of amplitude modulation. II. Spectral and temporal integration. *The Journal of the Acoustical Society of America*, Vol. 102, No. 5, pp. 2906–2919, 11 1997.
- [79] Brian C. J. Moore, Brian R. Glasberg, and T. Baer. A model of loudness applicable to time-varying sounds. *Audio Engineering Society*, Vol. 50, No. 5, pp. 331–342, 2002.
- [80] ISO 532-3: 2023, acoustics – methods for calculating loudness – part 3: Moore-glasberg-schlittenlacher method, 2023.
- [81] Roy D. Patterson, Ian Nimmo-Smith, Daniel L. Weber, Robert Milroy. The deterioration of hearing with age: Frequency selectivity, the critical ratio, the audiogram, and speech threshold. *The Journal of the Acoustical Society of America*, Vol. 72, No. 6, pp. 1788–1803, 12 1982.
- [82] Brian C.J. Moore and Brian R. Glasberg. A revised model of loudness perception applied to cochlear hearing loss. *Hearing Research*, Vol. 188, No. 1, pp. 70–88, 2004.
- [83] 難波精一郎, 桑野園子, 加藤徹. 時間的に変化する複合音の大きさ評価. 音楽知覚認知研究, Vol. 17, No. 1, pp. 13–33, 2011.

# 本研究に関する発表論文

## ジャーナル論文

- [1] Takuto Isoyama, Shunsuke Kidani, Masashi Unoki, “Computational models of sharpness and fluctuation strength using loudness models composed of gammatone and gammachirp auditory filterbanks,” *Journal of Signal Processing*, Vol. 25, No. 4, pp. 141–144, 2021, doi: 10.2299/jsp.25.141.
- [2] Takuto Isoyama, Shunsuke Kidani, Masashi Unoki, “Computational models of auditory sensations important for sound quality on basis of either gammatone or gammachirp auditory filterbank,” *Applied Acoustics*, Vol. 218, No. 15, 2024, doi: 10.1016/j.apacoust.2024.109914.

## 国際会議

(査読有・口頭発表)

- [3] Takuto Isoyama, Shunsuke Kidani, Masashi Unoki, “Computational models of sharpness and fluctuation strength using loudness models composed of gammatone and gammachirp auditory filterbanks,” *Proc. RISP International Workshop in Nonlinear Circuits, Communications and Signal Processing 2021 (NCSP'21)*, Online, Mar. 2021.

## (査読有・ポスター発表)

- [4] Takuto Isoyama, Shunsuke Kidani, Masashi Unoki, “Modeling of sound quality metrics using gammatone and gammachirp filterbanks,” *Proc. Forum Acusticum 2020*, pp. 2731–2735, Online, Dec. 2020.
- [5] Takuto Isoyama, Shunsuke Kidani, Masashi Unoki, “Computational model for predicting sound quality metrics using loudness model based on gammatone/gammachirp auditory filterbank and its applications,” *Proc. Inter Noise 2023*, 2-P-48, Chiba, Japan, Aug. 2023.

## 国内発表

### 研究会

- [6] 磯山拓都, 木谷俊介, 鵜木祐史, “聴覚フィルタバンクを利用した定常音に対するラウドネスモデルの構築,” 電子情報通信学会技術研究報告, Vol. 119, No. 440, pp. 273–278, 沖縄, 2020年3月
- [7] 磯山拓都, 木谷俊介, 鵜木祐史, “聴覚フィルタバンクを用いたラウドネスモデルとシャープネスモデルの構築,” 日本音響学会聴覚研究会資料, Vol. 50, No. 20, pp. 129–124, オンライン, 2020年5月
- [8] 磯山拓都, 木谷俊介, 鵜木祐史, “聴覚フィルタバンクを利用した音質評価指標の計算モデルの検討,” 日本音響学会聴覚研究会資料, Vol. 52, No. 6, pp. 441–446, 宮城, 2022年8月
- [9] 磯山拓都, 木谷俊介, 鵜木祐史, “聴覚フィルタバンクを用いた時変動音のラウドネス計算法の検討,” 日本音響学会聴覚研究会資料, Vol. 53, No. 6, pp. 283–288, 宮城, 2023年9月

## 国内学会

- [10] 磯山拓都, 木谷俊介, 鵜木祐史, “ガンマトーンフィルタバンクを用いたラウドネスモデルの構築,” 日本音響学会 2020 年春季研究発表会講演論文集, 1-Q-1, 埼玉, 2020 年 3 月
- [11] 磯山拓都, 木谷俊介, 鵜木祐史, “聴覚モデルを用いたシャープネスと変動強度の計算モデルの構築,” 日本音響学会 2021 年春季研究発表会講演論文集, 1-4P-2, オンライン, 2021 年 3 月
- [12] 磯山拓都, 木谷俊介, 鵜木祐史, “聴覚フィルタバンクを用いたラフネスモデルの構築,” 日本音響学会 2022 年秋季研究発表会講演論文集, 1-P-13, 北海道, 2022 年 9 月
- [13] 磯山拓都, 木谷俊介, 鵜木祐史, “聴覚フィルタバンクを用いた時変動音のラウドネス計算法の構築,” 日本音響学会 2023 年秋季研究発表会講演論文集, 1-R-5, 愛知, 2023 年 9 月



# 本研究以外の発表論文

## ジャーナル論文

- [1] Takuto Isoyama, Shunsuke Kidani, Masashi Unoki, “Blind speech watermarking method with frame self-synchronization based on spread-spectrum using linear prediction residue,” *Entropy*, Vol. 24, No. 5, 2022, doi: 10.3390/e24050677.

## 国際会議

### (査読有・口頭発表)

- [2] Shunsuke Kidani, Xiaoting Liu, Taiyang Guo, Takuto Isoyama, Junfeng Li, and Masashi. Unoki, “Study on the modulation frequency range that contributes to the perception of urgency,” *Proc. the 24th International Congress on Acoustics (ICA 2022)*, ABS-0547, 2022.
- [3] Aulia Adila, Candy Olivia Mawalim, Takuto Isoyama, and Masashi Unoki, “Study on Inaudible Speech Watermarking Method Based on Spread-Spectrum Using Linear Prediction Residue,” *Proc. RISP International Workshop in Nonlinear Circuits, Communications and Signal Processing 2024 (NCSP’ 24)*, pp. 55–58, 2024.

## 国内発表

### 研究会

- [4] 磯山拓都, 鵜木祐史, “線形予測残差を用いたスペクトル拡散型音声情報ハイディング法におけるフレーム同期の検討,” 電子情報通信学会技術研究報告, Vol. 121, No. 247, EMM2021-67, pp. 74–79, オンライン, 2021年11月
- [5] 磯山拓都, 小嶋徹也, 鵜木祐史, “線形予測残差を用いたスペクトル拡散型音声情報ハイディング法におけるフレーム同期の改良,” 電子情報通信学会技術研究報告, Vol. 122, No. 368, EMM2022-66, pp. 26–31, 宮城, 2023年1月
- [6] 木谷俊介, 劉小婷, 郭太陽, 磯山拓都, 李軍鋒, 赤木正人, 鵜木祐史, “音声に含まれる緊迫感の変調周波数帯域の検討,” 日本音響学会音声研究会資料, Vol.3, No.1, オンライン, 2023年1月

### 国内学会

- [7] 磯山拓都, 鵜木祐史, “線形予測残差を用いたスペクトル拡散型音声情報ハイディング法：フレーム同期とブラインド検出,” 日本音響学会 2021年秋季研究発表会講演論文集, 3-1Q-9, オンライン, 2021年9月
- [8] 木谷俊介, 劉小婷, 郭太陽, 磯山拓都, 李軍鋒, 赤木正人, 鵜木祐史, “緊迫感知覚に寄与する変調周波数帯域の検討,” 日本音響学会 2022年秋季研究発表会講演論文集, 3-10-7, 北海道, 2022年9月
- [9] 磯山拓都, 鵜木祐史, “線形予測残差を用いたスペクトル拡散型音声情報ハイディング法におけるフレーム同期,” 日本音響学会 2022年春季研究発表会講演論文集, 3-1P-8, オンライン, 2022年3月
- [10] 木谷俊介, 磯山拓都, 鵜木祐史, “謡曲の良さに寄与するスペクトル・時間変調情報の検討,” 日本音響学会 2023年秋季研究発表会講演論文集, 1-Q-4, 愛知, 2023年9月

- [11] 郭太陽, 磯山拓都, 木谷俊介, 鶴木祐史, “時間振幅包絡線の瞬時変調周波数成分が感情音声知覚に与える影響の検討,” 日本音響学会 2024 年春季研究発表会講演論文集, 1-3-4, 東京, 2024 年 3 月