

Title	Spectro-Temporal Modulation分析を用いた音質評価指標と聴覚的顕著性の対応関係の検討
Author(s)	田中, 聡一郎
Citation	
Issue Date	2022-12
Type	Thesis or Dissertation
Text version	author
URL	<a href="http://hdl.handle.net/10119/18169">http://hdl.handle.net/10119/18169</a>
Rights	
Description	Supervisor: 鶴木 祐史, 先端科学技術研究科, 修士(情報科学)

修士論文

Spectro-Temporal Modulation 分析を用いた音質評価指標と聴覚的顕著性の  
対応関係の検討

2010114 田中 聡一郎

主指導教員 鶴木 祐史

北陸先端科学技術大学院大学  
先端科学技術研究科  
(情報科学)

令和4年12月

## Abstract

In daily lives, humans rarely hear only one sound, instead, they hear multiple sounds coming from multiple sources simultaneously, i.e., they hear certain sounds and voices actively by directing their own attention to them. There is a phenomenon exists in which a particular sound is noticed, or the presence of a particular sound can be recognized simultaneously with the occurrence of the sound in a place where multiple sound sources are present. Thus, auditory saliency is introduced as the sounds that unintentionally attract the ear without actively directing attention to a particular sound.

Auditory saliency has been examined as passive auditory attention, focusing on various acoustic features. Nevertheless, since the correspondence between the two was investigated directly in a state of the manual investigation, the nature of the auditory saliency hasn't been clarified. When people hear sounds, differences in timbre are also relevant. Therefore, differences in timbre may also part in auditory saliency.

The relationship between acoustic features and auditory saliency has been investigated in a study by Kidokoro et al. based on Spectro-Temporal Modulation Spectrum (STMS) (2021). The results show that the degree of temporal variability of STMS is related to auditory saliency under conditions in which the loudness of the sound stimulus is controlled. Nevertheless, the direct investigation of the relationship between acoustic features and auditory saliency only explains the one-to-one relationship. Another study by Arnal et al. (2015) used STM analysis to investigate the characteristics of screams with high auditory saliency. In this study, STM analysis was shown to be able to analyze acoustic features and sound quality metrics (SQM) and to link acoustic features and SQM via STMS.

These results exhibit that features appearing in the STMS could be used to the explanation of the relationship between SQM and auditory saliency. This study aims to explain the relationship between SQMs (sharpness  $\cdot$  roughness) and auditory saliency via features appearing in the STMS.

Therefore, the following steps will be taken in this study.

1. STM analysis of stimuli of known auditory saliency
2. Analyze STMS using higher-order statistics
3. First, analyze the relationship between SQMs and auditory saliency. Next, analyze the relationship between features obtained from STMS and SQMs. Finally, analyze the relationship between features derived from STMS and auditory saliency
4. Consider the relationship between STMS-derived features and SQMs and auditory saliency

The filter bank used in the study was a constant band gammatone filter bank,

with a frequency bandwidth of 160 Hz per filter per channel based on previous studies and the definition of roughness, and a number of divisions of 551 channels. The time modulation region of 80 Hz and the spectral modulation region of 0 to 3.5 cycl/kHz were defined as the analysis range in the STM. In this study, focused on the distribution of STMS as a feature related to spectral modulation and temporal modulation. The mean and variance, skewness, and kurtosis of higher order statistics were obtained to examine the relationship between SQMs and auditory saliency. The results showed that roughness of the SQM could be related to auditory saliency via higher-order statistics of the STMS.

# 目次

<b>第1章</b>	<b>序論</b>	<b>1</b>
1.1	はじめに . . . . .	1
1.2	研究背景 . . . . .	2
1.3	本研究の目的 . . . . .	4
1.4	本研究の着想 . . . . .	4
1.5	本論文の構成 . . . . .	5
<b>第2章</b>	<b>関連研究</b>	<b>7</b>
2.1	聴覚的顕著性に関する研究 . . . . .	7
2.2	音質評価指標に関する研究 . . . . .	9
2.3	スペクトル変調・時間変調に関する研究 . . . . .	11
<b>第3章</b>	<b>対象刺激のSTM分析</b>	<b>14</b>
3.1	研究の方法論 . . . . .	14
3.2	STM分析の手順 . . . . .	16
3.3	分析条件 . . . . .	18
3.4	高次統計量 . . . . .	21
3.4.1	分析結果 . . . . .	22
3.4.2	考察 . . . . .	25
<b>第4章</b>	<b>STMSを介した音質評価指標と聴覚的顕著性の関係の検討</b>	<b>26</b>
4.1	分析方法 . . . . .	26
4.2	STMSの高次統計量と音質評価指標の関係 . . . . .	26
4.3	STMSの高次統計量と聴覚的顕著性の関係 . . . . .	35
4.4	音質評価指標と聴覚的顕著性の関係 . . . . .	40
4.5	考察 . . . . .	44
<b>第5章</b>	<b>全体考察</b>	<b>45</b>
<b>第6章</b>	<b>結論</b>	<b>52</b>
6.1	本研究で明らかにしたこと . . . . .	52
6.2	残された課題 . . . . .	52

# 図目次

1.1	聴覚的顕著性のイメージ図 . . . . .	3
1.2	本論文の構成 . . . . .	6
2.1	「The radio was playing too loudly」という文章の読み上げ音声におけるスペクトログラム (文献 [44] より引用) . . . . .	12
2.2	STMS 例と生理学的な関連 (文献 [11] より引用): 図中青色の変調領域には性別に係る情報, 緑色の変調領域には意味的情報, 橙色の変調領域には音の粗さに係る情報を示す . . . . .	13
3.1	本研究の方略 . . . . .	15
3.2	STM 分析のブロック図: (a) 入力信号, (b) パワー包絡線 (パワースペクトログラム), (c) STMS の対数表現 . . . . .	17
3.3	分析対象の音刺激 (文献 [20] より作成) . . . . .	19
3.4	顕著性スケール (文献 [20] より作成) . . . . .	20
3.5	STM 分析の結果: (a1-10): 対数パワースペクトル (対数パワー包絡線), (b1-10): STMS, (c1-10): STMS の分布. 各パネルは顕著性スケールの並びに沿って, 上から下 (1,2,...,10) に向かって聴覚的顕著性が低から高に配置 . . . . .	24
4.1	STMS 重心とラフネスの相関 . . . . .	27
4.2	STMS 分散とラフネスの相関 . . . . .	28
4.3	STMS 歪度とラフネスの相関 . . . . .	29
4.4	STMS 尖度とラフネスの相関 . . . . .	30
4.5	STMS 重心とシャープネスの相関 . . . . .	31
4.6	STMS 分散とシャープネスの相関 . . . . .	32
4.7	STMS 歪度とシャープネスの相関 . . . . .	33
4.8	STMS 尖度とシャープネスの相関 . . . . .	34
4.9	STMS 重心と顕著性スケールの相関 . . . . .	36
4.10	STMS 分散と顕著性スケールの相関 . . . . .	37
4.11	STMS 歪度と顕著性スケールの相関 . . . . .	38
4.12	STMS 尖度と顕著性スケールの相関 . . . . .	39
4.13	シャープネスと顕著性スケールの相関 . . . . .	42
4.14	ラフネスと顕著性スケールの相関 . . . . .	43

5.1	対数パワースペクトルの分布. 各パネルは顕著性スケールの並びに沿って配置されており, 上から下 (1,2,...,10) に向かって聴覚的顕著性が低から高に付置されている . . . . .	47
5.2	パワースペクトル歪度とラフネスの相関 . . . . .	48
5.3	パワースペクトル尖度とラフネスの相関 . . . . .	49
5.4	パワースペクトル歪度と顕著性スケールの相関 . . . . .	50
5.5	パワースペクトル尖度と顕著性スケールの相関 . . . . .	51

# 表 目 次

4.1	計算モデルより算出したシャープネスとラフネスの値 . . . . .	41
4.2	回帰統計 . . . . .	41
4.3	重回帰分析 . . . . .	41

# 第1章 序論

## 1.1 はじめに

ヒトが視覚・聴覚をはじめとした感覚機能（五感）を通じて得ることができる感覚情報は、生活を営む上でその用途や変化などに応じた主要な役割を果たす。五感の中では視覚が最も優位であり、伝達できる情報量が非常に多く、遠くからでも多くの情報を得られるが、能動的でなければその情報を得られない。一方で、聴覚では伝達できる情報量は視覚に比べると少ないが、無意識や睡眠中であっても情報を得られる。また、視覚的な情報が遮蔽物などで奪われた場合においても、音情報のみで周囲の状況を素早く察知できる。こうして我々の聴覚系は、音を通じて天敵の存在を察知したり、異性を見つけて繁殖を行う、また食糧となる獲物を見つけるために発達してきた歴史がある [1]。その主な役割は、危険察知や他者とのコミュニケーションをとることにある。このように、我々の聴覚は感覚器官のなかでも情報を取得できる範囲は大きく、日常生活において重要な役割を担う。

我々の聴覚系における情報処理について、普段聴取する音を考えると、周波数と時間（物理量）が重要であり、音色とも明確に区別される。物理的な気圧変化が音であり、時間的な変化で示したものが時間波形である。そして、時間波形の振幅およびスペクトルの変化によって音色が変化し、これらの変化を巧みな分析によって処理し、情報を取得している。聴覚における情報処理の観点でみると、聴覚系は外耳から聴神経に及ぶ聴覚末梢系における周波数分析、そしてその後の大脳皮質までの聴覚中枢系における再構築によって、混合音の中から単一の音として知覚することができる。普段の経験からもわかる通り、日常生活において複数の音源から生じる音を同時に聴取することが多い。この場合でも、ヒトは特定の音や音声のみを聞き取ることができる。この例に示した現象は、カクテルパーティー効果 [2] として名が知れており、自ら積極的に意識を向けること（能動的聴取）で実現される。

一方で、複数の音源がある場所において、特定の音が気にかかったり、音の発生をいち早く認識できる現象がある。例えば、一般に人々は、特別に指示がない限り、周囲の環境に対してほとんど注意を払わない [3] が、都心の雑踏の中における自動車のクラクションやヒトの悲鳴などは、その音や発生した方向に意識を向けなくてもすぐさまその音を認識できる。このように、能動的に特定の音に注意を向けずとも思わず耳を引き付けられる音を聴覚的顕著性 [4] という。

## 1.2 研究背景

聴覚的顕著性は、知覚的な驚きによってもたらされる聴覚における受動的注意である [4]。聴覚的顕著性が音のどのような特徴により生じるかは、知覚しやすい報知音の設計や音環境の設計に重要な要因である。聴覚は周囲 360 度を処理対象の範囲としているため、自動車の察知や危険を知らせるサイレンなどの周囲の状況の迅速な把握が求められる場合において、聴覚が果たす役割は大きい。

図 1.1 に聴覚的顕著性のイメージ図を示す。このように、ヒトの聴知覚は外界の監視を目的とした一種の監視システムと捉えることができる。そのため、突然に発せられる音を瞬時に知覚できる特性は、ヒトの聴知覚において重要な役割を持つ。目立ちやすい音に対するヒトの反応を解釈できれば、聴覚系における音響情報処理の理解に繋がる。すなわち、聴覚的顕著性を明らかにすることが、聴知覚メカニズムの解明に繋がる。

聴覚的顕著性は、ヒトの状態が変わらずとも音の物理的特性の違いにより変化すると考えられ、主に聴覚的顕著性に係る音響特徴が検討されてきた [5]。なかでも時間と共に変化するスペクトルに着目し、強度を色の濃淡として現したスペクトログラムがよく使用される。スペクトログラムでは、時間と周波数、時間・周波数方向へ変化する特徴が確認できる。特に、時間と周波数方向の変化を変調成分として捉える手法として、スペクトル変調・時間変調 (STM: Spectro-Temporal Modulation) 分析があり、STM 分析で得られる STM スペクトル (STMS) と聴覚的顕著性の関係が検討されている [6],[7]。しかし、単純に音の強度や周波数特性だけでは聴覚的顕著性を説明できず、ヒトにとってどのような音が顕著だと感じるかも不明である。

一方で、聴覚的顕著性は、音の不快さと大きく係り [8]、不快さに係る指標に音質評価指標の一つであるシャープネスがある [9]。音質評価指標は、音色に係る客観評価指標であり、様々な音響特徴と関連がある。また、音響特徴だけで聴覚的顕著性を説明する状況には至っていない。そのため聴覚的顕著性の本質を知るためには、音響特徴だけでなく音色からの検討も必要である。

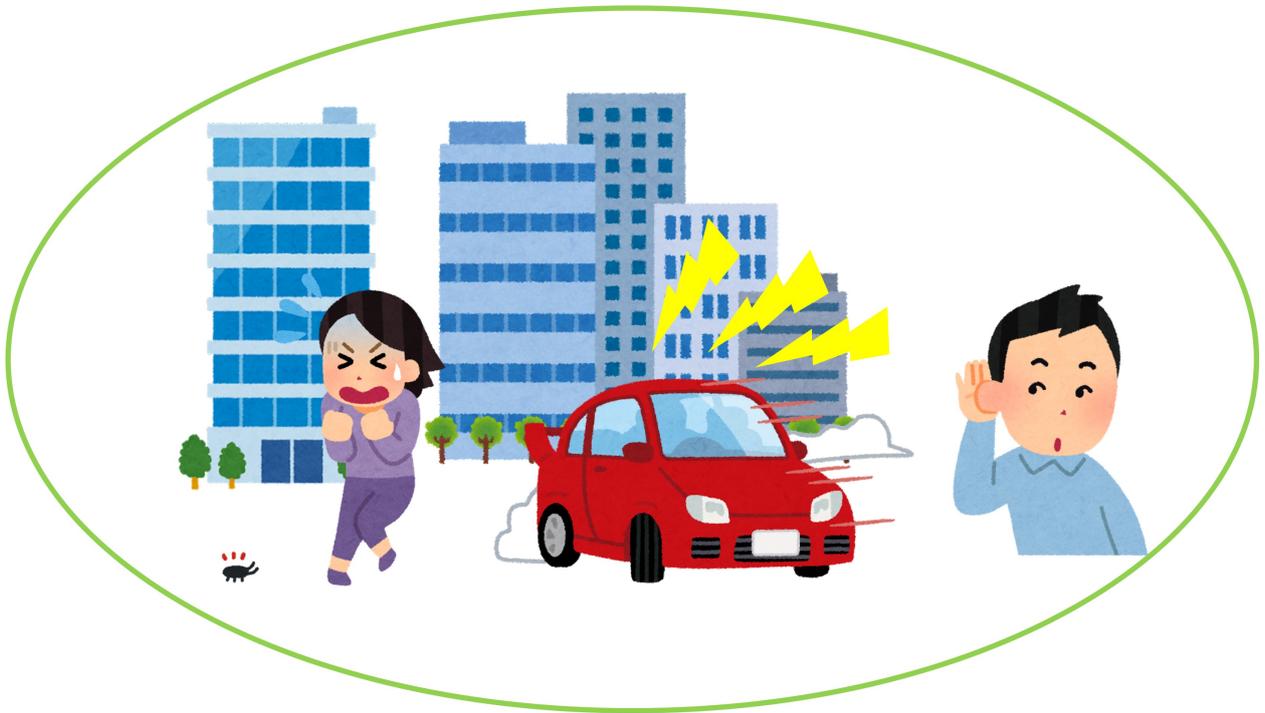


図 1.1: 聴覚的顕著性のイメージ図

### 1.3 本研究の目的

本研究では、STMスペクトル (STMS) から得られる特徴を介して、音質評価指標 (シャープネスおよびラフネス) と聴覚的顕著性の関係を明らかにすることを目的とした。音質評価指標は、ヒトが音を聴取した際の感覚を数値化した音色に係る客観評価指標 [9] のことで、変調知覚特性に基づく音響特徴に係っている。本研究では、聴覚的顕著性が既知である種々の音を対象として、STMS に現れる特徴を介すことで、音質評価指標と聴覚的顕著性の関係について検討する。

### 1.4 本研究の着想

聴覚的顕著性に関する研究では、音の物理的特徴の違いにより生じる聴覚の受動的注意として、聴覚的顕著性に係る種々の音響特徴が調査されている。木所らは、STMS を手がかりとして聴覚的顕著性を説明できるか検討し、刺激のラウドネスを揃えた条件下において、STMS の時間変化が聴覚的顕著性に係ることを示した [7]。しかし、STM から得られる特徴と聴覚的顕著性の対応関係を直接検討しており、STMS に現れるどのような特徴が聴覚的顕著性に係るかは明らかでない。

ここで、ヒトが音を聞く場合を考えると、音響特徴の違いといった物理的な要因の他にも、心理的要因も関係する [10]。また、聴覚的顕著性には音の不快さという聴覚要因に係る報告もある [8]。そのため聴覚的顕著性には、音色の違いも係ることが考えられる。音色に係る指標として音質評価指標があるため、音質評価指標を手がかりに聴覚的顕著性を説明できると考えられる。

本質的に顕著な音である「叫び声」を STM 分析し、叫び声を持つ特異性を調査した研究がある [11]。その結果、STMS の特定の時間変調領域に音質評価指標の一つであるラフネスに係る特徴が現れ、STMS を介して、音響特徴と音質評価指標を結びつけた。

これらの研究から、音質評価指標と聴覚的顕著性の間に STMS から得られる特徴を介すことで、音質評価指標と聴覚的顕著性に共通する関係性を説明できると考えた。

## 1.5 本論文の構成

本論文の構成を図 1.2 に示す.

### 第 1 章

本研究で扱う聴覚的顕著性に関する研究背景を述べることで目的を明確にし、本研究を行うにあたって得た着想を述べる.

### 第 2 章

聴覚的顕著性に関する研究と本研究で着目した音質評価指標に関する研究および、スペクトル変調・時間変調 (STM) に関する研究を紹介する.

### 第 3 章

STM 分析および STM スペクトル (STMS) の分析について述べる. まず, 聴覚的顕著性が既知である 10 個の刺激を STM 分析し, STMS を求める. その後, スペクトル変調と時間変調に係る特徴として STMS の分布に着目し, 高次統計量 (重心, 分散, 歪度, 尖度) を求める.

### 第 4 章

STMS の分布における高次統計量と計算モデルから算出したシャープネスおよびブラフネス, 聴覚的顕著性の関係について分析する. その後, シャープネスおよびブラフネスと聴覚的顕著性の関係を分析する.

### 第 5 章

前章までの聴覚的顕著性が既知である刺激の STM 分析結果および本研究のコンセプトの総合的な考察を論じる.

### 第 6 章

本研究で明らかになったことと残された課題について述べる.

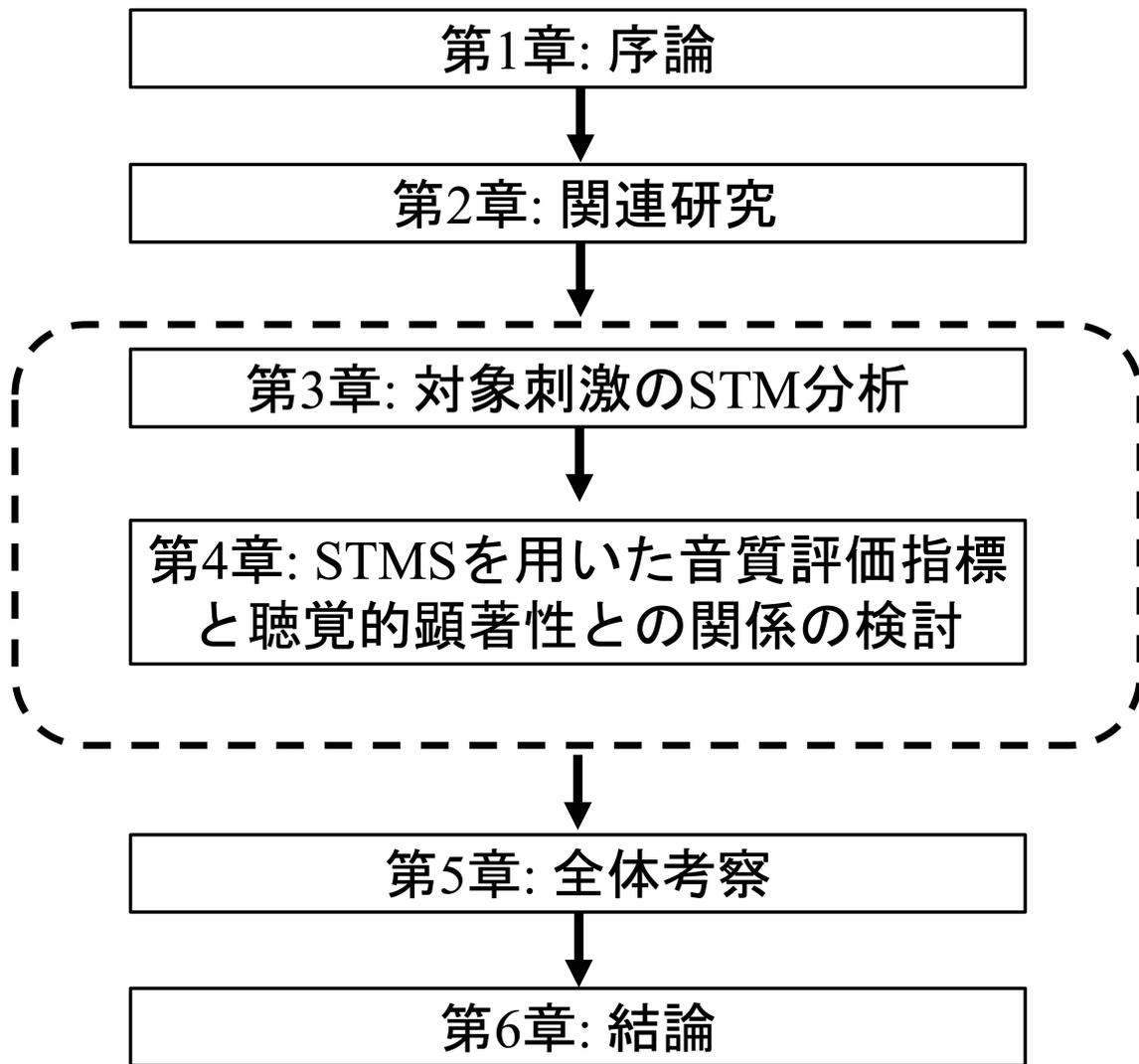


図 1.2: 本論文の構成

## 第2章 関連研究

### 2.1 聴覚的顕著性に関する研究

聴取者の状態によらない音の目立ちやすさの特性を聴覚的顕著性といい、これまでに様々な音刺激を対象として聴覚的顕著性を生起させる音響特徴が検討されてきた [4],[5]. また、聴覚的顕著性は、感覚的顕著性と意味的顕著性という2つの次元に分解できる. 感覚的顕著性は、特定の音響特徴に対する人間の聴覚系の反応によって決まる [4]. 一方で、意味的顕著性は、音と環境内の不一致を認識により生起される [12]. 前者は、行動を変化させる特徴を明示的に同定すること [4] や視覚分野の知見を土台として、スペクトログラムを分析していた [13].

このように、顕著性は、視覚の研究分野から発展してきた歴史があり [14], Kayserら [13] によって聴覚的顕著性へ発展された. この研究では、スペクトログラムにおける振幅スペクトルと時間コントラスト、周波数コントラストから顕著性マップを作成した. その結果、聴覚的顕著性と高い相関を示した. サウンドスケープでの聴覚的顕著性を調査した研究では、スペクトルと時間両方の音響特徴が聴覚的顕著性に係ることを示した [5].

音響特徴のほかに音色を取り入れた研究がある. Tordiniらは、音の明暗に係る Brightness, Tempo や Loudness を算出に係る音響特徴と共に聴覚的顕著性の係りを検討した [15],[16]. その結果、聴覚的顕著性とこれらの算出に用いるエンベロープにおける時間方向のモーメント重心である時間軸重心 (Temporal centroid) やスペクトル重心 (Spectral centroid), 音響的な周期性を表す Harmonicity の相関を示した. また, Kaya[17],[18] らや Taishih[19] らの研究でも、音色およびその算出に係る音響特徴を含めて聴覚的顕著性と相関が高いことが示された. しかし、聴覚的顕著性の評価では、被験者の主観に左右されるため、上述したボトムアップ (無意識的) による聴覚的顕著性の正確な評価が行えているかは不明である. また、聴覚的顕著性と音響特徴の関係がよくとりだたされるが、ヒトの生理学的・行動学的にも重要である聴覚的顕著性の検討のためには、知覚および各種神経系、またはそれらの相互作用も係ると考えられる.

この課題に対し、Liaoらは、聴覚的顕著性の評価にヒトの音を聴取した際の反応を生理指標として用いた [20]. 自然音や合成音に対する主観評価実験より、瞳孔の反応の強さと聴覚的顕著性の係りを示した. また、時間方向・スペクトル方向に非定常な音ほど顕著であることを突き止めた. さらに、ラウドネスを統制した際にも刺激間で聴覚的顕著性が異なること示唆した.

音のスペクトル方向・時間方向の変化を変調成分として捉える手法である Spectro-Temporal Modulation (STM) 分析がある。木所らは、STM 分析で得られる STM スペクトル (STMS) から得られる音響特徴を検討した [7]。結果より、ラウドネスを統制することで、STMS の時間変動の度合いが聴覚的顕著性に係ることを明らかにした。しかし、STMS に現れる特徴と聴覚的顕著性の対応を直接検討したため、その本質は明らかでない。

他方で、本質的に顕著的な音である叫び声を STM 分析して、叫び声を持つ特徴を調査した研究がある。結果より、0~150 Hz の振幅変調が不快な聴覚属性に係ることを明らかにし、聴覚属性の一つである音の粗さ (ラフネス) によって叫び声が他の信号と異なる特徴を有することを示した [11]。また、基本周波数 (F0) が高い泣き声は、他の発声よりも嫌悪感が高いことが示されている [21],[22]。

ここまでに概観した研究より、聴覚的顕著性には、ヒトが音を聴取した際の心理的要因 (音色) が関係し、複数の音響特徴の相互作用により生じると考えられる。しかし、聴覚的顕著性に係る研究の多くは、音響特徴と聴覚的顕著性の対応関係を直接検討していたため、その一部を説明するに留まり、聴覚的顕著性の本質は明らかでない。

## 2.2 音質評価指標に関する研究

音色に関する研究として、音質評価指標がある [9]。音質評価指標は、音を聴取した際のヒトの感覚量と物理量を結び付けたものである。家電製品や自動車のサウンドデザイン、または生活環境における音環境の快適さを評価する目的として使用されることが多く、感覚的な音の鋭さに係るシャープネスやザラザラ感といった粗さに係るラフネスや音の変動感に係る変動強度に代表される。

ラウドネスは、音圧レベル 40 dB である 1 kHz の純音が 1 sone と定義される。[23],[24],[25]。

$$N = \int_0^{24Bark} N'(z) dz \quad [\text{sone}] \quad (2.1)$$

ここで、Bark は臨界帯域、 $N'(z)$  は臨界帯域ごとのラウドネス、 $z$  は臨界帯域番号であり、音圧レベルのほかにも波形や継続時間、周波数にも依存する。

シャープネスは、音圧レベルが 60 dB、1 kHz を中心周波数として帯域幅が 160 Hz のとき 1 acum と定義される [9]。

$$S = 0.11 \frac{\int_0^{24Bark} N'(z) g'(z) dz}{\int_0^{24Bark} N'(z) dz} \quad [\text{acum}] \quad (2.2)$$

ここで、 $g(z)$  はシャープネスの重みを表す。

ラフネスは、音圧レベルが 60 dB、搬送波周波数が 1 kHz、変調周波数が 70 Hz のとき 1 asper と定義される。

$$R = 0.3 f_m \int_0^{24Bark} \Delta L(z) dz \quad [\text{asper}] \quad (2.3)$$

ここで、 $f_m$  は変調周波数、 $\Delta L$  は興奮レベルの山谷の差で、興奮レベルの変動に依存して音の粗さは変化する。ラフネスの計算では、臨界帯域ごとの  $\Delta L$  を計算し、全帯域で加算したものが、分析音に対する  $\Delta L$  として求めることができる。変動強度は、音の変動感や快適性と係りがあり、音圧レベルが 60 dB、搬送波周波数が 1 kHz、変調周波数が 4 Hz の音が 1 vacil と定義される [26]。

$$F = \frac{0.008 \int_0^{24Bark} \Delta L(z) dz}{(f_{mod}/4) + (4/f_{mod})} \quad [\text{vacil}] \quad (2.4)$$

音の変動感の強さは、 $\Delta L$  に比例し、 $f_m$  が 4 Hz のときに最大になる。

ここまでに紹介した音質評価指標は、主に音質評価やサウンドデザインで使用される。また、前節で紹介した顕著性と瞳孔の反応に関する研究 [20] でも示す通り、音質評価でもラウドネスの影響を排除した上で音質評価を行うことが多い。高

田ら [27] や Kwon ら [28],[29] は、ラウドネス統制した上で機械の作動音の音質劣化に関する要因を特定し、統制の前後で音の起伏といった音質の特徴がはっきりすると報告している。

重回帰分析により、機械作動音の音質を評価した研究も数多く存在する。You らは、説明変数としてラウドネスをはじめ四つの指標でモデルを構築した [30]。この研究では、各指標値と評価値が反比例すると報告した。変動強度と音の純音感に係るトナリティ [31] を用いた空調音の快適性評価では、変動強度が高く、トナリティが小さいほど、音を聴取した際に快適と感じることを示した [32]。一方で、Horvat らは、音質評価指標の統計量に基づく指標を用いた [33]。結果より、複雑な音の特徴を捉えるためには、音質評価指標以外の指標をモデルに含めることの重要性を示した。このように、異なるサウンドスケープ（ある人や人が文脈の中で知覚し、経験し、理解する音環境） [34] における音質評価では、音質評価指標が必須である。

音質評価に留まらず、音イベント検出法に音質評価指標を使った研究がある [35],[36]。この研究では、ヒトの音の聴覚的印象に着目し、音質評価指標とその統計量を用いた学習モデルによって検出精度を評価した。その結果、他の音の特徴を特徴量として採用した場合より、高い検出精度を示した。さらに畠山らは、年齢知覚の推定というヒトの聴覚メカニズムの解明に向け、シャープネス、ラフネス、変動強度に着目した [37]。その結果、基本周波数などのほかに声のざらつき感を年齢知覚の指標としている可能性を示唆した。

これら音質評価指標に係る研究より、音質評価指標は機械の音質評価だけでなく、ヒトの知覚メカニズム解明に向けた試みへの有効性が判明している。また、人間の感覚を種々の物理量と関連付けた指標であるため、音質評価指標を用いることで音の特徴を数値として見てとれ、聴覚的顕著性の高低に係る音の特徴が明確になることや聴覚的顕著性の操作が行えると考えられる。これらのことから、音質評価指標によって、ヒトの聴覚系における情報処理の記述、とりわけ聴覚的顕著性を解明できると考えられる。

## 2.3 スペクトル変調・時間変調に関する研究

図2.1に「The radio was playing too loudly」という文章の読み上げ音声のスペクトログラムを示す。この図より、音は時間方向と周波数方向、また時間と周波数の両方で変化するパワーエンベロープの山と谷（斜め方向への変化）が確認できる。この時間および周波数方向に対する変化を変調成分として捉える手法として、スペクトル変調・時間変調（STM: Spectro-Temporal Modulation）分析がある。STM分析で得られるSTMスペクトル（STMS）は、時間変調（Temporal Modulation: TM [Hz]）とスペクトル変調（Spectral Modulation: SM [cycles/kHz]）の2次元スペクトルとして表現したものである [38]。

音のSTM分析からスペクトル変調・時間変調を調査することで、その音が有する生理学・知覚的性質について知ることができる [39],[40]。Schonwiesnerらは、STMSと高次の変調知覚における神経発火に係るスペクトル・時間受容野（Spectro-Temporal Reception Field: STRF）の対応を示した [41],[42]。STRFは、刺激のニューロンに対する反応を表す。そのため、STMSにはヒトの知覚がある程度反映されると推測される。

これらの報告から、図2.1には、音の強度や周波数特性が現れるが、STMSには、神経発火に係る情報が現れるため、より詳細な音の特性の把握が期待できる。また、図2.2にSTMSからわかる特徴を示す。青色の変調領域には、話者の性別に係る情報が [43],[44]、緑色の変調領域には、語韻といった意味的情報が [45],[46]、オレンジ色の変調領域には、音の粗さに係る情報を示している [11]。ここに現れる特徴は、叫び声といった特異な音声に含まれる。

木所らは、STMSを手がかりとして聴覚的顕著性の解明に向けた検討を行い、ラウドネスが刺激間で同じ場合、STMSの時間変化と顕著性の相関が高いことを示した。また、音声了解度がSTMSにおける時間変調: 1~10 Hz、スペクトル変調: 1~3 cycl/kHzの範囲で伝達すると報告されており [47],[48]、[49] ヒトの聴覚システムが音声だけでなく複雑な信号の分析に、STMSに依存している生理学的・知覚的根拠がわかっている。以上の結果より、聴覚的顕著性と相関の高いSTMSに現れる特徴を用いることで、聴覚的顕著性を説明できると考えられる。

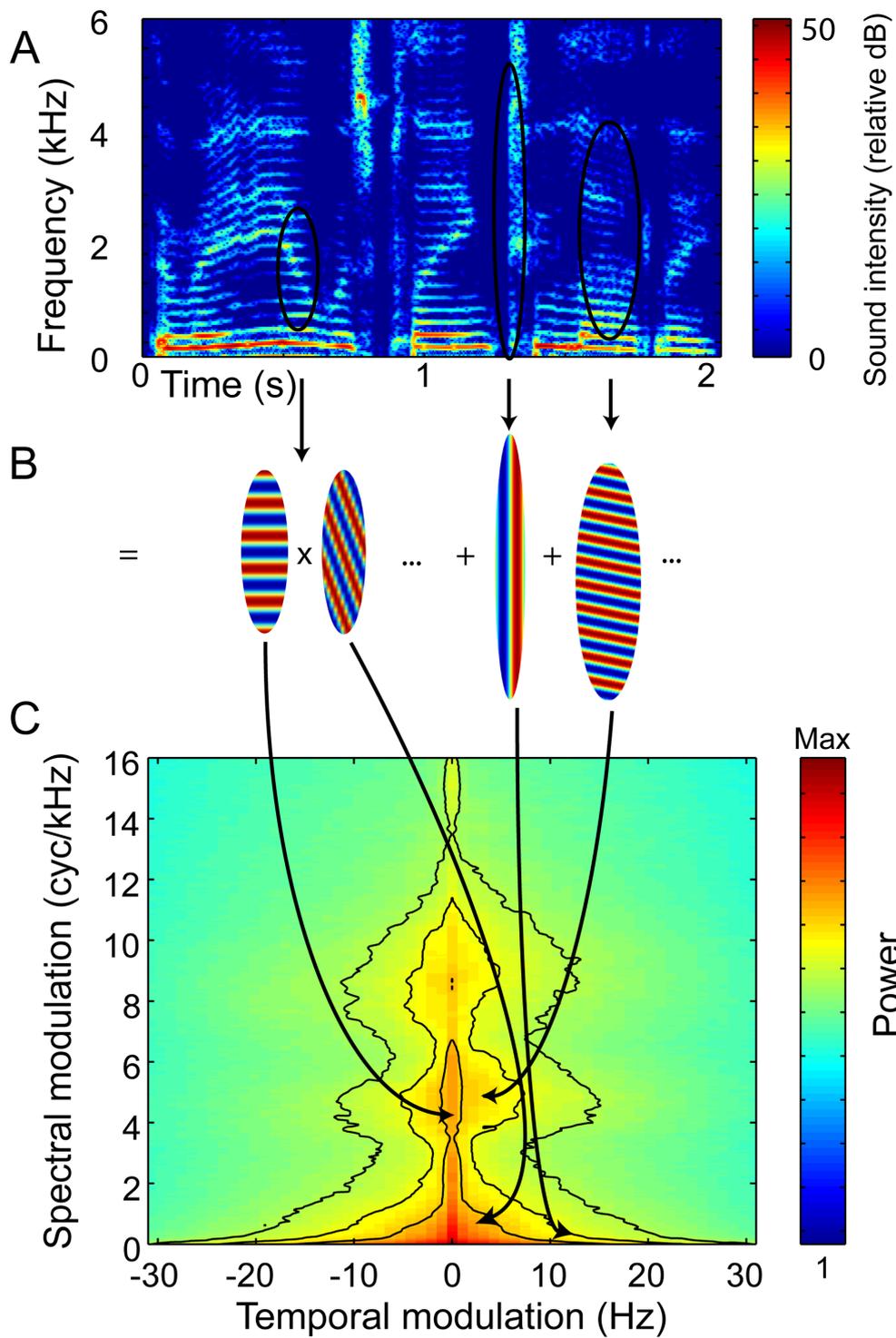


図 2.1: 「The radio was playing too loudly」という文章の読み上げ音声におけるスペクトログラム (文献 [44] より引用)

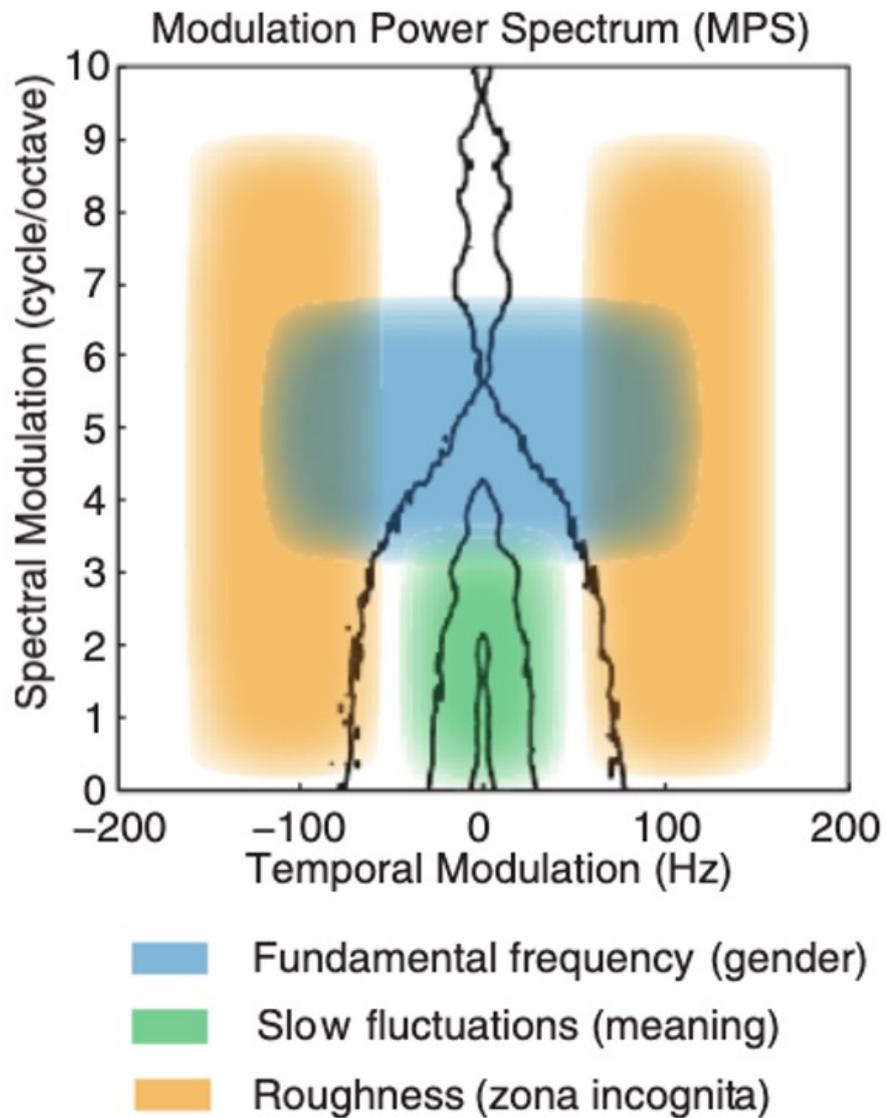


図 2.2: STMS 例と生理学的な関連 (文献 [11] より引用): 図中青色の変調領域には性別に係る情報, 緑色の変調領域には意味的情報, 橙色の変調領域には音の粗さに係る情報を示す

## 第3章 対象刺激のSTM分析

### 3.1 研究の方法論

本研究では、目的を達成するために、図3.1に示した方略に沿って研究を進める。

まず、聴覚的顕著性が既知の刺激をSTM分析する。STMSから得られる特徴を手がかりに聴覚的顕著性を説明できることがわかっている。また、STMSに現れる特徴と音の持つ性質の関係がわかっているため、スペクトル変調・時間変調を分析することにより、音質評価指標と聴覚的顕著性を結びつける特徴を検討する。

本研究では、スペクトル変調と時間変調に係る特徴として、STMSの分布に着目することで、ヒトの知覚に係る指標である音質評価指標と聴覚的顕著性の関係を検討する。また、STMSおよびSTMSの算出に用いたパワースペクトログラムに対して、高次統計量（重心、分散、歪度、尖度）を求める。

次に、計算モデル [50] を利用してシャープネスとラフネスを算出し、各刺激から求めた高次統計量と顕著性スケールの相関を求める。ここでは、相関が0.5以上かつ有意水準は0.05以下の場合に聴覚的顕著性と相関があるとした。

最後に、STMから得られる特徴を介して音質評価指標と聴覚的顕著性の関係を分析する。STMSの高次統計量を介した検討により、音質評価指標と聴覚的顕著性の関係が示せれば、聴覚メカニズムを解明することにつながり、聴覚的顕著性を音信号から予測することや音質評価指標の制御によって、聴覚的顕著性を操作できることが期待できる。

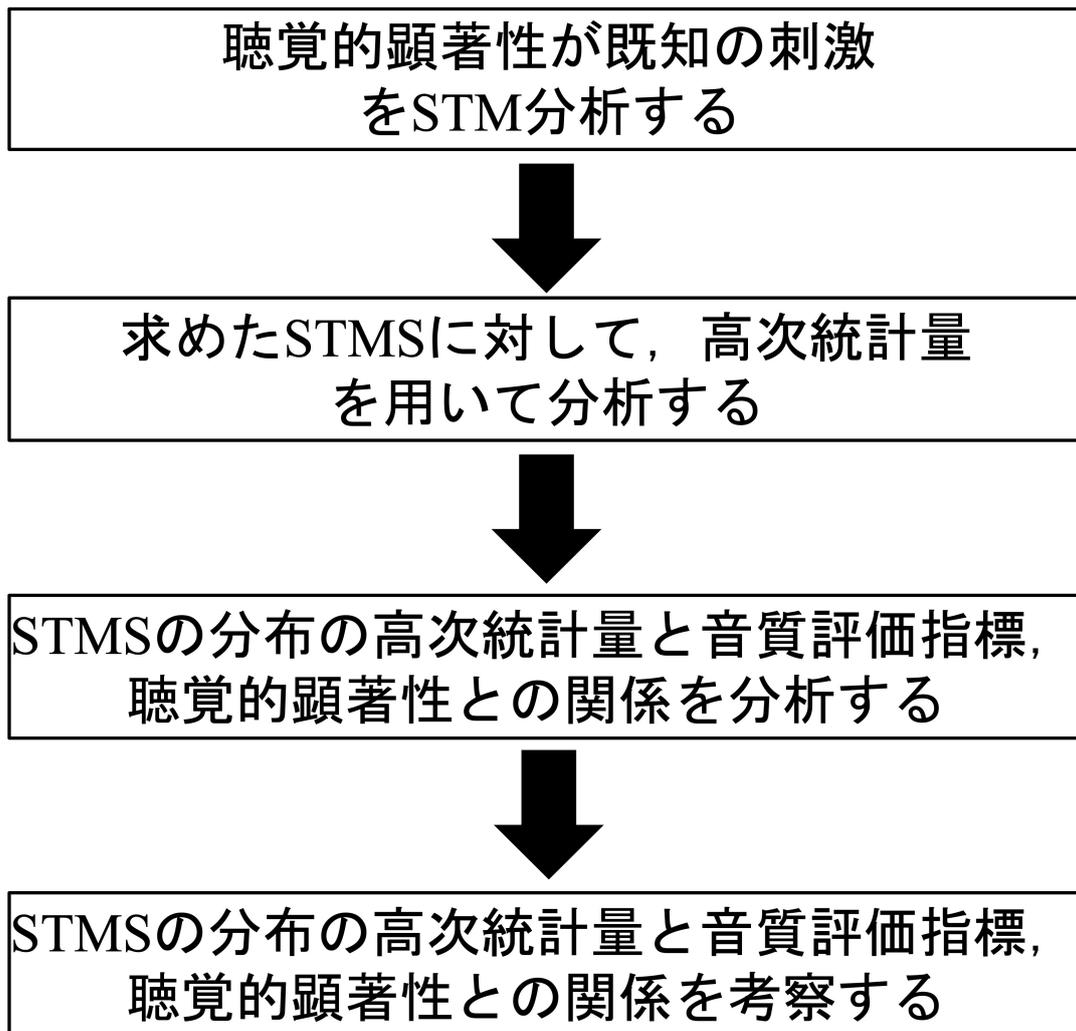


図 3.1: 本研究の方略

## 3.2 STM分析の手順

図3.2にSTM分析のブロック図を示す。STM分析では、まず定帯域ガンマトーンフィルタバンク（CGFB: Constant-band Gammatone FilterBank）のインパルス応答を利用して、周波数分析を行う。定義式を以下に示す。

$$g_k(n) = A(n\Delta t)^{(N-1)} \exp(-2\pi B n \Delta t) \cos(2\pi f_k n \Delta t + \Phi) \quad (3.1)$$

$$x_k(n) = g_k(n) * s(n) \quad (3.2)$$

ここで、 $x_k(n)$ はCGFBのチャンネル出力である。また、 $*$ は畳み込み演算子であり、 $A$ は振幅、 $t$ は時間、 $N$ はフィルタの次数、 $B$ （=160）は帯域幅、 $\Phi$ は位相を表す。また、添字の $k$ は定帯域ガンマトーンフィルタバンクのチャンネル番号であり、 $k$ （=551）個の定帯域ガンマトーンフィルタバンクが構築された。各フィルタは、フィルタ帯域幅でのゲインが $-3$  dBで交差している。

次に、フィルタバンク出力 $x_k(n)$ に対し、次式に示すヒルベルト変換および二乗処理により、各周波数帯域のパワー包絡線 $e_k^2(n)$ を求める。

$$e_k^2(n) = |x_k(n) + j \cdot \text{Hilbert}(x_k(n))|^2 \quad (3.3)$$

$\text{Hilbert}(\cdot)$ はヒルベルト変換、 $|\cdot|^2$ は二乗処理である。

最後に、パワー包絡線 $e_k^2(n)$ に対して、次式により、2次元Fourier変換によって、図中(c)に示すSTMスペクトル $\text{STMS}(m, l)$ が求まる。

$$\text{STMS}(m, l) = 2\text{DDFT}(\log e_k^2(n)) \quad (3.4)$$

ここで、 $m$ は時間変調周波数のインデックス（ $m = 1, 2, \dots, M$ ）、 $l$ はスペクトル変調のインデックス（ $l = 1, 2, \dots, L$ ）を表す。

本研究のSTMの範囲は、時間変調（TM）周波数 $f_{\text{TM}}$ で $\pm 80$  Hzの時間変調分解能（ $\Delta f_{\text{TM}}$ ）、スペクトル変調（SM）周波数 $f_{\text{SM}}$ で $0$  cycl/kHz $\sim 3.5$  cycl/kHzのスペクトル変調分解能（ $\Delta f_{\text{SM}}$ ）の変調領域とした。

ここで、本研究で定義した $f_{\text{TM}}$ と $f_{\text{SM}}$ は、 $f_{\text{TM}} = m\Delta f_{\text{TM}}$ と $f_{\text{SM}} = l\Delta f_{\text{SM}}$ で表される。

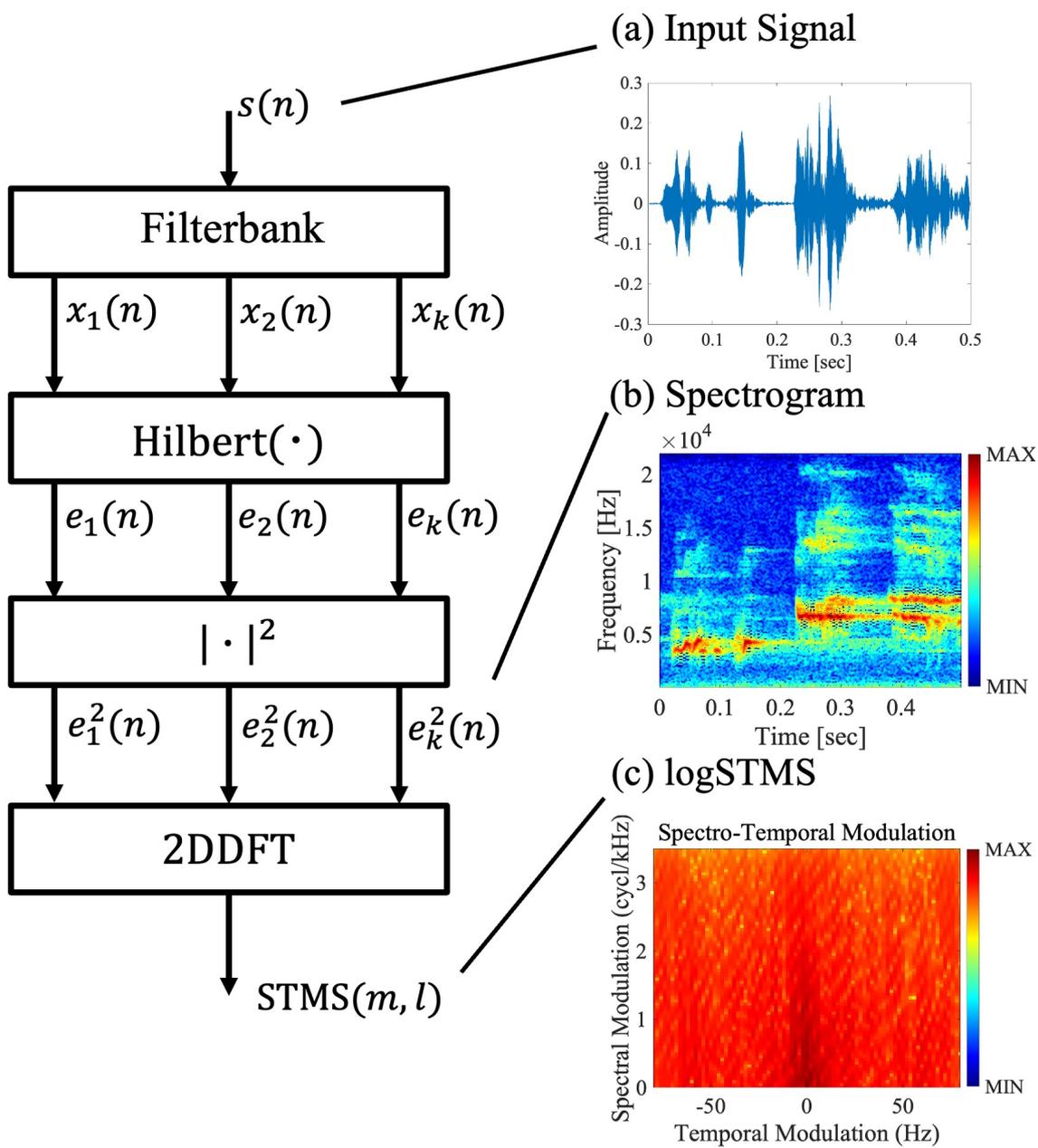


図 3.2: STM 分析のブロック図: (a) 入力信号, (b) パワー包絡線 (パワースペクトログラム), (c) STMS の対数表現

### 3.3 分析条件

本研究では，先行研究 [20] より，図 3.3 に示す自然音や人工音を含む 10 個の刺激を分析対象とした．この刺激は，サーストンスケールにおける相対的な聴覚的顕著性が判明している．

図 3.4 に顕著性スケールを示す．聴覚的顕著性の高さの順番に刺激が並んでおり，刺激長は 500 ms，サンプリング周波数は 44.1 kHz である．木所 [7] の報告より，Moore & Gracebarg のモデル [23] を使用し，刺激のラウドネスレベルの最大値を 65 phon として全刺激で音圧レベルを調整することでラウドネスレベルを統制した．

顯著性低



顯著性高

刺激名	刺激の内容
Tone	1 kHz の純音
Dog	犬の鳴き声
Crying	子供の泣き声
Noise	白色雑音
Laughter	子供の笑い声
Phone	電話の着信音
Scratch	黒板を引っ掻くような音
Beep	基本周波数500 Hz の調波複合音
Bird	鳥のさえずり
Chirp	200 Hz から8 kHz まで線形に 上昇するチャープ信号

図 3.3: 分析対象の音刺激 (文献 [20] より作成)

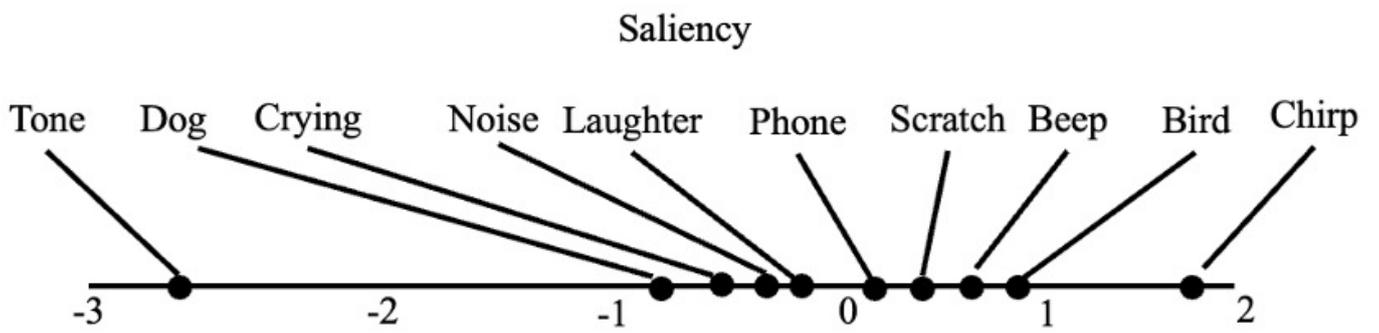


図 3.4: 顕著性スケール (文献 [20] より作成)

### 3.4 高次統計量

先行研究 [7] より，聴覚的顕著性を説明できる手掛かりが STMS に現れることがわかっており，ラフネスと係ることもわかっている [10]. ここまでに述べた STMS は，パワー包絡線を時間変調とスペクトル変調の関数としてスペクトル分析した結果である．そのため，時間変調とスペクトル変調に係る特徴として STMS の分布に着目することで，音質評価指標と聴覚的顕著性の関係が結びつくことが予想できる．

本研究では，STMS の分布に関する高次統計量として重心，分散，歪度，尖度を求めた．

(i) STMS 重心 ( $STM_{ME}$ )

$$STM_{ME} = \frac{1}{LM} \sum_{l=1}^L \sum_{m=1}^M \log STMS(m, l) \quad (3.5)$$

$M$  は TM 方向のデータ長， $L$  は SM 方向のデータ長であり，データ長はそれぞれ  $M = 441, L = 551$  である．STMS 重心は，STMS の平均を示す．

(ii) STMS 分散 ( $STM_{VA}$ )

$$STM_{VA} = v^2 = \frac{1}{LM} \sum_{l=1}^L \sum_{m=1}^M (\log STMS(m, l) - STM_{ME})^2 \quad (3.6)$$

STMS 分散は，分布のばらつきの程度を示す．

(ii) STMS 歪度 ( $STM_{SK}$ )

$$STM_{SK} = \frac{1}{LM} \sum_{l=1}^L \sum_{m=1}^M \frac{(\log STMS(m, l) - STM_{ME})^3}{v^3} \quad (3.7)$$

STMS 歪度は，正規分布に対する分布の歪み具合を示す高次統計量であり，非対称性に係る指標である．重心よりも低い方向に分布が偏る場合に正を，高い方向に偏っている場合に負の値となり，値が 0 となる場合は左右対称の分布となる．

(ii) STMS 尖度 ( $STM_{KU}$ )

$$STM_{KU} = -3+ = \frac{1}{LM} \sum_{l=1}^L \sum_{m=1}^M \frac{(\log STMS(m, l) - STM_{ME})^4}{v^4} \quad (3.8)$$

STMS 尖度は、正規分布に対する分布の先端部の尖りを示す。裾の広がりや尖りの程度を表す指標であり、分布が正規分布よりも尖る場合には正を、分布が扁平な場合に負の値をとる。

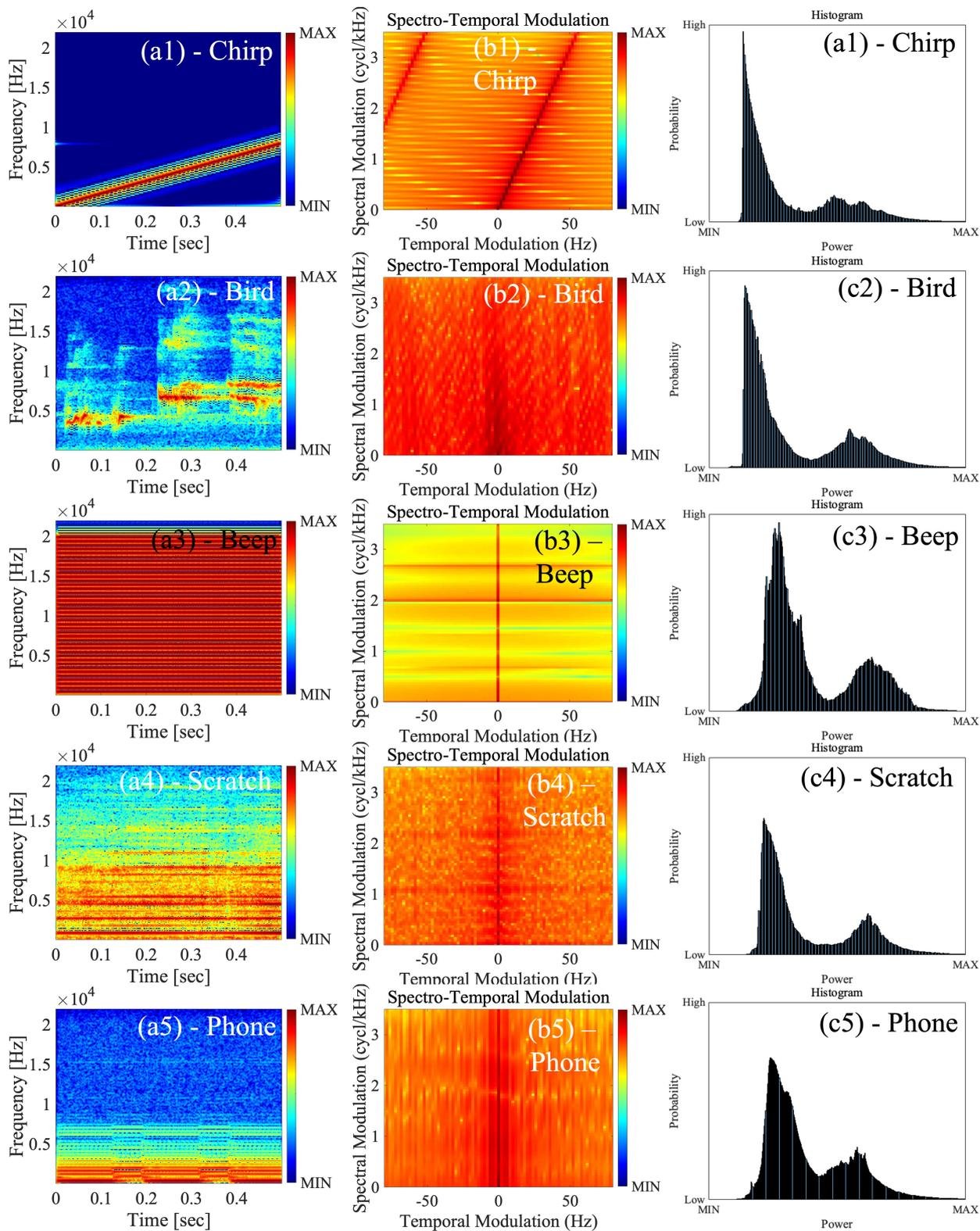
分析では、STMS の分布に対して一つの正規分布を仮定し、高次統計量を算出した。

### 3.4.1 分析結果

図 3.5 に結果を示す。まず、各音刺激に対する分析結果について確認する。(a) は各刺激の対数パワースペクトルである。(a) 横軸時間 (単位: sec) , 縦軸周波数 (単位: Hz) を表す。(b) に STMS を示す。単位と各軸の範囲は図 3.2 と同じである。

図 3.5 (b) より、動物の音声 (Bird, Dog, Crying, Laughter) では、 $\pm 20$  Hz までの時間変調における低いスペクトル変調領域に特徴が集中していた。この結果は、緑色の変調領域 [9] と同じ変調領域に特徴があることを示す。人工音では、Chirp において、時間と共に斜めに変化する STMS が顕著に確認できた。これは、特徴が STM を横断して現れており、音の粗さに係る変調領域を有していた。

図 3.5 (c) は STMS の分布であり、横軸が STMS の値、縦軸がその頻度である。全刺激を通じて STMS の値の小さい Min 方向と大きい Max 方向で二つのピークが確認でき、STMS の値が小さい Min 方向のピークの方が顕著だった。顕著性スケールに着目すると、Tone のように聴覚的顕著性が低い刺激では二つのピークの間隔が広く、聴覚的顕著性の高い刺激になるにつれて、分布における Min 方向の頻度のピークが高く、Max の頻度のピークとの間隔も狭く変化する傾向が確認できた。



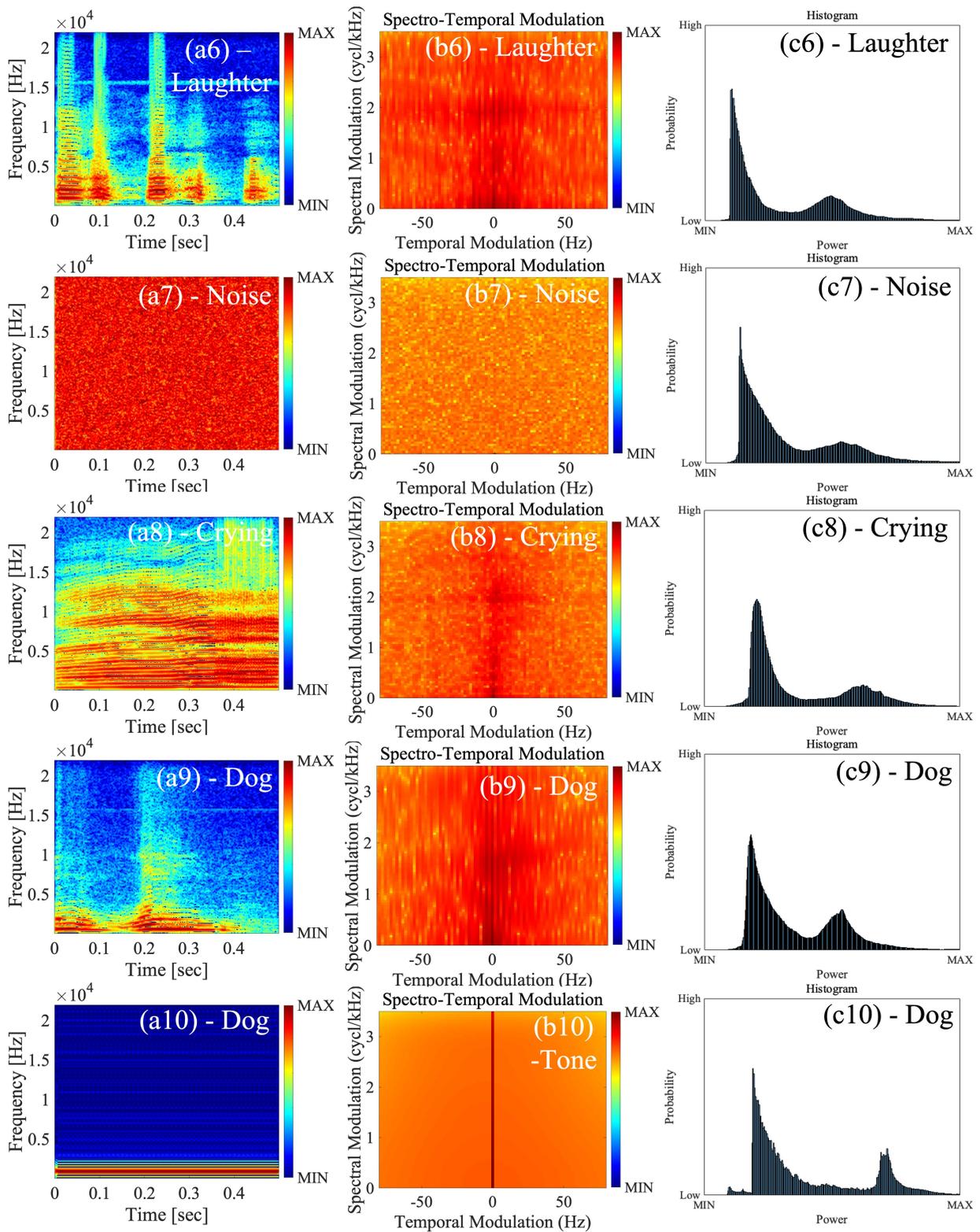


図 3.5: STM 分析の結果: (a1-10): 対数パワースペクトル (対数パワー包絡線), (b1-10): STMS, (c1-10): STMS の分布. 各パネルは顕著性スケールの並びに沿って, 上から下 (1,2,...,10) に向かって聴覚的顕著性が低から高に配置

### 3.4.2 考察

分析結果より、以下のように考察した.

- ・ STMS の歪度に着目した場合

全刺激を通して、歪度が小さい場合、ピークの間隔が離れており、この場合には顕著性が低い傾向があると考えられる。一方で、二つのピークの間隔が狭まる刺激では、顕著性と高い相関となると考えられる。

- ・ STMS の尖度に着目した場合

全刺激を通して、尖度が小さい場合、Min 方向のピークが小さくなり、この場合に顕著性と低い相関となると考えられる。

## 第4章 STMSを介した音質評価指標と聴覚的顕著性の関係の検討

### 4.1 分析方法

前章までに、STMから得られる特徴として、STMSの分布に着目し、高次統計量を算出した。本節では、音質評価指標と聴覚的顕著性の関係を説明するために、STMSの分布における高次統計量を介した検討を行った。STMSの高次統計量と音質評価指標、聴覚的顕著性の関係について、相関を求めることで検討した。音質評価指標では、音刺激の特性を考慮して、シャープネスとラフネスを採用し、計算モデル [50] より求めた。

### 4.2 STMSの高次統計量と音質評価指標の関係

図 4.1~4.8 に各刺激から求めた高次統計量とシャープネス、ラフネスの値の散布図を示す。ここで、横軸が説明変数とした各高次統計量、縦軸が目的変数であるシャープネスとラフネスである。

分析結果より、重心と分散とシャープネス、ラフネスの間には、それぞれ  $\text{Corr} = 0.25$  で  $p < 0.49$ ,  $\text{Corr} = -0.13$  で  $p < 0.73$ ,  $\text{Corr} = 0.42$  で  $p < 0.23$ ,  $\text{Corr} = -0.19$  で  $p < 0.60$  となり、相関は確認できなかった。一方で、歪度・尖度とラフネスにおいて、それぞれ  $\text{Corr} = 0.66$  で  $p < 0.04$ ,  $\text{Corr} = 0.63$  で  $p < 0.05$  となり、有意な相関が確認でき、シャープネスは  $\text{Corr} = 0.46$  で  $p < 0.26$ ,  $\text{Corr} = 0.55$  で  $p < 0.10$  となり、相関は確認できなかった。刺激間での傾向として、聴覚的顕著性が高い（顕著性スケールの並びの半分以上の刺激になる）と、各音質評価指標の値も高かった。

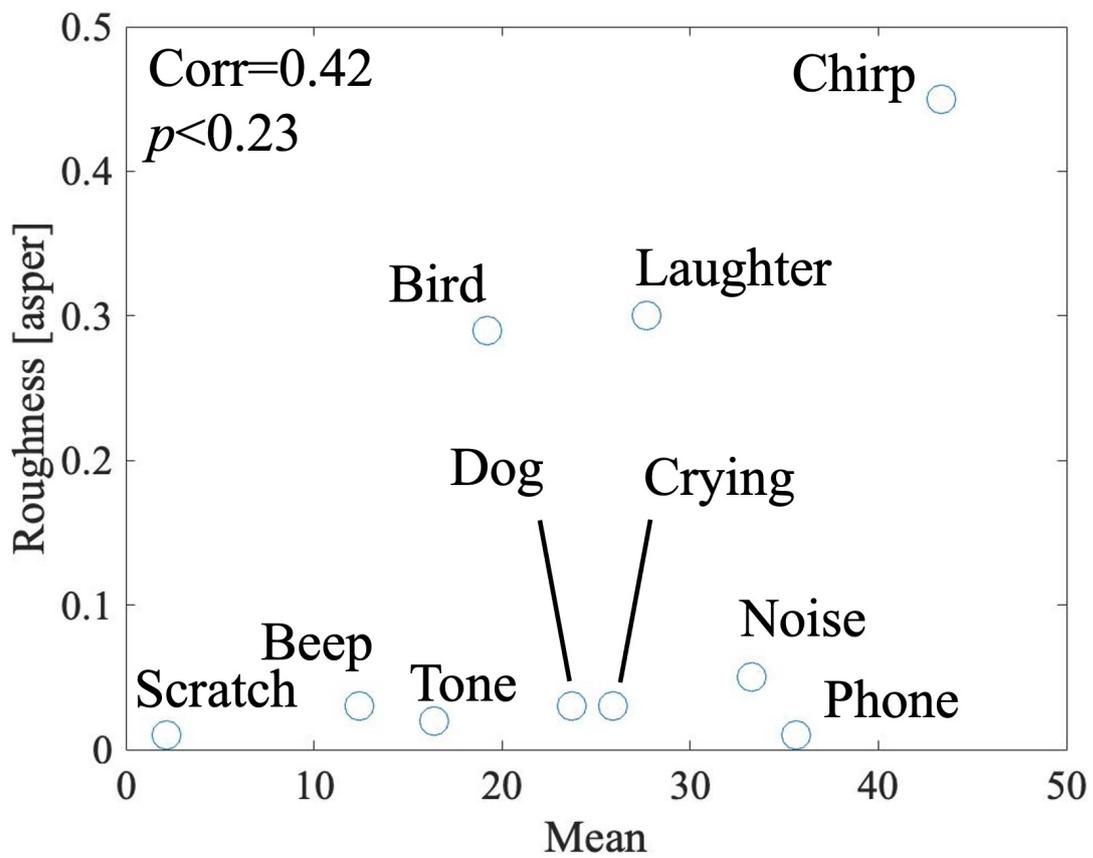


図 4.1: STMS 重心とラフネスの相関

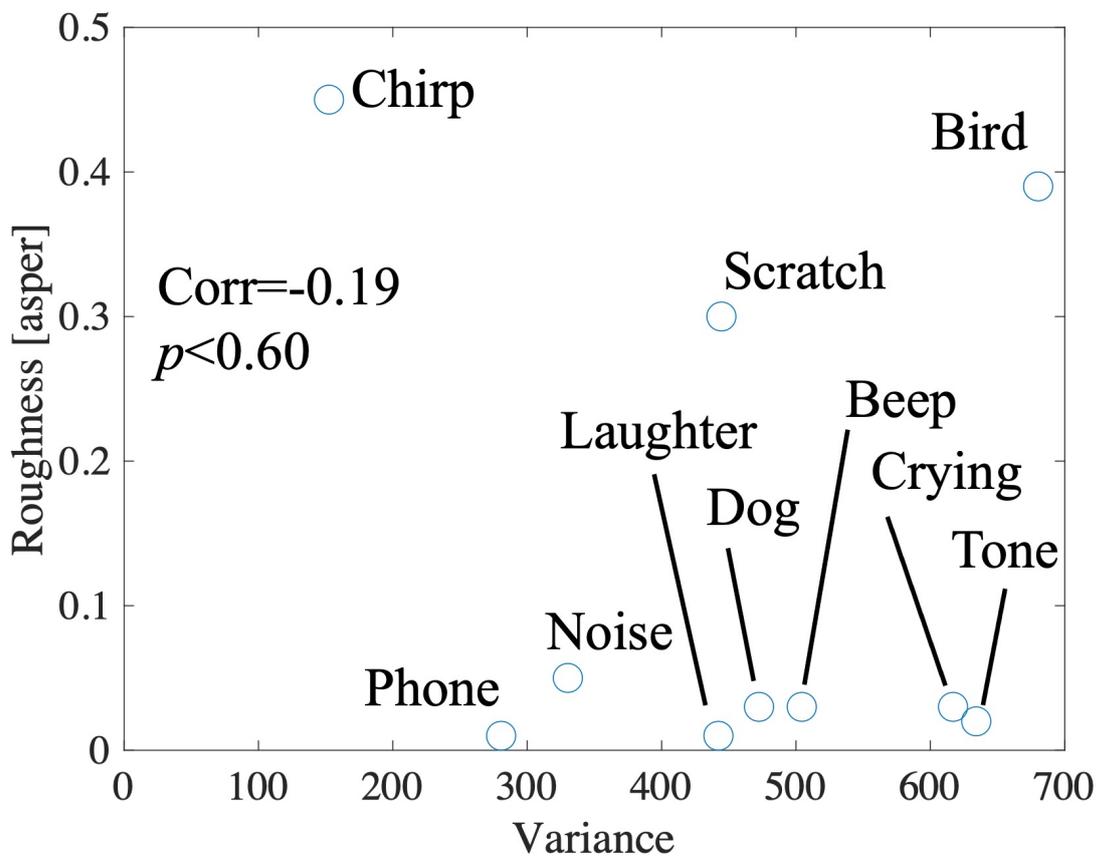


図 4.2: STMS 分散とラフネスの相関

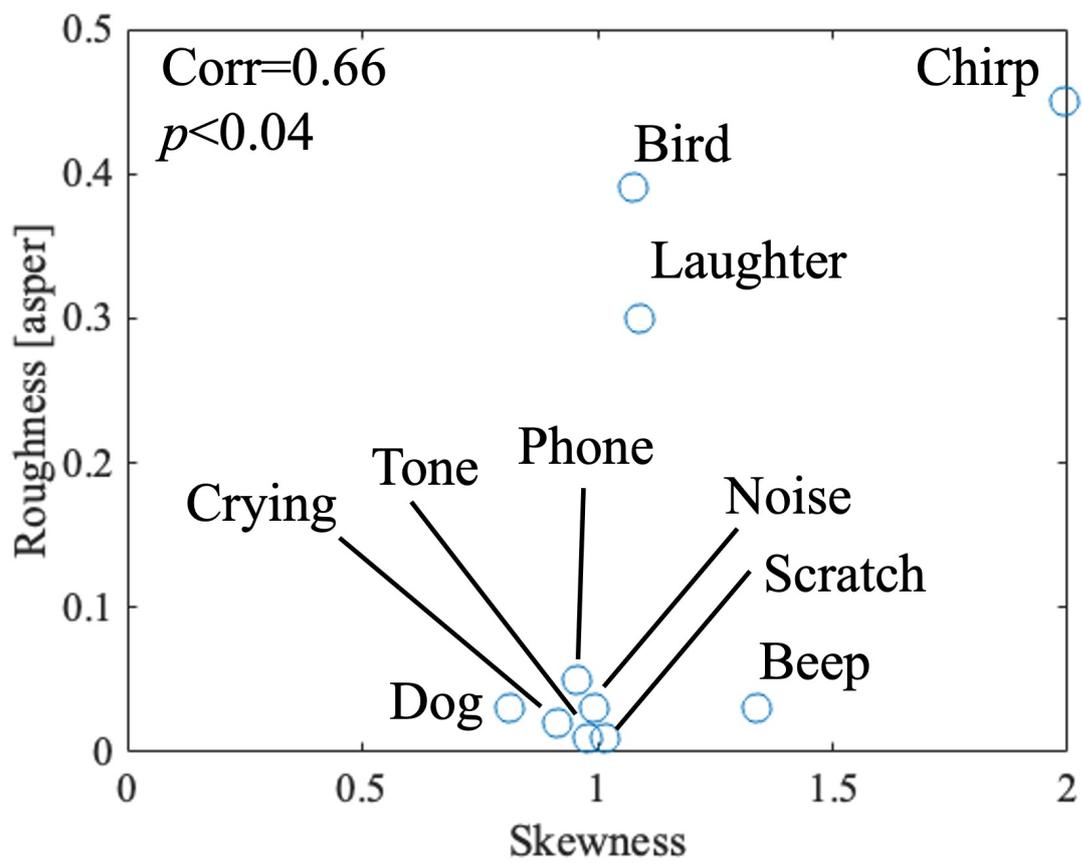


図 4.3: STMS 歪度とラフネスの相関

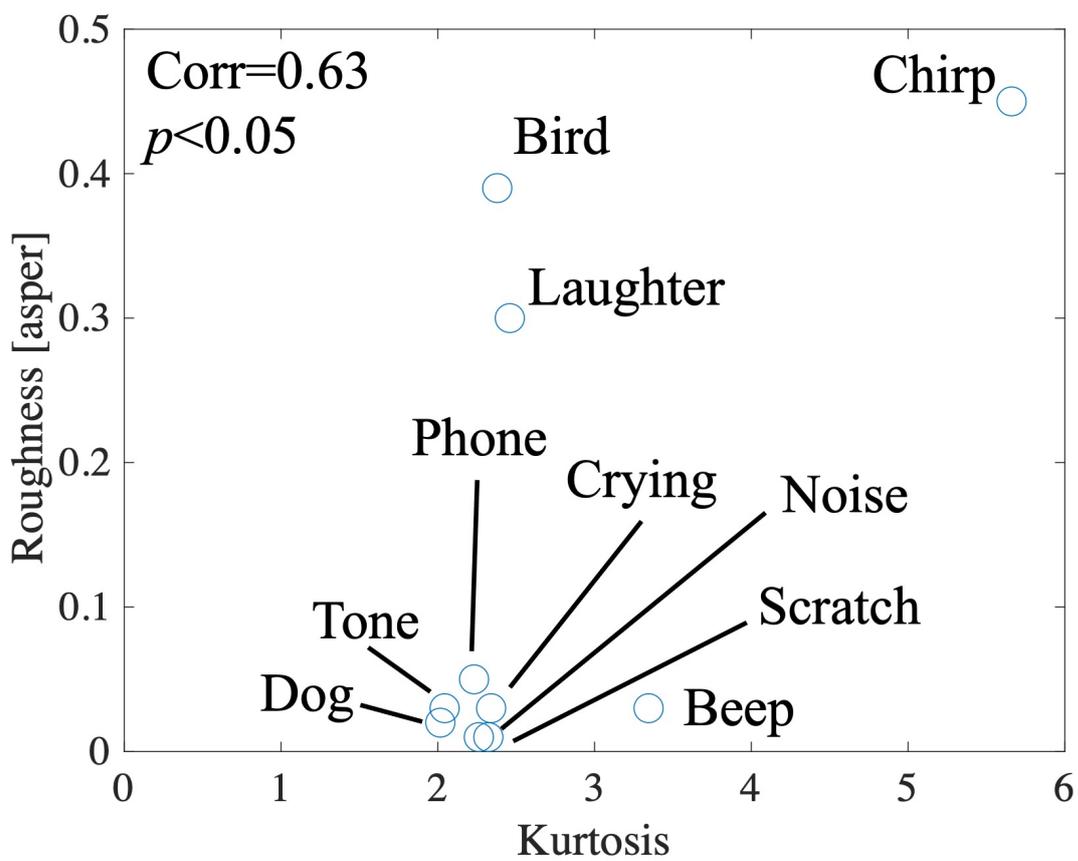


図 4.4: STMS 尖度とラフネスの相関

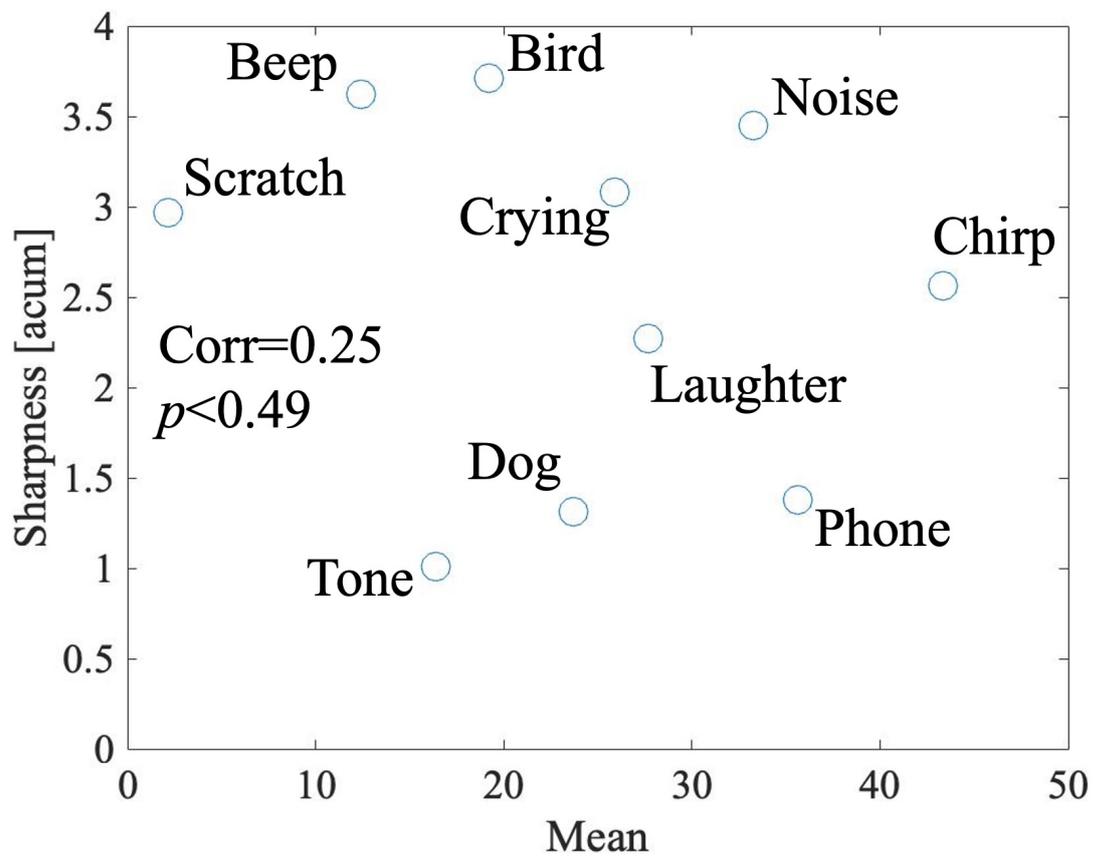


図 4.5: STMS 重心とシャープネスの相関

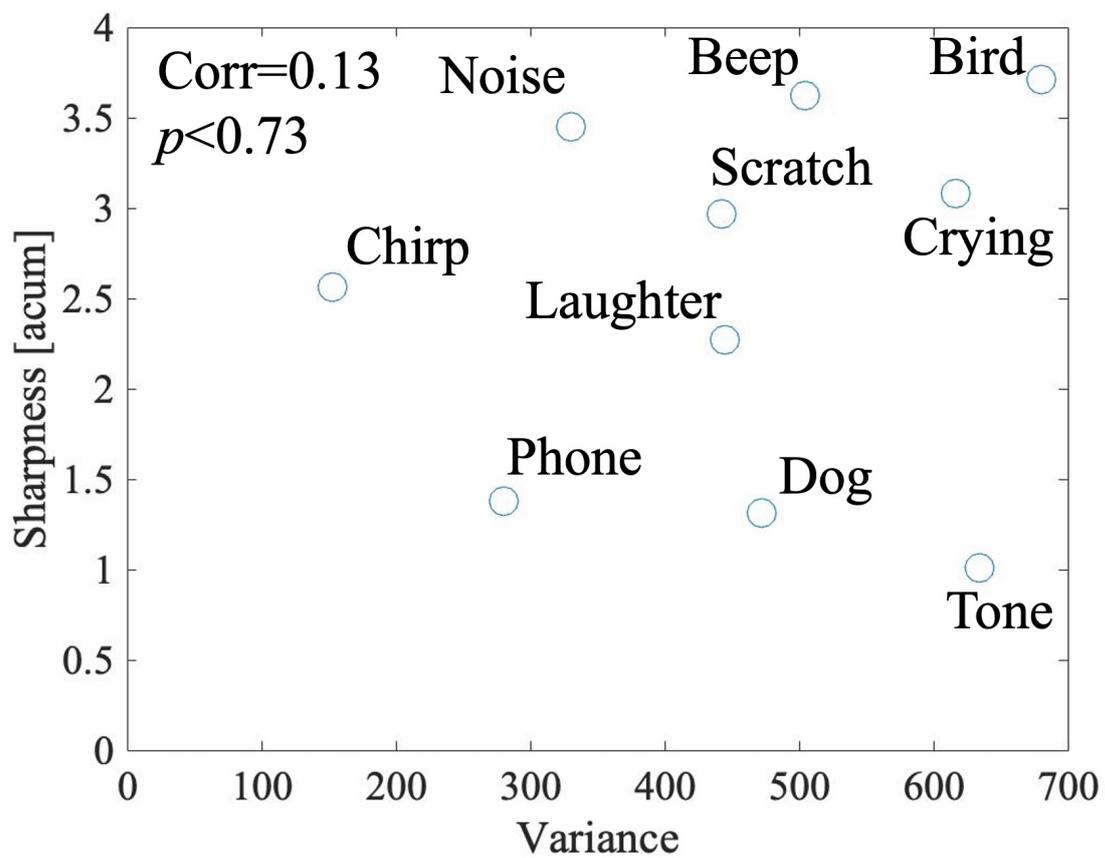


図 4.6: STMS 分散とシャープネスの相関

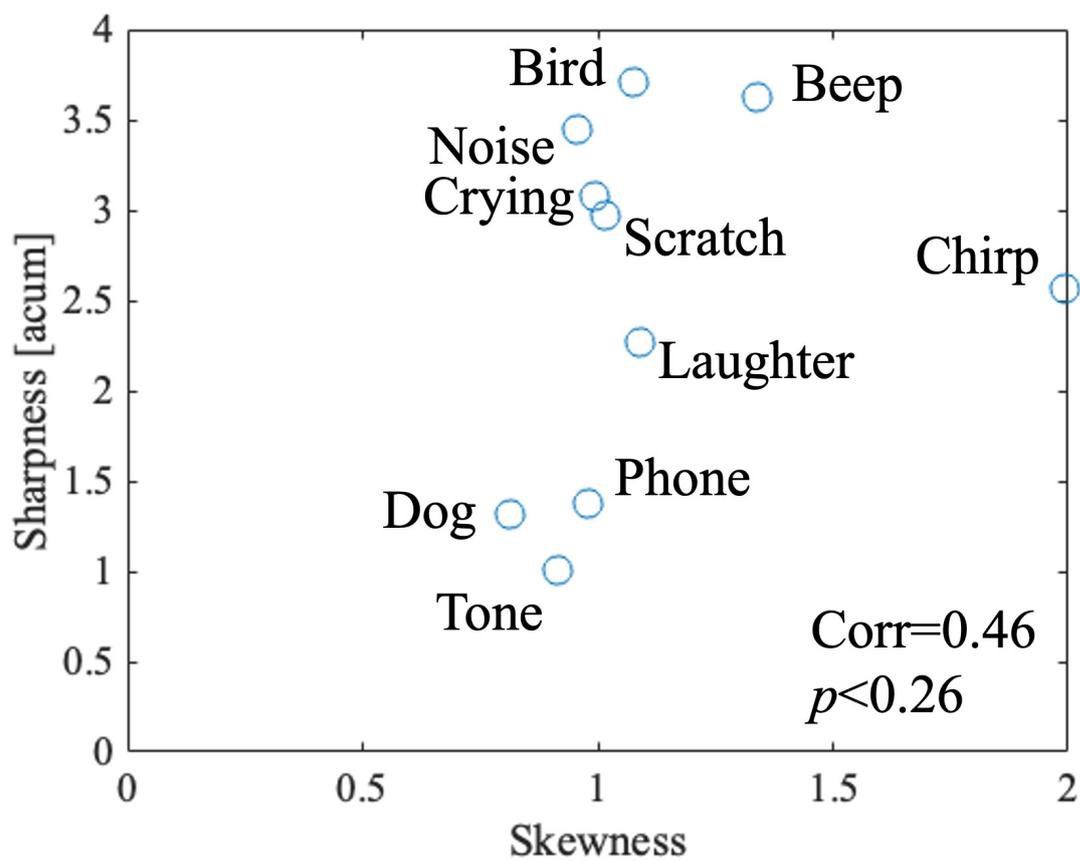


図 4.7: STMS 歪度とシャープネスの相関

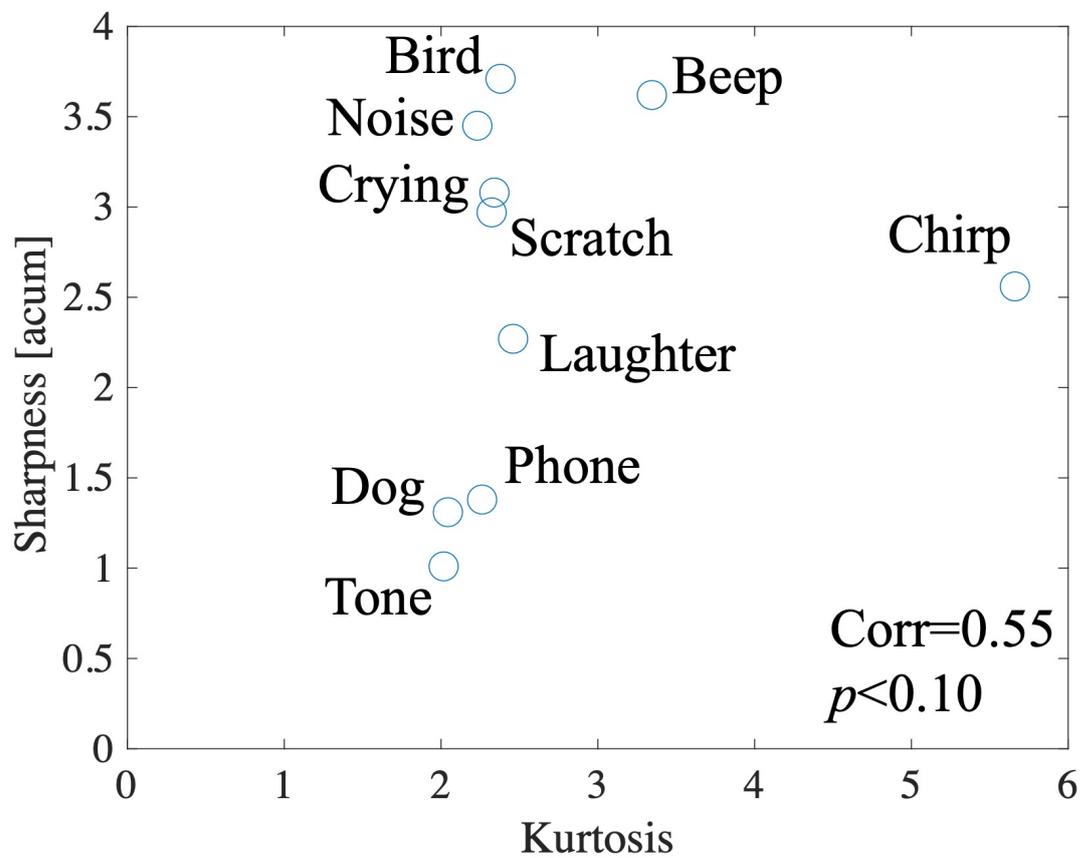


図 4.8: STMS 尖度とシャープネスの相関

### 4.3 STMS の高次統計量と聴覚的顕著性の関係

図 4.9, 4.10, 4.11, 4.12 に, STMS の分布から求めた高次統計量と顕著性スケールの値の散布図を示す. ここで, 横軸が説明変数とした各高次統計量, 縦軸が目的変数である顕著性スケールである. 結果より, 重心と分散と顕著性スケールでは, 重心において  $\text{Corr} = 0.24$  で  $p < 0.51$ , 分散で  $\text{Corr} = -0.52$  で  $p < 0.13$  となり, 相関は確認できなかったが, 分散では Chirp では分散が小さく, 聴覚的顕著性の最も低い Tone では分散は最も大きいという傾向があった. 一方で, 歪度において  $\text{Corr} = 0.75$  で  $p < 0.01$ , 尖度では  $\text{Corr} = 0.70$  で  $p < 0.02$  と有意な相関が確認できた.

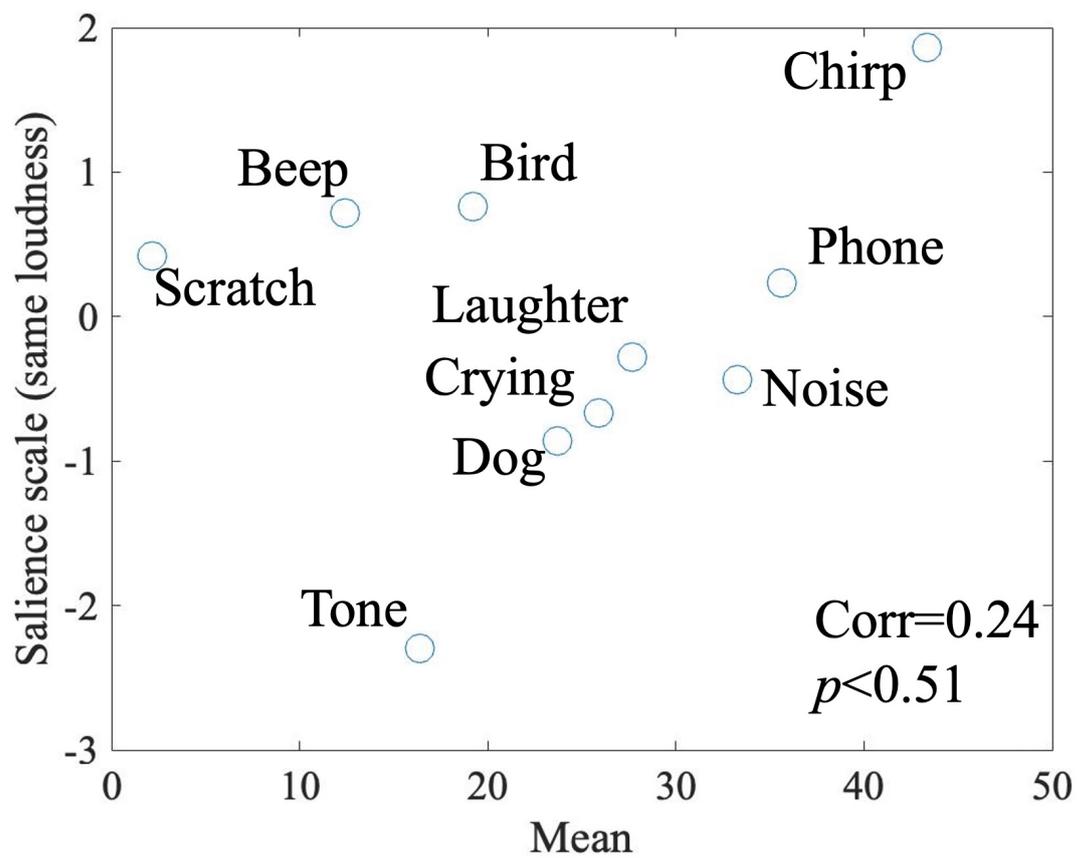


図 4.9: STMS 重心と顕著性スケールの相関

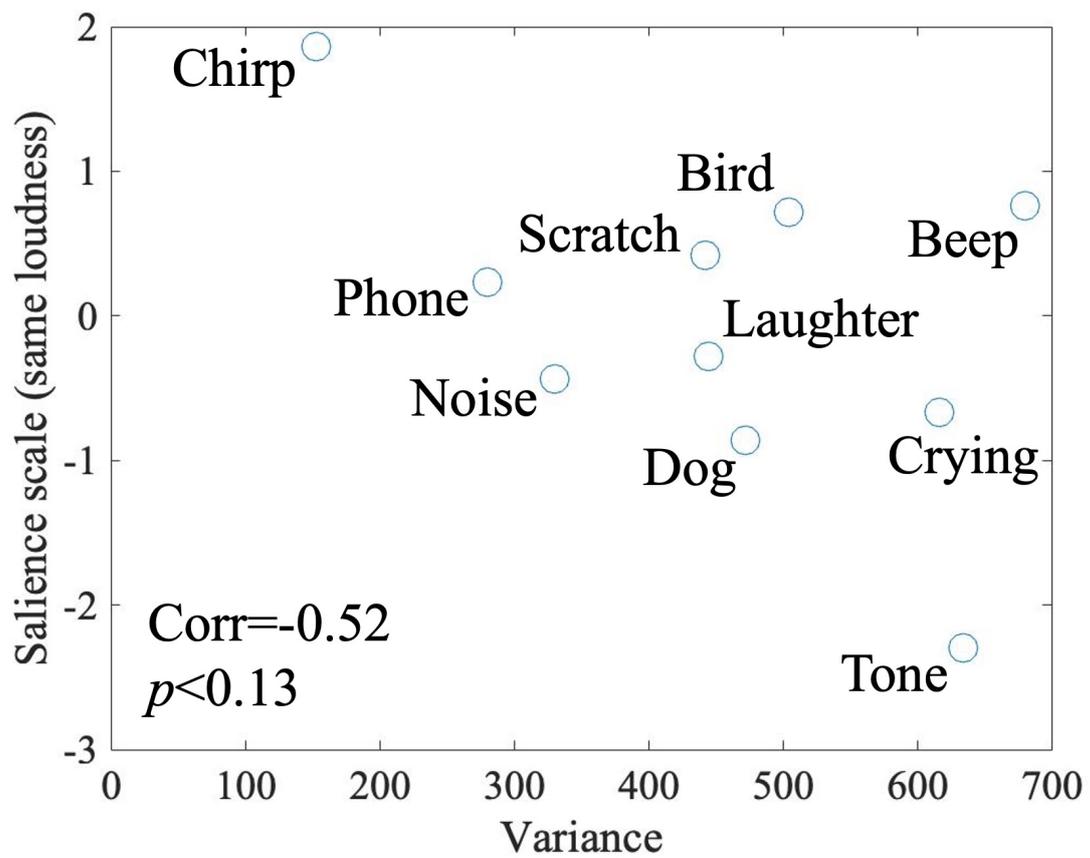


図 4.10: STMS 分散と顕著性スケールの相関

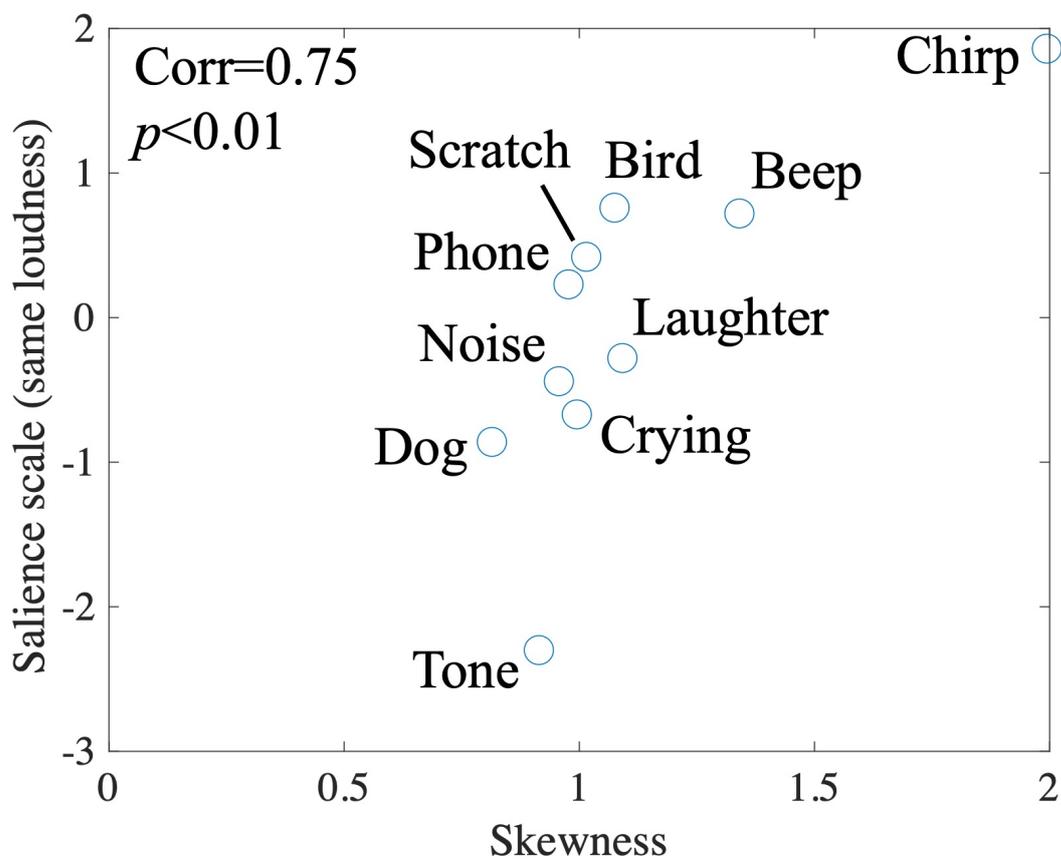


図 4.11: STMS 歪度と顕著性スケールの相関

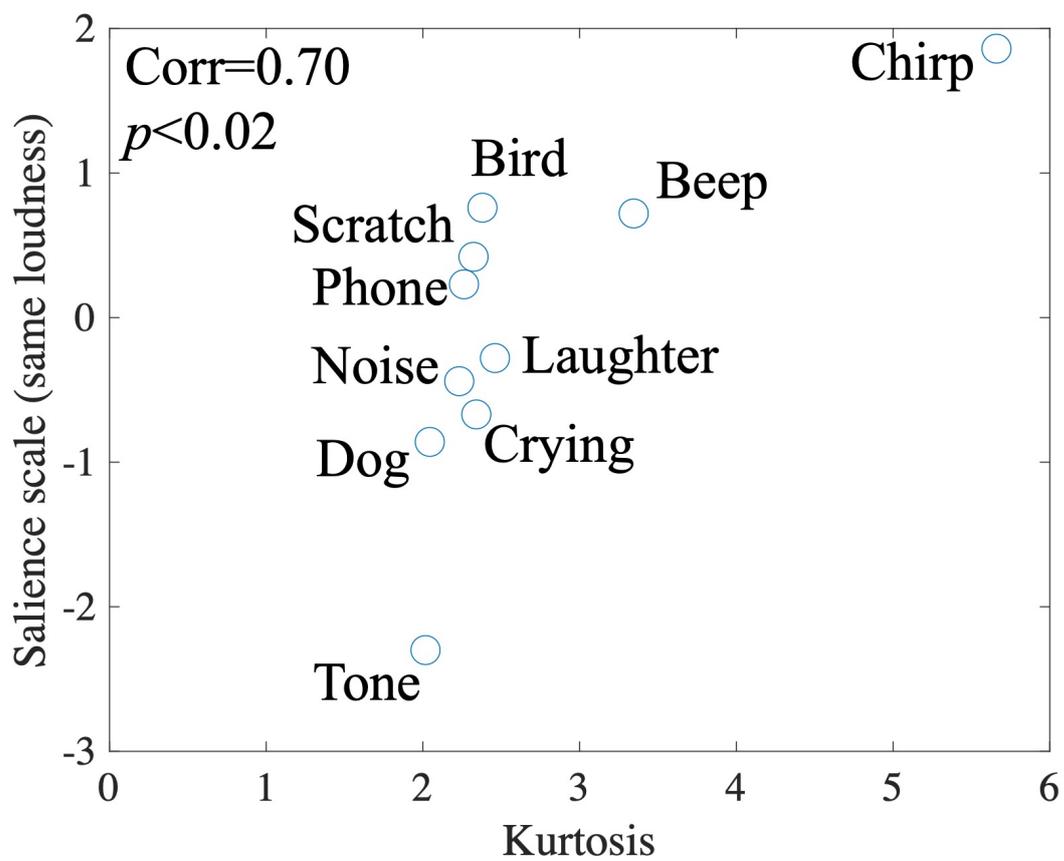


図 4.12: STMS 尖度と顕著性スケールの相関

## 4.4 音質評価指標と聴覚的顕著性の関係

表4.1に計算モデルで算出した各音質評価指標の値を示し，図4.13, 4.14に，シャープネス，ラフネスと顕著性スケールの相関を示す．ここで，横軸が説明変数とした各音質評価指標の値，縦軸が目的変数である顕著性スケールである．結果より，音質評価指標と聴覚的顕著性の間には，シャープネスで  $\text{Corr} = 0.56$  で  $p < 0.11$ ，ラフネスで  $\text{Corr} = 0.58$  で  $p < 0.06$  と，有意傾向だった．

次に，この二つの指標と聴覚的顕著性の関係について，説明変数にシャープネス，ラフネスを，従属変数を聴覚的顕著性として重回帰分析し，表4.2, 4.3および以下に示す式を求めた．ここで，回帰式中  $W_R$  はラフネスの係数， $W_S$  はシャープネスの係数である．結果より，単回帰分析と同様に相関はなく，有意傾向に留まった．また，顕著性が高い刺激では，ラフネスも高いという刺激間での傾向があった．

$$\text{聴覚的顕著性} = -1.825968346 + 3.740328523 \times W_R + 0.532749034 \times W_S \quad (4.1)$$

表 4.1: 計算モデルより算出したシャープネスとラフネスの値

	Tone	Dog	Crying	Noise	Laughter	Phone	Scratch	Beep	Bird	Chirp
シャープネス [accum]	1.01	1.31	3.08	3.45	2.27	1.38	2.97	3.62	3.71	2.56
ラフネス [asper]	0.02	0.03	0.03	0.05	0.30	0.01	0.01	0.03	0.39	0.45

表 4.2: 回帰統計

重相関	0.73432926
重決定	0.53923946
標準誤差	0.87018944
観測数	10
有意 F	0.06639929

表 4.3: 重回帰分析

	回帰係数	標準誤差	t	$p$ 値
聴覚的顕著性	-1.8259683	0.79058654	-2.3096375	0.05421574
$W_R$	3.74032852	1.91895977	1.9491438	0.09228386
$W_S$	0.53274903	0.29058963	1.8333381	0.10940861

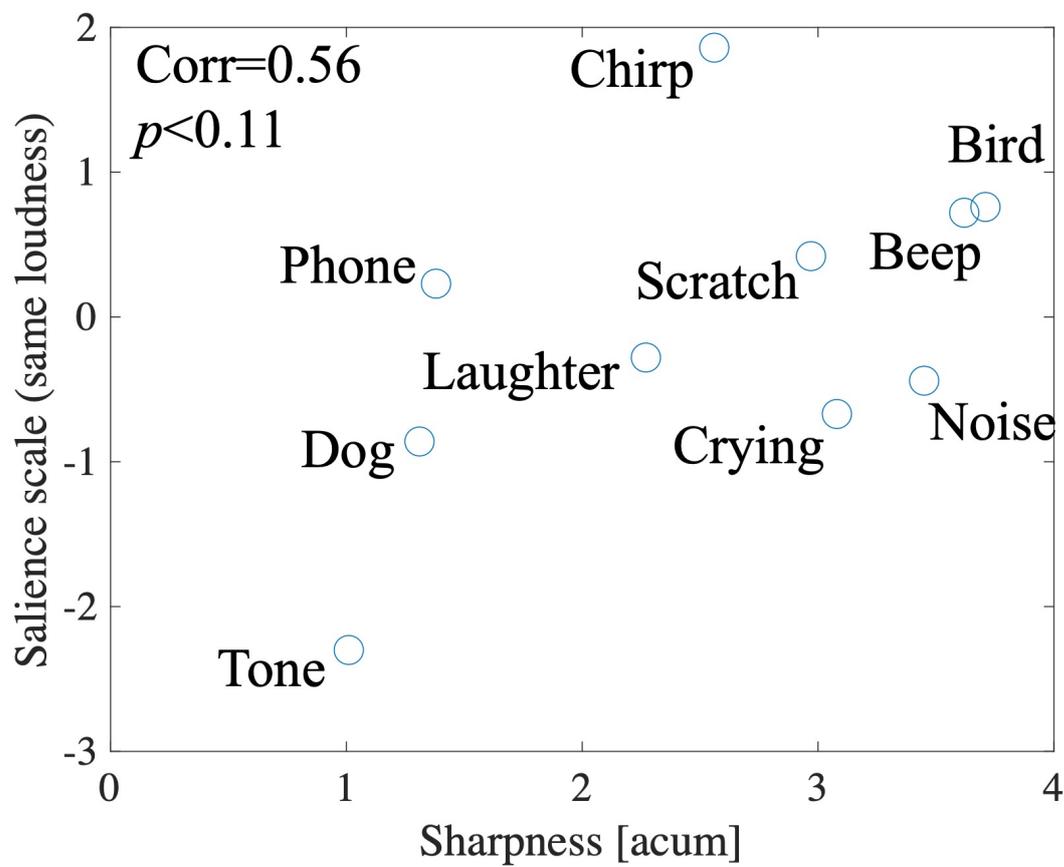


図 4.13: シャープネスと顕著性スケールの相関

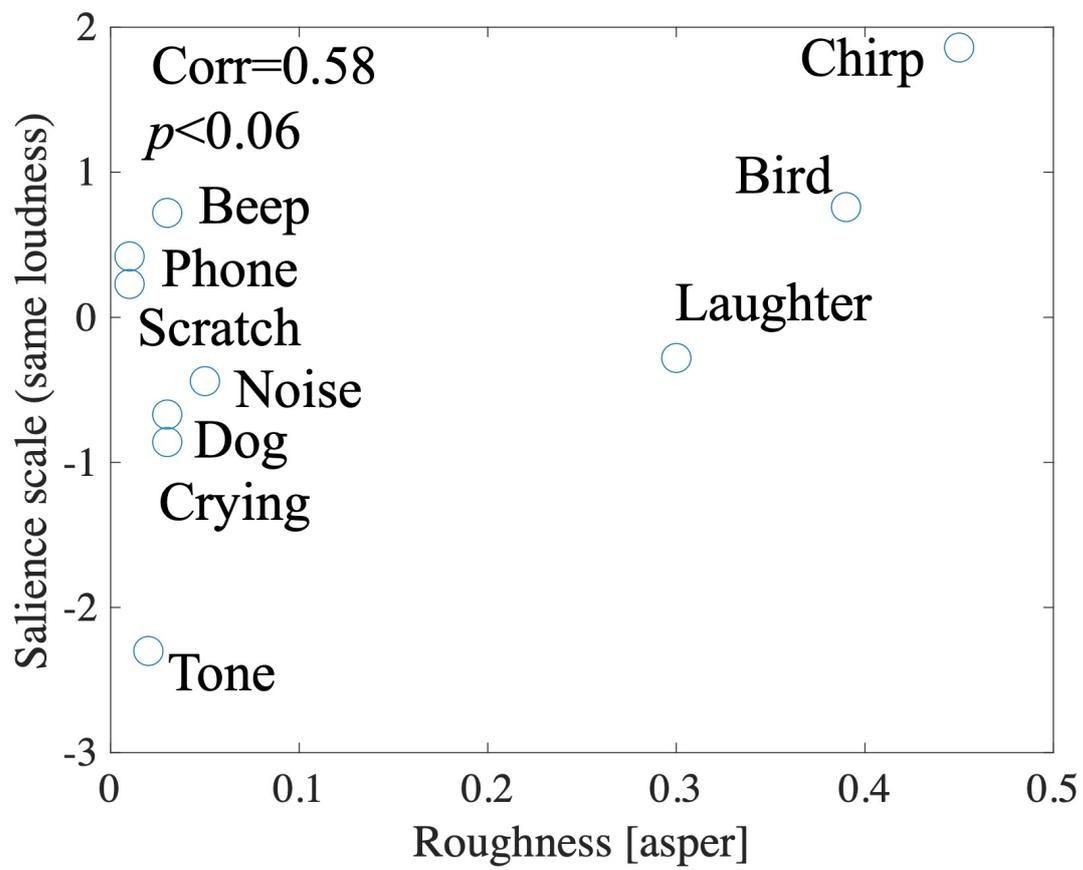


図 4.14: ラフネスと顕著性スケールの相関

## 4.5 考察

STM から得られる特徴と音質評価指標の関係について、歪度・尖度とラフネスで高い相関があった。しかし、シャープネスにおいて相関はなかった。提案法において、各変調領域の情報を統合的に分析したが、各音質評価指標に係る変調領域での分布を求めることで、説明性が上がると考えられる。また、音質評価指標と聴覚的顕著性で有意傾向を示した結果と STMS の分布における高次統計量と音質評価指標の相関結果からも、他の説明する要因を増やすことで、音質評価指標のみでも聴覚的顕著性を説明できることが考えられる。

## 第5章 全体考察

本研究では、聴覚的顕著性が既知の10個の刺激をSTM分析し、求めたSTMSの分布に対して高次統計量を求め、高次統計量と音質評価指標、聴覚的顕著性の相関を求めた。結果より、歪度・尖度と聴覚的顕著性と相関が高く、音質評価指標の中では最もラフネスと相関が高かった。また、分布における二つのピークの間隔が狭い刺激の方が聴覚的顕著性が高い傾向となった。

STM分析の有効性を確認するために、STMSの算出の基になるパワースペクトログラムに関するパワースペクトル重心 ( $\text{Spec}_{\text{ME}}$ )、パワースペクトル分散 ( $\text{Spec}_{\text{VA}}$ )、パワースペクトル歪度 ( $\text{Spec}_{\text{SK}}$ )、パワースペクトル尖度 ( $\text{Spec}_{\text{KU}}$ ) に対しても分析した。結果を図5.1, 5.2, 5.3, 5.4, 5.5に示す。結果より、対数パワースペクトルの値が大きい方向よりも対数パワースペクトルの値が小さい方向に歪んだ分布となる傾向があった。顕著性スケールとラフネスの関係について、相関および聴覚的顕著性の高さと同数パワースペクトルの分布において規則的な変化傾向はなかった。また、歪度・尖度と顕著性スケールでも相関はなかった。これらの結果より、STM分析の有効性および、STMSの分布における高次統計量を介して、音質評価指標と聴覚的顕著性を結び付けられると考えられる。

対数パワースペクトルの分析と比較して、STMSを分析した場合のほうがラフネス、顕著性スケールと相関が高くなった理由について、対数パワースペクトルとSTMSで表現される特徴（時間変調・スペクトル変調）の違いが係ることが考えられる。対数パワースペクトログラムで表現されるものは、音の強度および周波数特性であるのに対して、STMSはそれよりも高次のヒトの変調知覚における神経発火に係るスペクトル・時間受容野 (Spectro-temporal receptive field: STRF) と対応が良い。そのため、聴覚的顕著性には変調知覚が係っており、変調知覚に係るニューロンの反応と対応が良いSTMSの高次統計量と聴覚的顕著性とで相関が高くなったと考えられる。

また、STMSにおける分布を確認すると、全ての刺激で変調成分が多かった。Chirpが最も顕著性スケールの値が高く、歪度・尖度と顕著性スケールに高い正の相関が、ラフネスにおいて有意傾向を示した。先行研究 [7] でも、STMSが時間と共に変化する刺激で顕著性が高いことがわかっている。一方で、先行研究 [11] と同じ変調領域の特徴が顕著ではない刺激においても聴覚的顕著性が高かった。そのため、多様な変調成分を有する刺激よりも、STMSにおける変調成分が強く現れる刺激の方が聴覚的顕著性が高いことが考えられる。対数パワースペクトルの分布を確認すると、刺激間においてSTMSのような傾向はなかった。よって、前

述の考察と関連付けると、STMS の高次統計量を分析した場合のほうがラフネス、顕著性スケール共に相関が高かったと考えられる。

そのため、STMS の分布における高次統計量を特徴として捉えることで、音質評価指標と聴覚的顕著性の関係を説明できる。さらには、これまでにどういう音が顕著かは明らかでなく、顕著性スケールが既知の刺激が、あらゆる音の中でどこに位置しているのかもわかっていなかった。しかし、分析結果から、あらゆる刺激に対してラフネスの高さが聴覚的顕著性に寄与することが明らかとなり、ヒトの知覚に基づく指標である音質評価指標と聴覚的顕著性の関係が説明できると考えられる。

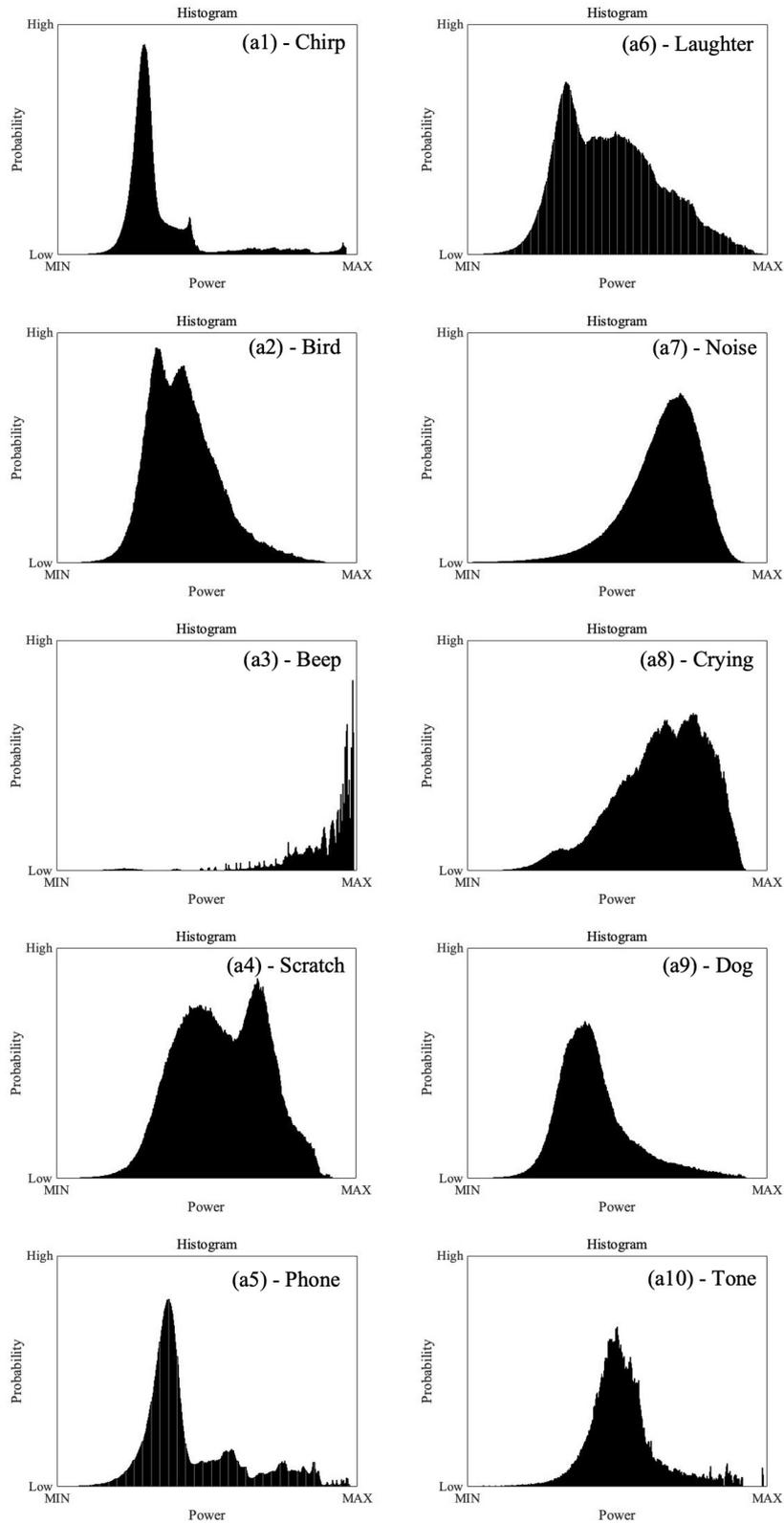


図 5.1: 対数パワースペクトルの分布. 各パネルは顕著性スケールの並びに沿って配置されており, 上から下 (1,2,...,10) に向かって聴覚的顕著性が低から高に付置されている

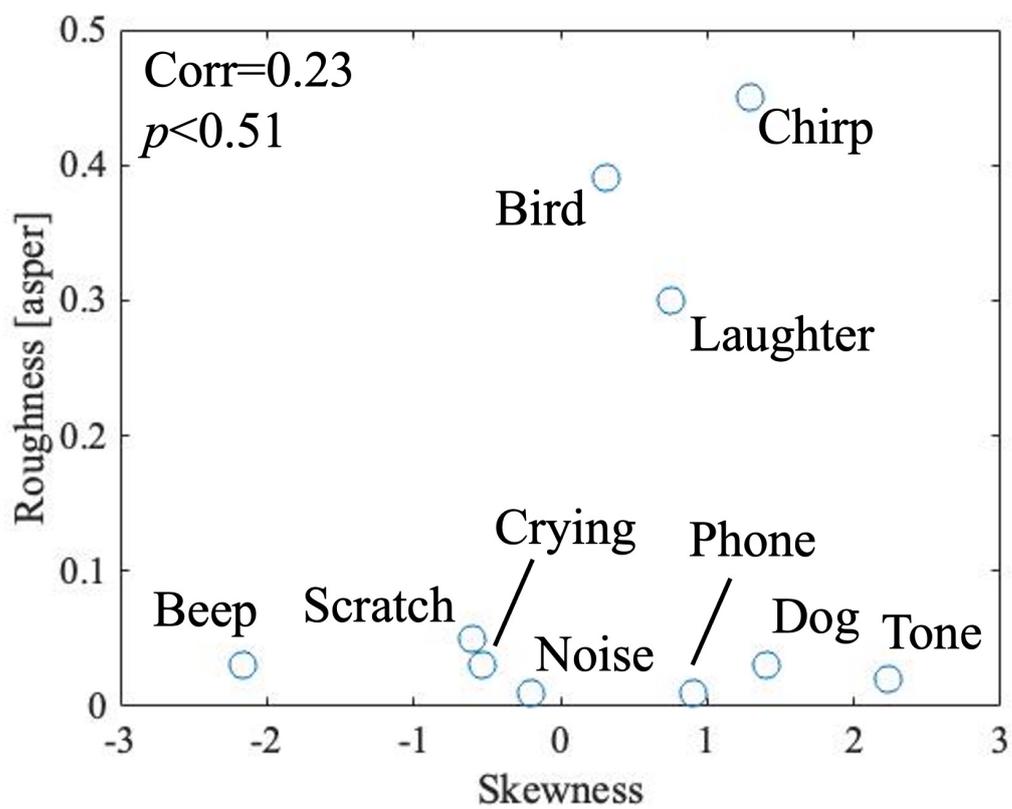


図 5.2: パワースペクトル歪度とラフネスの相関

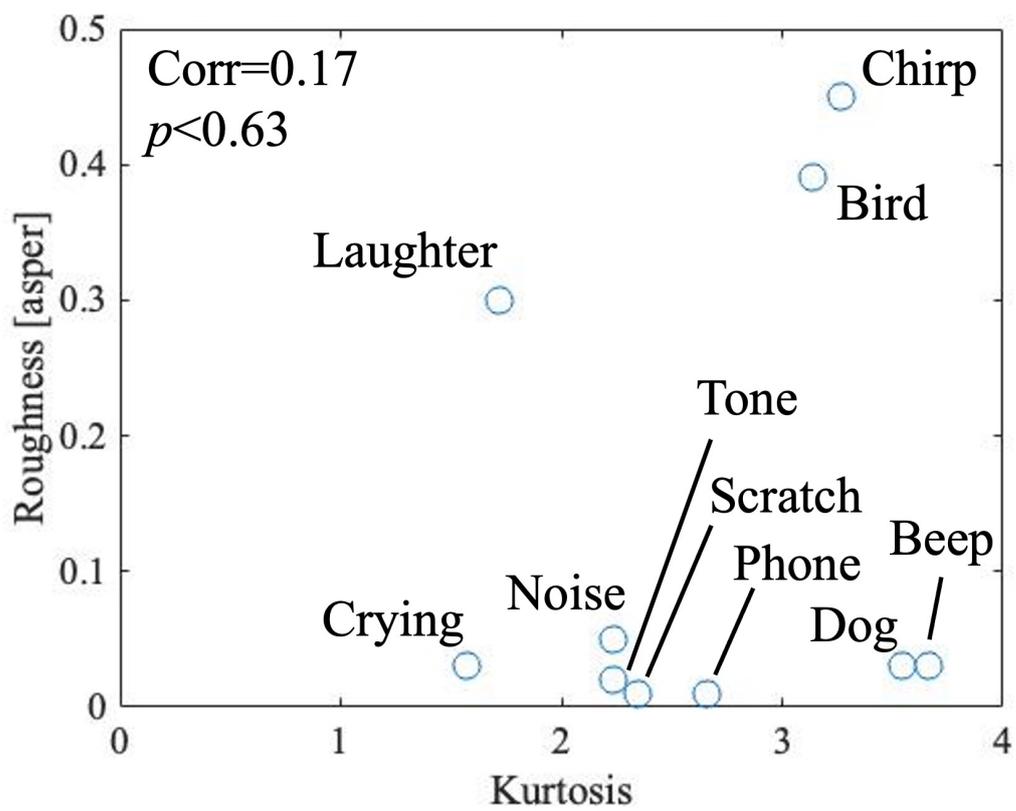


図 5.3: パワースペクトル尖度とラフネスの相関

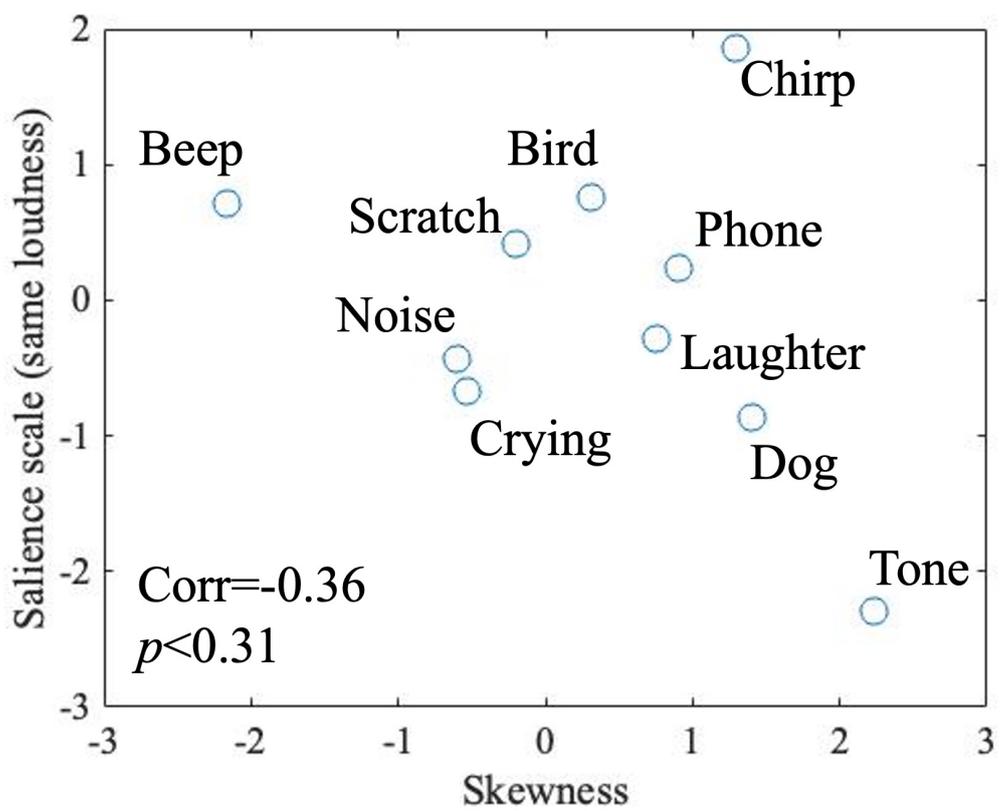


図 5.4: パワースペクトル歪度と顕著性スケールの相関

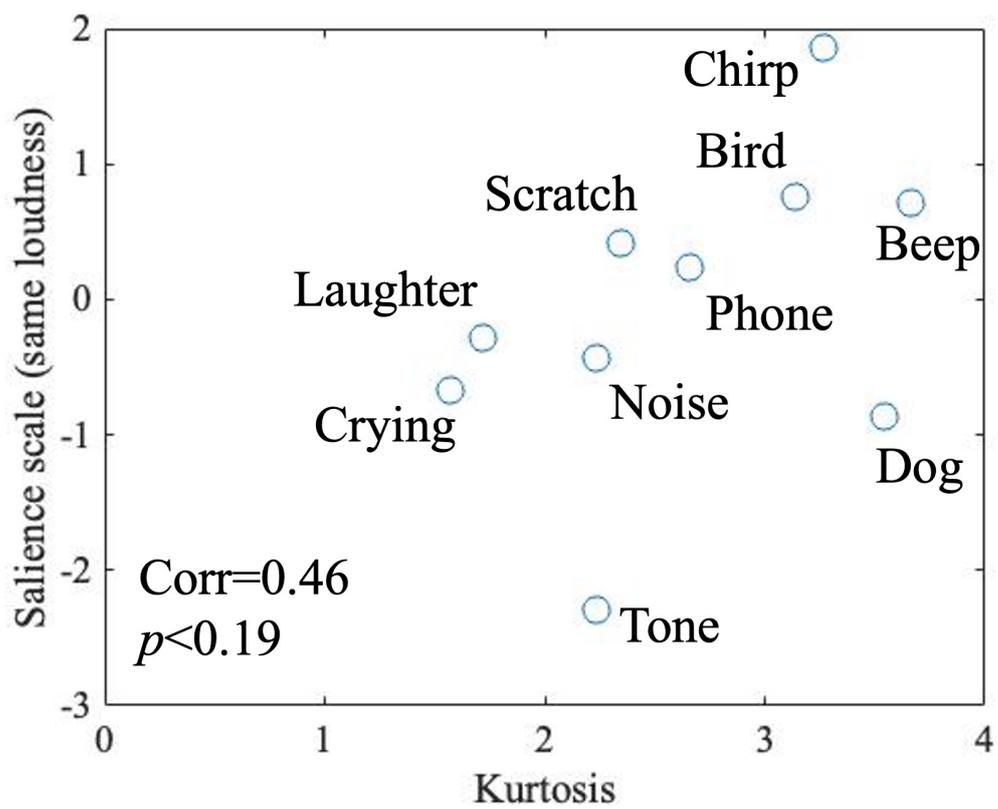


図 5.5: パワースペクトル尖度と顕著性スケールの相関

## 第6章 結論

### 6.1 本研究で明らかにしたこと

本研究では、Spectro-Temporal Modulation (STM) 分析で得られる STM スペクトル (STMS) に現れる特徴を介して、音質評価指標と聴覚的顕著性の関係を明らかにすることを目的とした。そのため、まず聴覚的顕著性が判明しており、刺激のラウドネスを揃えた 10 個の刺激を STM 分析した。そして、STMS の分布における高次統計量 (重心, 分散, 歪度, 尖度) を求め、音質評価指標と顕著性スケールの相関を求めることで、音質評価指標と聴覚的顕著性の関係を検討した。その結果、以下のことが明らかとなった。

- ・ STMS の歪度・尖度は、音質評価指標の一つであるラフネスと有意な正の相関がある。
- ・ STMS の歪度・尖度は、聴覚的顕著性と有意な正の相関がある。
- ・ 音質評価指標のシャープネスおよびラフネスと聴覚的顕著性は有意傾向がある。

以上のことより、STMS の歪度と尖度が聴覚的顕著性と高い相関をもち、ラフネスとも相関が高いことがわかった。以上の結果を関連付けると、STMS の歪度と尖度をもって音質評価指標のラフネスと聴覚的顕著性を結び付けられることが示された。

### 6.2 残された課題

- ・ 音色属性およびその算出に係る音響特徴も含めた検討

本研究では、STMS に現れる特徴を手がかりに音色に係る音質評価指標と聴覚的顕著性の関係を結び付けられると考え、シャープネス・ラフネスと聴覚的顕著性の関係を探った。その結果、シャープネス・ラフネスと聴覚的顕著性では有意傾向に留まったが、説明変数を増やすことで、相関が高くなることを見込まれた。そのため、音色に係る指標として変動強度や音色属性 [51] およびそれらの算出に係る音響特徴も含めた検討により、ラフネス以外の指標と聴覚的顕著性の関係を結び付けられると考えられる。

- ・ STMS の分析範囲の検討

本研究では，STMS 全体に対する高次統計量を求めた．STMS には，音の違いによって異なる特徴が現れることがわかっている [11] が，本研究ではどの変調領域に直接聴覚的顕著性が係るかわかっていない．よって，今後はどの変調領域が聴覚的顕著性に直接寄与するか，高次統計量の分析範囲を絞り込む必要がある．

## 参考文献

- [1] 平原達也, “音声工学と聴覚科学 –機械の方略と人間の方略–,” 日本音響学会誌, Vol. 53, No. 9, pp. 714–719, 1997.
- [2] 赤木正人, “カクテルパーティ効果とそのモデル化,” 電子情報通信学会誌, Vol. 78, No. 5, pp. 450–453, 1995.
- [3] J. Droll and M. Eckstein, “Gaze control and memory for objects while walking in a real world environment,” *Visual Cognition*, Vol. 17, No. 7, pp. 1159–1184, 2009.
- [4] K. Emine and E. Mounya, “Modeling auditory attention: A review,” *PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B BIOLOGICAL SCIENCES*, Vol. 372, No. 1714, pp. 1–10, 2017.
- [5] N. Huang and M. Elhilali, “Auditory salience using natural soundscapes,” *The Journal of the Acoustical Society of America*, Vol. 141, pp. 2163–2176, 2017.
- [6] 木所晃利, “聴覚的顕著性とスペクトル・時間変調情報の関係,” 北陸先端科学技術大学院大学修士論文, 2021.
- [7] 木所晃利, 木谷俊介, 鶴木祐史, “聴覚的顕著性に寄与する Spectro-Temporal Modulation 情報の検討,” 日本音響学会誌, 1-4-6, 2021.
- [8] 古川茂人, “聴覚的顕著性に基づく, 音響情の選択的協調技術の創出,” *SCOPE*, 2015.
- [9] 岩宮眞一郎, “音質評価指標-入門とその応用-,” 日本音響学会誌, Vol. 66, No. 12, pp. 603–609, 2010.
- [10] 大串健吾, “音響聴覚心理学,” 誠信書房, 2019.
- [11] A. Luc, F. Adeen, K. Andreas, G. Anne and P. David., “Human screams occupy a privileged niche in the communication soundscape,” *Current Biology*, Vol. 25, No. 15, pp. 2051–2056, 2015.

- [12] A. Zlatintsi, P. Koutras, G. Evangelopoulos, N. Malandrakis, N. Efthymiou, K. Pastra, A. Potamianos and P. Maragos., “COGNIMUSE: a multimodal video database annotated with saliency, events, semantics and emotion with application to summarization,” *EURASIP Journal on Image and Video Processing*, Vol. 2017, No. 1, pp. 54, 2017.
- [13] C. Kayser, C. Petkov, M. Lippert and N. Logothetis., “Mechanisms for allocating auditory attention: an auditory saliency map,” *Current Biology*, Vol. 15, No. 21, pp. 1943–1947, 2005.
- [14] A. Borji and L. Itti, “State of the art in visual attention modeling,” *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE*, Vol. 35, No. 1, pp. 185–207, 2013.
- [15] F. Tordini, A. Bredman, J. Cooperstock, A. Ankolekar and T. Sandholm., “Toward an improved model of auditory saliency,” *International Conference on Auditory Display*, Vol. 6, No. 10, pp. 189–196, 2013.
- [16] F. Tordini, A. Bregman and A. Cooperstock, “The loud bird doesn ’ t (always) get the worm: why computational salience also needs brightness and tempo,” *International Conference on Auditory Display*, Graz Austria, Vol. 8, No. 10, pp. 236–243, 2015.
- [17] E. Kaya and M. Elhilali, “A temporal saliency map for modeling auditory attention,” *46th Annual Conference on Information Sciences and Systems*, Vol. 1, No. 6, 3146–3152, 2012.
- [18] E. Kaya and M. Elhilali, “Investigating bottom-up auditory attention,” *Frontiers in Human Neuroscience*, Vol. 8, No. 327, pp. 1–12, 2014.
- [19] C. Taishih, R. Powen and S. Shihab, “Multiresolution spectrotemporal analysis of complex sounds,” *J Acoust Soc Am*, Vol. 118, No. 10, pp. 887–906, 2005.
- [20] H. Liao, S. Kidani, M. Yoneya, M. Kashino and S. Furukawa., “Correspondences among pupillary dilation response, subjective salience of sounds, and loudness,” *Psychon Bull Rev*, Vol. 23, No. 2, pp. 412–425, 2015.
- [21] R. Bisping, H. Steingrueber, M. Oltmann and C. Wenk, “Adults’ tolerance of cries: an experimental investigation of acoustic features,” *Child Development*, Vol. 61, No. 4, pp. 1218–1229, 1990.

- [22] G. Gustafson and J. Green, “On the importance of fundamental frequency and other acoustic features in cry perception and infant development,” *Child Development*, Vol. 60, No. 4, pp. 772–780, 1989.
- [23] International Standard, ISO 532-2:2017 Acoustics —Methods for calculating loudness— Part 2: Moore-Glasberg method (2017).
- [24] H. Fastl and E. Zwicker, “PSYCHOACOUSTICS,” Springer, Chapter. 4, 1990.
- [25] 高田正幸, “音質評価指標の計算法と適用事例,” *日本音響学会誌*, Vol. 75, No. 10, pp. 582–589, 2019.
- [26] 橋本竹夫, “機械騒音の音質改善の手法,” *日本音響学会誌*, Vol. 53, No. 6, pp. 450–455, 1997.
- [27] M. Takada, K. Iida, S. Tsunekawa and S. Iwamiya, “Sound quality evaluation of noise emitted from brush cutters,” *Proc. Inter Noise*, No. 1464, 2018.
- [28] G. Kwon, H. Jo and Y. Kang, “Model of psycho acoustic sportiness for vehicle interior sound: Excluding loudness,” *Applied Acoustics*, Vol. 136, pp. 16–25, 2018.
- [29] M. Khan, O. Johansson and U. Sunduack, “Development of an annoyance index for heavy-duty diesel engine noise using multivariate analysis,” *Noise Control Engineering Journal*, Vol. 45, pp. 157–167, 1997.
- [30] P. Ambika, N. Harper and J. Schnupp., “Exploring the distribution of statistical feature parameters for natural sound textures,” *PLoS ONE*, Vol. 16, No. 6, pp.0–20, 2020.
- [31] J. Jeon, J. You, C. Jeong, S. Kim and M. Jho., “Varying the spectral envelope of air conditioning sounds to enhance indoor acoustic comfort,” *Building and Environment*, Vol. 46, No. 3, pp. 739–746, 2011.
- [32] J. Jeon, J. You and H. Chang, “Sound radiation and sound quality characteristics of refrigerator noise in real living environments,” *Applied Acoustics*, Vol. 68, No. 10, pp. 1118–1134, 2007.
- [33] M. Horvat, M. Domitrovic and K. Jambrosic, “Sound quality evaluation of hand-held power tools,” *Acta Acustica united with Acustica*, Vol. 98, No. 3, pp. 487–504, 2012.

- [34] International Organization for Standardization, “ISO 12913-1:2014 Acoustics –Soundscape– Part 1: Definition and conceptual framework,” 2014.
- [35] 水野滉介, “音質評価指標を用いた異常音検知の研究,” 北陸先端科学技術大学院大学修士論文, 2020.
- [36] S. Tanaka, S. Morita and M. Unoki, “Study on laughter detection based on auditory attributes,” RISP International Workshop on Nonlinear Circuits, Communications and Signal Processing, 3-2-4, 2021.
- [37] 畠山達也, “聴知覚メカニズムに着目した音声の年齢知覚に関する研究,” 北陸先端科学技術大学院大学修士論文, 2021.
- [38] N. Singh and F. Theunissen, “Modulation spectra of natural sounds and ethological theories of auditory processing,” *The Journal of the Acoustical Society of America*, Vol. 114, No. 6, pp. 3394–3411, 2003.
- [39] N. Singh and F. Theunissen, “Modulation spectra of natural sounds and ethological theories of auditory processing,” *The Journal of the Acoustical Society of America*, Vol. 114, issue. 6, No. 10, pp. 3394–3411, 2003.
- [40] K. Patil<sup>1</sup>, D. Pressnitzer, S. Shamma and M. Elhilali<sup>1</sup>, “Music in Our Ears: The Biological Bases of Musical Timbre Perception,” *PLOS Computational Biology*, Vol. 9, No. 10, 2013.
- [41] E. Monty and R. Heather, “Representation of spectro temporal sound information in the ascending auditory pathway,” *Biol Cybern*, Vol. 89, No. 5, pp. 350–362, 2003.
- [42] M. Schonwiesner and J. Zatorre, “Spectro-temporal modulation transfer function of single voxels in the human auditory cortex measured with high-resolution fMRI,” *PNAS*, Vol. 106, No. 34, pp. 14611–14616, 2009.
- [43] M. Elliott and F. Theunissen, “The modulation transfer function for speech intelligibility,” *PLoS COMPUTATIONAL BIOLOGY*, vol. 5, No. 3, pp. 1–14, 2009.
- [44] E. Monty and R. Heather, “Representation of spectrotemporal sound information in the ascending auditory pathway,” *Biological Cybernetics*, Vol. 89, No. 5, pp. 350–362, 2003.
- [45] G. Fant, “Acoustic Theory of Speech Production: With Calculations Based on X-Ray Studies of Russian Articulations,” Vol. 2, 1971.

- [46] S. Rosen, “Temporal information in speech: acoustic, auditory and linguistic aspects,” *Biological Sciences*, Vol. 336, pp. 367–373, 1992.
- [47] A. Sabin and D. Eddins, “Perceptual Learning Evidence for Tuning to Spectrotemporal Modulation in the Human Auditory System,” *The Journal of Neuroscience*, Vol. 32, No. 19, pp. 6542–6549, 2012.
- [48] M. Elhilali, T. Chi and S. Shamma, “A spectro-temporal modulation index (STMI) for assessment of speech intelligibility,” *Speech Communication*, Vol. 41, pp. 331–348, 2003.
- [49] T. Chi, Y. Gao, M. Guyton, P. Ru, S. Shamma., “Spectro-temporal modulation transfer functions and speech intelligibility,” *The Journal of the Acoustical Society of America*, Vol. 106, No. 5, pp. 2719–2732, 1999.
- [50] 磯山拓都, 木谷俊介, 鶴木祐史, “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,” *聴覚研究会資料*, Vol. 50, No. 2, pp. 129–134, 2020.
- [51] A. Pearce, T. Brookes and R. Mason, “Timbral attributes for sound effect library searching,” *Audio Engineering Society Conference on Semantic Audio*, Erlangen, pp. 22–24, 2017.