

Title	聴覚的顕著性とスペクトル・時間変調情報の関係
Author(s)	木所, 晃利
Citation	
Issue Date	2021-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/17092
Rights	
Description	Supervisor: 鷓木 祐史, 先端科学技術研究科, 修士 (情報科学)

修士論文

聴覚的顕著性とスペクトル・時間変調情報の関係

1910079 木所 晃利

主指導教員 鷗木 祐史
審査委員長主査 鷗木 祐史
審査委員 赤木 正人
党 建武
吉高 淳夫

北陸先端科学技術大学院大学
先端科学技術研究科
(情報科学)

令和3年2月

Abstract

Sound, as a physical phenomenon, is a kind of wave. Humans perceive these waves through the auditory system, starting from the auricle, one of the auditory organs, and ending with the auditory nervous system and the auditory cortex of the cerebrum. The waves provide us with a great deal of information necessary for daily life. For example, the sirens that announce emergencies and the time signals on the radio that tell the time. In this way, our daily life can be said to be lived with sound. When listening to sounds, our ability to consciously select and listen to sounds is known as active auditory attention or the cocktail party effect.

On the other hand, there are sounds that we can recognize even if we are unconscious of their existence. The degree of conspicuousness of these sounds that can be recognized unconsciously is called auditory saliency, and it has been studied as an aspect of passive auditory attention by focusing on various acoustic features. Auditory saliency The first model of auditory saliency was developed by focusing on Intensity, Temporal contrast, and Spectro contrast in the sound spectrogram. They reported that the model was able to explain the results of listening experiments, indicating that acoustic features related to intensity, temporal contrast, and frequency contrast were related to saliency. Subsequent studies of the model extended the first model to include Temporal Modulation (TM), which is a spectrum that varies along the time axis in the spectrogram, Spectro Modulation (SM), which is a spectrum that changes in the direction of the frequency axis, and Spectro-Temporal Modulation (STM), which is a spectrum in which both of them exist and change simultaneously. The results of the listening experiments showed that there was a significant correlation between the stimuli judged to be salient by the listeners and those judged to be salient by the model. This indicates that the acoustic features related to SM, TM, and some parts of STM are related to saliency. On the other hand, while investigating the relationship between auditory saliency and pupil diameter response, some studies have focused on the relationship between psychological quantity of sound and saliency, and investigated the relationship between loudness and saliency. The results showed that loudness and saliency are correlated, and it is thought that acoustic features related to loudness are also related to saliency. Following this study, other studies have focused on the relationship between acoustic features and saliency and investigated the relationship between loudness, duration of acoustic features, and spectral structure and salient stimuli in environmental sounds. They suggested that spectral structure and duration of acoustic features also contribute to saliency. They also suggested that a single acoustic feature alone cannot account for saliency. However, even in the various studies presented so far, no consistently explanatory acoustic feature

for auditory saliency has been identified.

The purpose of this study was to investigate what acoustic features appear in STM that can be observed, including the interaction between SM and TM, and to clarify how these features are related to auditory saliency. In this study, we examined the acoustic features that appear in STM including SM and TM. Therefore, this study focuses on STM analysis, which integrates SM and TM, and investigates the relationship between auditory saliency and acoustic features obtained from the results of STM analysis. In this study, the following steps are taken for this purpose: 1. prepare stimuli whose saliency is already known from previous studies. 2. perform STM analysis on the prepared stimuli to obtain STM. 3. analyze the acoustic features that are considered to be related to saliency based on the results of STM analysis and the findings of previous studies, such as average power, frequency spectrum spread, harmonicity, and temporal modulation. 4. Analyze to what extent the features obtained in 3 contribute to the saliency. 5. consider the contribution of the acoustic features obtained from the STM to the saliency based on the analysis results obtained in step 4.

The STM analysis used in this study is based on a method used in speech recognition and described in other review papers. The method is based on a two-dimensional Fourier transform of the spectrogram obtained from the filter bank. The advantage of this method is that it is easy to quantitatively measure the acoustic features from the final STM, and it is also possible to create quantitatively controlled stimuli by performing 2-D filtering and inverse transformation on the STM obtained by this method. The filter bank used for the STM analysis in this study is a constant-band gammatone filter bank, and adjacent filters intersect at a point of -3 dB. In this STM analysis, the filter bank and the frequency bandwidth per channel are related to the resolution of the SM. The narrower the filter, the higher the resolution of the SM, but the lower the resolution of the TM due to the effect on the amplitude envelope in the low frequency range. Based on the trade-off between these two factors and the characteristics of the stimuli used in this study, the frequency bandwidth of the filter per channel was set to 80 Hz. For the above reasons, the range of analysis was -40 Hz to $+40$ Hz on the TM axis and 0 cyc/kHz to 6.5 cyc/kHz on the SM axis.

In this study, the research strategy was to conduct STM analysis on the stimuli whose saliency was already known, and to further analyze the acoustic features that were thought to contribute to the saliency from the obtained STM. In this research, we need to understand the acoustic features that contribute to the STM. For this purpose, we created amplitude-modulated signals, harmonic complex tones, spectral structures, and frequency-modulated signals, and analyzed their STM. As a result, the information that appears in the STM analysis when the amplitude

modulation signal, harmonic complex, spectral structure, and frequency modulation signal are input was found. Next, to investigate the relationship between auditory saliency and STM, 10 stimuli with known saliency at the Thurston scale were subjected to STM analysis. After that, we analyzed the results of the STM analysis and the findings of previous studies for mean power, frequency spectral spread, harmonicity, and temporal modulation, which were considered to be related to saliency. Correlation coefficients were calculated between the results obtained from the analysis and the prominence at the Thurston scale. As a result, the acoustic features obtained only from the SMTM did not have a significant correlation with the saliency. In order to investigate the correlation coefficient between the acoustic features on the STM plane and the saliency, we integrated the features from only SM and TM and calculated the correlation coefficient with the saliency scale. As a result, a weak correlation trend was observed. However, there were no acoustic features that were completely uncorrelated. These results indicate that the acoustic features related to SM and TM on the STM plane are related to auditory saliency. These results suggest that it is possible to examine auditory saliency using the STM analysis used in this study.

目次

第1章	序論	1
1.1	背景	1
1.2	聴覚的顕著性	2
1.3	目的	3
1.4	本研究の着目点	3
1.5	本研究の方略	4
1.6	本論文の構成	6
第2章	STM 情報分析	8
2.1	STM 情報分析の求め方	8
2.2	音響特徴に対する STM 情報分析の性質	10
2.2.1	STM 情報と振幅変調信号	10
2.2.2	STM 情報と調波複合音	14
2.2.3	STM 情報とスペクトル構造	18
2.2.4	STM 情報と振幅変調調波複合音	28
2.2.5	STM 情報と周波数変調信号	32
2.3	STM 情報分析の性質のまとめ	36
第3章	検討対象とした刺激と聴覚的顕著性の関係の分析	37
3.1	平均パワーの分析	41
3.1.1	着目理由	41
3.1.2	算出方法	41
3.1.3	分析結果	41
3.1.4	考察	41
3.2	周波数スペクトルの広がり	43
3.2.1	着目理由	43
3.2.2	算出方法	43
3.2.3	分析結果	43
3.2.4	考察	43
3.3	調波性の分析	46
3.3.1	着目理由	46
3.3.2	算出方法	46

3.3.3	分析結果	46
3.3.4	考察	47
3.4	TM情報の分析	51
3.4.1	着目理由	51
3.4.2	算出方法	51
3.4.3	分析結果	51
3.4.4	考察	52
第4章	全体考察	56
第5章	結論	58
5.1	本研究により明らかにされたこと	58
5.2	残された課題	58
5.2.1	STM情報と顕著性の分析	58
5.2.2	時間変化による音響特徴の変化も含めた分析	59

目 次

1.1	本研究のアプローチ	5
1.2	本研究の構成	7
2.1	STM 情報分析のイメージ図	9
2.2	振幅変調周波数 (0 Hz) での STM 情報	11
2.3	振幅変調周波数 (15 Hz) での STM 情報	12
2.4	振幅変調周波数 (30 Hz) での STM 情報	13
2.5	基本周波数 (200 Hz) の調波複合音での STM 情報	15
2.6	基本周波数 (300 Hz) の調波複合音での STM 情報	16
2.7	基本周波数 (600 Hz) の調波複合音での STM 情報	17
2.8	スペクトル傾斜のある信号の作成手法	19
2.9	スペクトル構造を正弦波で変化させた信号の作成手法	20
2.10	スペクトル傾斜 (-3 dB/oct) のある信号での STM 情報	21
2.11	スペクトル傾斜のない信号 (ホワイトノイズ) での STM 情報	22
2.12	スペクトル傾斜 (+3 dB/oct) のある信号での STM 情報	23
2.13	SM 情報の変調周波数 (0 Hz) での STM 情報	24
2.14	SM 情報の変調周波数 (1 Hz) での STM 情報	25
2.15	SM 情報の変調周波数 (3 Hz) での STM 情報	26
2.16	SM 情報の周波数 (6 Hz) での STM 情報	27
2.17	振幅変調周波数 (0 Hz) の調波複合音の STM 情報	29
2.18	振幅変調周波数 (15 Hz) の調波複合音の STM 情報	30
2.19	振幅変調周波数 (30 Hz) の調波複合音の STM 情報	31
2.20	周波数変調周波数 (1 Hz) での STM 情報	33
2.21	周波数変調周波数 (3 Hz) での STM 情報	34
2.22	周波数変調周波数 (6 Hz) での STM 情報	35
3.1	分析対象とした刺激の顕著性 (横軸はサーストンスケール)	39
3.2	(a) 各刺激のサウンドスペクトログラム. (b) 各刺激の STM 情報. (c) TM 情報 0 Hz での SM 情報. (d) SM 情報 0 cyc/kHz での TM 情報.	40
3.3	各刺激の平均パワー	42
3.4	SM 情報極小範囲で求めた各刺激の傾き	45

3.5	SM 情報の自己相関	48
3.6	SM 情報の自己相関より求めた各刺激のピークの高さ	49
3.7	SM 情報の自己相関より求めたスペクトルのピークの幅	50
3.8	TM 情報の自己相関	53
3.9	TM 情報の自己相関より求めた各刺激のピークの高さ	54
3.10	TM 情報の自己相関より求めたスペクトルのピークの幅	55
4.1	STM 情報の音響特徴と顕著性スケールの相関	57

表 目 次

3.1 本研究で用いた顕著性の判明している刺激	38
-----------------------------------	----

第1章 序論

1.1 背景

音は、物理現象としては波の一種であり、ヒトはその波を聴覚器官の一つである耳介からはじまり、最終的に聴覚神経系及び大脳の聴覚皮質にいたる、聴覚系で認識している [1]。私たちはこの認識から得た、様々な情報を日々の生活で用いている。この情報というのは、音の有無や言葉の理解といった部分に留まらず、音に含まれる感情や性別、音源の種類、方向など生物として生きていくための必要な情報を音から得ている。例としては、作業の開始、終了を知らせるコンピューターの「ピッ」という音や、コミュニケーションのための人々の話し声、異常や緊急を知らせるサイレン音、時刻を知らせるラジオの時報などがある。このように、私たちは、日々の生活を音と共に生きていても良い。そして、これらの音たちというのは、防音室や無響室などの特殊な環境を除いては単独で存在することはありえず、いくつもの音が同時に存在する混じり合った音（混合音）である。最初に述べた聴覚系は、この混合音を極めて正確に分析し、必要な情報を抽出する、一種の分析器と捉えることもできる。もし、私たちがこの聴覚系という優れた分析器を有しない、または不完全形で有していた場合、私たちの日々の生活は大いに不便を被ることになり、今日の豊かな営みは成り立たなかつたであろう。

この優れた分析器がどのように混合音から個々の情報を得ているのかという謎に対して、多くの研究がなされて来た。その中でヒトは、混合音中のすべての音を処理しているのではなく、一部の音にのみ、意識的に注目して聴取し、処理していることが明らかになってきた。このような聴覚系の働きは、聴覚の能動的注意、カクテルパーティー効果 [2, 3]、などからも知られている。

一方で、こういった混合音の中において、（例えば、防災無線のサイレン音や自動車のクラクションなど）思わず耳を引き付けられるような音の存在を感じることもある。このような音は、自らの意思でもって聞こうとしてする音とは異なり、無意識であってもその存在を認識することができる。これら、無意識に引き付けられる音は他の音に対して目立つ音と、捉えることができる。この、目立つ音の目立ちやすさの度合いを聴覚的顕著性といい聴覚の受動的注意の側面として様々な音響特徴に着目し研究されている [4]。

1.2 聴覚的顕著性

本節では聴覚的顕著性に関する研究を述べていく。最初に音の顕著性に着目した，聴覚的顕著性モデルを作成したのはKayser[5]らである。KayserらはItti[6]らの視覚的顕著性モデルを基に聴覚分野へ応用した。Kayserのモデルはサウンドスペクトログラムを一枚の画像と捉えスペクトログラム上の強度(Intensity)，時間コントラスト(Temporal contrast)，周波数コントラスト(Spectro contrast)を特徴として聴覚的顕著性モデルを構築した。彼らの報告によると，モデルが聴取実験の結果を説明できたことから，強度，時間コントラスト，周波数コントラストにかかわる音響特徴が顕著性に関係することを示した。Kalinli & Narayanan[7]はKayserらのモデルを拡張し時間コントラスト，周波数コントラストの両者が存在する，スペクトログラム上の斜め方向の成分とピッチを含めた。その結果，Kayserらのモデルよりもノイズなどに対してよりロバスト性の高いモデルとなったことから，時間コントラスト，周波数コントラストの両者が存在する斜め方向の成分に関わる音響特徴も顕著性に関係することを示した。Duangudom & Anderson[8]のモデルはKalinli & Narayananらのモデルで利用した音響特徴をより高次元に表現する手法を含めた形での拡張を行った。彼らが利用した特徴はスペクトログラム上の時間軸方向に変化するスペクトルである時間変調情報(TM情報)と周波数軸方向に変化するスペクトルであるスペクトル変調情報(SM情報)，この両者が同時に存在し変化するスペクトルであるスペクトル・時間変調情報(STM情報)であった。この，Duangudom & Andersonのモデルで用いたSTM情報は高TM情報部に着目していた。聴取実験での結果から，聴取者が顕著性のあると判断した刺激とモデルによって顕著性がある判断された刺激との間に有意な相関があることを示した。このことからSM情報，TM情報，STM情報の一部，に関わる音響特徴が顕著性に関わると関係すると考えられる。一方で，Liaoら[9]は聴覚的顕著性と瞳孔径反応の関係を調べる中で，音の心理量と顕著性の関係に着目しラウドネスと顕著性の関係を調査した。その結果，ラウドネスと顕著性に相関があることを示したことから，ラウドネスに関わる音響特徴も顕著性に関係すると考えられる。Huang & Elhilali[10]らも音響特徴と顕著性の関係に着目し，ラウドネス，音響特徴の持続時間，スペクトル構造と環境音中の顕著な刺激との関係を調査した。彼らの報告からこれらスペクトル構造，音響特徴の持続時間も顕著性に寄与することを示唆した。また，Tordiniら[11, 12]は心理量に着目したモデルを作成し，その中でもラウドネス，テンポ，ブライツネスに着目してモデルを作成した。その結果，これら心理量の算出に関わるtemporal centroid, spectro centroid, harmonicityも含めてラウドネス，テンポ，ブライツネスは聴覚的顕著性と相関があることを示した。更にKaya & Elhilali[13, 14]は顕著性モデルを作成するに当たり音響特徴の時間的変化も含めて顕著性との関係に着目した。このモデルではShammaら[15]のヒトの聴知覚を多次元に表現する手法を採用し，時々刻々と変化するラウドネス，音色といった心理量を特徴として採用した。この研究の報告

によると、聴取実験の結果からラウドネスやブライトネスといった心理量に関わる音響特徴と顕著性に関わる音響特徴には相関があるが単一の音響特徴のみでは説明できないことを示唆した。Wang ら [16] のモデルは Kaya & Elhilali らの結果を裏付ける形の研究となった。Wang らは (1) メル周波数ケプストラム係数から求めた時間的に変化する音響特徴, (2) パワースペクトル密度, (3) Kayser らのモデル, これら 3 つが並列化しているモデルを作成した。その結果, Kayser らのモデルよノイズなどに対して高いロバスト性を示した。

これらの研究 [9, 11, 12, 13, 14, 16] の報告により聴覚的顕著性の生起要因は単一の音響特徴のみによって定まるものではなく, 複数の音響特徴の相互に作用し合うことで起こるものであると考えられる。また, ここまでに述べてきた研究はいずれも周波数軸方向の変化, 時間軸方向の変化を直接的ないし間接的に用いていることから, SM 情報, TM 情報, STM 情報に関わる音響特徴が顕著性に大きく寄与することも考えられる。しかし, ここまで紹介した様々な研究においても聴覚的顕著性に対する一貫して説明可能な音響特徴は明らかになっていない。

1.3 目的

前節において聴覚的顕著性に関する研究を紹介してきた。これまでの研究で聴覚的顕著性には複数の音響特徴の相互作用が寄与することが示されている。また, SM 情報, TM 情報に関わる音響特徴が聴覚的顕著性に大きく寄与する可能性も示唆されている。この SM 情報と TM 情報の相互作用も観察できる STM 情報分析がある。STM 情報分析を用いることで聴覚的顕著性の関与する音響特徴を明らかにできると考えられる。本研究の目的は SM 情報と TM 情報の相互作用も含めて観察できる STM 情報分析でどのような音響特徴が現れるか検討し, 現れた音響特徴がどのように聴覚的顕著性に関係するか明らかにすることである。

1.4 本研究の着目点

他の研究においても STM 情報を用いていたが, SM 情報と TM 情報を個別に算出し統合する, STM 情報の一部のみを利用する [8] に留まっていた。SM 情報と TM 情報を統合して分析し, STM 情報全体からの得られた音響特徴の観点から顕著性との関係も検討する必要がある。

先行研究の知見では SM 情報, TM 情報に関わる音響特徴が顕著性に寄与すること, 聴覚的顕著性は一つの音響特徴のみでは定まらないことも示唆されている。この 2 点から SM 情報と TM 情報を一体化した STM 情報分析に関わる音響特徴が顕著性に関係すると考えられる。本研究では SM 情報と TM 情報を統合した STM 情報に着目し, STM 情報分析の結果から得られる音響特徴をてがかりに聴覚的顕著性との関係を検討する。

1.5 本研究の方略

本節では前節の目的を達成するための具体的な研究方略を説明する．本研究では Elliott & Theunissen[17] にて利用され古川のレビュー論文 [18] にて紹介されている STM 情報分析を用いる．この STM 情報分析についての詳細な説明および性質は次章にて行う．本研究では図 1.1 のような流れで研究を進める．上から

1. 先行研究等で顕著性の既に判明している刺激を用意する．
2. 用意した刺激に対して STM 情報分析を行い，STM 情報を求める．
3. 2 で求めた STM 情報分析の結果と先行研究の知見から顕著性に関係すると考えられる音響特徴である平均パワー，周波数スペクトルの広がり，調波性，時間変動情報，に対し，分析を行う．
4. 3 で求めた特徴がどの程度，顕著性に寄与するか分析する．
5. 4 までに求めた分析結果から STM 情報から得られた音響特徴の顕著性への寄与を考察する．

この 1 から 5 までの手法を実際に行っているのが本論文における 4 章であり，3.2 の STM 情報分析の性質は 4 章の STM 情報分析の結果と音響特徴との関係を予め把握するための節である．

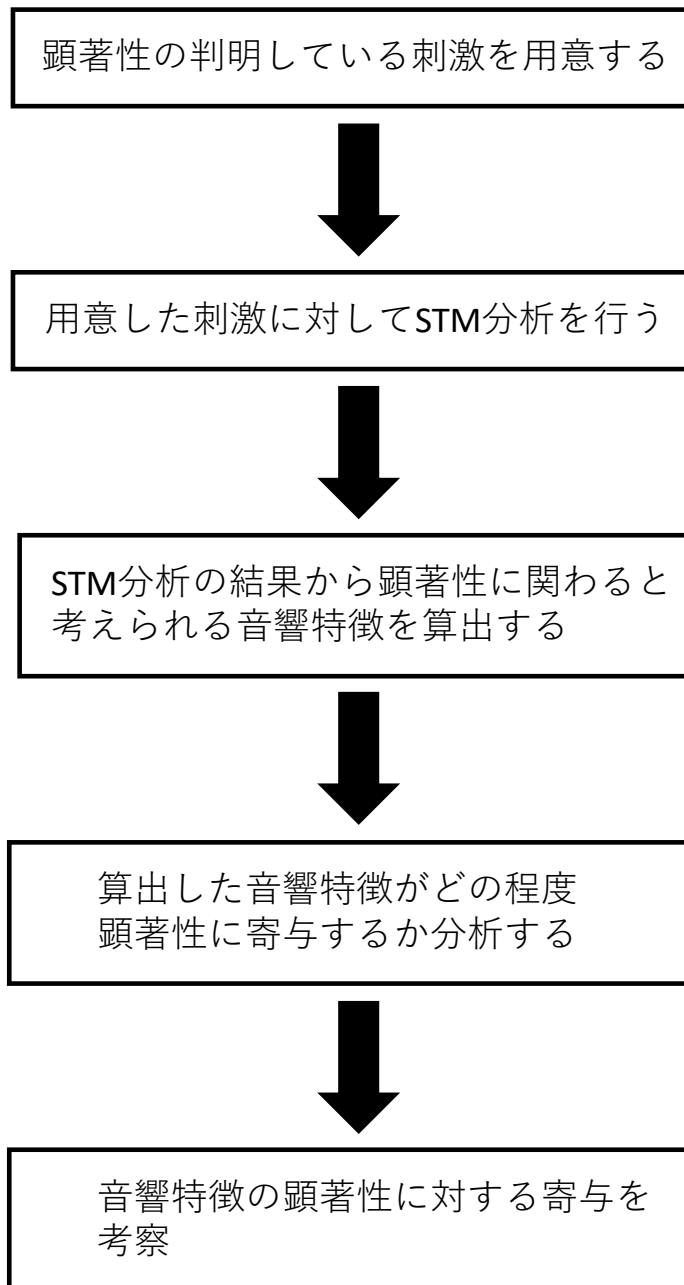


図 1.1: 本研究のアプローチ

1.6 本論文の構成

前節までに本研究の背景，目的を述べた．本節では論文全体の構成を図1.2に示す．1章は序論及び背景，目的であり，本研究のテーマである聴覚的顕著性に関する研究の現在に至るまでの知見を示す．その後，本研究の着想点を述べ本研究の目的を述べる．

2章は本研究の方略について述べる．前章までの背景，目的で述べた先行研究の知見と，目的から本研究での方法論を述べる．

3章では本研究で用いた分析手法であるSTM情報分析について述べる．その次にSTM情報分析の性質を説明する．

4章では3章で得られたSTM情報分析の性質と先行研究の知見からSTM情報分析の結果より得られる音響特徴と聴覚的顕著性の関係を分析する．

5章では4章での分析結果に対しての全体的な考察を述べる．

6章は結論とし本研究において明らかにできた部分，残された課題について述べる．

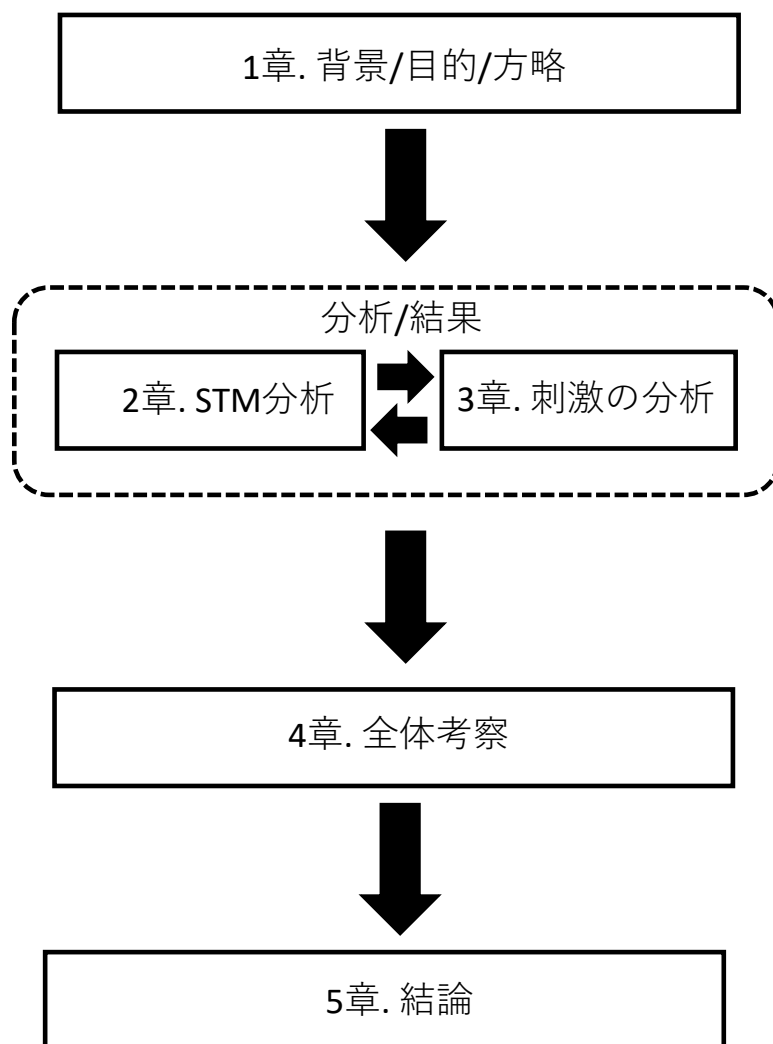


図 1.2: 本研究の構成

第2章 STM情報分析

本章では前章までにおいて言及したSTM情報および本研究で用いたSTM情報分析について説明する

2.1 STM情報分析の求め方

本研究に用いたSTM情報分析のブロックダイアグラムを図2.1に示す。図中のサウンドスペクトログラムにも見られるように、一般的な音のスペクトログラムは時間軸方向の変化，周波数軸方向の変化，さらに周波数軸・時間軸の両方に対して変化するスペクトルを観察できる。サウンドスペクトログラム上の時間軸方向の変化がTM情報であり，周波数軸方向での変化がSM情報，そして時間軸と周波数軸の両方向の変化がSTM情報である。本研究に用いたSTM情報の算出方法はElliott & Theunissen[17, 19]らの手法を参考にした。まず，原信号をフィルタバンクに入力し得られた出力から[21, 33]を参考にヒルベルト変換を利用して各周波数帯域ごとの振幅包絡線を求める。求めた振幅包絡線からスペクトログラムを作成した。作成したスペクトログラムに対して2次元離散フーリエ変換を行い，求めた2次元のスペクトルがSTM情報となる。本手法の利点は最終的に求められるSTM情報から音響特徴を定量的に計りやすいことである，また本手法にて求めたSTM情報に対して2次元フィルタ処理を行い逆変換を行うことで定量的に制御された刺激を作成することも可能である。本研究のSTM情報分析で用いたフィルタバンクは[30]を参考にした定帯域ガンマトーンフィルタバンクであり隣り合うフィルタは -3 dBの点で交差している。本手法でのSTM情報分析において，フィルタバンク，1チャンネルあたりの周波数帯域幅はスペクトル変調情報の分解能と関係している。狭帯域のフィルタを用いたほうが高分解能のSM情報を得られるが，低周波数帯域において振幅包絡線に影響を与えTM情報の分解能が低下する。この両者のトレードオフの関係と本研究で用いた刺激の特性から，1チャンネルあたりのフィルタの周波数帯域幅は 80 Hzとした。また，分析した範囲は上記の理由からTM情報で -40 Hz $+40$ Hz，SM情報で 0 cyc/kHz 6.5 cyc/kHzとした。

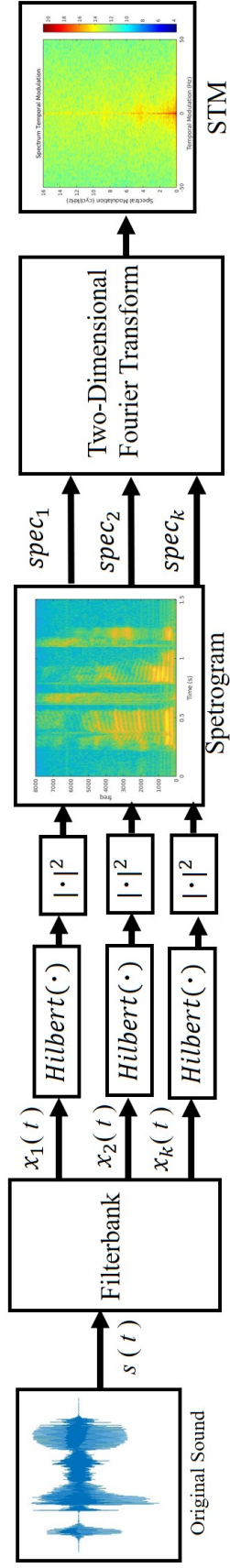


図 2.1: STM 情報分析のイメージ図

2.2 音響特徴に対する STM 情報分析の性質

本研究では、すでに顕著性が判明している刺激に対して STM 情報分析を行い、得られた STM 情報から顕著性に寄与していると考えた音響特徴を更に分析するという形での研究アプローチを行っている。そのため刺激を分析した際の STM 情報がどのような音響特徴によって生み出されるかを把握する必要がある。本節は人工的に音響特徴を持たせた刺激を作成、音響特徴を変化させていくことで STM 情報と音響特徴の関係を把握していく。

2.2.1 STM 情報と振幅変調信号

振幅の変化に伴う STM 情報の変化を観察するために振幅変調信号 (AM 信号) を作成し分析した。作成した AM 信号は [30] を参考に式 2.1 を利用した。式中の搬送波周波数 f_c は 5 kHz の正弦波とし、振幅変調周波数 f_a のみを 0 Hz, 15 Hz, 30 Hz に変化させ振幅の変化と STM の関係を観察した。各振幅変調周波数ごとの分析結果を図 2.2, 図 2.3, 図 2.4 に示す、結果より振幅変調周波数が増加すると TM 情報が STM 情報の縦方向のスペクトルとして変化し、振幅変調と TM 情報の対応関係がわかった。

$$\text{AM signal} = (1 + \sin(2\pi f_a t)) \cdot \sin(2\pi f_c t). \quad (2.1)$$

$$f_a = 0 \text{ [Hz]}$$

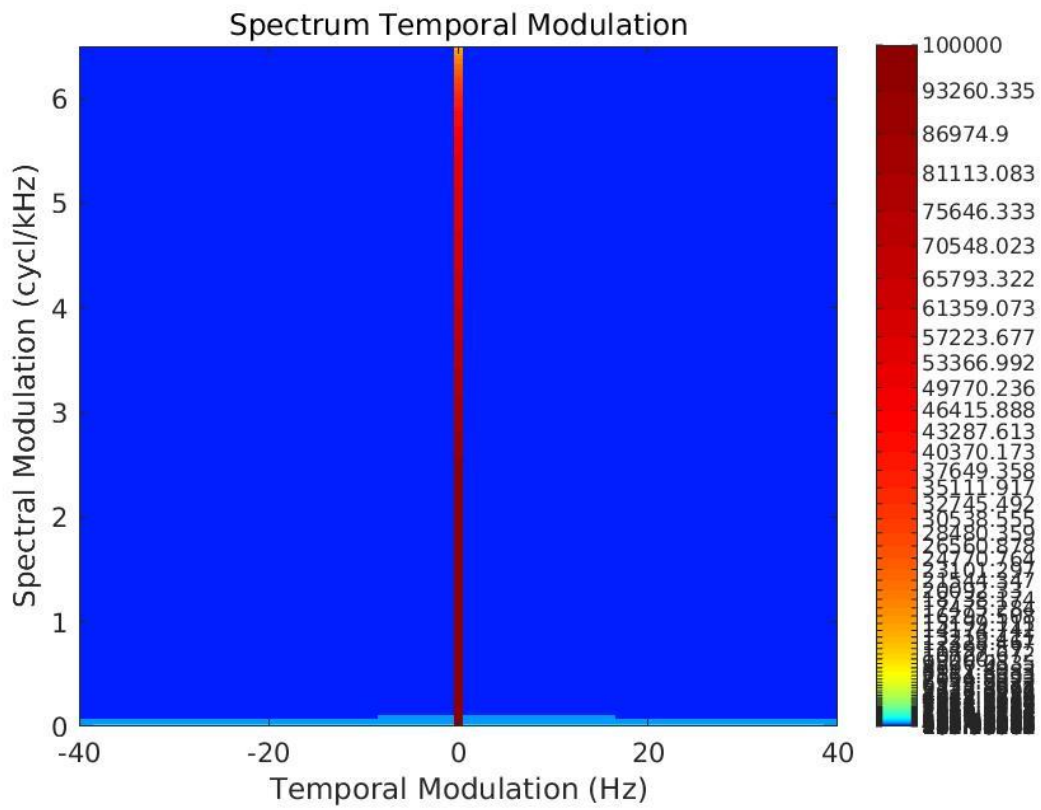


図 2.2: 振幅変調周波数 (0 Hz) での STM 情報

$$f_a = 15 \text{ [Hz]}$$

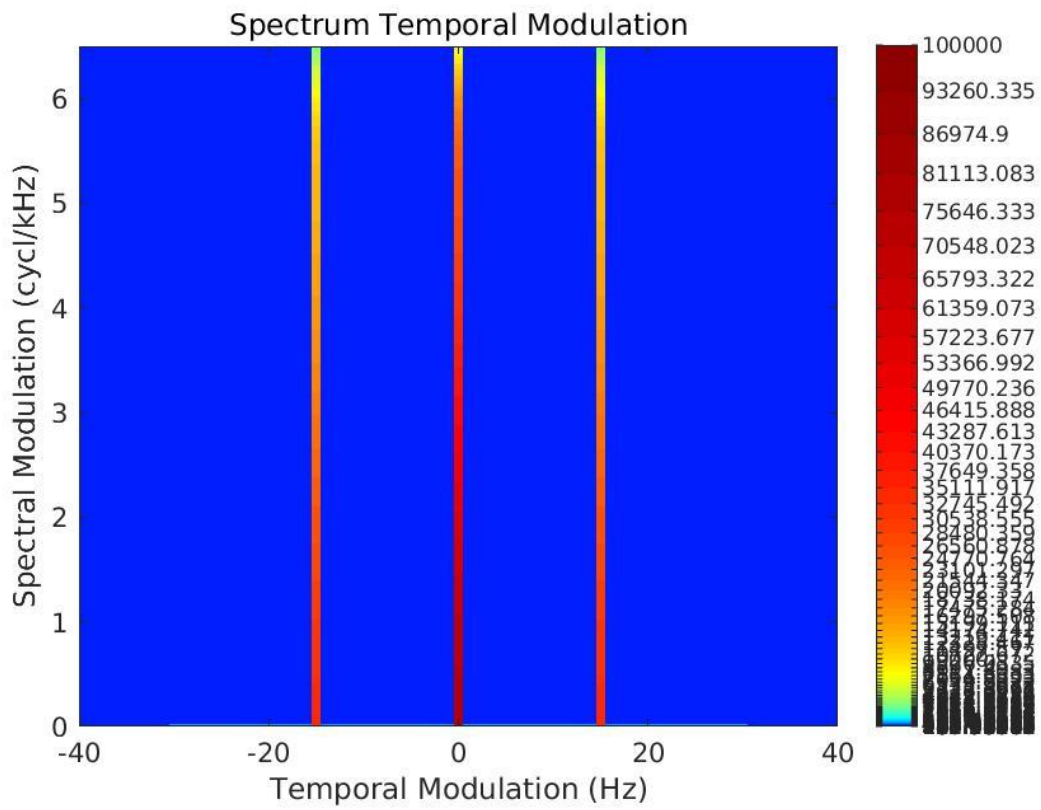


図 2.3: 振幅変調周波数 (15 Hz) での STM 情報

$$f_a = 30 \text{ [Hz]}$$

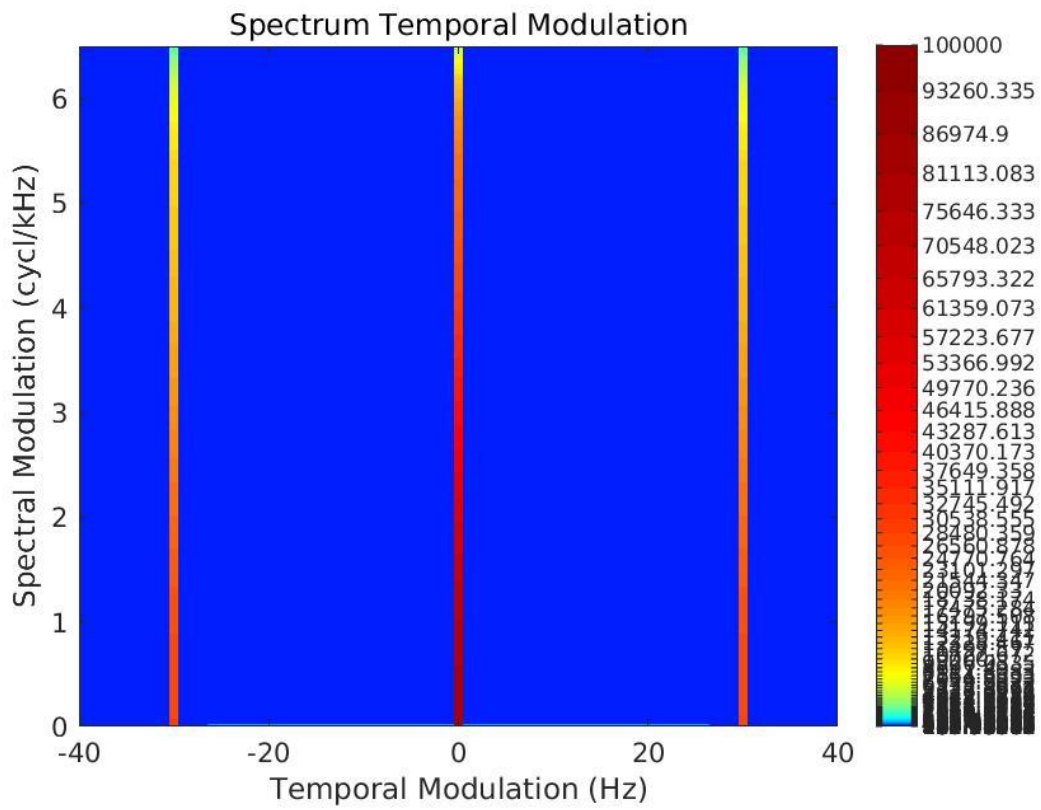


図 2.4: 振幅変調周波数 (30 Hz) での STM 情報

2.2.2 STM 情報と調波複合音

スペクトル構造の変化に伴う STM 情報の変化を観察するために倍音構造を持つ調波複合音 (HCT) を分析した. 調波複合音 (HCT) は [23, 29] を参考に式 2.2 より, 基本周波数 f_0 とその倍音にあたる, 周波数 $f_1 \dots f_i$ の正弦波の足し合わせとし, 基本周波数 f_0 を 200 Hz, 300 Hz, 600 Hz に変化させスペクトル構造と STM 情報の関係を観察した. 式中の i は調波複合音の次数である. 各基本周波数のごとの調波複合音の分析結果を図 2.5, 図 2.6, 図 2.7 に示す, 調波複合音の基本周波数を変化させると STM 情報の SM 情報が赤い点として SM 情報軸上に現れ変化した. また, 赤い点として現れた SM 情報の値は調波複合音の基本周波数の逆数と一致する. 結果より倍音構造を持つ調波複合音 (HCT) と STM 情報の対応関係がわかった.

$$\text{HCT} = \sum_{i=0}^n \sin(2\pi f_i t). \quad (2.2)$$

$$f_0 = 200 \text{ [Hz]}$$

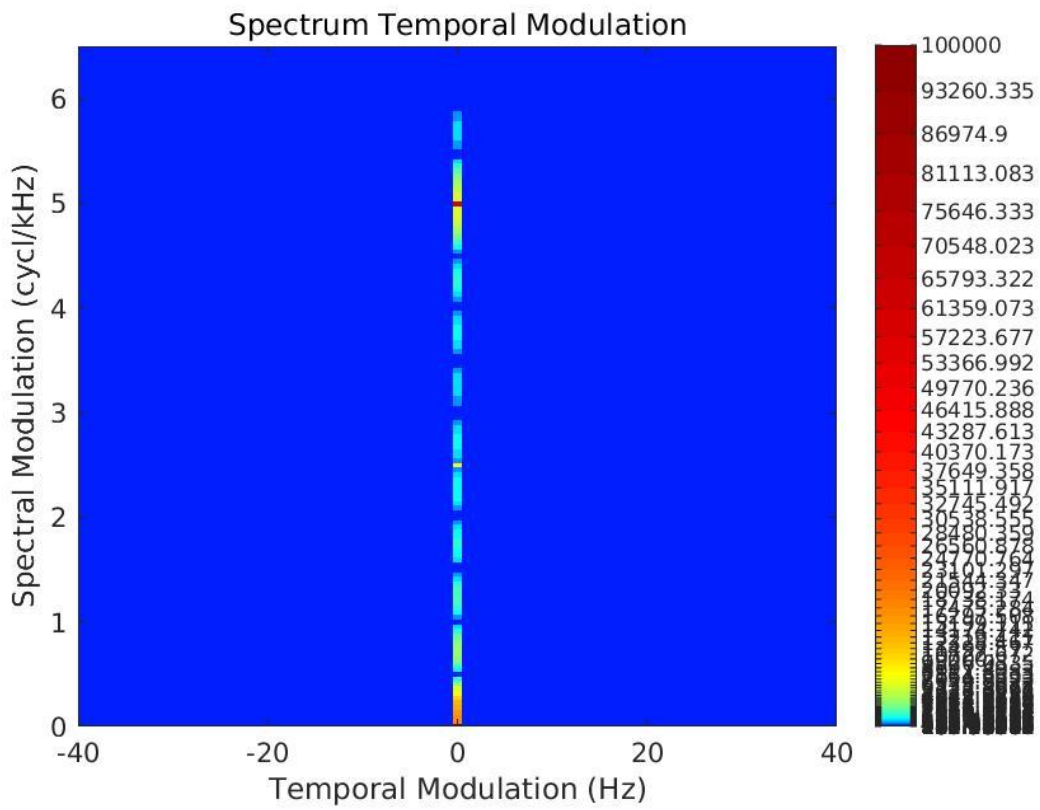


図 2.5: 基本周波数 (200 Hz) の調波複合音での STM 情報

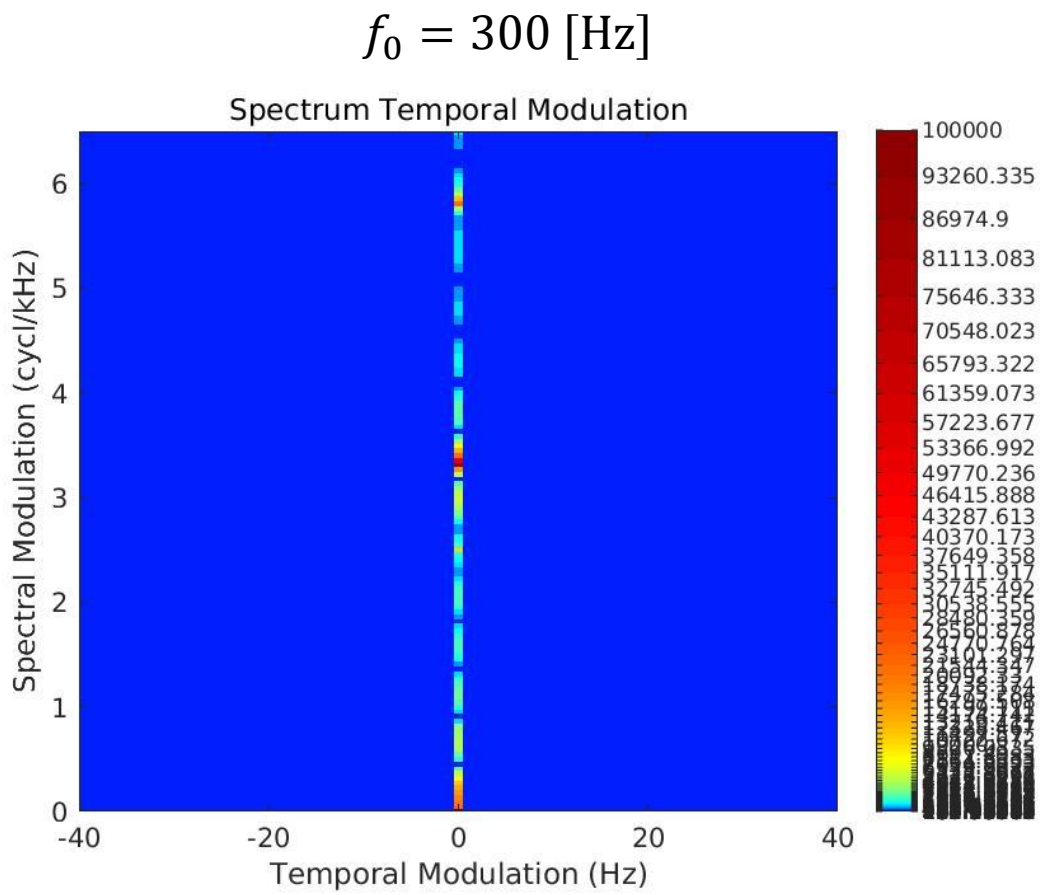


図 2.6: 基本周波数 (300 Hz) の調波複合音での STM 情報

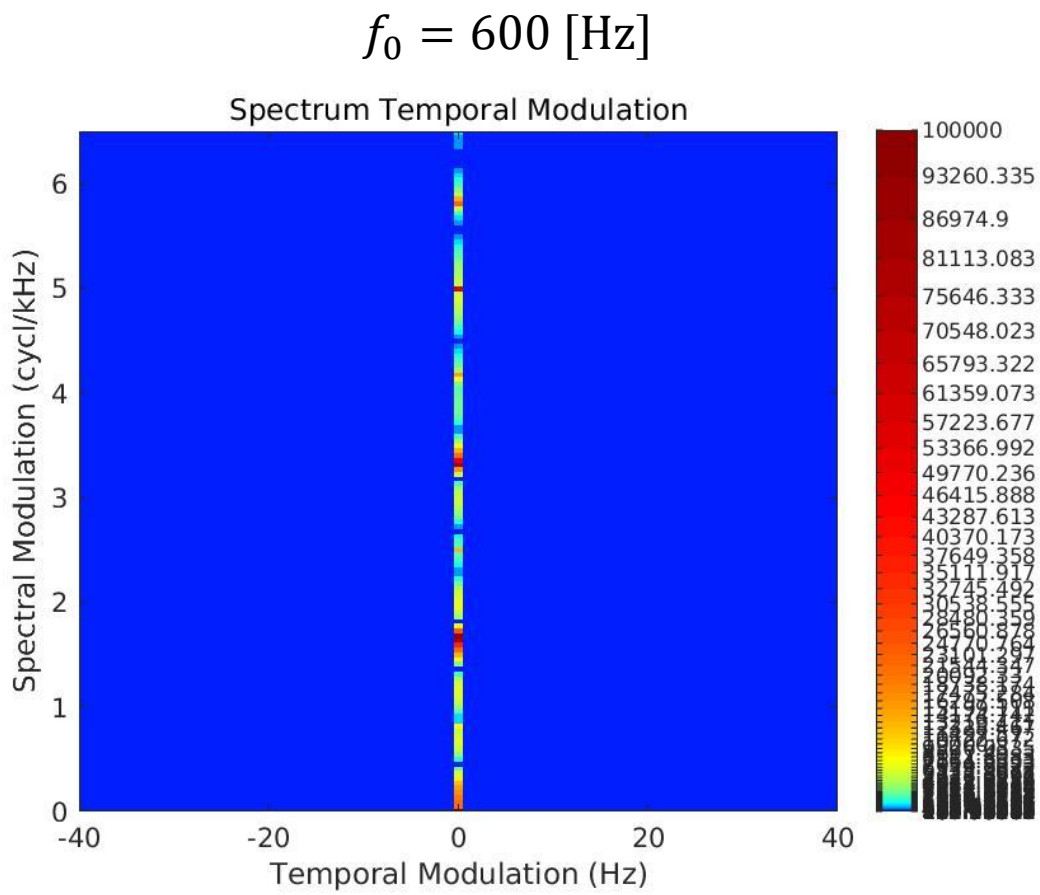


図 2.7: 基本周波数 (600 Hz) の調波複合音での STM 情報

2.2.3 STM情報とスペクトル構造

スペクトル構造の変化に伴うSTM情報の変化を観察するためにホワイトノイズのスペクトル構造を2種類に変化させてSTM情報分析した。ホワイトノイズのスペクトル構造を変化させた手法を図2.8と図2.9のブロックダイアグラムにて示す。図2.8はスペクトル傾斜のあるノイズの作成手法であり[24, 25, 26, 27, 28]を参考に、ホワイトノイズに対し+3 dB/octで変化するピンクノイズと-3 dB/octで変化するブルーノイズを作成しSTM情報分析した。図2.9はホワイトノイズのスペクトル構造を正弦波で変化させた信号の作成手法である。具体的にはホワイトノイズを離散フーリエ変換し、変換後に得られた周波数スペクトルに対して正弦波を乗算する。その後、離散フーリエ逆変換を行うことでスペクトル構造を変化させた信号を作成した。使用したホワイトノイズはガウス分布を示すホワイトノイズを利用し、乗算またはフィルタ処理されるホワイトノイズすべて同じホワイトノイズである。図2.9中の f_m はSM情報の変調周波数であり、0 Hz, 1 Hz, 3 Hz, 6 Hzで作成した。各スペクトル傾斜ごとの分析結果を図2.10, 図2.11, 図2.12に示す。各スペクトル構造ごとの分析結果を図2.13, 図2.14, 図2.15, 図2.16に示す。図2.10, 図2.11, 図2.12よりホワイトノイズではTM情報0 Hz, SM情報0 cyc/kHzのみに点スペクトルが現れ、ブルーノイズ, ピンクノイズではTM情報0 Hz, SM情報0 cyc/kHzからのスペクトルの広がり方が変化した。図2.13, 図2.14, 図2.15, 図2.16より $f_m = 0$ つまり、ホワイトノイズでは図2.11と同様の結果となった、正弦波にてスペクトル構造を変化させた $f_m = 1$ ではSM情報1 cyc/kHzに点スペクトルが現れた、その後の値においても f_m と点が現れる値が一致した。結果よりスペクトル構造とSTM情報の対応関係がわかった。

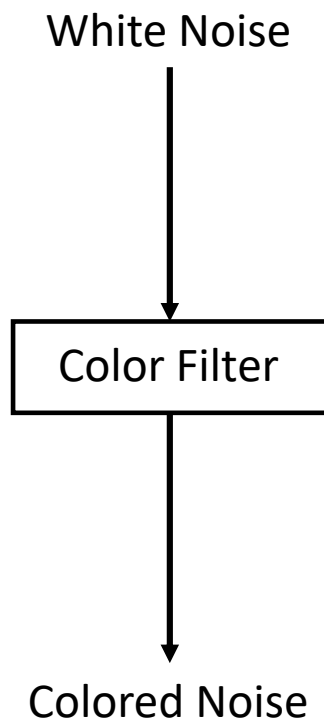


図 2.8: スペクトル傾斜のある信号の作成手法

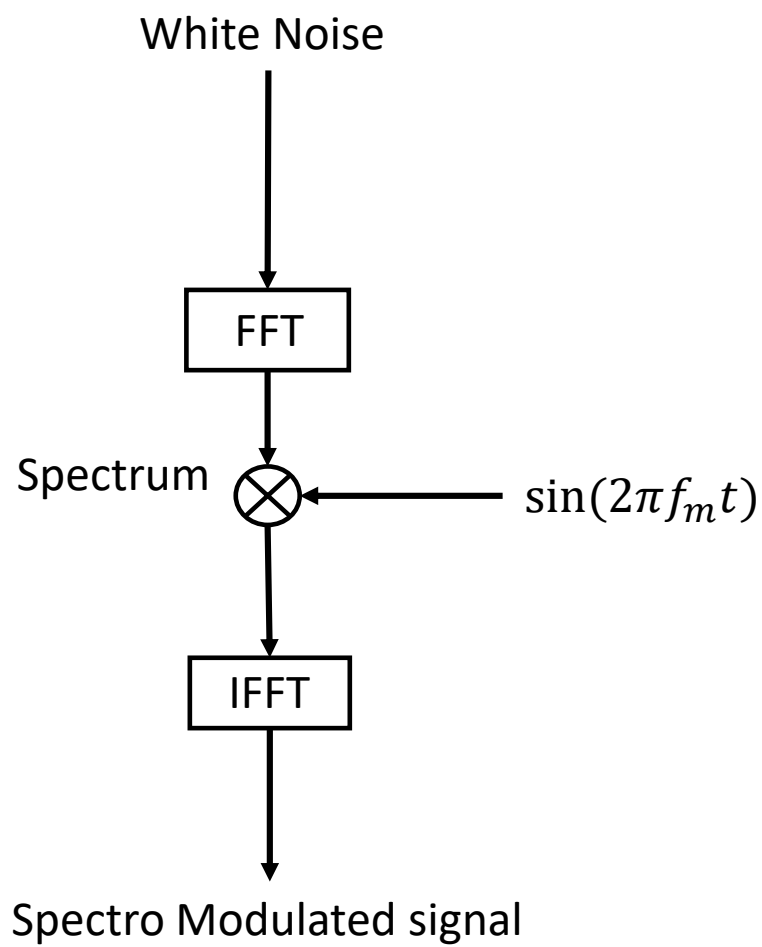


図 2.9: スペクトル構造を正弦波で変化させた信号の作成手法

ピンクノイズ (-3 dB/oct)

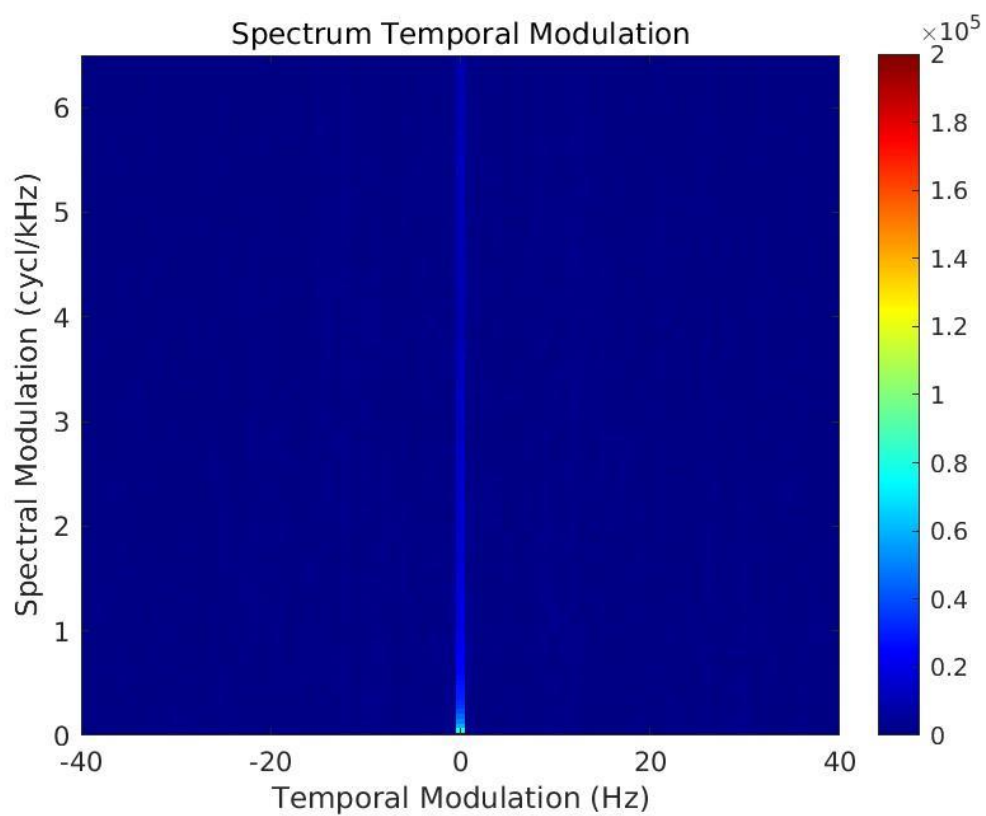


図 2.10: スペクトル傾斜 (-3 dB/oct) のある信号での STM 情報

ホワイトノイズ（スペクトル傾斜なし）

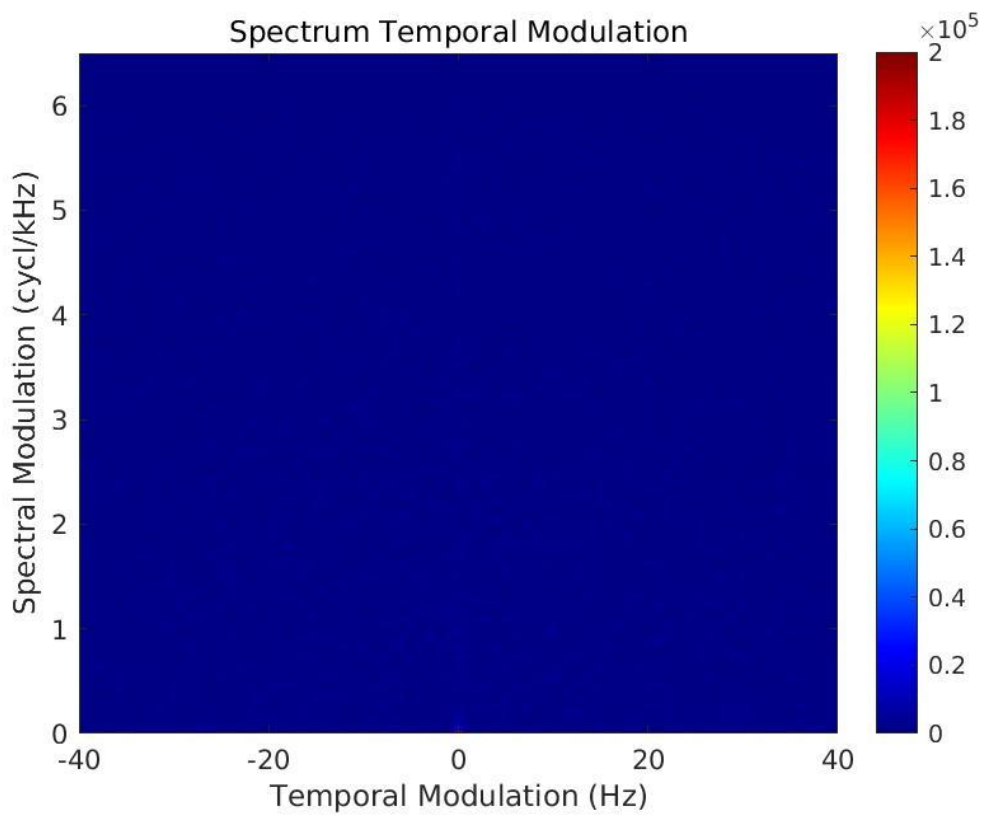


図 2.11: スペクトル傾斜のない信号（ホワイトノイズ）でのSTM情報

ブルーノイズ (+3 dB/oct)

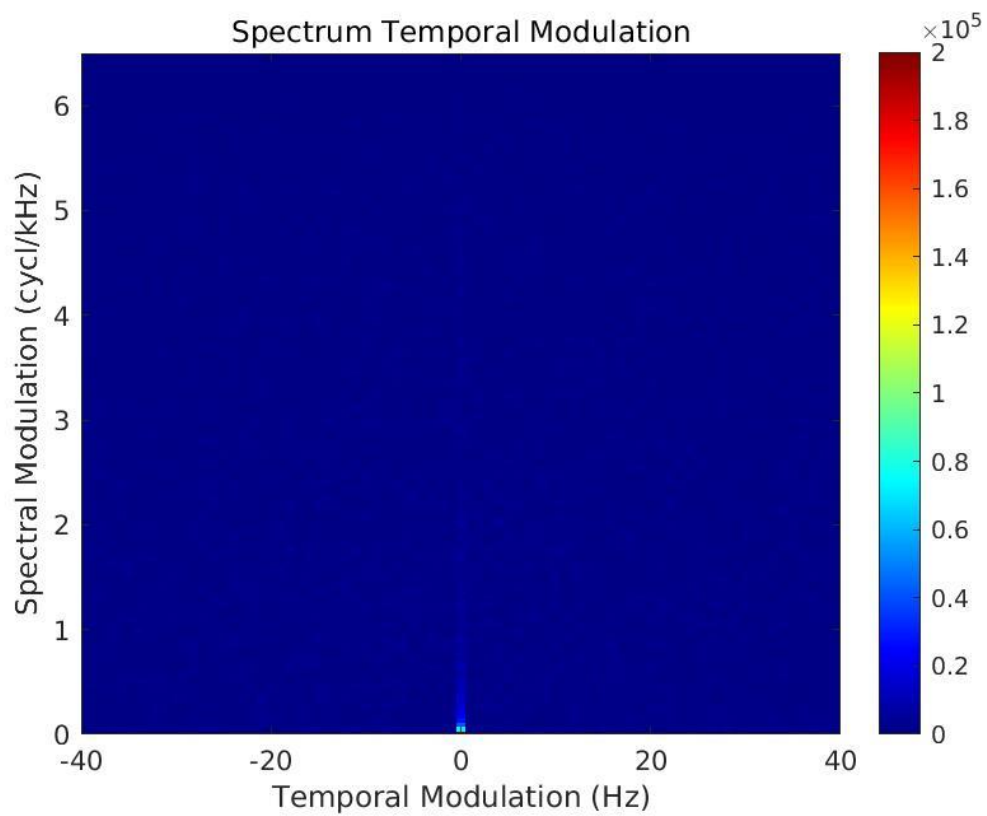


図 2.12: スペクトル傾斜 (+3 dB/oct) のある信号での STM 情報

$$f_m = 0 \text{ [Hz]}$$

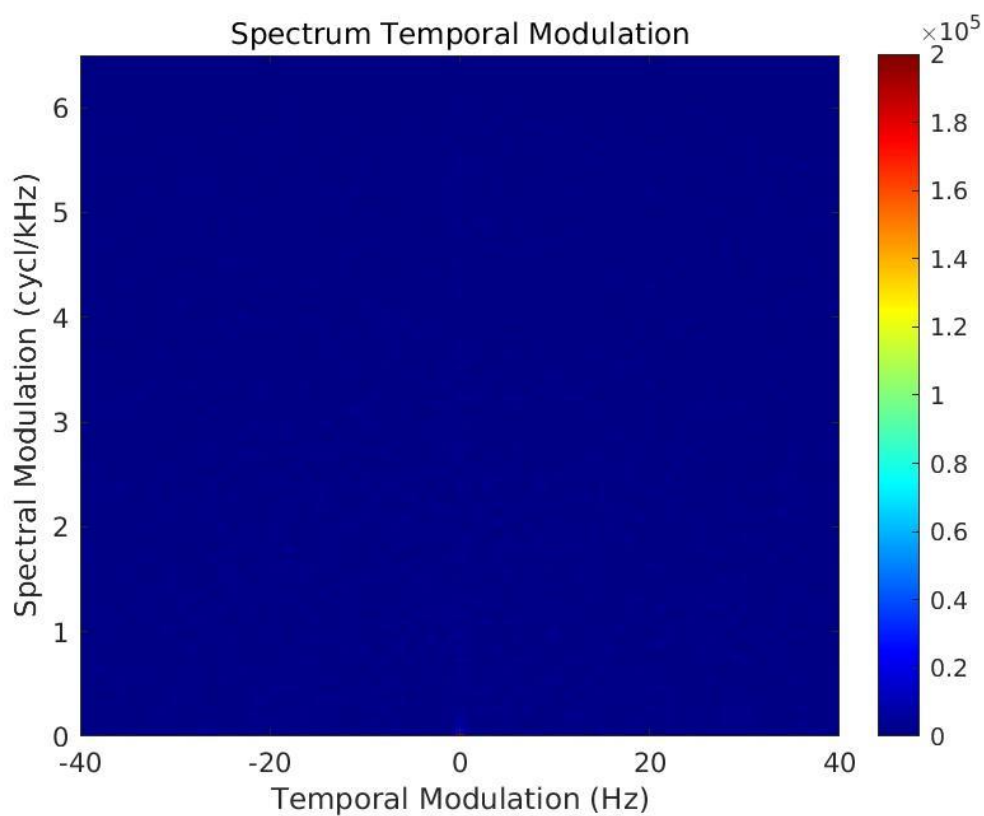


図 2.13: SM 情報の変調周波数 (0 Hz) での STM 情報

$$f_m = 1 \text{ [Hz]}$$

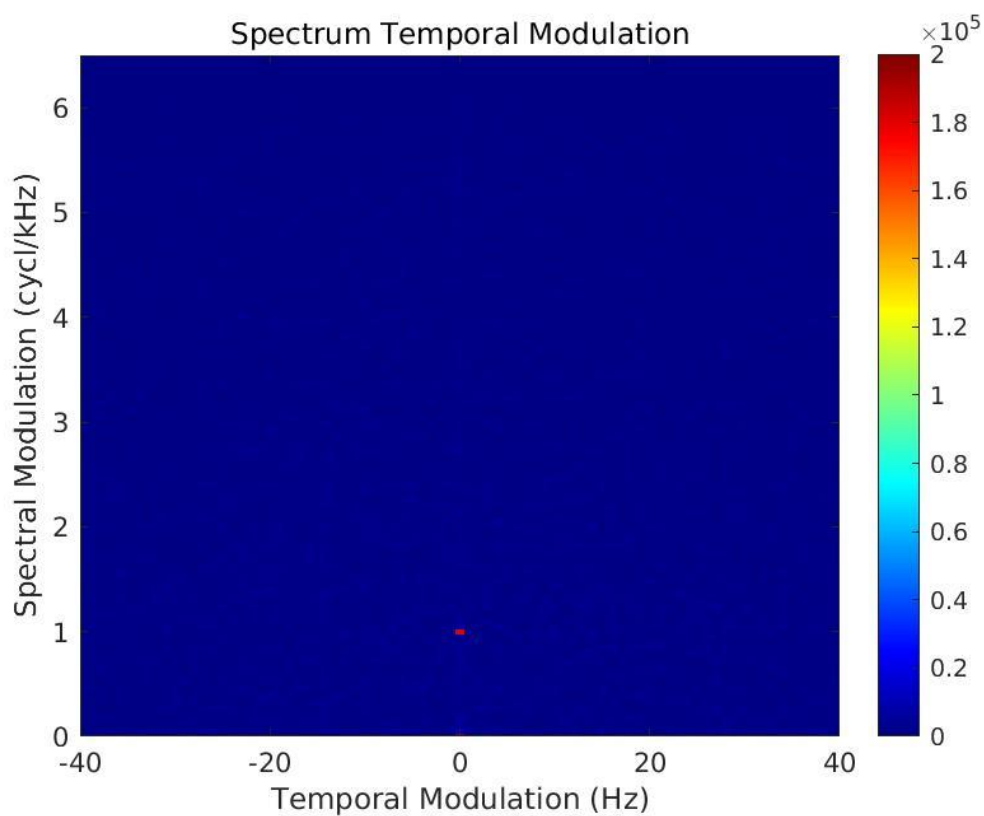


図 2.14: SM 情報の変調周波数 (1 Hz) での STM 情報

$$f_m = 3 \text{ [Hz]}$$

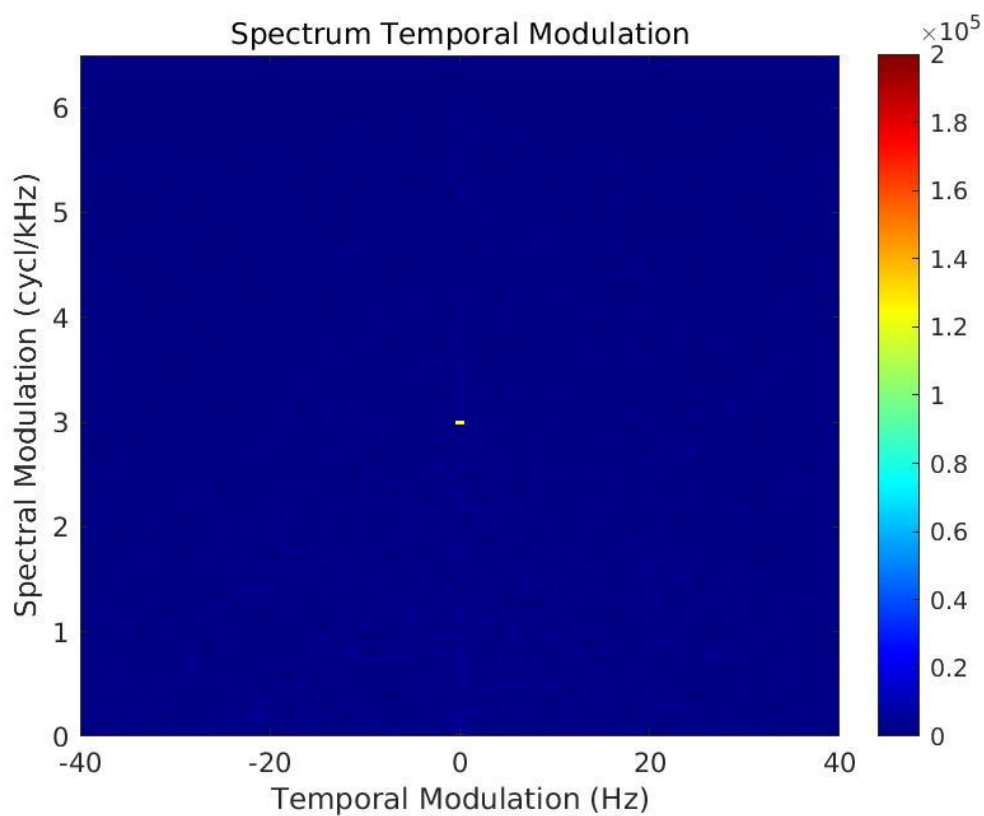


図 2.15: SM 情報の変調周波数 (3 Hz) での STM 情報

$$f_m = 6 \text{ [Hz]}$$

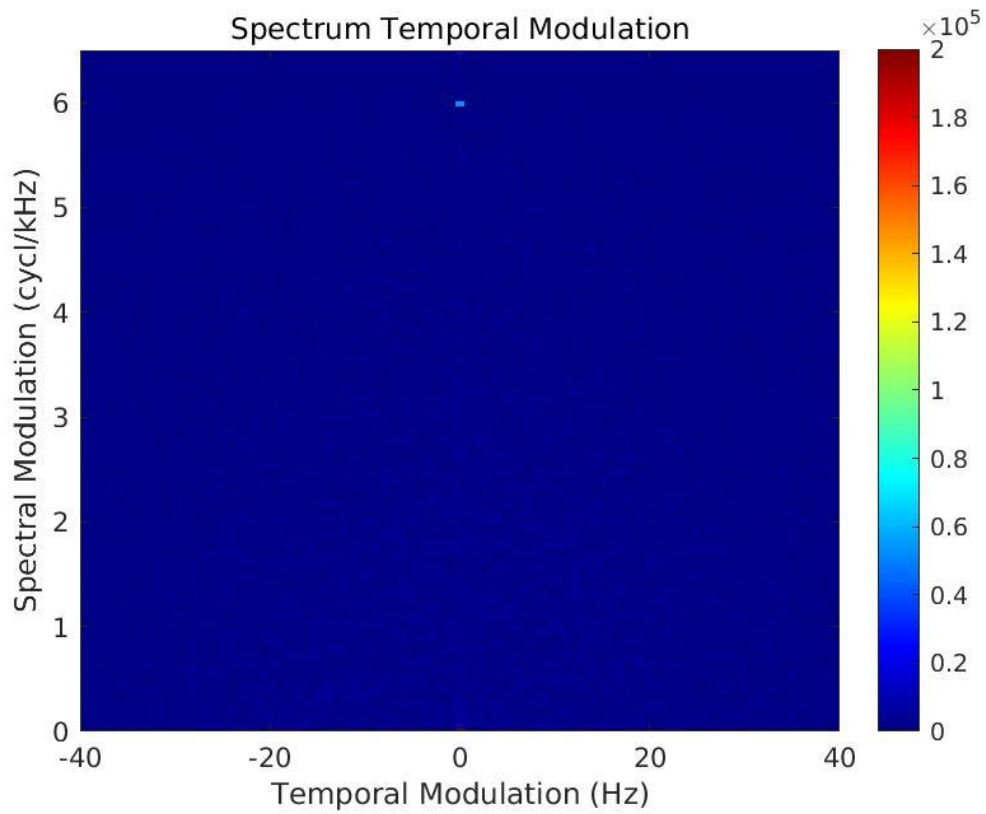


図 2.16: SM 情報の周波数 (6 Hz) での STM 情報

2.2.4 STM 情報と振幅変調調波複合音

スペクトル構造が存在する状態で振幅変調周波数の変化に伴う STM 情報の変化を観察するために、倍音構造のスペクトル構造をもつ調波複合音を振幅変調させた振幅変調調波複合音（AM 調波複合音（AM HCT））を STM 情報分析した。搬送波となる調波複合音は式 2.3 より、基本周波数 f_0 とその倍音の周波数 $f_1 \dots f_n$ の正弦波の足し合わせとし、基本周波数を 300 Hz とした。式中の n は調波複合音の次数である。振幅変調波は式 2.3 より、振幅変調周波数 f_a を 0 Hz, 15 Hz, 30 Hz の正弦波とし作成、分析した。各振幅変調周波数ごとの調波複合音の分析結果を図 2.17, 図 2.18, 図 2.19 に示す、分析結果を観察すると SM 情報のスペクトルと TM 情報のスペクトルの交わる座標にスペクトルが現れた。このスペクトルは SM 情報に対しては調波の基本周波数の逆数と対応しており、TM 情報に対しては振幅変調周波数が対応していた。よって、観察されたスペクトルは SM 情報と TM 情報の相互作用により現れた STM 情報の変化であると考えられる。結果より STM 情報と AM 調波複合音の対応関係がわかった。

$$\text{AM HCT} = \sum_{i=0}^n (1 + \sin(2\pi f_a t) \cdot \sin(2\pi f_i t)) \quad (2.3)$$

$$f_a = 0 \text{ [Hz]}$$

$$f_0 = 300 \text{ [Hz]}$$

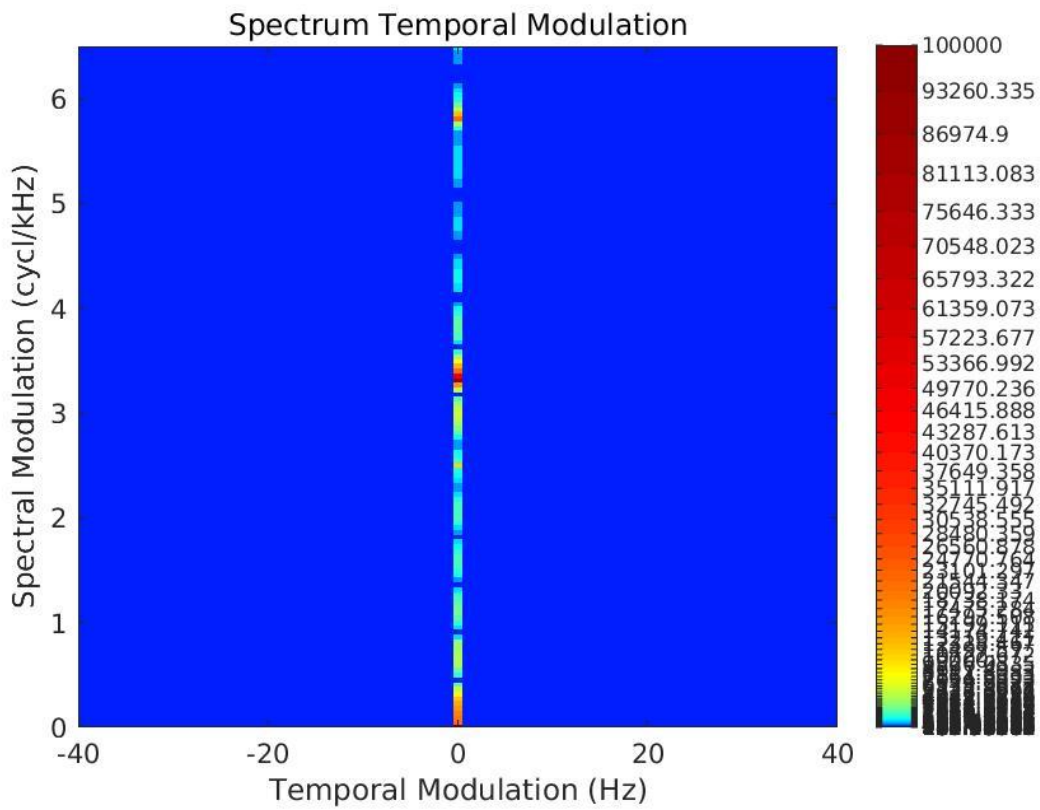


図 2.17: 振幅変調周波数 (0 Hz) の調波複合音の STM 情報

$$f_a = 15 \text{ [Hz]}$$

$$f_0 = 300 \text{ [Hz]}$$

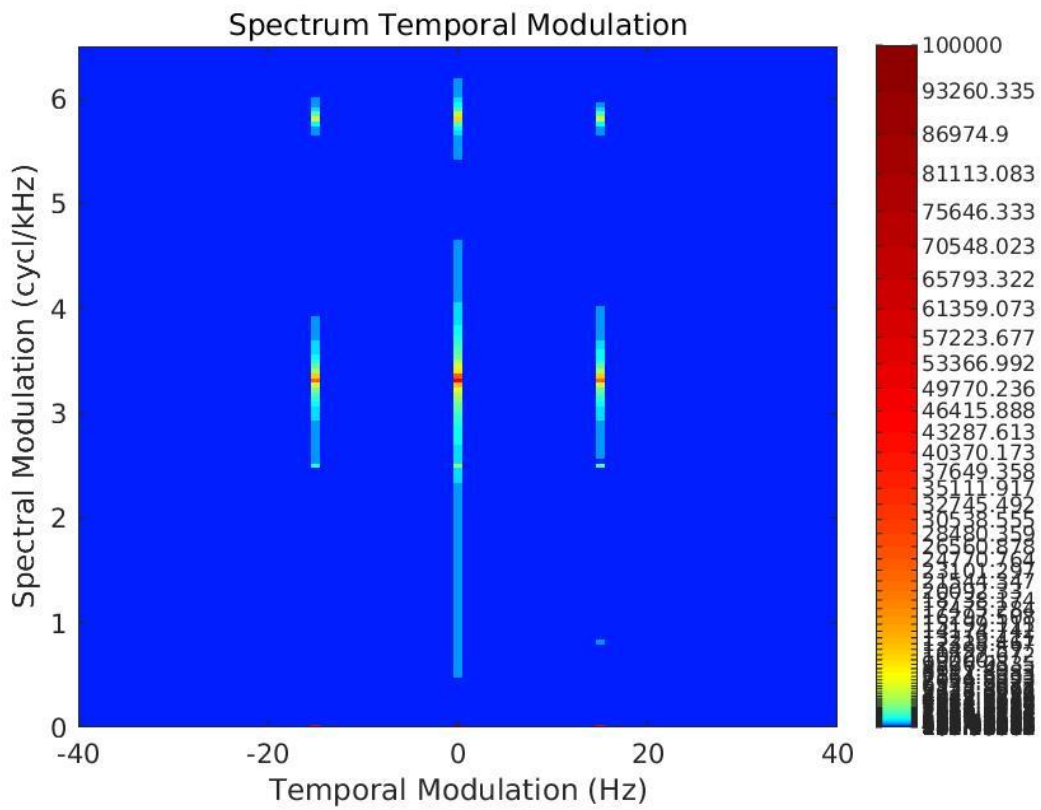


図 2.18: 振幅変調周波数 (15 Hz) の調波複合音の STM 情報

$$f_a = 30 \text{ [Hz]}$$

$$f_0 = 300 \text{ [Hz]}$$

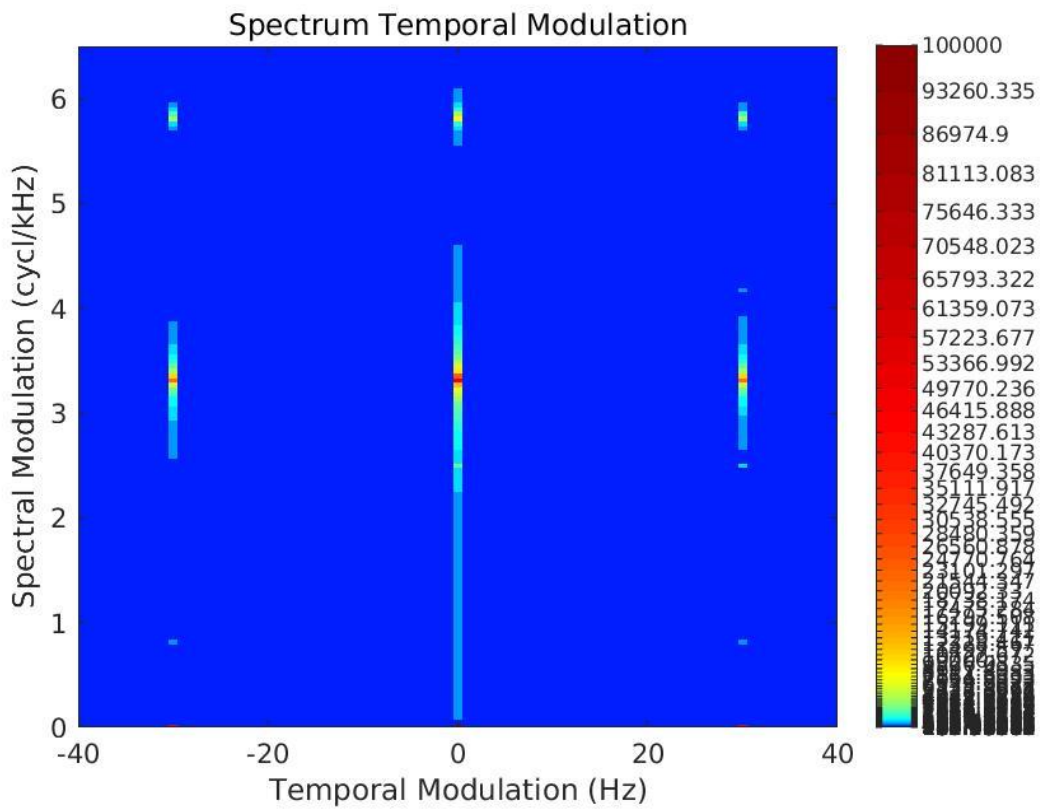


図 2.19: 振幅変調周波数 (30 Hz) の調波複合音の STM 情報

2.2.5 STM 情報と周波数変調信号

周波数の変化に伴う STM 情報の変化を観察するために周波数変調信号 (FM 信号) を分析した. FM 信号は [34] を参考に式 2.4 より導出した, 式中の I は変調指数であり, 式 2.5 を用いて作成した. 搬送波周波数 f_c は 5 kHz の正弦波とし, 周波数偏差 f_{dev} は 2 kHz とした. この条件下において周波数変調周波数 f_m のみを 1 Hz, 3 Hz, 6 Hz の正弦波で変調させ STM 情報を観察した. 各周波数変調周波数ごとの分析結果を図 2.20, 図 2.21, 図 2.22 に示す, STM 情報が V 字型に点で並ぶスペクトルが現れた. この, FM 信号の周波数変調周波数を変化させると, STM 情報の V 字型のスペクトルの角度と点の広がり方が変化する結果より, STM 情報と FM 信号の対応関係がわかった.

$$\text{FM signals} = \sin(2\pi f_c t + I \sin(2\pi f_m t)), \quad (2.4)$$

$$I = \frac{f_{dev}}{f_m}. \quad (2.5)$$

$$f_m = 1 \text{ [Hz]}$$
$$f_{dev} = 2000 \text{ [Hz]}$$

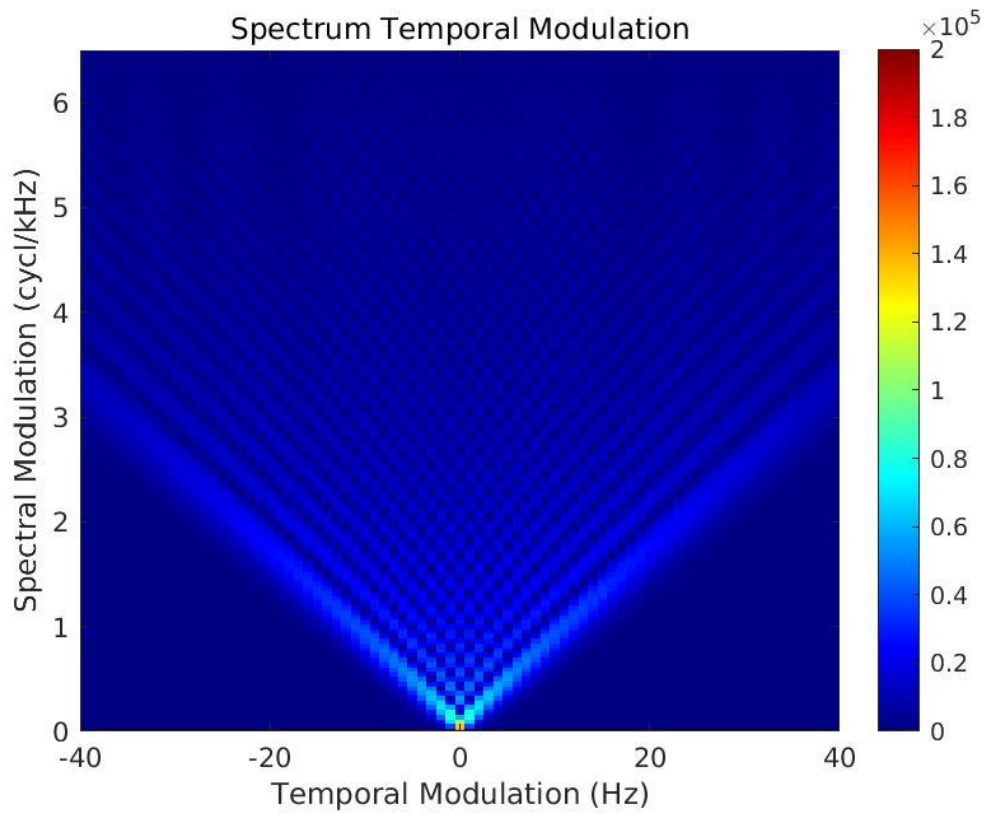


図 2.20: 周波数変調周波数 (1 Hz) での STM 情報

$$f_m = 3 \text{ [Hz]}$$
$$f_{dev} = 2000 \text{ [Hz]}$$

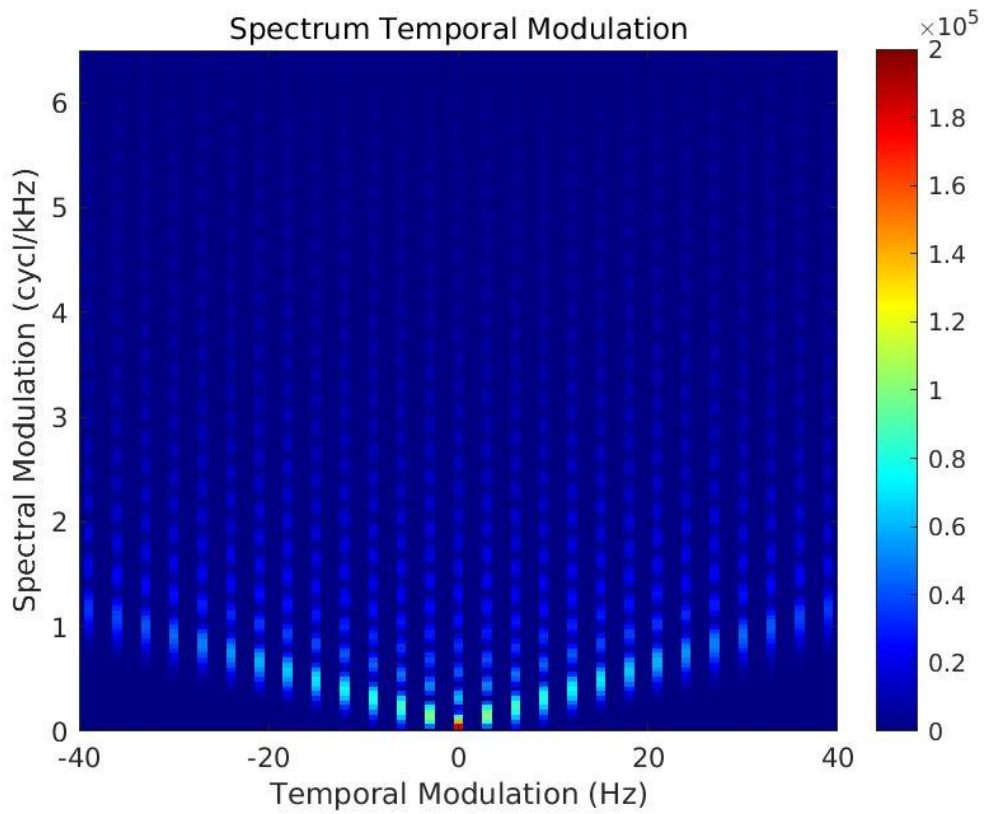


図 2.21: 周波数変調周波数 (3 Hz) での STM 情報

$$f_m = 6 \text{ [Hz]}$$
$$f_{dev} = 2000 \text{ [Hz]}$$

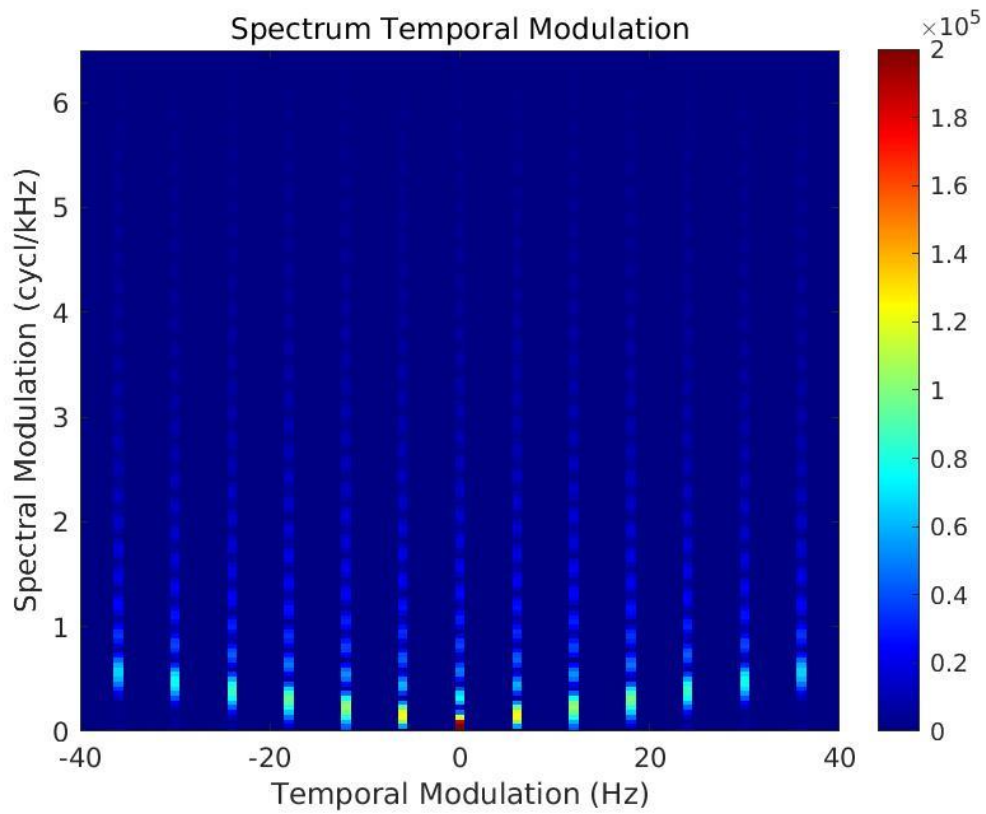


図 2.22: 周波数変調周波数 (6 Hz) での STM 情報

2.3 STM 情報分析の性質のまとめ

本節において、ここまで分析してきた STM 情報分析と各音響特徴との関係をまとめるとめる。

I, 振幅変調周波数が増加すると STM 情報分析において時間軸方向の変化を表す, TM 情報が変化する。

II, 調波複合音の基本周波数 (スペクトル構造) を変化させると STM 情報分析においてスペクトル軸方向の変化を表す, SM 情報が変化する。

III, ホワイトノイズのスペクトル構造を変化させると STM 情報分析においてスペクトル軸方向の変化を表す, SM 情報が変化する。

IV, AM 調波複合音の振幅変調周波数を変化させると STM 情報分析においてスペクトル軸方向と時間軸方向, 相互作用による変化を表す, STM 情報が変化する。

V, FM 信号の周波数変調周波数をスペクトル・時間変調情報分析においてスペクトル軸方向と時間軸方向, 相互作用による変化を表す, スペクトル・時間変調情報情報が変化する。

以上が STM 情報分析と各音響特徴との関係である, ここまでに求めてきた STM 情報分析の性質をもとに次章より, 実際に聴覚的顕著性と STM 情報の関係を検討していく。

第3章 検討対象とした刺激と聴覚的 顕著性の関係の分析

前章までに STM 情報分析と各音響特徴の関係を調査した。本章においては実際に STM 情報分析から求めた SM 情報と TM 情報に関わる音響特徴と聴覚的顕著性の関係を調査する。このために先行研究 [9] においてサーストンスケールでの顕著性が判明している 10 種類の刺激を分析対象とした。図 3.1 に分析対象とした刺激の聴覚的顕著性のサーストンスケール（顕著性スケール）を示す。

各刺激の内容は表 3 に示す。刺激の長さは全て 0.5 sec であり，サンプリング周波数は 44.1 kHz である。

図 3.2 (a) に分析した各刺激のサウンドスペクトログラムを示す，横軸は時間 (sec)，縦軸は周波数 (Hz) である。このスペクトログラムを基に図 3.2 (b) の STM 情報を算出した，横軸は TM 情報 (Hz)，縦軸は SM 情報 (cyc/kHz) である。図 3.2 (c) は (b) の STM 情報のより TM 情報 0 Hz での SM 情報を取り出したものであり，横軸は SM 情報 (cyc/kHz)，縦軸はパワーを示す。図 3.2 (d) は (b) の STM 情報のより SM 情報 0 cyc/kHz での TM 情報を取り出したものであり，横軸は TM 情報 (Hz)，縦軸はパワーを示す。これらの結果と先行研究において示された知見より TM 情報ないし SM 情報において顕著性にかかわると考えられる特徴に対し分析を行った。

表 3.1: 本研究で用いた顕著性の判明している刺激

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

Saliency

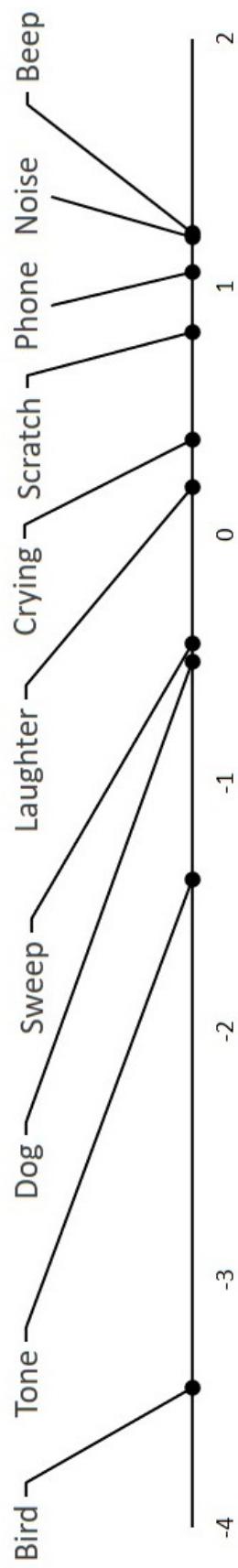


図 3.1: 分析対象とした刺激の顕著性 (横軸はサーストンスケール)

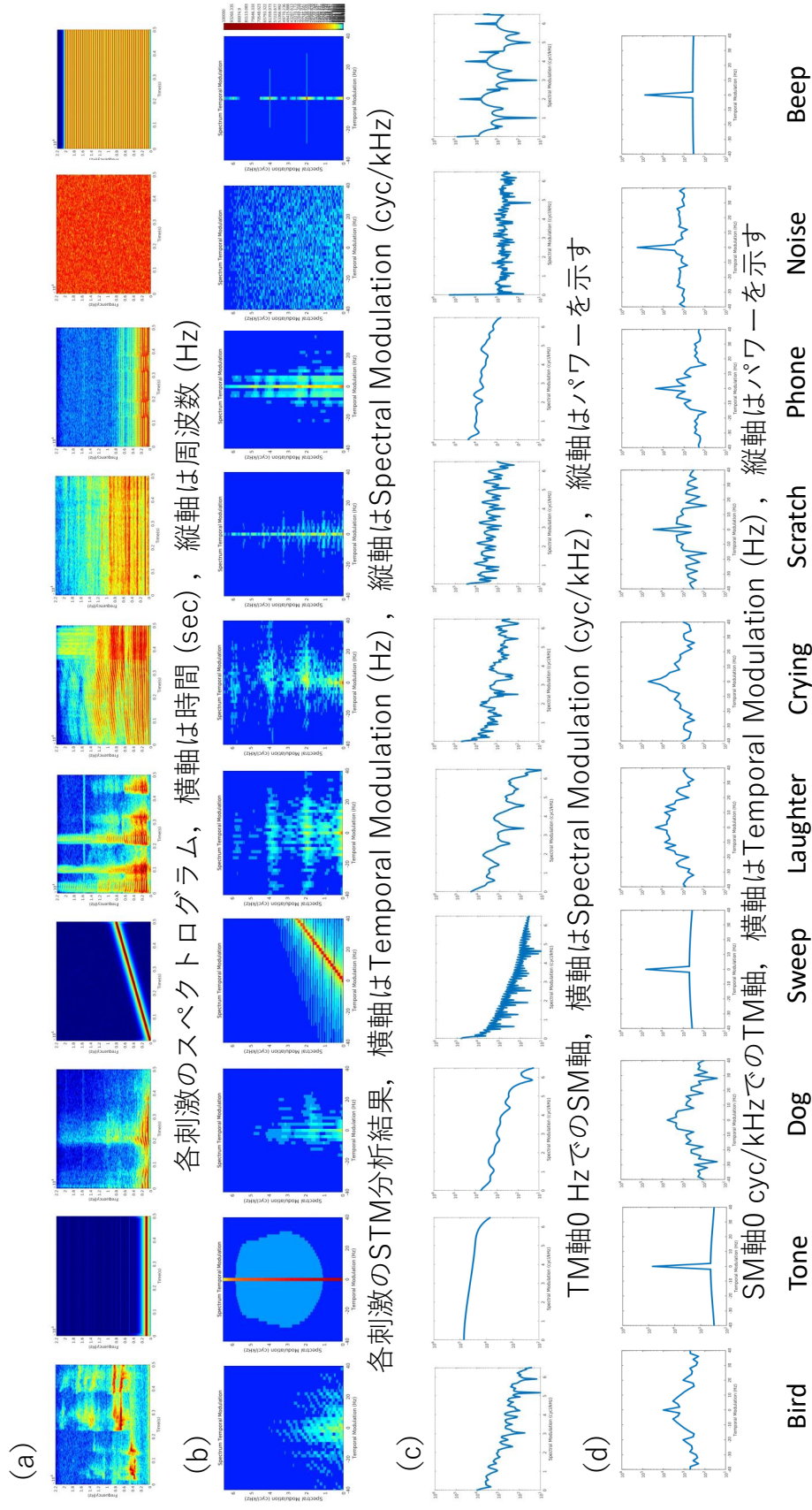


図 3.2: (a) 各刺激のサウンドスペクトログラム. (b) 各刺激のSTM情報. (c) TM情報0 HzでのSM情報. (d) SM情報0 cyc/kHzでのTM情報.

3.1 平均パワーの分析

3.1.1 着目理由

Kayser らを始めとする聴覚的顕著性モデル [5, 7, 8] ではスペクトログラム上の強度 (intensity) を特徴量の一つとして利用していた。また, Wang らもパワースペクトル密度を用いてモデルを作成したことからパワーと顕著性の間に関係があると考え, 調べた。STM 情報分析においては強度は平均パワーに対応するため平均パワーの大きさと顕著性の相関を調べた。

3.1.2 算出方法

STM 情報分析においては TM 情報 0 Hz, SM 情報 0 cyc/kHz に平均パワー現れることから, 図 3.2 (c) に示した TM 情報 0 Hz での SM 情報より, SM 情報 0 cyc/kHz でのスペクトルのパワーを求め, [31] を参考に顕著性との相関係数を調査した。

3.1.3 分析結果

図 3.3 に各刺激の平均パワーと顕著性スケール上の値について各刺激をプロットした散布図を示す。横軸は顕著性スケール, 縦軸は平均パワーである。散布図中の r , p は, 顕著性と調査した特徴との相関係数と p 値である。各プロット点での刺激名の下線は 3.1 で示した顕著性スケールにおいてマイナス側の刺激を表している。図 3.3 より顕著性の高い Noise と Beep はパワーも大きかった, 逆に顕著性の低い Bird と Dog はパワーも小さかった。これら以外の刺激の結果では Laughter, Phone, Scratch が一か所に集まる傾向が見られた。平均パワーと顕著性の相関係数は $r = 0.39$, $p = 0.19$ となり弱い相関が見られた。

3.1.4 考察

3.1 節の平均パワーに対する分析結果において, 平均パワーは聴覚的顕著性モデル [5, 8, 16] の強度ないしパワースペクトル密度と対応することから, 平均パワーが大きいと顕著性が高いと考えた。結果は, 平均パワーの大きい Noise, Beep, 平均パワーの小さい Bird, Dog は顕著性との関係が見られ, 弱い相関がみられた。これは, 平均パワーでは顕著性すべてを説明することはできないものの, Kayser らが特徴の一つとして挙げたように, 聴覚的顕著性を構成する特徴の一つであることが改めて示唆している。

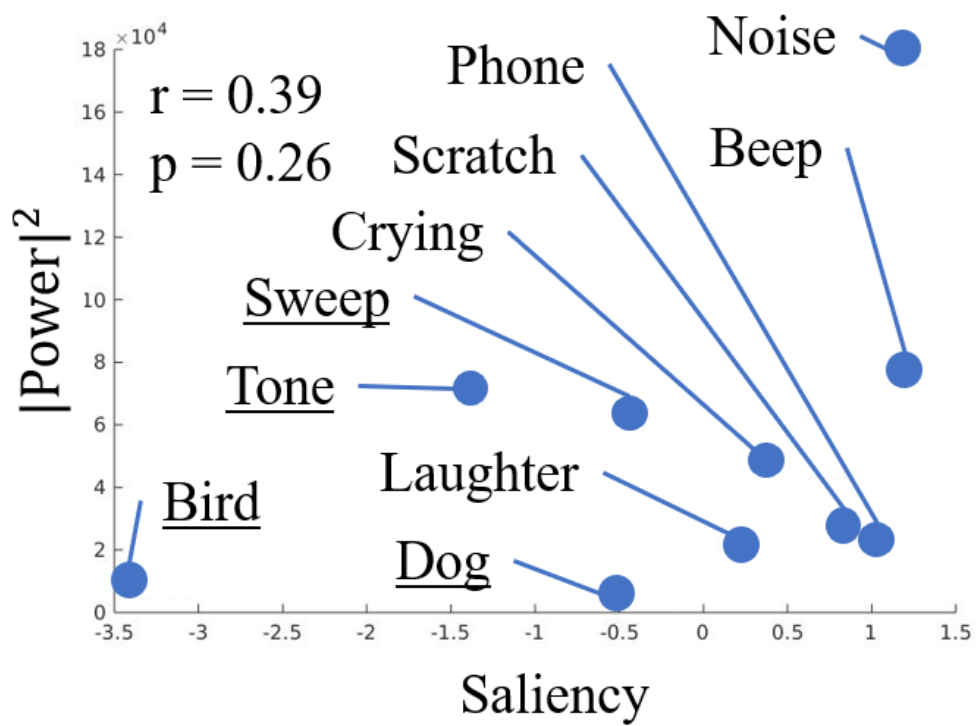


図 3.3: 各刺激の平均パワー

3.2 周波数スペクトルの広がり分析

3.2.1 着目理由

Huang & Elhilali[10] は刺激全体の周波数スペクトルの構造が顕著性にかかわることを示している。また, Tordini[11, 12] とも周波数スペクトルの構造に着目してモデルを構築している, STM 情報分析では分析対象の信号の周波数スペクトルが広帯域の場合は SM 情報 0 cyc/kHz に鋭いスペクトルが現れ, 刺激の周波数スペクトルが狭帯域の場合は SM 情報 0 cyc/kHz に鈍いスペクトルが現れる。よって, STM 情報分析のスペクトルの鋭さを分析することで刺激の周波数スペクトル構造の一つとしての周波数スペクトルの広がり顕著性の関係を検討できると考えた。

3.2.2 算出方法

先に述べたとおり刺激の周波数スペクトルの広がり STM 情報分析においてスペクトルの鋭さとして現れる。本研究では, 周波数スペクトルの広がりを調べるために, スペクトルの鋭さをスペクトルの傾きとして捉え, 図 3.2 (c) に示した TM 情報 0 Hz での SM 情報のうち SM 情報のかかっている部分の極小範囲での傾きを求めた。傾きを求めるための SM 情報の極小範囲はサンプリング周波数とフィルタバンクの幅から SM 情報 0 0.01 cyc/kHz とし, この区間内の点で差分を求め, 差分の平均を傾きとした。

3.2.3 分析結果

図 3.4 に求めた傾きと顕著性の散布図を示す。横軸は顕著性スケール, 縦軸は傾きの大きさを表す。求めた傾きはすべて負の傾きであったため絶対値を求め, 大きさを比較した。図の見方は図 3.3 と同様である。図 3.4 の結果より顕著性が高い Noise, Beep は傾きも大きく (スペクトルが広く), 逆に顕著性が低い Sweep, Dog, Tone, Bird は傾きが小さかった (狭帯域のスペクトルであった)。結果全体では Noise, Beep 以外の刺激は傾きが小さい傾向であった。傾きの大きさと顕著性との相関係数では $r = 0.45$, $p = 0.19$ となり弱い相関が見られた。

3.2.4 考察

3.2 節の SM 情報に対する分析結果において, Huang & Elhilali[10] らは刺激の周波数スペクトルの帯域幅と顕著性にかかわることを示したこと, Tordini[11, 12] とも Spectro centroid を利用したことから, スペクトルの構造と顕著性に関係があると考えた。そして, スペクトルの構造の一つである周波数スペクトルの広がりに着目した。Huang & Elhilali[10] の知見から周波数スペクトルの広がりが広いと

顕著性が高いと考えた。結果より、スペクトルが広い Noise, Beep, スペクトルの狭い Sweep, Tone は顕著性との関係が見られた。顕著性スケールとの相関を見ると、スペクトルの広がりとは弱い相関がみられた。このことは、スペクトルの広がりは、Huang & Elhilali が示すように聴覚的顕著性を構成する特徴の一つであることが示唆している。

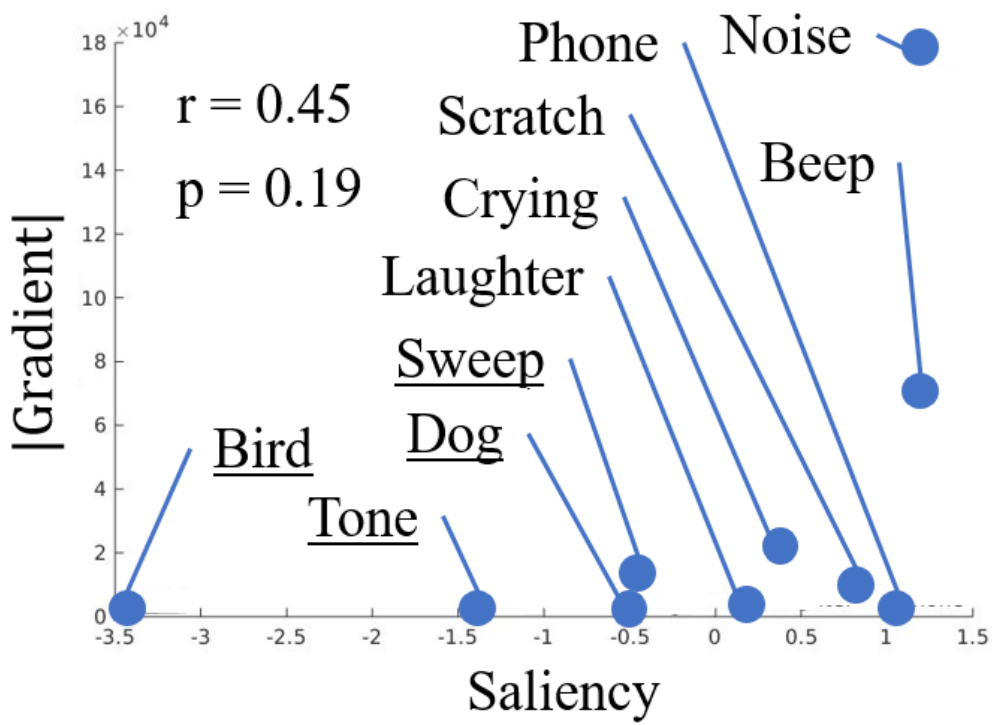


図 3.4: SM 情報極小範囲で求めた各刺激の傾き

3.3 調波性の分析

3.3.1 着目理由

Huang & Elhilali [10] は周波数スペクトル構造が顕著性にかかわることを示している。STM 情報分析では分析対象が倍音構造などの調波性のあるスペクトル構造を有していた場合、SM 情報において 0 cyc/kHz と調波の基本周波数に当たる値に 2つのスペクトルのピークが現れる。よって、SM 情報のスペクトルのピークを分析することで刺激の周波数スペクトル構造の一つとしての調波性と顕著性の関係を検討できると考えた。

3.3.2 算出方法

先に述べたとおり刺激の調波性は、SM 情報において 0 cyc/kHz と調波の基本周波数に当たる値に 2つのスペクトルのピークとして現れる。また、高次まで有する調波の場合は SM 情報で、スペクトルのピークの幅が狭くなり、低次の調波では SM 情報のスペクトルのピークの幅が広がる。このことから図 3.2 (c) に示した TM 情報 0 Hz での SM 情報を用いて、SM 情報で 0 cyc/kHz 以上の点においてピークの有無とピークの高さ、幅を調べた。SM 情報のピークを求めるために [22] を参考に TM 情報 0 Hz での SM 情報に対して正規化された自己相関を求めた。次に Lag 0 以上で閾値以上かつピークの高さが最も高いピークを調波のピークとして分析した。また、ピークの幅はピークの高さの半分の位置での幅とした。分析時のピークの閾値は自己相関の平均値とした。

3.3.3 分析結果

図 3.5 に各刺激の SM 情報の自己相関の結果および、各刺激の閾値（平均値）を横の直線で示す、横軸は Lag, 縦軸は自己相関係数を表す。図 3.6 に SM 情報の自己相関より求めたピークの高さと顕著性の散布図を示す、横軸は顕著性スケール、縦軸はピークの高さを表す。図 3.7 にピークの幅と顕著性の散布図を示す、横軸は顕著性スケール、縦軸はスペクトルのピークの幅の大きさを表す。図 3.6 及び図 3.7 の見方は、図 3.3 の見方と同様である。また図 3.5 より Tone はピークを検出することができなかったため高さ 0, 幅 0 となっている。図 3.6 の結果より、最も顕著性が高い Beep はピークも高く、顕著性の低い Tone, Dog はピークを検出できないまたはピークが低かった。結果全体では顕著性の低い刺激がピークの低い方に集まる傾向が見られた。ピークの高さと顕著性との相関係数では $r = 0.41$, $p = 0.24$ となり弱い相関が見られた。図 3.7(d) の結果では顕著性の高い Beep, Noise は幅が小さかった。結果全体では顕著性の高低に関係なくまばらに点が現れた。ピー

クの幅と顕著性との相関係数では $r = -0.13$, $p = 0.71$ となりほとんど相関傾向が見られなかった。

3.3.4 考察

3.3 節の SM 情報に対する分析結果において, Huang & Elhilali[10] らは刺激の周波数スペクトルの倍音構造と顕著性にかかわることを示したこと, Tordini[11, 12] らも Spectro centroid を利用したことから, スペクトルの構造と顕著性に関係があると考えた。そして, スペクトルの構造の一つである周波数スペクトルの調波性に着目した。Huang & Elhilali[10] の知見から周波数スペクトルが調波性を有し, 調波の次数が高ければ顕著性が高いと考えた。結果より, 調波に対する結果では調波性を有し次数の高い Beep と調波性をもたない Tone は顕著性との関係が見られた。顕著性スケールとの相関を見ると, 調波性は弱い相関があり, 調波の次数はほとんど相関傾向が見られなかった。これらのことは, 調波性は, Huang & Elhilali が示すように聴覚的顕著性を構成する特徴の一つであることが示唆している。一方, 調波の次数は聴覚的顕著性にほとんど寄与していないと考えられる。

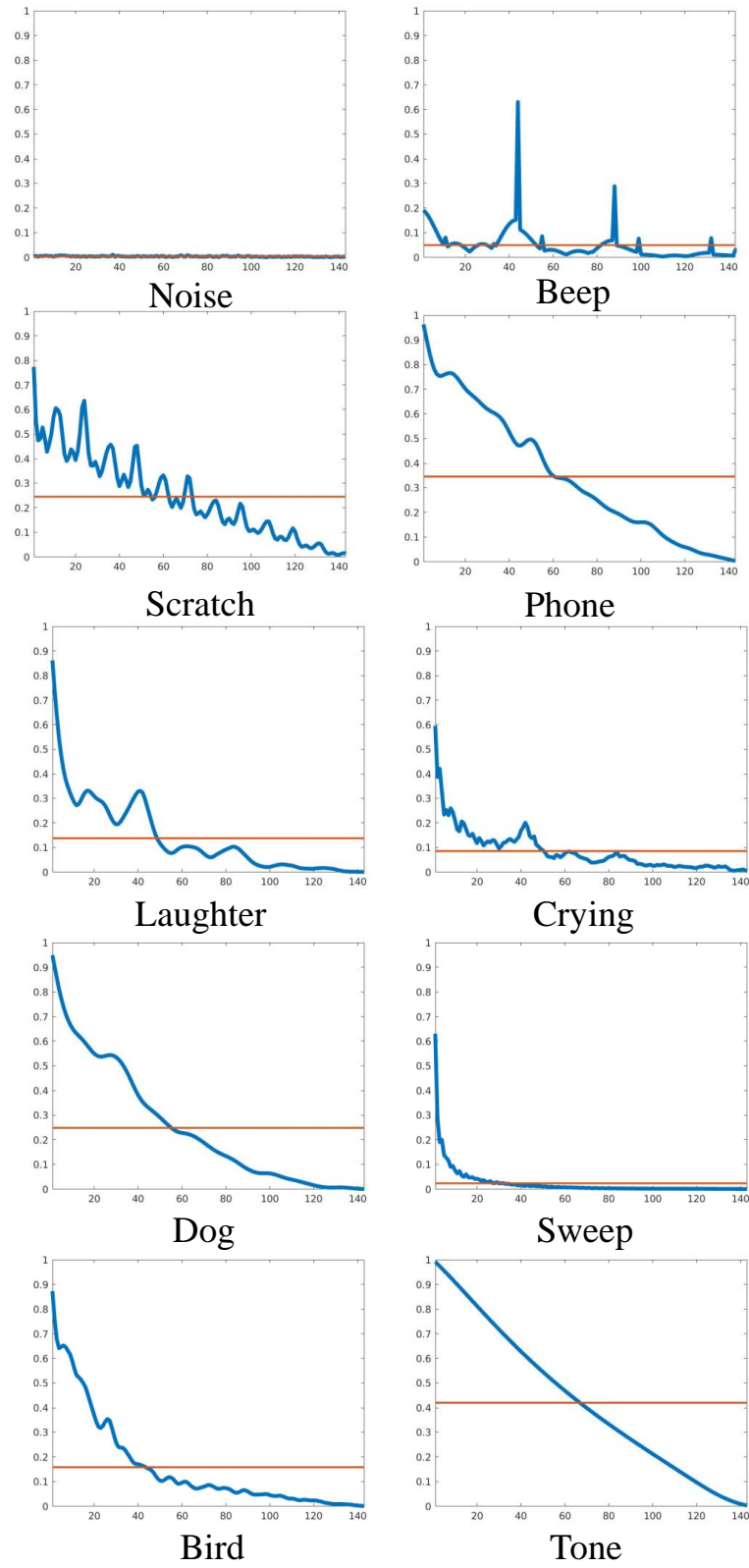


図 3.5: SM 情報の自己相関

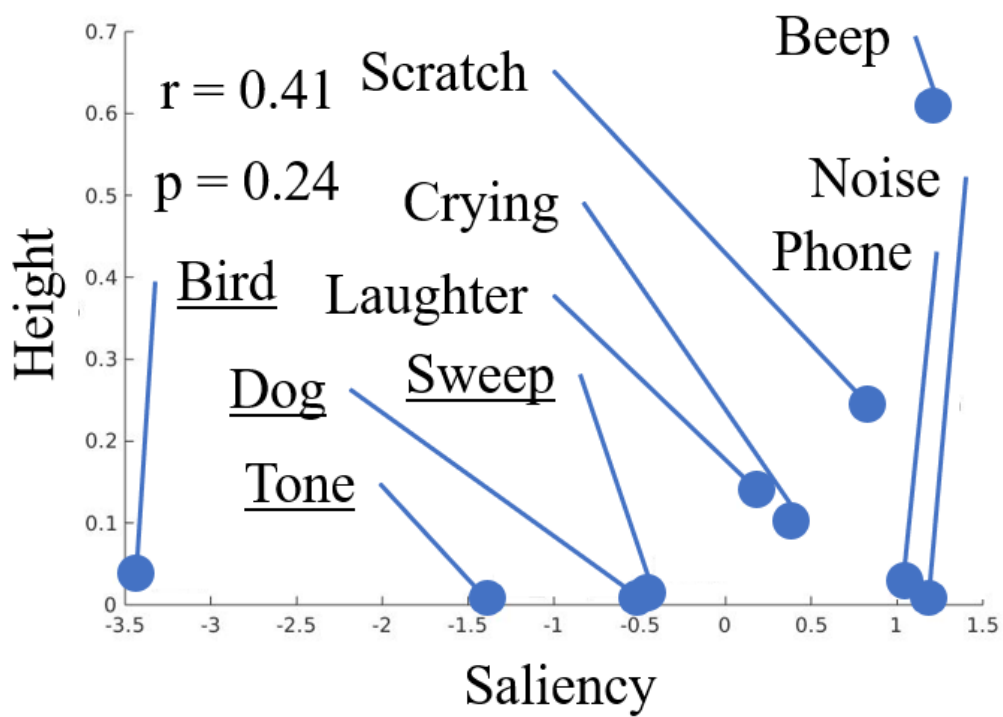


図 3.6: SM 情報の自己相関より求めた各刺激のピークの高さ

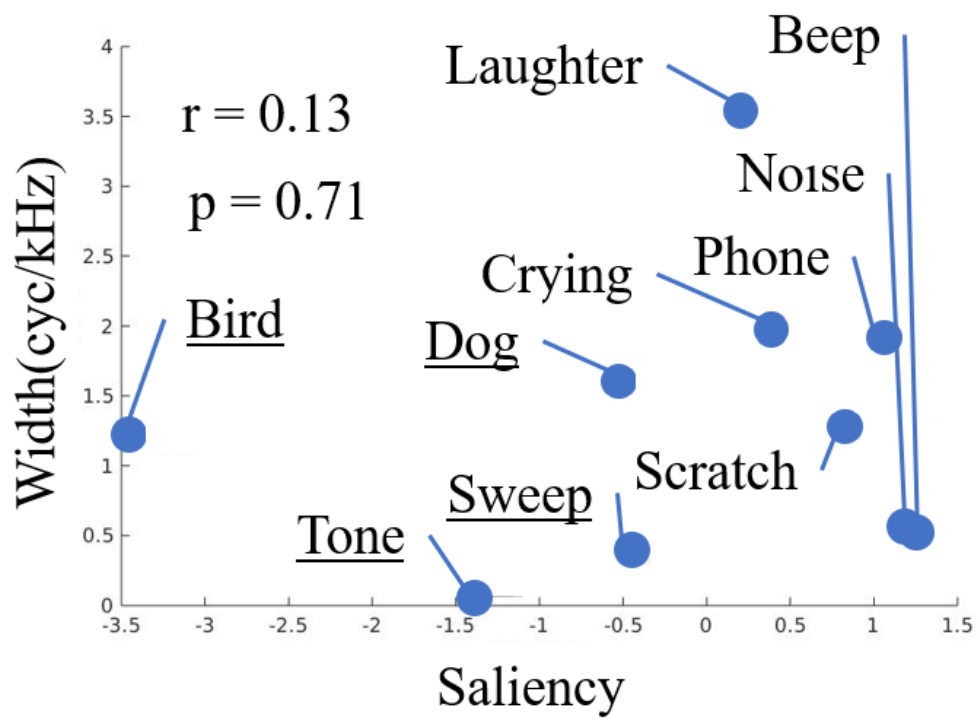


図 3.7: SM 情報の自己相関より求めたスペクトルのピークの幅

3.4 TM情報の分析

3.4.1 着目理由

Duangudom & Anderson はモデルにおいて TM 情報を特徴の一つとして採用した [8] . Tordini らも時間変調に関わる特徴の一つである Temporal centroid を利用してモデルを構築した [11, 12]. また Asemi ら振幅包絡線の時間的な変化の有無が目的音検知の手がかりになる, すなわち目立ちやすいことを示した [20]. STM 情報分析では振幅変調音などの TM 情報を持つ信号を分析した場合, TM 情報において 0 Hz と信号の変調周波数に当たる値の二箇所にスペクトルのピークが現れる. よって, 本節では TM 情報および刺激中の TM 情報の継続時間についてスペクトルのピークを分析することで顕著性の関係について検討できると考えた.

3.4.2 算出方法

先に述べたとおり, 刺激の TM 情報は TM 情報のスペクトルのピークとして現れる. また, TM 情報の継続時間が長い信号では TM 情報のスペクトルのピークの幅が狭くなり, 短い信号では TM 情報のスペクトルのピークの幅が広がる. このことから図 3.2 (d) に示した SM 情報 0 cyc/kHz での TM 情報より, TM 情報で 0 Hz 以上の点におけるスペクトルのピークの有無と高さ, 幅を調べた. TM 情報のピークを求めるために [22] を参考に SM 情報 0 cyc/kHz での TM 情報に対して正規化された自己相関を求めた. 求めた自己相関から Lag 0 以上で閾値以上かつピークの高さが最も高いピークを TM 情報のスペクトルのピークとして分析した. また, ピークの幅はピークの高さの半分の位置での幅とした. 分析時のピークの閾値は自己相関の平均値とした.

3.4.3 分析結果

図 3.8 に各刺激の TM 情報の自己相関の結果および, 各刺激の閾値 (平均値) を横の直線で示す, 横軸は Lag, 縦軸は自己相関係数を表す. 図 3.9 に自己相関の結果より求めたピークの高さと顕著性の散布図を示す, 横軸はサーストンスケールでの顕著性, 縦軸は自己相関でのピークの高さを表す. 図 3.10 にピークの幅と顕著性の散布図を示す. 横軸は顕著性スケール, 縦軸はピークの幅の大きさを表す. 図 3.9 及び図 3.10 の見方は図 3.3 の見方と同様である. また図 3.8 より Beep, Sweep, Tone では TM 情報のスペクトルのピークが検出されなかったため, 高さ 0, 幅 0 となっている. 図 3.9 より顕著性の高い Phone, Scratch はピークが高く. 顕著性の低い Sweep, Tone ではピークが低かった. 全体の結果では Laughter が最も高く, ピークを検出できなかった Beep, Sweep, Tone を除くと Crying が最も低い結果となった. ピークの高さが 0.1 以上のグループと 0.05 以下の高さのグループに分かれ

る傾向がみられた。ピークの高さと顕著性の相関係数では $r = -0.12$, $p = 0.60$ となり相関傾向がほぼ見られなかった。図 3.10 の結果より、顕著性のピークを求めることができなかった Beep, Sweep, Tone を除くと、顕著性が高い Noise, Scratch, Phone はピークの幅が広く、顕著性の低い Dog のピーク幅が小さくなった。幅の大きさでは最も幅の大きい Noise が他の刺激と大きく離れており、次いで Scratch, Phone と続き、Laughter 以降は集まる傾向がみられた。ピークの幅と顕著性との相関係数では $r = 0.38$, $p = 0.28$ となり、弱い相関が見られた。

3.4.4 考察

3.4 節の刺激の TM 情報の分析において、Asemi [20] らは振幅包絡の時間的な変化の有無が目的音検知にかかわることを示した。また Tordini [11, 12] らも Temporal centroid をモデルに利用していた。このことから TM 情報を有し、TM 情報の継続時間が長いと顕著性が高いと考えた。結果は、TM 情報を有し継続時間の長い Scratch, Phone は顕著性との関係が見られた。相関係数において TM 情報のピークの幅、ピークの高さは顕著性と弱い相関または、きわめて弱い相関傾向であった。これらのことは、TM 情報の有無は聴覚的顕著性にほとんど寄与していないと考えられる。しかし、TM 情報のある刺激においては継続時間、すなわち瞬間的な刺激の動きが聴覚的顕著性にかかわる特徴であることを示唆している。

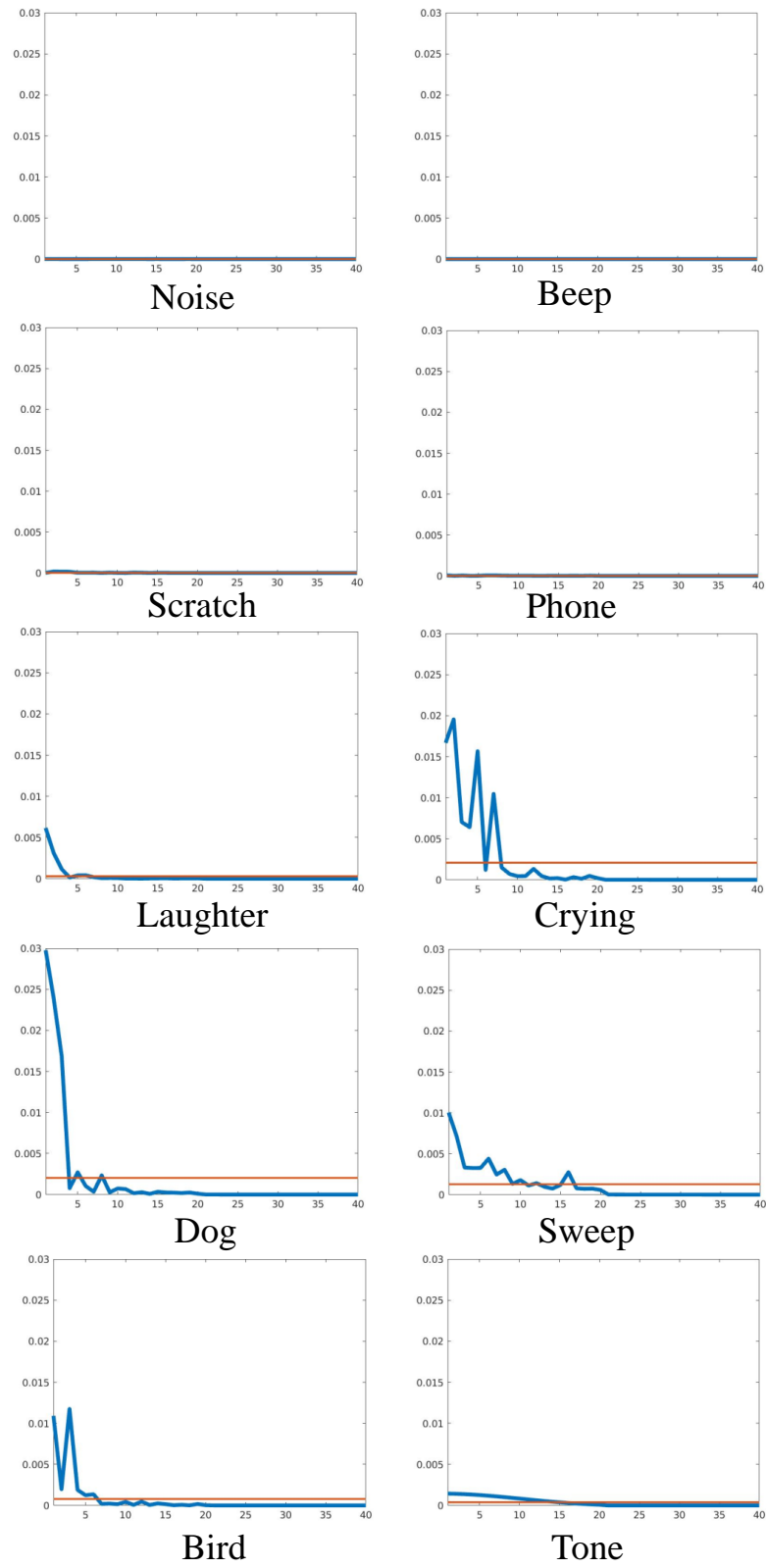


図 3.8: TM 情報の自己相関

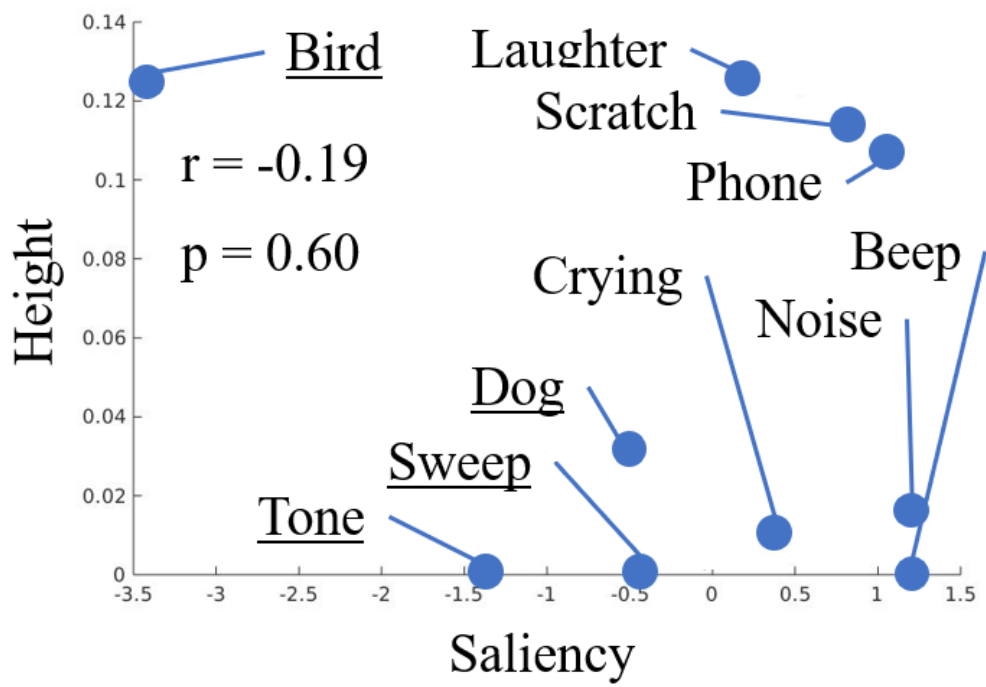


図 3.9: TM 情報の自己相関より求めた各刺激のピークの高さ

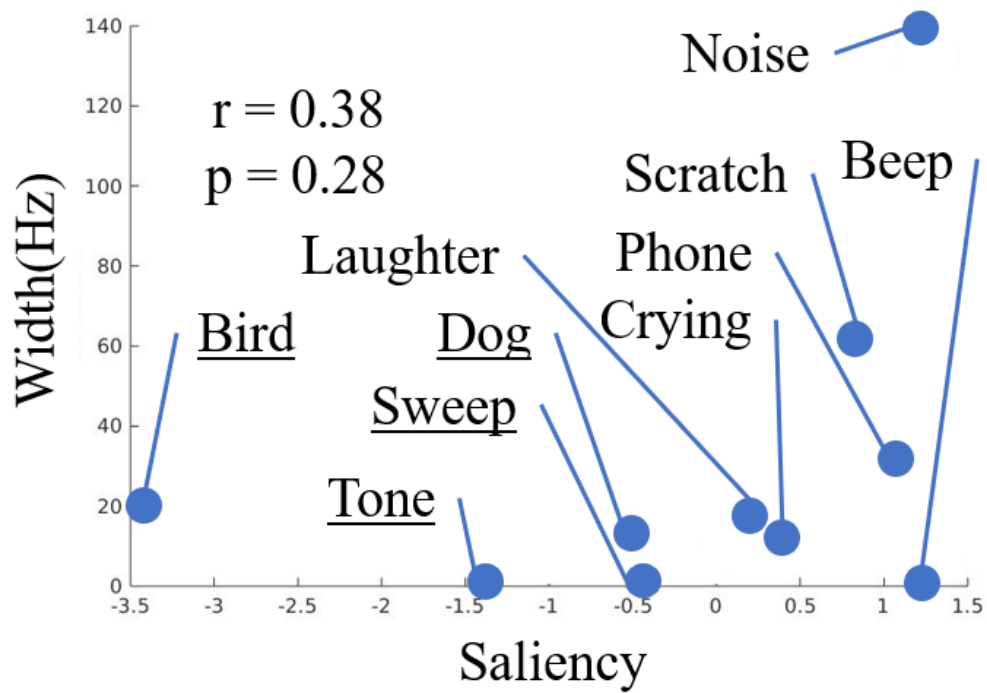


図 3.10: TM 情報の自己相関より求めたスペクトルのピークの幅

第4章 全体考察

本研究では聴覚顕著性が判明している刺激を用いて、STM 情報分析の結果から得られる音響特徴である平均パワー、周波数スペクトルの広がり、調波性、TM 情報と顕著性スケールとの相関係数を調査した。調査した音響特徴はすべて SM 情報ないし、TM 情報のみから求めたものである。Kayser ら [5] のモデルでは、音響特徴をスペクトログラム上の Intensity, Temporal contrast, Spectro contrast から個別に求めて、統合していた。Duangudom & Anderson[8] のモデルでも、音響特徴を Intensity, SM 情報, TM 情報を個別に求めてから統合していた。Wang ら [16] のモデルにおいても Kayser らの特徴量に加え、メル周波数ケプストラムとパワースペクトル密度を並列化して求め統合していた。これらの研究に基づくと、Intensity, SM 情報, TM 情報を統合すると顕著性が高くなると考えた。そこで、本研究では、SM 情報から求めた特徴と、TM 情報から求めた特徴、これに Intensity に相当する平均パワーを統合して、STM 情報と聴覚的顕著性との関係を検討した。そのために、本研究では、これら音響特徴を足し合わせ、顕著性スケールとの相関係数を求めた。その結果を図 4.1 に示す、図の縦軸は足し合わせた音響特徴 (Integrated Feature)、横軸は Liao ら [9] の顕著性スケールである。散布図中の r , p は、顕著性と統合した音響特徴との相関係数、及び p 値である。結果は、顕著性の高い Beep, Noise, 顕著性の低い Bird, Dog で顕著性との関係がみられた。散布図全体では Beep が突出しており、Noise 以下の刺激は比較的低い値で集まる傾向がみられた。相関係数において、統合した音響特徴と顕著性スケールとの関係は、弱い相関傾向であった。散布図中、Integrated Feature が比較的大きい値は Tone, Sweep, Beep, Tone となった。対して、Integrated Feature が比較的小さい値は Bird, Dog, Laughter, Crying であった。これらの音は 3.2 (a) のスペクトログラムより、人工的な音と自然な環境に分けて考えることができる。Huang & Elhilali[10] の研究においては、自然な環境音のみを対象として検討していたことから、人工的な音と自然な環境音は分けて検討した方が、それぞれで高い相関を示した可能性が考えられる。

一方で、Kayser ら、Wang ら、Duangudom & Anderson の全てモデルにおいてパワーに関する音響特徴が含まれていた。本研究において相当する、平均パワーは、TM 情報において 0 Hz の SM 情報、SM 情報において 0 cyc/kHz の TM 情報から得た値であり、両軸の交差する点から得た値として 1 点のみであるが SM 情報と TM 情報の合わさった STM 情報の音響特徴の一つと考えられる。

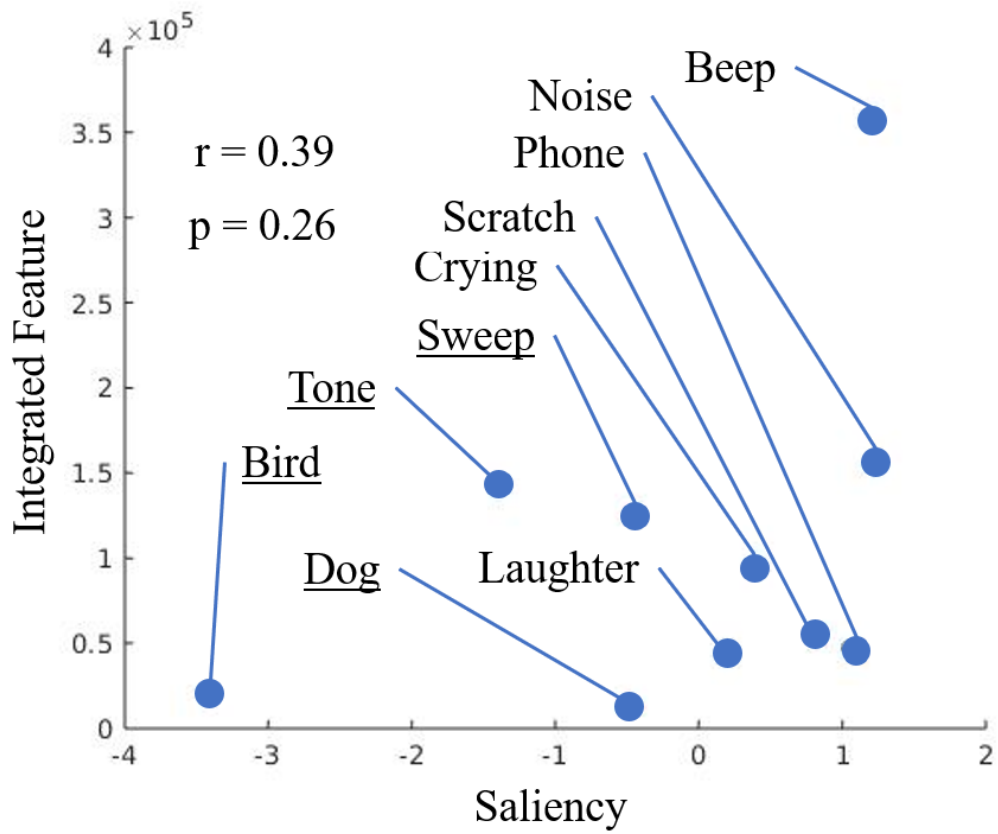


図 4.1: STM 情報の音響特徴と顕著性スケールの相関

第5章 結論

5.1 本研究により明らかにされたこと

本研究の目的は SM 情報と TM 情報の相互作用も含めて観察できる STM 情報にどのような音響特徴が現れるか検討し，現れた音響特徴がどのように聴覚的顕著性に関係するか明らかにすることである．そのため，最初に，人工的に音響特徴をもたせた刺激を作成し，STM 情報分析と音響特徴の関係を調べた．その結果，振幅変調周波数と TM 情報，調波複合音及びスペクトル構造と SM 情報，AM 調波複合音及び FM 変調周波数と STM 情報の関係を明らかにした．

次に，聴覚的顕著性が既に判明している刺激を STM 情報分析し，STM 情報の結果から得られた音響特徴と顕著性との関係を相関係数をもって検討した．その結果，調査した個々の音響特徴と聴覚的顕著性との間で求めた相関係数は最も高いもので周波数スペクトルの広がりとの関係を調べた Spectro Modulation の傾きでの $r = 0.45$, $p = 0.19$ であった．対して最も小さいものは，調波性との関係を調べたスペクトル変調情報軸のスペクトルのピークの幅であり $r = 0.13$, $p = 0.71$ であった．STM 情報分析から求めた，それぞれの音響特徴と顕著性の相関係数は異なるものの，聴覚的顕著性に全く関係のない音響特徴は無いという結果を得られた．この結果から，聴覚的顕著性に関わる音響特徴は平均パワー，周波数スペクトルの広がり，調波性，TM 情報が関係することを明らかにした．

これらの結果より本研究において用いた STM 情報分析で聴覚的顕著性を検討することが可能であることが示唆された．

5.2 残された課題

5.2.1 STM 情報と顕著性の分析

本研究においては聴覚的顕著性に寄与する音響特徴として SM 情報と TM 情報に関係する音響特徴が聴覚的顕著性に寄与することが明らかにされた．しかし，この結果においてはこの2つの情報を一体にして分析してはいたものの相関係数を求めるために利用した特徴量自体はそれぞれ個別の情報から算出したものであった．そのため，2つの情報を一体にした特徴量と聴覚的顕著性との相関を求めることも必要であると考えられる．

5.2.2 時間変化による音響特徴の変化も含めた分析

本研究で用いたSTM情報分析は入力された刺激を一枚のサウンドスペクトログラムとして処理していた。そのため、[13, 14, 10]らが聴覚的顕著性に関わると示したラウドネス、テンポ、調波性といった時間の流れと共に時々刻々と変化し続けている音響特徴は考慮されていない、しかし、これら研究の知見からの時間ともに変化する特徴も考慮する必要があると考えられる。このため、刺激全体をスペクトログラムとせず、短時間ごとに分割しそれぞれのスペクトログラムを求めてからSTM情報分析を行うといった手法を用いることで、時々刻々と変化する特徴も考慮したSTM情報と顕著性に関係を検討可能であると考えられる。

参考文献

- [1] 大串健吾, 音響聴覚心理学, 誠信書房, 東京, 2019.
- [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] 日本音響学会, 音響キーワードブック, コロナ社, 東京, 2016.
- [4] E. M. Kaya, M. Elhilali, “Modeling auditory attention A review,” *Philos. Trans. R. Soc. B: Biol. Sci*, vol. 372, no. 1714, pp. 1–10, 2017.
- [5] C. Kayser, C. Petkov, M. Lippert and N. K. Logothetis, “Mechanisms for allocating auditory attention: an auditory saliency map,” *Curr. Biol*, vol. 15, no. 21, pp. 1943–1947, 2005.
- [6] A. Borji, L. Itti, “State-of-the-art in visual attention modeling,” *IEEE Trans. Pattern Anal.*, vol. 35, no. 1, pp. 185–207, 2013.
- [7] O. Kalinli, S. Narayanan, “A saliency-based auditory attention model with applications to unsupervised prominent syllable detection in speech,” *Interspeech-2007*, pp. 1941 - 1944, 2007.
- [8] V. Duangudom, D. V. Anderson, “Using auditory saliency to understand complex auditory scenes,” 15th European Signal Processing Conf, Poznan, Poland, no. 15109600, pp. 1206 - 1210, 2007.
- [9] H. Liao, S. Kidani, M. Yoneya, M. Kashino, S. Furukawa, “Correspondences among pupillary dilation response, subjective salience of sounds, and loudness,” *Psychon. Bull.* no. 10.3758/s13423-015-]0898-0, 2015.
- [10] N. Huang, M. Elhilali, “Auditory salience using natural soundscapes,” *J. Acoust. Soc. Am.*, vol. 141, no. 10. 1121, 1. 4979055, pp. 2163 - 2176, 2017.
- [11] F. Tordini, AS. Bregman, A. Cooperstock. JR, A. Ankolekar, T. Sandholm, “Toward an improved model of auditory saliency,” In *Proc. of the 19th Int. Conf. on Auditory Display*, Lodz, Poland, 2013.

- [12] F. Tordini, A.S. Bregman, A. Cooperstock. JR, “The loud bird doesn’t (always) get the worm: why computational salience also needs brightness and tempo,” In Proc. of the 21st Int. Conf. on Auditory Display, Graz, Austria: Institute of Electronic Music and Acoustics, University of Music and Performing Arts Graz, 2015.
- [13] E. M. Kaya, M. Elhilali, “A temporal saliency map for modeling auditory attention,” 46th Annu. Conf. on Information Sciences and Systems, 2012.
- [14] E. M. Kaya, M. Elhilali, “Investigating bottom-up auditory attention,” *Front. Hum. Neurosci.*, vol. 8, no. 327, pp. 1 - 12, 2014.
- [15] C. Taishih, R. Powen, S. A. Shihab, “Multiresolution spectrotemporal analysis of complex sounds,” *J. Acoust. Soc. Am.*, vol. 118, no. 10. 1121, 1. 1945807, pp. 887–906, 2005.
- [16] J. Wang, K. Zhang, K. Madani, C. Sabourin, “2015 Salient environmental sound detection framework for machine awareness,” *Neurocomputing* vol. 152, pp. 444–454.
- [17] N. C. Singh, F. E. Theunissen, “Modulation spectra of natural sounds and ethological theories of auditory processing,” *J. Acoust. Soc. Am.*, vol. 114, issue. 6, no. 10. 1121, 1.1624067, pp. 3394–3411, 2003.
- [18] S. Hurukawa, “Processing of temporal information in the auditory system,” *Audiology Japan.*, vol. 59 no. 6, pp. 615–622, 2016.
- [19] T. M. Elliott, F. E. Theunissen, “The modulation transfer function for speech intelligibility,” *PLoS Comput. Biol.*, vol. 5, no. 3, pp. 1–14, 2009.
- [20] N. Asemi, Y. Sugita, Y. Suzuki, “Auditory search asymmetry between pure tone and temporal fluctuating sounds distributed on the frontal - horizontal plane,” *Acoust. Sci. & Tech.*, vol. 24, no. 3, pp. 145–147, 2003.
- [21] 松尾博, やさしいフーリエ変換, 森北出版, 東京, 1986.
- [22] 青木直史, デジタルサウンド処理入門, CQ 出版社, 東京, 2006.
- [23] 青木直史, “はじめての音声信号処理とサウンドプログラミング,” *日本音響学会誌*, vol. 73, no. 4, pp. 230–238, 2017.
- [24] H. Zhivomirov, “A Method for Colored Noise Generation,” *Romanian Journal of Acoustics and Vibration*, vol. 15, No. 1, pp. 14–19, 2018.

- [25] P. Handel, A. Chung, “Noise in physical systems and 1/f fluctuations,” New York, AIP, 1993.
- [26] E. Milotti, “1/f noise A pedagogical review,”
Online at : <http://arxiv.org/abs/physics/0204033>, Last accessed on Nov. 20th, 2020.
- [27] B. Shepard, Refining sound: A practical guide to synthesis and synthesizers, Oxford, Oxford University Press, 2013.
- [28] M. Schroeder, C. Fractals, Power Laws: Minutes from an Infinite Paradise, New York, Dover Publications, 2009.
- [29] 実吉純一, 電気音響工学, コロナ社, 東京, 1957.
- [30] R. D. Patterson, I. Nimmo. Smith, J. Holdsworth, P. Rice, “An efficient auditory filterbank based on the gammatone function,” IOC Speech Group on Auditory Modelling at RSRE, vol. 2, no. 7, pp. 2 - 33, 1987.
- [31] L. R. Rabiner, G. Bernard, Theory and application of digital signal processing, PRENTICE-HALL, New Jersey, 1975.
- [32] 岸源也, “変調方式の基礎,” 自動制御, vol. 5, no. 1, pp. 21–28, 1958.
- [33] R. Drullman, “Temporal envelope and fine structure cues for speech intelligibility,” J. Acoust. Soc. Am., Vol. 97, No. 1, pp. 585 - 592, 1995.
- [34] 石井聡, 無線通信とデジタル変復調技術, CQ 出版, 東京, 2005.

研究業績

国内発表

1. 木所晃利, 木谷俊介, 鷗木祐史, “Spectro-Temporal Modulation 分析を利用した聴覚的顕著性の検討,” 日本音響学会聴覚研究会, Vol. 50, No. 6, pp. 383-388, 2020.
2. 木所晃利, 木谷俊介, 鷗木祐史, “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 2021 年会春季研究発表会, 2021.

謝辞

本研究を行うに当たり、指導教員として常に心熱くも厳しいご指導を賜り、ときに励ましのお言葉をくださった、北陸先端科学技術大学院大学の鷓木 祐史教授には深く感謝申し上げます。同じく北陸先端科学技術大学院大学の赤木 正人教授には研究室での会議などにおいて多くの有益なご助言をいただき、深く感謝申し上げます。木谷 俊介助教には本研究での分析対象とした刺激の提供をはじめ日頃から熱心な議論と多くのご助言を頂き、多くの時間を割いて本研究にご協力頂いたことをここに深く感謝申し上げます。

鳥谷さん、磯山さんをはじめとする諸先輩方の皆様には研究や研究室運営などで多くのご協力を頂き感謝申し上げます。また、田中さん、佐藤さんをはじめとする M1 メンバーとの多くの面で支えていただき感謝申し上げます。そして、鷓木・赤木研究室の同期である藤田さん、倉さん、坂本 貴望さん、坂本 湧暉さん彼らと過ごした時間は研究室での生活において楽しく有意義でありました、この場において深く感謝します。他研究室の友人方や JAIST のスタッフの方々などこの 2 年間の石川県での生活において関わった全ての方々に対して深く感謝申し上げます。

最後に、遠く銀嶺の山々を超えて石川の地にまで応援をしてくれた父、母、妹、友人たちに心より御礼申し上げます。