

Title	雑音駆動合成を用いた調波複合音のピッチ知覚の検討
Author(s)	寶坂, 友希菜
Citation	
Issue Date	2020-03
Type	Thesis or Dissertation
Text version	author
URL	<a href="http://hdl.handle.net/10119/16419">http://hdl.handle.net/10119/16419</a>
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

雑音駆動合成を用いた調波複合音のピッチ知覚の検討

1810168 寶坂 友希菜

主指導教員 鵜木祐史  
審査委員主査 鵜木祐史  
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02年02月

## Abstract

The aims of this study is to investigate the possibility of pitch perception using the amplitude envelope information. The first experiment was conducted by Thurston's paired comparison to investigate whether or not the pitch perceptual scales of harmonic complex tones mimicking musical instruments and its noise-vocoded sounds were placed comparable, respectively. Here, to use the musical pitch as a measure of the pitch of the sound, the harmonic complex tones mimicking musical instruments as the original sound. As results, it was found that the correct rates of pitch perception of noise-vocoded sounds are almost the same as those of the harmonic complex tones. And the pitch perceptual scales of the noise-vocoded sounds and harmonic complex tones were almost identical by conducting a Thurston's paired comparison to investigate whether the pitch perceptual scales of harmonic complex tones mimicking musical instruments. However, we have not yet revealed what kind of cues can play an important role in pitch perception. This point will be analyzed in the next experiments.

Next, the second experiment was conducted by the same ways in the first experiment to investigate whether or not the results of the first experiments are affected under spectral-tilt conditions by using both stimuli of harmonic complex tones and noise-vocoded sounds with three spectral tilts of growing down, flat, and growing up. As results, it was found that the pitch perceptual scales of noise-vocode sounds are relatively affected by spectral-tilt conditions while those of harmonic complex tones are not affected by spectral-tilt conditions. Hence, these suggest that cues derived from the temporal amplitude envelope play an important role for pitch perception although these are related to spectral-title conditions.

In addition, the third experiment was conducted by the same ways in the first and second experiment using conducted pitch discrimination studies of noise-vocoded sounds with an increased cutoff frequency when extracting the amplitude envelope to investigate cues for pitch discrimination of noise-vocoded sounds. As results, it was found that the increase in the information of the amplitude envelope caused by increasing the cutoff frequency does not affect the pitch discrimination of the noise-vocoded sounds.

Furthermore, to investigate cues for pitch discrimination of noise-vocoded sounds, the relationship between the excitation pattern and modulation spectrum and the pitch discrimination results was investigated. As a result, the pitch scales from the analysis of the excitation patterns and the modulation spectrum did not match the pitch scales of the noise-vocoded sounds by the Thurston's paired comparison. Therefore, it was suggested that there was no relation between the excitation patterns and modulation spectrum and the pitch discrimination results.

These results suggest that the pitch perception of the noise-vocoded sounds

is affected by the low-pass filter when extracting the spectral gradient and the amplitude envelope, but not by the increase in the cutoff frequency.

From the results of this experiment and analysis, it cannot be said that important cues for pitch perception is contained the amplitude envelope information. However, in Chapters 4 and 5, it was suggested that pitch discrimination was possible under the right-down condition. In this study, it was not possible to clarify the mechanism of pitch perception of amplitude envelope information and discuss the strategy of acquiring pitch perception for cochlear implant users.

In this study, the pitch discrimination of noise-vocoded sounds was investigated, but it is not clear what pitch the noise-vocoded sounds have. Therefore, to investigate what pitch is perceived from the noise-vocoded sounds using the adjustment method, etc., it is possible to further examine the cues of pitch discrimination of the noise-vocoded sounds. In addition, if it becomes clear that cues for pitch discrimination of noise-vocoded sounds, the investigation of pitch discrimination by cochlear implant users using the sound emphasizing the cues will help to obtain the pitch perception of the cochlear implant users. It is possible to the strategy can be discussed.

# 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	研究目的	3
1.4	論文構成	3
<b>第2章</b>	<b>ピッチ知覚のメカニズムと振幅包絡線の役割</b>	<b>5</b>
2.1	ピッチ知覚のメカニズム	5
2.2	音声知覚における振幅包絡線の役割	5
2.2.1	振幅包絡線と言語情報知覚	5
2.2.2	振幅包絡線と非言語情報知覚	6
2.2.3	振幅包絡線とパラ言語情報知覚	6
<b>第3章</b>	<b>本研究の着目点</b>	<b>7</b>
3.1	ピッチ知覚における振幅包絡線の役割	7
3.2	雑音駆動合成音の作成方法	7
3.3	研究計画	9
<b>第4章</b>	<b>雑音駆動合成を用いた調波複合音のピッチ知覚の可能性</b>	<b>10</b>
4.1	実験目的	10
4.2	実験刺激	10
4.3	実験手続き	11
4.4	実験結果	11
4.4.1	正答率	11
4.4.2	サー斯顿の一対比較法による音階の配置	13
4.5	考察	15
<b>第5章</b>	<b>雑音駆動合成音のピッチ知覚におけるスペクトル傾斜の影響</b>	<b>17</b>
5.1	実験目的	17
5.2	実験刺激	17
5.3	実験手続き	20
5.4	実験結果	21
5.4.1	正答率	21

5.4.2	サーストンの一対比較法による音階の配置	23
5.5	エキサイテーションパターンの分析	25
5.5.1	分析方法	25
5.5.2	分析結果	26
5.6	変調スペクトルの分析	31
5.6.1	分析方法	31
5.6.2	分析結果	31
5.7	考察	37
5.7.1	実験結果の考察	37
5.7.2	エキサイテーションパターンと実験結果	39
5.7.3	変調スペクトルと実験結果	40
<b>第6章</b>	<b>雑音駆動合成音のピッチ知覚の手がかりの検討</b>	<b>41</b>
6.1	実験目的	41
6.2	実験刺激	41
6.3	実験手続き	43
6.4	実験結果	43
6.4.1	正答率	43
6.4.2	サーストンの一対比較法による音階の配置	46
6.5	エキサイテーションパターンの分析	51
6.5.1	エキサイテーションパターンの形状	51
6.5.2	エキサイテーションパターンから算出した F0 による音階の配置	53
6.6	変調スペクトルの分析	58
6.6.1	変調スペクトル	58
6.6.2	変調スペクトルから算出した F0 による音階の配置	62
6.7	考察	73
6.7.1	実験結果の考察	73
6.7.2	エキサイテーションパターンと実験結果	74
6.7.3	変調スペクトルと実験結果	74
<b>第7章</b>	<b>全体考察</b>	<b>76</b>
7.1	雑音駆動合成音のピッチ弁別の可能性	76
7.2	雑音駆動合成音のピッチ弁別の手がかり	77
<b>第8章</b>	<b>結論</b>	<b>79</b>
8.1	本研究で明らかにしたこと	79
8.2	残された課題	80
	<b>謝辞</b>	<b>80</b>

参考文献	81
研究業績	84

# 目 次

1.1	本論文の構成	4
3.1	雑音駆動合成音の作成方法	9
4.1	実験刺激 (C4, 原音)	10
4.2	実験刺激 (C4, NVS)	11
4.3	サーストンの一対比較法による音階の配置 (原音)	13
4.4	サーストンの一対比較法による音階の配置 (NVS)	13
4.5	サーストンの一対比較法による正しい音階の配置	14
4.6	変調スペクトル (C4)	16
5.1	実験刺激 (C4 原音, 右下り)	17
5.2	実験刺激 (C4 原音, 平坦)	18
5.3	実験刺激 (C4 原音, 右上り)	18
5.4	実験刺激 (C4NVS, 右下り)	19
5.5	実験刺激 (C4NVS, 平坦)	19
5.6	実験刺激 (C4NVS, 右上り)	20
5.7	サーストンの一対比較法による音階の配置 (原音)	23
5.8	サーストンの一対比較法による音階の配置 (NVS)	24
5.9	エキサイテーションパターン (G3, 右下り)	27
5.10	エキサイテーションパターン (G3, 平坦)	27
5.11	エキサイテーションパターン (G3, 右上り)	28
5.12	図 3.1 に示すフィルタバンクにおいて 1.5 kHz 以下の帯域のエキサイテーションパターンから求めたサーストンの一対比較法による音階の配置 (NVS)	29
5.13	原音で F0 を含む帯域のエキサイテーションパターンから求めたサーストンの一対比較法による音階の配置 (NVS)	30
5.14	変調スペクトル (C4, 右下り)	32
5.15	変調スペクトル (C4, 平坦)	33
5.16	変調スペクトル (C4, 右上り)	34
5.17	図 3.1 に示すフィルタバンクにおいて 1.5 kHz 以下の帯域の変調スペクトルから求めたサーストンの一対比較法による音階の配置	35

5.18	第 16 帯域の変調スペクトルから求めたサーストンの一対比較法による音階の配置	36
5.19	変調スペクトル (C4, 第 3 帯域, 右下り)	38
5.20	変調スペクトル (C4, 第 3 帯域, 平坦)	38
5.21	変調スペクトル (C4, 第 3 帯域, 右上り)	39
6.1	実験刺激 (C4, 右下り, $F_c = 64\text{Hz}$ )	41
6.2	実験刺激 (C4, 平坦, $F_c = 64\text{Hz}$ )	42
6.3	実験刺激 (C4, 右上り, $F_c = 64\text{Hz}$ )	42
6.4	サーストンの一対比較法による音階の配置 ( $F_c = 64\text{ Hz}$ )	48
6.5	サーストンの一対比較法による音階の配置 ( $F_c = 128\text{ Hz}$ )	49
6.6	サーストンの一対比較法による音階の配置 ( $F_c = 256\text{ Hz}$ )	50
6.7	サーストンの一対比較法による音階の配置 ( $F_c = 1024\text{ Hz}$ )	51
6.8	エキサイテーションパターン ( $F_c = 64\text{ Hz}$ , 右下り)	52
6.9	エキサイテーションパターン ( $F_c = 64\text{ Hz}$ , 平坦)	52
6.10	エキサイテーションパターン ( $F_c = 64\text{ Hz}$ , 右上り)	53
6.11	図 3.1 に示すフィルタバンクにおいて 1.5 kHz 以下の帯域のエキサイテーションパターンから求めたサーストンの一対比較法による音階の配置 ( $F_c = 64\text{ Hz}$ )	54
6.12	原音で F0 を含む帯域のエキサイテーションパターンから求めたサーストンの一対比較法による音階の配置 ( $F_c = 64\text{ Hz}$ )	55
6.13	原音で F0 を含む帯域のエキサイテーションパターンから求めたサーストンの一対比較法による音階の配置 ( $F_c = 128\text{ Hz}$ )	56
6.14	原音で F0 を含む帯域のエキサイテーションパターンから求めたサーストンの一対比較法による音階の配置 ( $F_c = 256\text{ Hz}$ )	57
6.15	原音で F0 を含む帯域のエキサイテーションパターンから求めたサーストンの一対比較法による音階の配置 ( $F_c = 1024\text{ Hz}$ )	58
6.16	変調スペクトル ( $F_c = 64\text{ Hz}$ )	59
6.17	変調スペクトル ( $F_c = 128\text{ Hz}$ )	60
6.18	変調スペクトル ( $F_c = 256\text{ Hz}$ )	61
6.19	変調スペクトル ( $F_c = 1024\text{ Hz}$ )	62
6.20	図 3.1 に示すフィルタバンクにおいて 1.5 kHz 以下の帯域の変調スペクトルから求めたサーストンの一対比較法による音階の配置 ( $F_c = 64\text{ Hz}$ )	64
6.21	図 3.1 に示すフィルタバンクにおいて 1.5 kHz 以下の帯域の変調スペクトルから求めたサーストンの一対比較法による音階の配置 ( $F_c = 128\text{ Hz}$ )	65

6.22	図 3.1 に示すフィルタバンクにおいて 1.5 kHz 以下の帯域の変調スペクトルから求めたサーストンの一対比較法による音階の配置 ( $F_c=256$ Hz) . . . . .	66
6.23	図 3.1 に示すフィルタバンクにおいて 1.5 kHz 以下の帯域の変調スペクトルから求めたサーストンの一対比較法による音階の配置 ( $F_c=1024$ Hz) . . . . .	67
6.24	原音で $F_0$ を含む帯域の変調スペクトルから求めたサーストンの一対比較法による音階の配置 ( $F_c=64$ Hz) . . . . .	69
6.25	原音で $F_0$ を含む帯域の変調スペクトルから求めたサーストンの一対比較法による音階の配置 ( $F_c=128$ Hz) . . . . .	70
6.26	原音で $F_0$ を含む帯域の変調スペクトルから求めたサーストンの一対比較法による音階の配置 ( $F_c=256$ Hz) . . . . .	71
6.27	原音で $F_0$ を含む帯域の変調スペクトルから求めたサーストンの一対比較法による音階の配置 ( $F_c=1024$ Hz) . . . . .	72
6.28	第 16 帯域の変調スペクトルから求めたサーストンの一対比較法による音階の配置 ( $F_c=64$ Hz) . . . . .	73
7.1	サーストンの一対比較法による音階の配置 (第 4 章 NVS 再実験) . . . . .	77

# 表 目 次

4.1	全刺激対に対する全実験参加者のピッチ弁別平均正答率 . . . . .	12
4.2	ピッチ弁別が可能であると判断した刺激対数 . . . . .	12
5.1	全刺激対に対する全実験参加者のピッチ弁別平均正答率 . . . . .	21
5.2	ピッチ弁別が可能であると判断した刺激対数 (全刺激) . . . . .	22
5.3	ピッチ弁別が可能であると判断した刺激対数 (同傾斜同士) . . . . .	22
6.1	全刺激対に対する全実験参加者のピッチ弁別平均正答率 . . . . .	44
6.2	ピッチ弁別が可能であると判断した刺激対数 (全刺激) . . . . .	45
6.3	ピッチを逆に弁別可能であると判断した刺激対数 (全刺激) . . . . .	45
6.4	ピッチ弁別が可能であると判断した刺激対数 (同傾斜同士) . . . . .	46

# 第1章 序論

## 1.1 はじめに

音とは、音波またはそれにより引き起こされる聴覚的感觉である [1]. 音には純音と複合音が存在する. 純音は一つの正弦波で構成される音であり, 正弦波の周期  $T_s$  の逆数, つまり周波数  $f$  Hz が純音の音の高さとして知覚される. 複合音は周波数の異なる複数の純音で構成され, 音の高さが明確で周期的な波形をもつ音は楽音と呼ばれる. 周波数が  $f, 2f, 3f, \dots, nf$  の正弦波を合成した音は調波複合音であり,  $f$  Hz が基本周波数 (F0) と呼ばれ, 調波複合音の音の高さに対応する. 基本周波数以外の  $2f, 3f, \dots, nf$  は高調波 (倍音) と呼ばれる. 自然界に存在する音は殆どが複合音である. 例えば, テレビやラジオの時報で使用される電子音は純音であり, バイオリン等の楽器音は調波複合音である.

ヒトは日常生活において, 純音や複合音を問わずさまざまな音を聴いている. 例えば, ヒトの話し声や動物の鳴き声, 楽器の音, 自動車のクラクション, コンピュータのブープ音等, 自然界に存在する音から人工的に作られた電子音まで, 様々な音が存在する. そして, これらの音には様々な情報が含まれている. ヒトの話し声からは発話者の感情や声の高低, 楽器の音からは楽器の種類や音の高さ, クラクションやブープ音からは危険性等, 音を聴くだけで様々な情報を得ることができる.

音には音の大きさ (Loudness), 音の高さ (Pitch), 音色 (Timbre) の三つの心理的な属性が存在する [1]. 上述した様々な情報を得るために, ヒトはこれらの三つの属性を重要な手がかりとしている. ヒトの話し声の感情や楽器音が奏でる旋律を理解し, コンピュータのブープ音を知覚するには, 三つの属性の中でも音の高さ (ピッチ) を知覚することが重要である.

音には時間微細構造と原信号の緩やかな変化 (概形) を抽出した振幅包絡線が存在する [2]. 音の振幅変調 (AM) では, 時間微細構造はキャリア, 振幅包絡線は変調波に対応する. 時間微細構造は, ピッチ知覚や音像定位に重要である. 一方, 振幅包絡線は, Fujisaki による音声の意味分類 [3] における言語情報の知覚に重要である [2,4].

## 1.2 研究背景

ピッチが知覚できないとすると，楽器音による音楽等の娯楽が楽しむことができなだけでなく，ヒトの声の抑揚やクラクション等の危険性を知覚することができない．ヒトとの会話や，クラクション等の危険を知らせる報知音を知覚することができないことは，日常生活に支障をきたすことになる．

日本では現在，補聴器の装用効果が得られない高度の感音性難聴と呼ばれる重度の聴覚障がい者がおよそ 35 万人いると言われている [5]．補聴器の装用効果が得られない患者に対し，聴覚の獲得を目的として装用される人工内耳と呼ばれる人工臓器がある [6]．人工内耳では，体外で収集された音が蝸牛内に埋め込まれた電極に伝送され，蝸牛の聴神経を刺激し，脳に到達することで音として認識される．この際，言語情報の知覚に重要な音の時間的な振幅包絡線情報は正確に伝送されるが，ピッチ知覚に重要な時間微細構造は伝送されない [7]．そのため，ピッチ知覚の理論の一つである，聴神経発火の時間パターンを手がかりとする時間説の観点からピッチ知覚が困難である [8]．また，蝸牛に埋め込む電極数にも限りがあるため，聴神経の発火位置を手がかりとする場所説の観点からもピッチ知覚が困難であるという問題を抱えている [9]．これまでに，人工内耳装用者のピッチ知覚について，人工内耳の性能やピッチ知覚のメカニズムからその困難さが報告されてきた [10–12]．人工内耳装用者は，補聴器の装用者よりもピッチ知覚が困難 [13,14] であり，リズム弁別は良好である [10] が，1 オクターブ離れた音のピッチ弁別も困難な例が多い [11] ことが明らかになっている．また，人工内耳装用者の周波数弁別閾に基づき音階そのものを作成する検討がされてきた [12]．しかし，ピッチ知覚の改善という本質的な問題については検討されていない．そのため，人工内耳装用者は，日常生活における音楽の知覚や音声の韻律の知覚，危険を知らせる報知音等の知覚といった場面で問題を抱えている．これまでに，人工内耳装用者の聴こえを模擬した音の一つである雑音駆動合成音 [15] を用いた検討が行われてきた [16,17]．音声の振幅包絡線情報に，Fujisaki [3] による音声の意味分類における言語情報だけでなく個人性や感情といった非言語情報 [16]，緊迫感（パラ言語情報）が含まれていることが明らかにされてきた．音声の感情や緊迫感の知覚にはピッチ等の韻律的特徴が必要である [18] こと，振幅包絡線に音声の韻律を必要とする情報（個人性や感情，緊迫感）が含まれていること [16,17] から，振幅包絡線情報にピッチに係わる特徴が含まれていると考えられる．

さらに，これまでに振幅包絡線によるピッチ知覚の検討が行われてきた [19,20]．Shamma ら [19] によるスペクトル情報と時間情報を組み合わせたピッチ知覚の生物学的モデルの検討や，Shofner ら [20] による，ピッチ知覚における倍音構造の役割の検討が行われてきた．これらのことから，振幅包絡線情報にピッチに係わる特徴が含まれていると考えられる．

正常聴力を有するヒトも，人工内耳装用者も知覚することができる音の振幅包絡線情報によるピッチ知覚の可能性を明らかにすることができれば，ピッチ知覚の

メカニズム解明や人工内耳装用者のピッチ知覚の獲得および QoL (Quality of life) の向上に繋がると考えられる。

### 1.3 研究目的

本研究では, Zhu ら [16] や Unoki [17] らの検討と同様に, 雑音音駆動合成音を用いたアプローチにより, 音の時間的な振幅包絡線情報を用いてピッチ知覚が可能かどうかを検討することを目的とする。

はじめに, 振幅包絡線情報によるピッチ弁別の可能性を検討するため, 振幅包絡線情報のみが保持された雑音駆動合成音では既報 [9, 12] のようにピッチ弁別が困難であるかどうかを調査する。次に, 既報と異なり雑音駆動合成音のピッチ弁別が可能であると仮定し, 雑音駆動合成音のピッチ弁別が正しくピッチを知覚したことによるものか, ピッチではなく刺激のスペクトル傾斜を音色として知覚したこと [21, 22] によるものかを検討する。これは, ピッチと音色の知覚が相互に作用するため [23] である。さらに, 雑音駆動合成音のピッチ弁別が可能であると仮定し, 雑音駆動合成音のピッチ弁別の手がかりを検討するため, カットオフ周波数を増加させることで振幅包絡線の情報を増加させた場合にピッチ弁別の成績が向上するかどうかを検討する。最後に, ピッチ知覚の理論における場所説および時間説の観点から雑音駆動合成音のピッチ弁別の手がかりを検討することで, 雑音駆動合成音のピッチ知覚について考察する。これらの検討を通じて, 振幅包絡線情報に関するヒトのピッチ知覚メカニズムの解明および人工内耳装用者のピッチ知覚の獲得の方略を議論することを目指す。

### 1.4 論文構成

本論文は, 8 章で構成される。図 1.1 に本論文の構成を示す。

第 1 章では, 序論として本研究の研究背景と研究目的について述べる。

第 2 章では, ピッチ知覚のメカニズムと音声知覚における振幅包絡線の役割について述べる。

第 3 章では, 本研究の着目点および振幅包絡線情報によるピッチ知覚を検討するための研究計画について述べる。

第 4 章では, 振幅包絡線情報によるピッチ知覚を検討するため, 雑音駆動合成音のピッチ弁別の可能性を明らかにする。

第 5 章では, 雑音駆動合成音のピッチ弁別が正しくピッチを知覚したことによるものか, スペクトル傾斜を知覚したことによるものかを検討するため, 雑音駆動合成音のピッチ弁別へのスペクトル傾斜の影響を明らかにする。

第6章では、雑音駆動合成音のピッチ弁別の手がかりを検討するため、雑音駆動合成音のピッチ弁別とエキサイテーションパターンおよび変調スペクトルとの関係性を明らかにする。

第7章では、第4章–第6章での検討結果から雑音駆動合成音のピッチ弁別の可能性と手がかりについて述べる。また、振幅包絡線情報によるピッチ弁別について考察する。

第8章では、結論として本研究で明らかになったこと、残された課題について述べる。

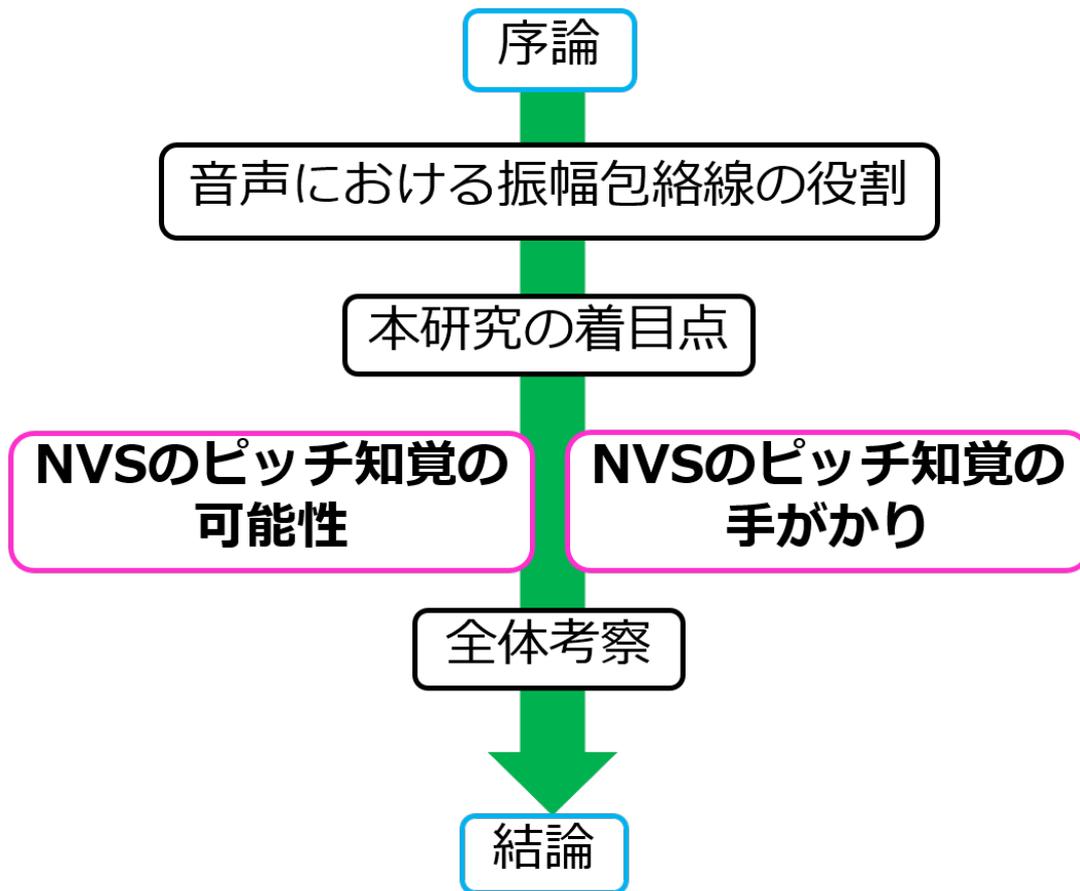


図 1.1: 本論文の構成

## 第2章 ピッチ知覚のメカニズムと振幅包絡線の役割

### 2.1 ピッチ知覚のメカニズム

ピッチ知覚の理論には、時間説と場所説の二つが並行して考えられてきた [1,24]. 時間説は、音の高さは音によって引き起こされた聴神経発火の時間パターンに関係するとされる理論である。聴神経の性質として、位相同期と呼ばれる現象が存在する。これは、聴神経発火が音の波形の特定の位相で生じることで、聴神経発火の間隔が音の波形周期の整数倍に近くなる現象である。そのため、聴神経発火の時間パターンを手がかりにピッチを知覚しているとされる。

一方で、場所説は、異なる周波数の音が蝸牛の異なった場所で異なる特徴周波数をもった聴神経を発火させることで音の高さを知覚するといった理論である。

今日においても、これらの理論のどちらが優勢かといった結論は出されていない。そのため、本研究では、時間説および場所説どちらの観点からも振幅包絡線情報によるピッチ知覚の可能性を検討する。

### 2.2 音声知覚における振幅包絡線の役割

Fujisaki [3] は、音声伝える意味および内容を言語情報、非言語情報、パラ言語情報の三つに分類した。この分類では、個人性や感情は非言語情報に、話者の意図や態度はパラ言語情報に分類される。本節では、分類された言語情報、非言語情報、パラ言語情報の知覚における振幅包絡線の役割について述べる。

#### 2.2.1 振幅包絡線と言語情報知覚

人工内耳の模擬音の一つである雑音駆動合成音声 (Noise-vocoded speech) は、音声を帯域分割し、各帯域の振幅包絡線で白色雑音を振幅変調した合成音である。そのため、音の振幅包絡線情報のみを保持し、時間微細構造を持たない。これまでに、雑音駆動合成音声を用いた検討から、音声の振幅包絡線情報が言語情報知覚において重要な役割を果たしていることが明らかにされた。4帯域以上の雑音駆動合成音声であれば言語情報を知覚できることが報告されている [15].

## 2.2.2 振幅包絡線と非言語情報知覚

Zhu ら [16] は、音声の振幅包絡線情報が非言語情報の知覚に影響するかどうかを検討するため、雑音駆動合成音声を用いて音声の振幅包絡線の変調周波数を制限することによる話者および感情認識への影響を調査した。その結果、変調周波数が低くなると話者認識および感情認識率が低下することが明らかにされた。また、非言語情報の知覚に重要な変調周波数は、言語情報の知覚に重要な変調周波数よりも高いことが明らかにされた。したがって、音声の振幅包絡線情報が個人性や感情といった非言語情報の知覚に重要であることが明らかにされた。

## 2.2.3 振幅包絡線とパラ言語情報知覚

Unoki ら [17] は、音声の振幅包絡線情報がパラ言語情報の知覚に影響するかどうかを検討するため、雑音駆動合成音声を用いて音声の振幅包絡線の変調周波数を制限することによる緊迫感知覚への影響を調査した。その結果、原音声と雑音駆動合成音声の緊迫感が類似していること、変調周波数 6 Hz – 8 Hz が緊迫感の知覚に重要な手がかりであることが明らかにされた。したがって、音声の振幅包絡線情報がパラ言語情報（緊迫感）の知覚に重要であることが明らかにされた。

## 第3章 本研究の着目点

### 3.1 ピッチ知覚における振幅包絡線の役割

ピッチ知覚には時間微細構造の存在が重要である [2,7] といわれている。しかし、第2章で述べた通り、振幅包絡線に音声の韻律を必要とする情報（個人性や感情、緊迫感）が含まれていることが明らかになった。さらに、これまでに振幅包絡線によるピッチ知覚の検討が行われてきた [19,20,26]。Shamma ら [19] は、スペクトル情報と時間情報を組み合わせたピッチ知覚の生物学的モデルを検討した。その結果、振幅変調と Unresolved harmonics および Resolved harmonics [25] のテンプレートマッチングを用いて全てのピッチ知覚のスペクトル-時間調波テンプレートの作成方法を示した。Shofner [20] らは、ピッチ知覚における倍音構造の役割を検討するため、雑音駆動した調波複合音を用いたピッチ変化知覚を調査した。その結果、雑音駆動の帯域が8あれば1オクターブ離れた刺激のピッチ変化の方向を正しく知覚できることが明らかにされた。このことから、振幅包絡線情報にピッチに係わる特徴が含まれていると考えられる。そこで、本研究では振幅包絡線情報を用いたピッチ知覚に着目した。

### 3.2 雑音駆動合成音の作成方法

本研究では、Zhu ら [16] や Unoki ら [17] の研究と同様に、雑音駆動音声の合成法を用いて雑音駆動合成音を作成した。雑音駆動合成音の作成方法を図 3.1 に示す。はじめに、入力信号（原音）を聴覚フィルタバンクによって帯域分割した。聴覚フィルタには6次の IIR 型（Infinite Impulse Response）の Butterworth 帯域通過フィルタを利用した。聴覚フィルタバンクにおけるフィルタは、次式で定義される  $ERB_{N-number}$  尺度に沿って配置された。

$$ERB_{N-number} = 21.4 \log_{10} \left( \frac{4.37f}{1000} + 1 \right) \quad (3.1)$$

これは、 $ERB_{N-number}$  を1とする周波数軸上に低域側から順番に並べたものである。ここで、 $N$  は健聴者（Normal Hearing）の特性、 $f$  は周波数（Hz）である。また、 $ERB_N$  は、健聴者の聴覚フィルタの帯域幅を表す等価矩形帯域幅（Equivalent Rectangular Bandwidth）である。本研究では、現在の人工内耳のチャンネル数（12 - 24）と対応を取るため、帯域通過フィルタの帯域幅を  $2ERB_N$  として、3か

ら  $35ERB_N$  までの 16 帯域に分割した。次に，帯域制限信号から Hilbert 変換と低域通過フィルタを利用して時間的な振幅包絡線を抽出した。低域通過フィルタのカットオフ周波数は 64 Hz とした。最後に，各帯域において振幅包絡線に狭帯域雑音を乗じることで得られた帯域信号を全周波数領域で足し合わせることで雑音駆動合成音を作成した。なお，狭帯域雑音は，白色雑音を入力信号として，同じ聴覚フィルタバンクによって帯域分割して得られた。

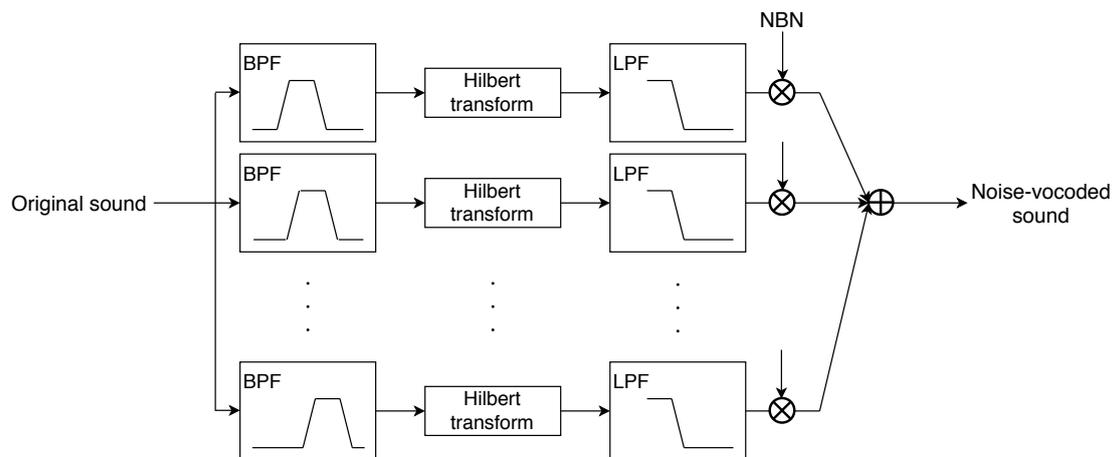


図 3.1: 雑音駆動合成音の作成方法

### 3.3 研究計画

本研究では、振幅包絡線情報によるピッチ知覚の可能性を検討することを目的とする。はじめに、健聴者が雑音駆動合成音のピッチを知覚することができるかどうかを調査するため、楽音を模した調波複合音とその雑音駆動合成音を用いた聴取実験を行う。加えて、雑音駆動合成音のピッチの判別ができたとすると、それが正しくピッチを知覚したことによるものか、音色によるものかどうかを調査するため、三つのスペクトル傾斜を持つ調波複合音とその雑音駆動合成音による聴取実験を行う。なお、音楽的ピッチを音のピッチの尺度とするため、原音には楽音を模した調波複合音を用いる。さらに、雑音駆動合成音のピッチ弁別の手がかりを検討するため、エキサイテーションパターンおよび変調スペクトルの分析を行う。最後に、雑音駆動合成音のピッチ弁別の手がかりを検討するため、振幅包絡線抽出時のカットオフ周波数を増加させた雑音駆動合成音のピッチ弁別調査を行い、雑音駆動合成音のピッチ弁別の手がかりについて考察する。

# 第4章 雑音駆動合成を用いた調波複合音のピッチ知覚の可能性

## 4.1 実験目的

振幅包絡線情報によるピッチ弁別の可能性を検討することを目的とする。本章では、健聴者が雑音駆動合成音の音のピッチを知覚することができるかどうかを調査するため、調波複合音とその雑音駆動合成音を用いた聴取実験を行った。

## 4.2 実験刺激

実験刺激として、C3 - C5 (130.8 Hz - 523.3 Hz) の白鍵のみ 15 種の基本周波数と、10 kHz までの倍音成分を持つ調波複合音を作成した。作成した調波複合音に -6 dB/Oct. のスペクトル傾斜を持たせ、実験刺激とした。また、これらを原音とする雑音駆動合成音も作成し、実験刺激は原音 15 種、雑音駆動合成音 15 種の計 30 種とした。原音の刺激の一例を図 4.1 に、雑音駆動合成音の一例を図 4.2 に示す。

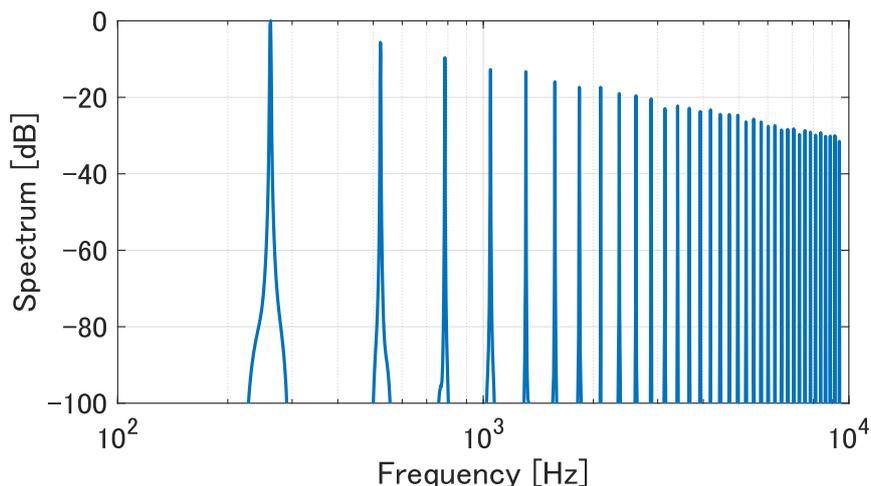


図 4.1: 実験刺激 (C4, 原音)

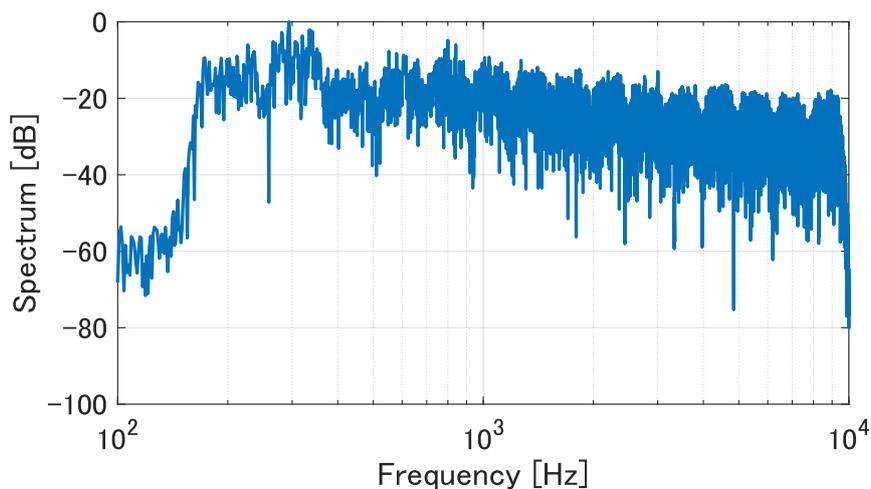


図 4.2: 実験刺激 (C4, NVS)

### 4.3 実験手続き

実験にはサーストンの一対比較法 [27], [28] を用いた. 同音同士を除いた刺激対を実験参加者にランダムに呈示し, 音が高いと感じる方を強制的に判断させた. 各刺激の呈示は 1 s とし, 刺激と刺激の間隔は 0.5 s であった. 実験は 1 セクション 105 対とし, 原音 2 セクション, 雑音駆動合成音 2 セクションの順に計 420 対行った. 実験参加者は正常聴力を有する大学院学生 15 名 (22 – 28 歳, うち女性 4 名) であった.

実験は無響室で行われ, 刺激の呈示には PC (Windows10, MATLAB), オーディオインターフェース (Fireface UCX), ヘッドフォンアンプ (AudioTechnica AT-HA21), ヘッドフォン (SENNHEISER HDA-200) を使用し, 実験刺激のサンプリング周波数は 44.1 kHz とした. 刺激は, 人工耳 (B&K Artificial Ear Type 4153), マイク (B&K Type 4291), 騒音計 (B&K Sound Level Meter Type 2250) を用いて, ヘッドフォンからの出力が A 特性音圧レベル 70 dB となるように設定された.

## 4.4 実験結果

### 4.4.1 正答率

振幅包絡線抽出時のカットオフ周波数毎に, 全刺激対の平均正答率を算出した. 算出した結果を表 4.1 に示す. 原音および雑音駆動合成音のどちらにおいても正答率が 90% を超えていた.

さらに, 各刺激対に対する全実験参加者の平均正答率を算出した. ここでは, チャンスレベルが 50% であるため, 平均正答率が 75% 以上でピッチ弁別が可能で

あると判断した。ピッチ弁別が可能と判断された刺激対数と、全刺激対に対するピッチ弁別が可能と判断された刺激対の割合を表 4.2 に示す。その結果、原音では全刺激対においてピッチ弁別が可能であった。雑音駆動合成音では、105 対中 94 対と 90.0%の刺激対でピッチ弁別が可能であり、原音と同程度に弁別可能であることがわかった。

表 4.1: 全刺激対に対する全実験参加者のピッチ弁別平均正答率

刺激種類	全刺激対に対する 全実験参加者の平均正答率 (%)
原音	97.0
雑音駆動合成音	92.1

表 4.2: ピッチ弁別が可能であると判断した刺激対数

刺激グループ	ピッチ弁別可能な 刺激対数	ピッチ弁別可能な 刺激対の割合 (%)
原音	105	100
NVS	94	90.0

#### 4.4.2 サーストンの一対比較法による音階の配置

サーストンの一対比較法により、各刺激に対して実験参加者が知覚した音階の配置を算出した。原音の配置を図 4.3 に、雑音駆動合成音の配置を図 4.4 に示す。図中の赤字および下線は、並びが入れ替わって知覚されていることを示している。図 4.5 に示すように、C3-C5 の各刺激が音楽音階通りに等間隔で並んでいれば、ピッチを正しく知覚できていると言える。本来、音楽音階において白鍵の並びは等間隔ではない。しかし、本研究では白鍵のみを実験刺激として用いること、サーストンの一対比較法を用いることから、各刺激が等間隔に並んでいれば正しくピッチを知覚できていると定義した。その結果、原音であっても音階の並びが入れ替わっている部分が見られた。また、中音域において刺激間の間隔が狭くなっており、等間隔に並んでいない部分が見られた。雑音駆動合成音においては、並びが入れ替わっている部分が多く見られた。音階の並びの誤りが E3 - F4 の低-中音域に見られたが、G4 - C5 の高音域では原音と同様の配置となった。刺激間の間隔では、G3 - F3 および D4 - G4, G4 - A4 で間隔が広く、その他の間隔は狭くなっていた。

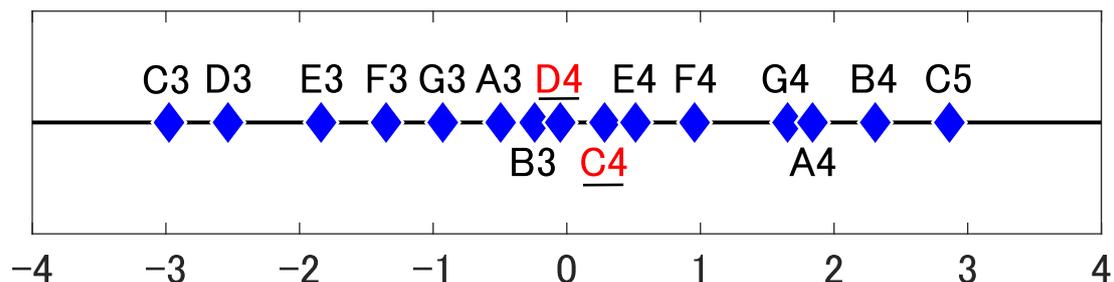


図 4.3: サーストンの一対比較法による音階の配置（原音）。図中の赤字および下線は、並びが入れ替わって知覚されていることを示している。

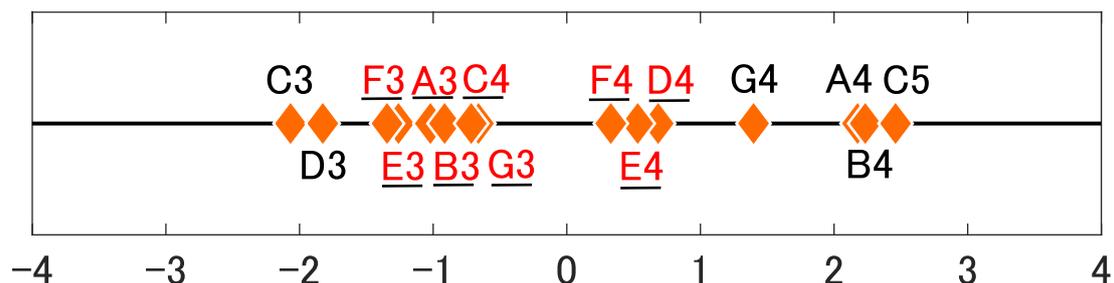


図 4.4: サーストンの一対比較法による音階の配置（NVS）。図の見方は図 4.3 と同様である。

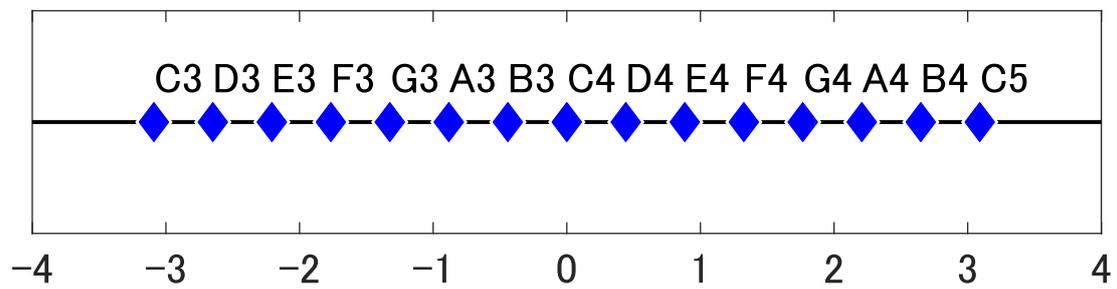


図 4.5: サーストンの一対比較法による正しい音階の配置

## 4.5 考察

雑音駆動合成音では，全刺激対に対する全実験参加者の平均正答率が92.1%，ピッチ弁別が可能な刺激対の割合が90 %の対で能であり，原音とほぼ同定度であった．このことから，既報 [8-14] とは異なり，振幅包絡線情報のみを持つ雑音駆動合成音であっても，ピッチを知覚できる可能性があると考えられる．

サーストンの一対比較法による音階の配置では，原音および雑音駆動合成音のどちらにおいても，音階の並びの誤りは隣り合ったところでのみ見られた．間隔変動は，C4 - D4，G4 - A4 で広く，他の刺激では狭くなっていた．また，図 4.4 から，雑音駆動合成音は G4 - C5 の高音域ほど原音に似た知覚ができると考えられる．

雑音駆動合成音作成時に抽出した振幅包絡線のスペクトル（変調スペクトル）を算出した．例として，C4 の変調スペクトルを図 4.6 に示す．雑音駆動合成音は時間微細構造を持たないため，基本周波数に関する調波性（時間微細構造）を持たない．しかし，図 4.6 に示す変調スペクトルには基本周波数に関わる調波性が見られたことから，時間的な振幅包絡線情報に基本周期が見られたといえる．変調スペクトルに見られた特徴がどの程度ピッチ弁別に影響しているかどうかは明らかでなかったため，次章で分析を行う．

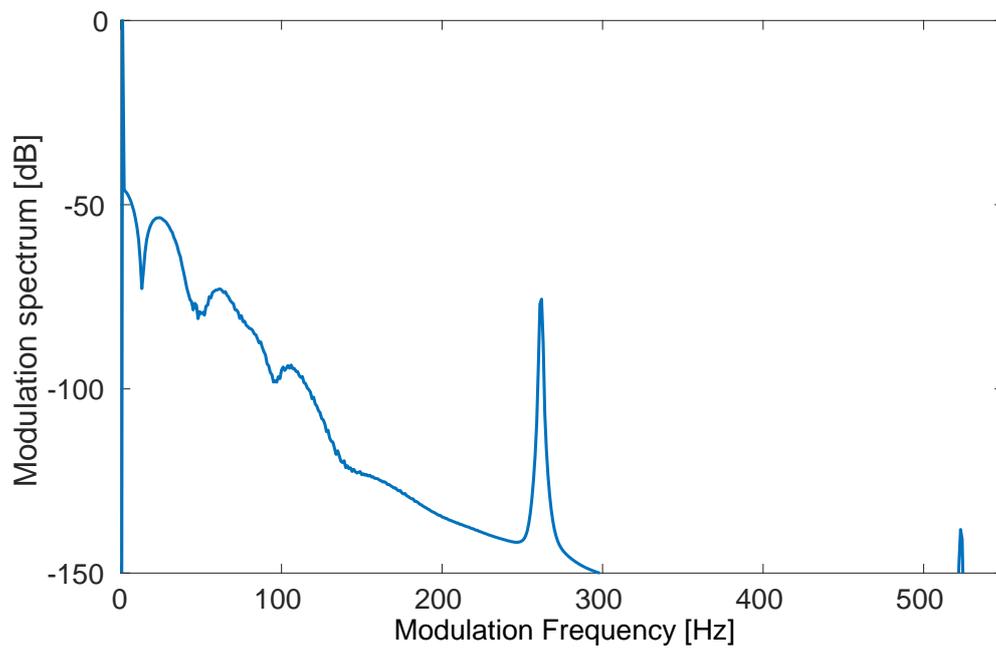


図 4.6: 変調スペクトル (C4)

# 第5章 雑音駆動合成音のピッチ知覚 におけるスペクトル傾斜の 影響

## 5.1 実験目的

第4章における雑音駆動合成音のピッチ弁別が、振幅包絡線情報を手がかりとしてピッチを正しく知覚したことによるものか、スペクトル傾斜を知覚したことによるものかどうかを検討することを目的とする。また、雑音駆動合成音のピッチ弁別の手がかりについても検討する。

## 5.2 実験刺激

第4章で作成した調波複合音に-6 dB/Oct. (右下り), 0 dB/Oct. (平坦), 6 dB/Oct. (右上り) の三つの異なるスペクトル傾斜を持たせた45種の調波複合音を作成した。また、これらを原音とする雑音駆動合成音45種も作成し、実験刺激は原音45種, 雑音駆動合成音45種の計90種とした。原音の刺激の例を図5.1 - 図5.3に, 雑音駆動合成音の刺激の例を図5.4 - 図5.6に示す。

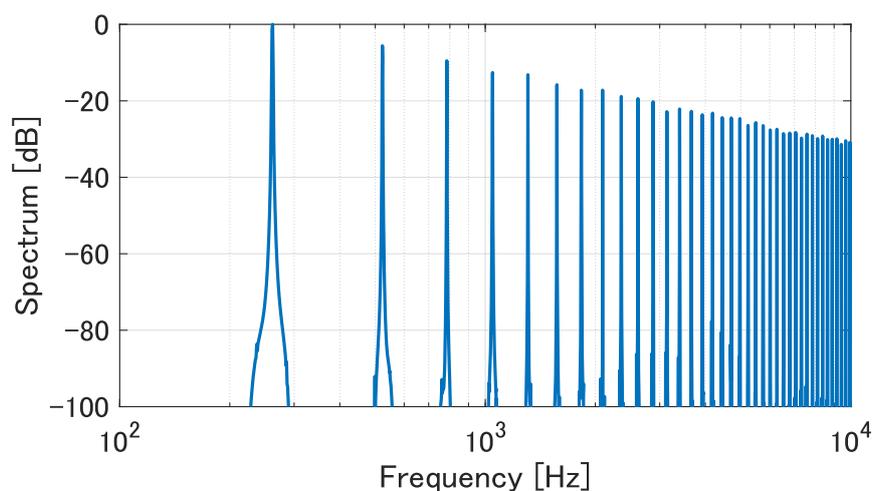


図 5.1: 実験刺激 (C4 原音, 右下り)

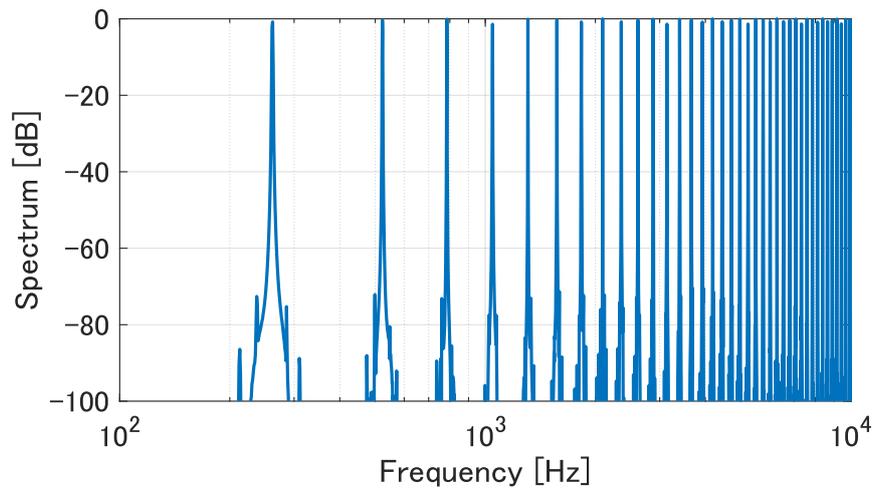


図 5.2: 実験刺激 (C4 原音, 平坦)

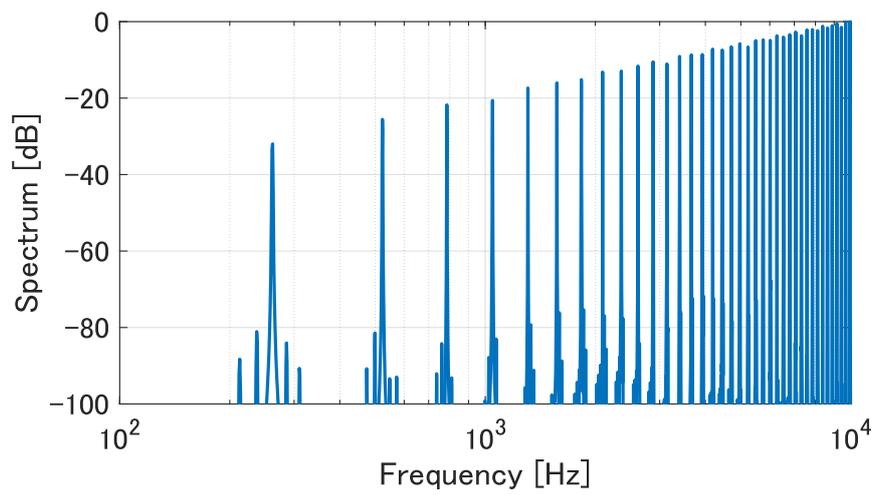


図 5.3: 実験刺激 (C4 原音, 右上り)

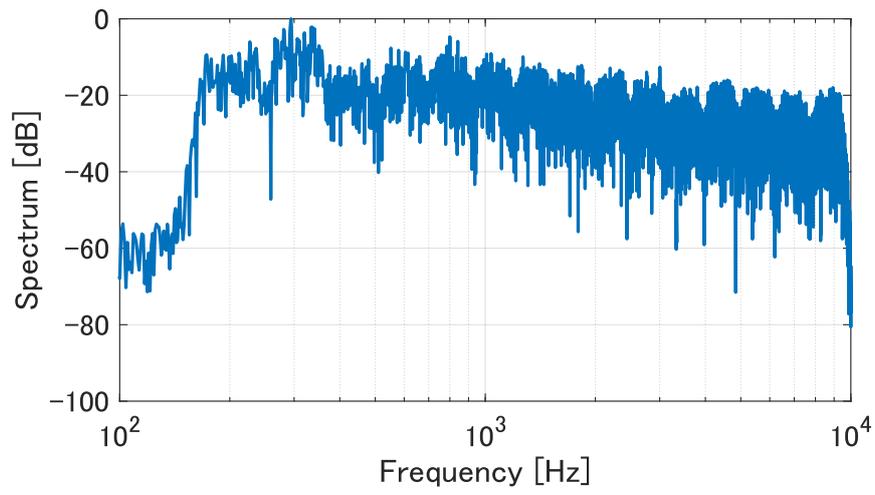


図 5.4: 実験刺激 (C4NVS, 右下り)

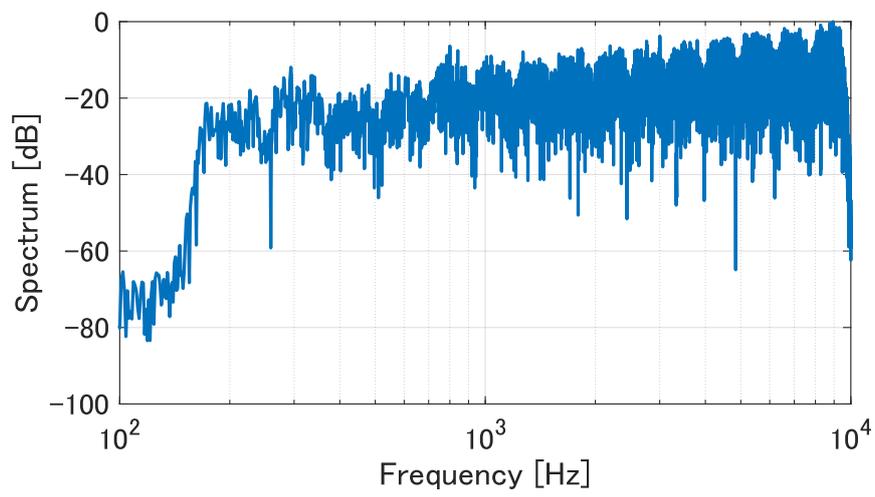


図 5.5: 実験刺激 (C4NVS, 平坦)

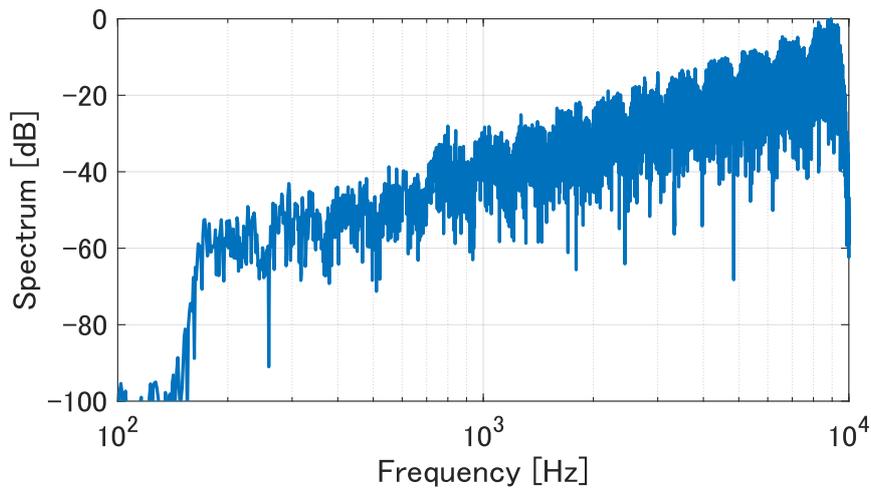


図 5.6: 実験刺激 (C4NVS, 右上り)

### 5.3 実験手続き

第4章と同様に、実験にはサーストンの一対比較法を用いた。同音同士を除いた刺激対を実験参加者にランダムに呈示し、音が高いと感じる方を強制的に判断させた。各刺激の呈示は1 sとし、刺激と刺激の間隔は0.5 sであった。ここでは、実験刺激のスペクトル条件を問わず総当たりで呈示した。実験は1セッション135対とし、原音7セッション、雑音駆動合成音7セッションの順に計1890対行った。実験参加者は正常聴力を有する成人14名(22–28歳、うち女性3名)であった。

実験は無響室で行われ、刺激の呈示にはPC(Windows10, MATLAB)、オーディオインターフェース(Fireface UCX)、ヘッドフォンアンプ(AudioTechnica AT-HA21)、ヘッドフォン(SENNHEISER HDA-200)を使用し、実験刺激のサンプリング周波数は44.1 kHzとした。刺激は、人工耳(B&K Artificial Ear Type 4153)、マイク(B&K Type 4291)、騒音計(B&K Sound Level Meter Type 2250)を用いて、ヘッドフォンからの出力がA特性音圧レベル70 dBとなるように設定された。

## 5.4 実験結果

### 5.4.1 正答率

第4章と同様にして，振幅包絡線抽出時のカットオフ周波数毎に，全刺激対の平均正答率を算出した．算出した結果を表6.1に示す．第4章と比較して，雑音駆動合成音の正答率が低くなっていた．

さらに，各刺激対に対する実験参加者の平均正答率を算出した．表5.2に示すように，ピッチ弁別可能な刺激対は，原音では945対中898対，雑音駆動合成音では945対中550対であった．原音では90%以上，雑音駆動合成音では50%以上のパターンでピッチ弁別が可能であった．雑音駆動合成音では，原音と比較して正答率が大幅に低下した．

表5.3に示すように，同じスペクトル傾斜を持つ刺激対のみの結果では，原音ではスペクトル傾斜が右下りの条件で105対中103対，平坦の条件で105対中103対，右上りの条件では105対中102対であり，いずれも90%以上の刺激対でピッチ弁別が可能であった．一方，雑音駆動合成音では右下りの条件で105対中92対，平坦の条件で105対中72対，右上りの条件で105対中22対であった．右下りの条件ではピッチ弁別可能な刺激対の割合が全体の87%と最も成績が良く，平坦では68.6%，右上りの条件では21.0%であり，成績が大幅に悪くなっていた．

表 5.1: 全刺激対に対する全実験参加者のピッチ弁別平均正答率

刺激種類	全刺激対に対する 全実験参加者の平均正答率 (%)
原音	93.9
雑音駆動合成音	72.6

表 5.2: ピッチ弁別が可能であると判断した刺激対数 (全刺激)

刺激グループ	刺激種類	ピッチ弁別可能な刺激対数	ピッチ弁別可能な刺激対の割合 (%)
全刺激 (945 対中)	原音	898	95.0
	NVS	550	58.2

表 5.3: ピッチ弁別が可能であると判断した刺激対数 (同傾斜同士)

刺激グループ	刺激種類		ピッチ弁別可能な刺激対数	ピッチ弁別可能な刺激対の割合 (%)
同傾斜 (105 対中)	原音	右下り	103	98.1
		平坦	103	98.1
		右上り	102	97.1
	NVS	右下り	92	87.6
		平坦	72	68.6
		右上り	22	21.0

### 5.4.2 サーストンの一対比較法による音階の配置

第4章と同様に、サーストンの一対比較法により、各刺激に対して実験参加者が知覚した音階の配置を求めた。ここでは、同じスペクトル傾斜を持つ刺激同士の間隔を算出した。原音の音階の配置を図5.7に、雑音駆動合成音の音階の配置を図5.8に示す。図中の赤字および下線は、並びが入れ替わって知覚されていることを示している。音階の正しい知覚を示す図4.5と比較して、原音の右下りの条件では中音域で音階の並びの誤りが発生し、刺激間隔が短くなっている。平坦の条件では、音階の並びは右下りの条件の結果と概ね同様であったが、刺激間隔の変動は特にD3-E3やG3-D4の低-中音域で見られた。右上りの条件では、音階の並びは平坦の条件と同様であったが、刺激間隔は右下りのスペクトル条件での音階の配置に近くなっていた。なお、音階の並びの誤りは、右下りの条件では1箇所、平坦の条件・右上りの条件では2箇所と少なかった。

雑音駆動合成音では、スペクトル傾斜が右下りの条件においてD3-E3, B3-C4, F4-A4で刺激の間隔が広がっていた。音階の並びの誤りはF3-A3の低音域, E4-B4の中-高音域で見られた。平坦の条件では刺激全体の間隔が狭く、特にC3-C4の1オクターブにおいて並びの誤りが多く見られた。右上りの条件では、平坦の条件と比較して更に刺激全体の間隔が狭くなり、C3-F4の広い範囲で並びの誤りが多く見られた。

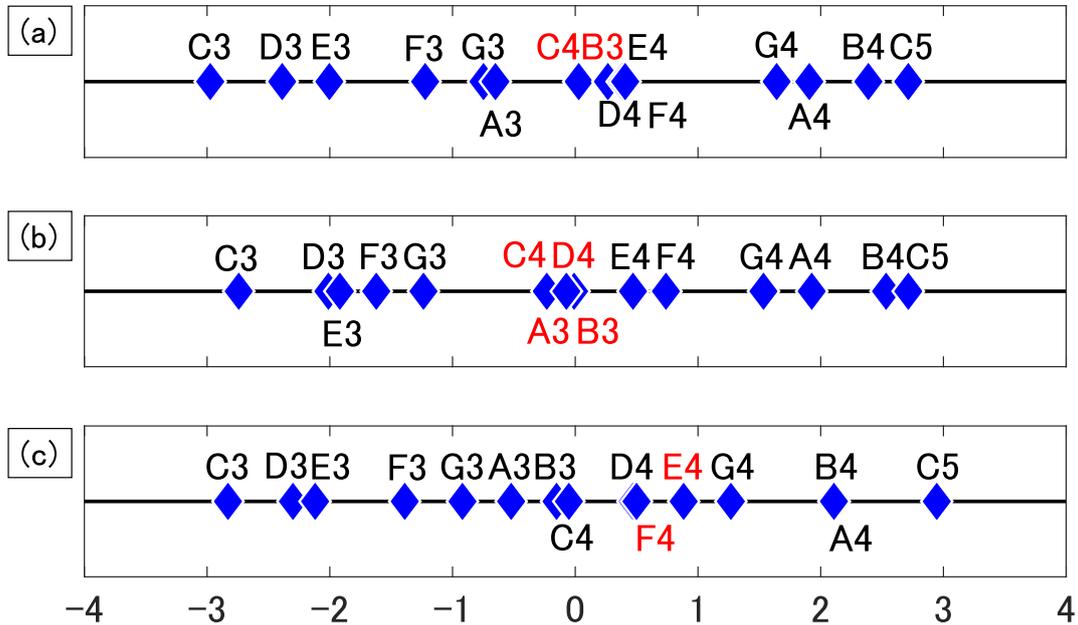


図 5.7: サーストンの一対比較法による音階の配置 (原音)。(a) は右下り, (b) は平坦, (c) は右上りのスペクトル傾斜を示す。図中の赤字および下線は、並びが入れ替わって知覚されていることを示す。

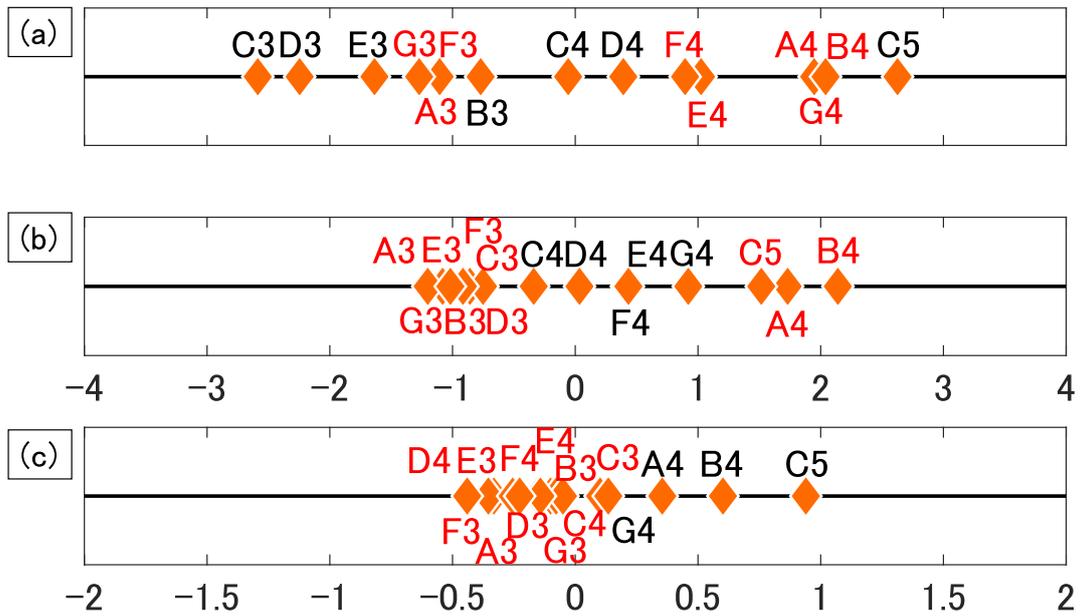


図 5.8: サーストンの一対比較法による音階の配置 (NVS) . 図の見方は図 5.7 と同様である.

## 5.5 エキサイテーションパターンの分析

### 5.5.1 分析方法

サーストンの一対比較法による音階の配置が，場所説または時間説どちらの手がかりを使って得られたかを検討するため，場所説の観点からエキサイテーションパターンに着目した．ここでは，実験刺激の雑音駆動合成音 45 種を分析対象とし，エキサイテーションパターンおよび変調スペクトルから得られる F0 を分析した．

エキサイテーションパターンの算出には，ERB 尺度に基づいたガンマトーンフィルタバンクを用いた [29]．フィルタバンクの聴覚フィルタは， $ERB_{N-number}$  が 1.8 Cam から 38.9 Cam まで 0.1 Cam 刻みに並ぶよう配置された 372 帯域であった．ガンマトーンフィルタバンクの各帯域の出力信号に半波整流および二乗処理を行い，カットオフ周波数 600 Hz の低域通過フィルタの出力の定常区間 0.5 – 1.0 s を時間平均してエキサイテーションパターンを算出した．その後，図 3.1 に示すフィルタバンクにおいて位相同期が起こる上限の周波数である [30]，[31] 中心周波数 1.5 kHz 以下に該当し，原音で倍音成分を含む帯域のエキサイテーションに対し，Lag 窓と Clipping 処理を組み込んだ変形自己相関法 [32] を利用して算出した F0 を音階として配置した．さらに，原音において F0 が含まれる帯域のエキサイテーションの時間変化を算出し，同様にして F0 の算出を行った．算出された F0 を音階として配置した．なお，調波性を検出できなかった刺激の F0 は 0 Hz として配置した．

## 5.5.2 分析結果

算出したエキサイテーションパターンの例として、右下り、平坦、右上りのスペクトル条件における G3 のエキサイテーションパターンを図 5.9 – 図 5.11 に示す。図中の点線は原音のエキサイテーションパターンを、実線は雑音駆動合成音のエキサイテーションパターンを示す。原音では 1.5 kHz より低域側のエキサイテーションパターンで倍音構造が見られたのに対し、雑音駆動合成音のエキサイテーションパターンでは、顕著な倍音構造が見られなかった。一部、倍音構造のようなものが見られるが、これらは、倍音成分が含まれない聴覚フィルタが存在するため、その帯域にエキサイテーションパターン上の谷が形成され、見かけ上のピークが生じたことによるものである。

次に、図 3.1 に示すフィルタバンクにおいて中心周波数 1.5 kHz 以下に該当し、原音で倍音成分を含む帯域のエキサイテーションパターンから算出された F0 を利用して各刺激の配置を求めた。その結果を図 5.12 に示す。三つのスペクトル傾斜の条件において、0 Hz に配置された刺激が多く見られた。右下りの条件では、C3 – C4 における半数以上の刺激の F0 が算出されず、G4 – C5 の刺激においても並びは音楽音階と同様であるが、間隔変動が大きかった。平坦の条件では F0 算出することができた刺激は A4 のみであった。実際の F0 と比較して僅かに低い算出値であった。右上りの条件ではどの刺激においても調波性を検出できなかったため、音階の配置として並べることができなかった。

原音の F0 が含まれる帯域におけるエキサイテーションの時間変化から算出された F0 を音階として配置したものを図 5.13 に示す。右下りの条件では、音楽音階に似た配置となっているものの、1 オクターブ離れている C4, C5 が隣接して配置されているなどの並びの誤りが見られた。平坦の条件では F0 を算出できなかった刺激が主に低音域で多く見られた。右上りの条件では、並びは右下りの条件と同様であるが、主に低音域および中音域で間隔変動が見られた。全ての条件において、高音域では原音の F0 が含まれる帯域における雑音駆動合成音のエキサイテーションの時間変化から算出された音階が音楽音階に近い配置となったが、右下りおよび平坦の条件において C5 の配置が異なった。これらの結果は、サーストンの一対比較法による音階の配置とは一致していない。

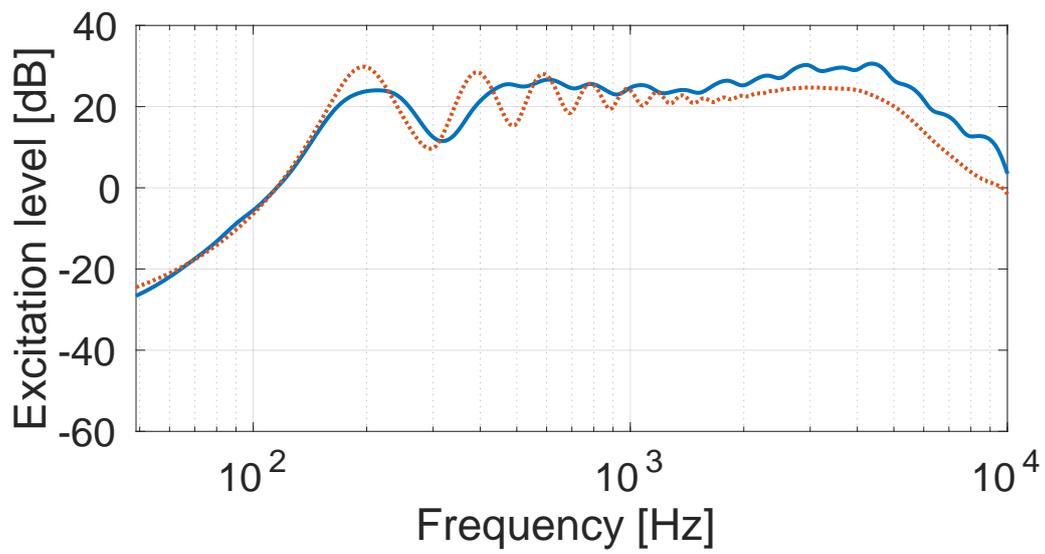


図 5.9: エキサイテーションパターン (G3, 右下り)

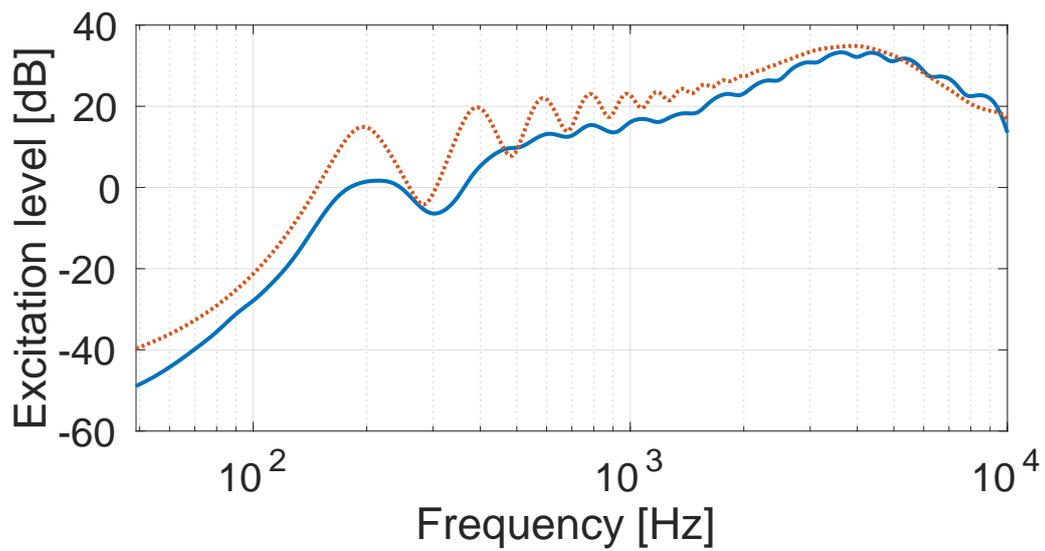


図 5.10: エキサイテーションパターン (G3, 平坦)

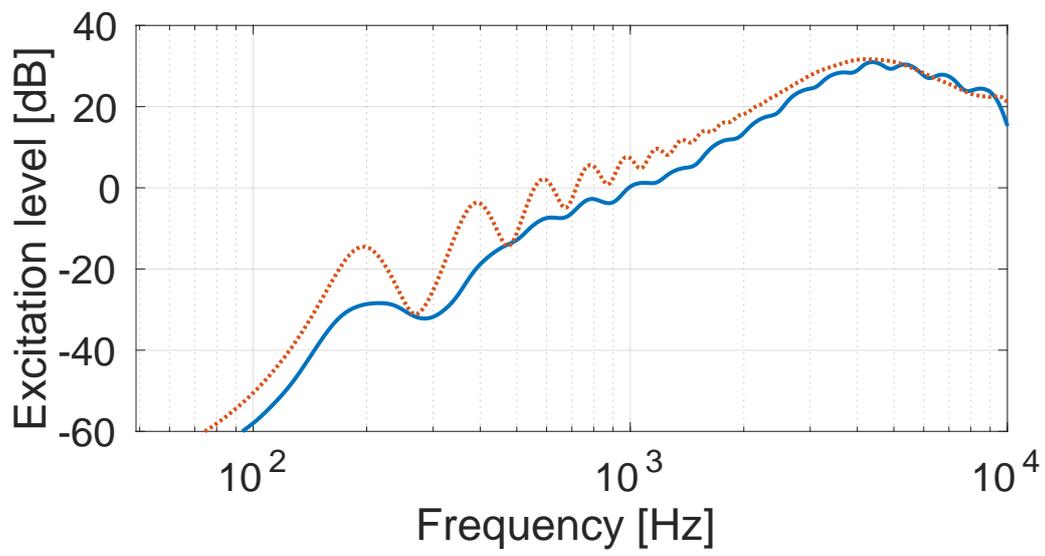


図 5.11: エキサイテーションパターン (G3, 右上り)

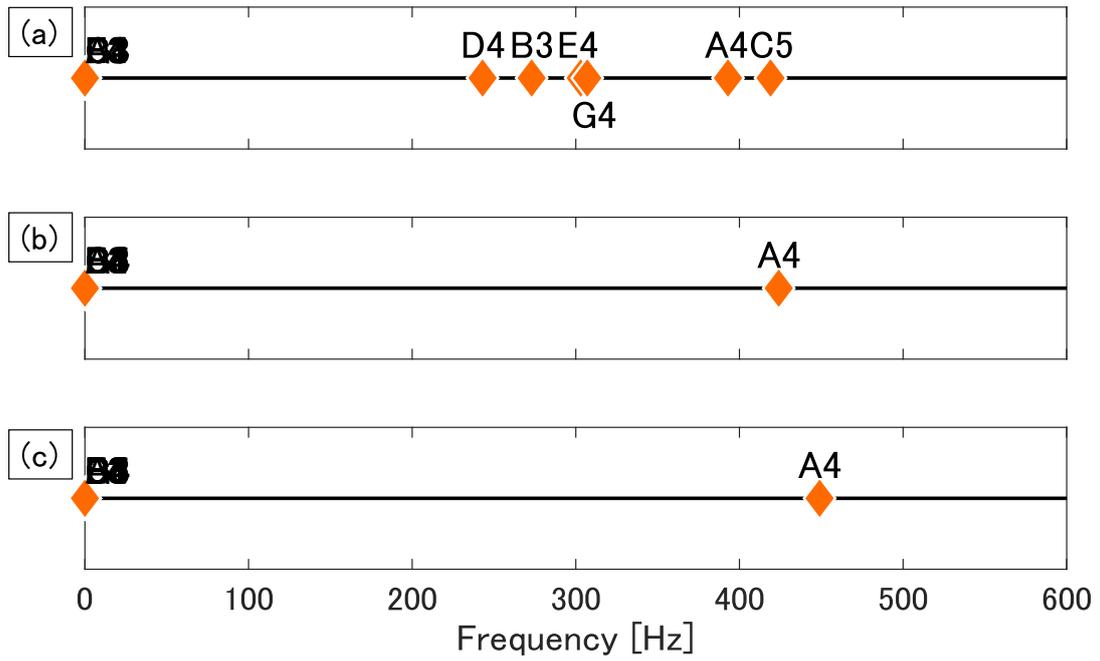


図 5.12: 図 3.1 に示すフィルタバンクにおいて 1.5 kHz 以下の帯域のエキサイテーションパターンから求めたサーストンの一対比較法による音階の配置 (NVS)

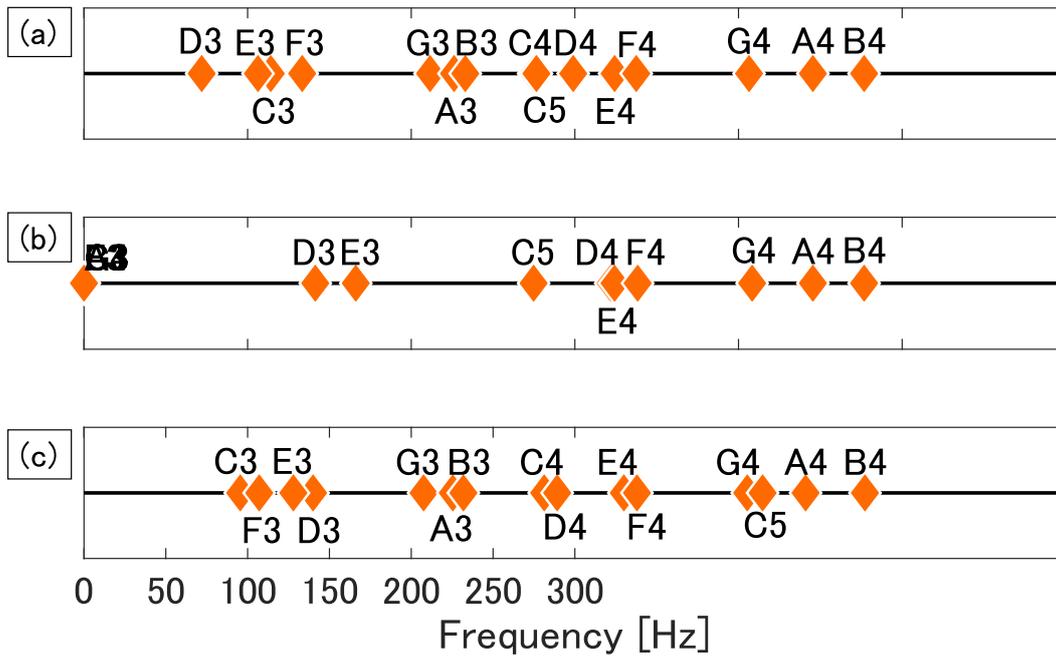


図 5.13: 原音で F0 を含む帯域のエキサイテーションパターンから求めたサーストンの一対比較法による音階の配置 (NVS)

## 5.6 変調スペクトルの分析

### 5.6.1 分析方法

サーストンの一対比較法による音階の配置が，場所説または時間説どちらの手がかりを使って得られたかを検討するため，時間説の観点から振幅包絡線情報のスペクトルである変調スペクトルに着目した．変調スペクトルの算出には，図 3.1 に示す各帯域での振幅包絡線情報を利用した．ここでは，図 3.1 に示すフィルタバンクにおいて 1.5 kHz 以下に該当する帯域の振幅包絡線情報に，Lag 窓と Clipping 処理を組み込んだ変形自己相関法 [32] を利用して算出した F0 を音階として配置した．

### 5.6.2 分析結果

例として，算出した C4 の変調スペクトルを図 5.14-5.16 に示す．変調スペクトル上に F0 に対応する調波性が見られた．原音において C4 の F0 を含む帯域である第 3 帯域では調波性が見られず，ピーク値も低くなっていた．一方，第 8 帯域以上の帯域では調波性が見られ，ピーク値も高くなっていた．

次に，図 3.1 に示すフィルタバンクにおいて，中心周波数が 1.5 kHz 以下の帯域に限定して変調スペクトルを算出し，自己相関法によって F0 算出を行った．算出された F0 を周波数軸上に配置したものを図 5.17 に示す．いずれの条件においても C3—B3 の低音域において F0 算出が可能であった．音階の配置は，音楽音階とは大幅に異なった．平坦および右上りの条件では間隔変動が少なく，比較的等間隔に配置され，図 5.7 と図 5.8 に示すサーストンの一対比較法による音階の配置とは並び方が大幅に異なった．

原音において F0 が含まれる帯域の変調スペクトルでは，自己相関法によって F0 の算出ができなかったため，第 16 帯域の変調スペクトルから算出された F0 を周波数軸上に配置したものを図 5.18 に示す．第 16 帯域の変調スペクトルでは，いずれの条件においても音楽音階に似た音階の配置となった．A4 および B4 は F0 算出ができないか，A4 では 1 オクターブ下の A3 と同じ周波数で算出された．図 5.7 および図 5.8 に示すサーストンの一対比較法による音階の配置とは，右下りのスペクトル条件では比較的似ているが，高音域の配置が大きく異なった．また，平坦および右上りのスペクトル条件では，低音域の配置が大幅に異なった．

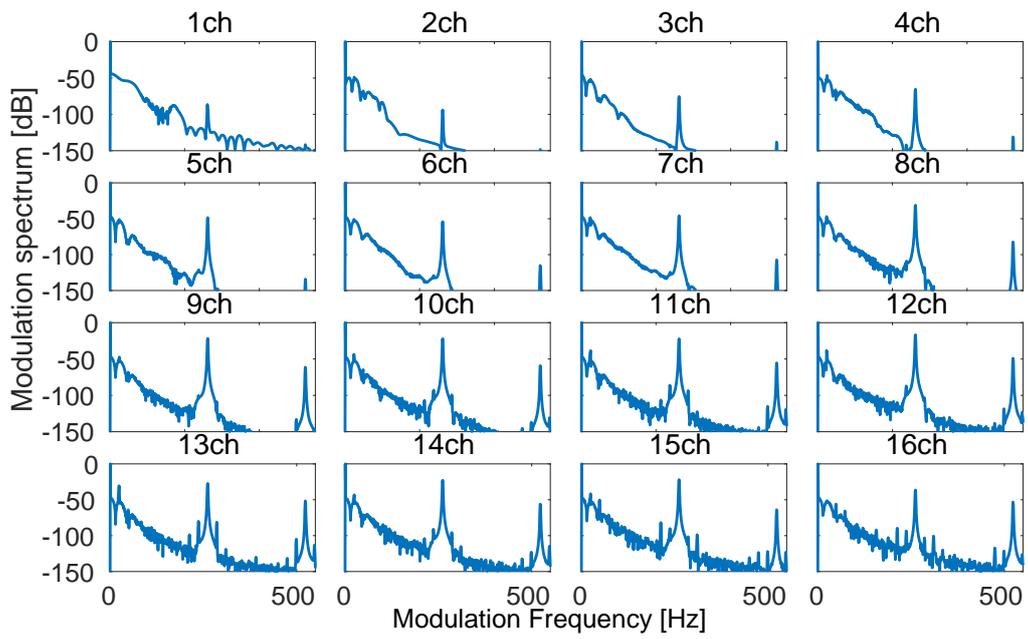


図 5.14: 変調スペクトル (C4, 右下り)

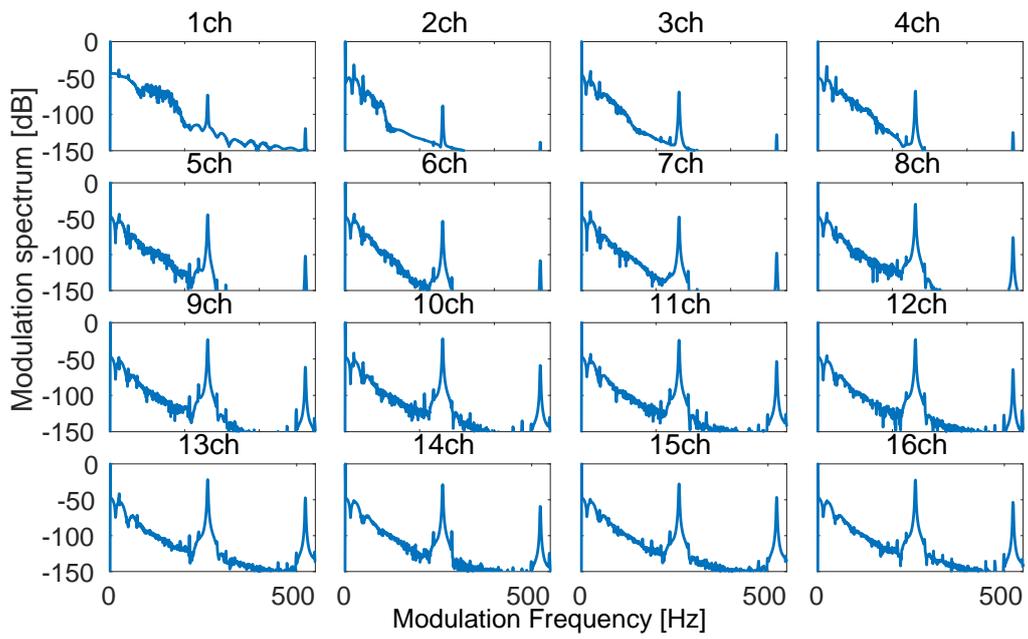


図 5.15: 変調スペクトル (C4, 平坦)

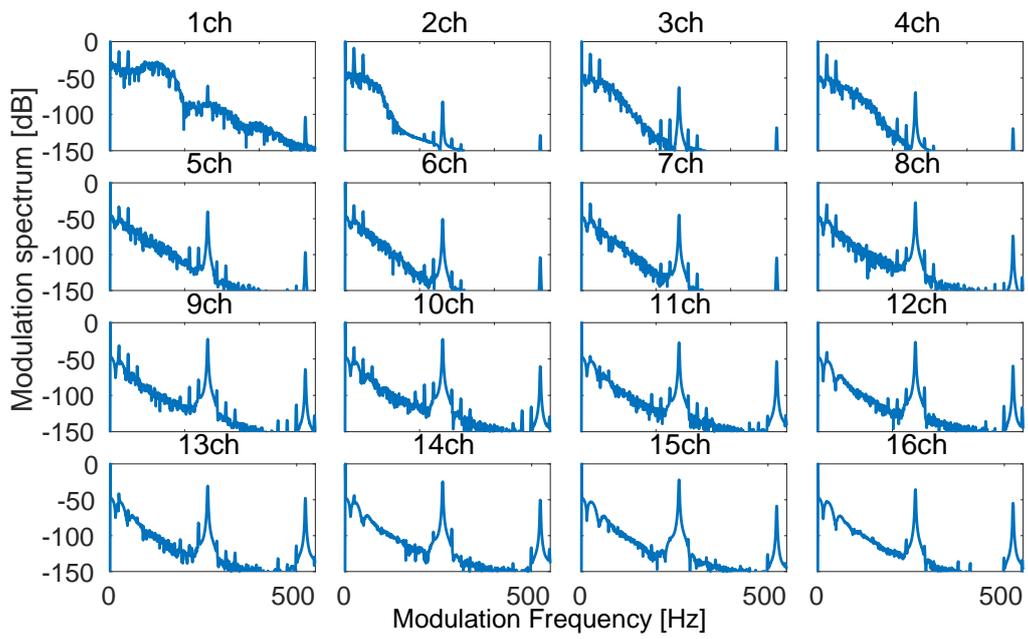


図 5.16: 変調スペクトル (C4, 右上り)

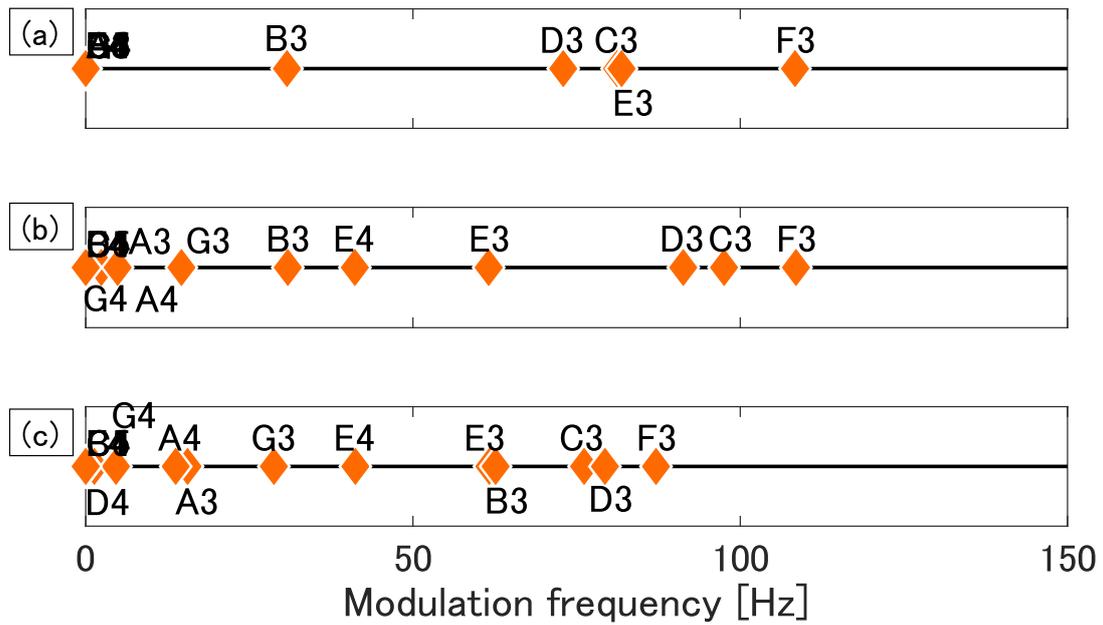


図 5.17: 図 3.1 に示すフィルタバンクにおいて 1.5 kHz 以下の帯域の変調スペクトルから求めたサーストンの一対比較法による音階の配置

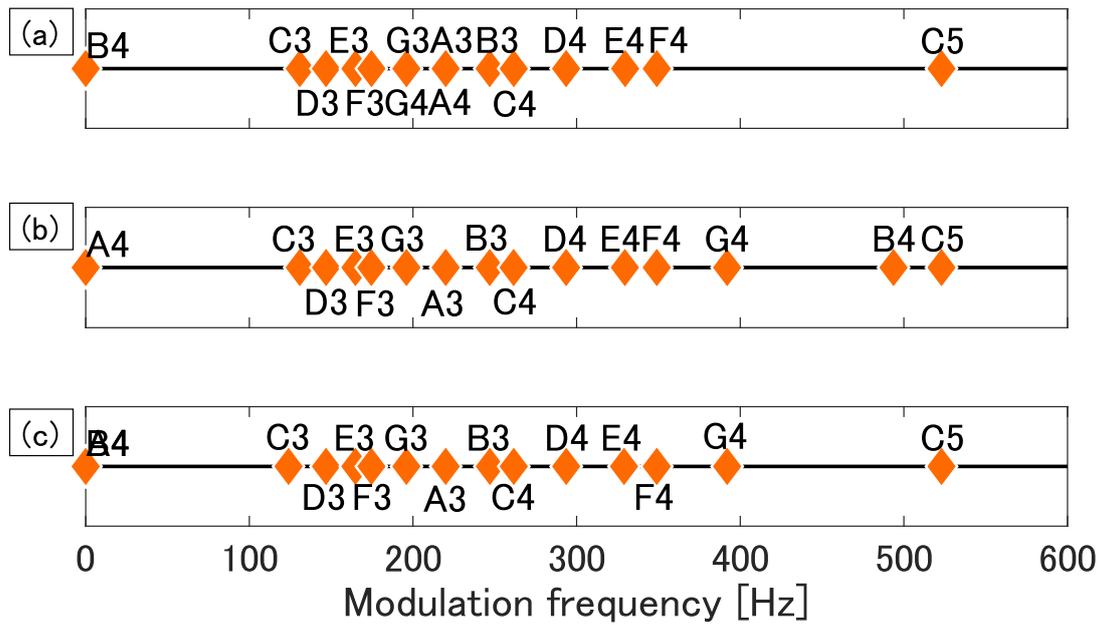


図 5.18: 第 16 帯域の変調スペクトルから求めたサーストンの一対比較法による音階の配置

## 5.7 考察

### 5.7.1 実験結果の考察

スペクトル傾斜が右下りの条件において、雑音駆動合成音では87.6%の刺激対でピッチ弁別が可能であった。このことから、スペクトル傾斜が右下りの条件であればピッチを知覚できる可能性があると考えられる。スペクトル傾斜が平坦の条件および右上りの条件では、原音での結果と異なり、音階の並びに多くの誤りが見られた。さらに、これらの条件では刺激全体の音階の配置の間隔が短くなっていったため、同じスペクトル傾斜内では刺激が似て聴こえていたと考えられる。そのため、平坦の条件および右上りの条件では、既報 [9] – [12] と同様にピッチ知覚は困難であると考えられる。サーストンの一対比較法による音階の配置では、図 5.8(b), (c) から、雑音駆動合成音は高音ほど原音に近い知覚ができると考えられる。さらに、他のスペクトル傾斜の条件と比較して、右下りの条件の刺激では、音階を正しく弁別できていると考えられる。これは、原音が低域の調波成分を多く含むことから、基本周波数成分を手がかりとしてピッチ弁別をしている可能性があると考えられる。また、スペクトル傾斜の条件によってピッチ弁別の結果が異なったことから、スペクトル傾斜も手がかりとしてピッチ知覚をしている可能性があると考えられる。しかし、サーストンの一対比較法による刺激同士の間隔がピッチ弁別にどの程度影響するかまでは明らかにできなかった。

第4章と同様に、変調スペクトルに基本周波数に関係した調波性が見られた。例として、C4の図 3.1 に示す帯域通過フィルタの3番目のフィルタ出力の変調スペクトルを図 5.19 – 図 5.21 に示す。これらのピークは261.6 Hzと523.2 Hz、図 5.21 の右上りの条件では784.8 Hzに見られた。右下りの条件から平坦の条件、右上りの条件にかけて、基本周波数に対応する変調スペクトルのピークでは約6 dB、第2倍音に対応する変調スペクトルのピークでは約10 dBほど、スペクトルレベルが強くなっていることがわかった。これらは、ピッチ知覚の手がかりと利用されている可能性があるが、変調知覚の手がかりとしてどの程度関係しているか明らかでない。この点については、次項にて考察を行う。

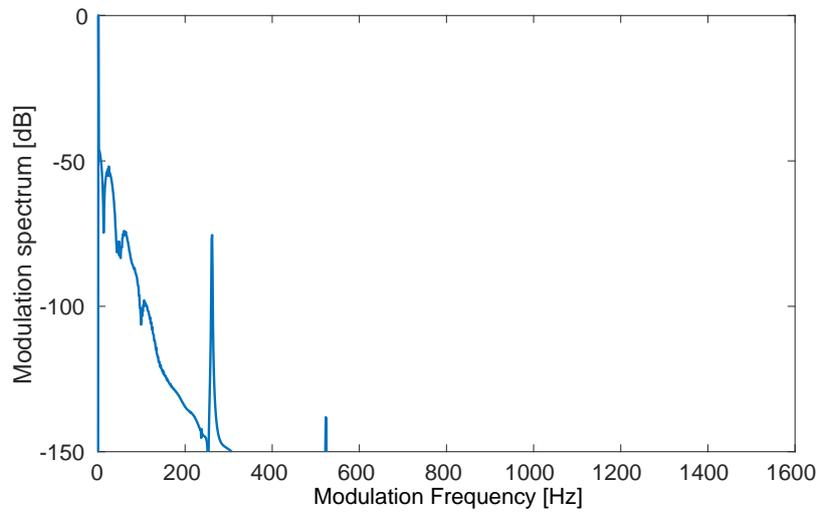


図 5.19: 変調スペクトル (C4, 第3帯域, 右下り)

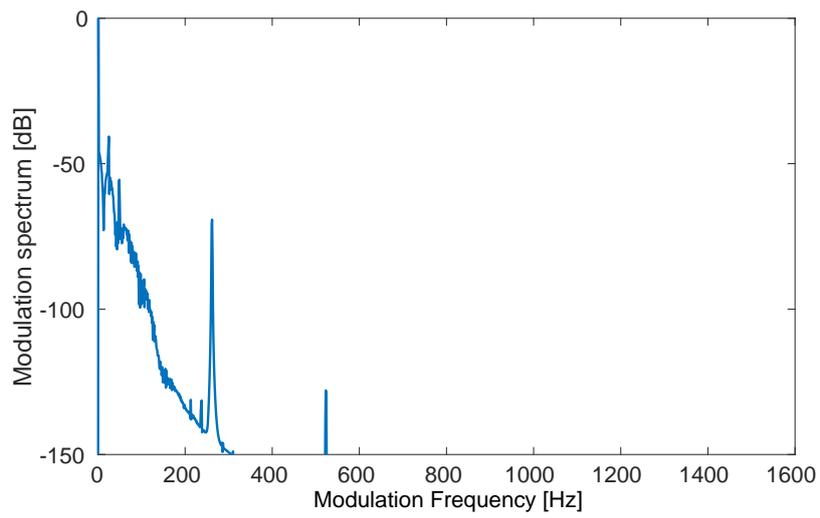


図 5.20: 変調スペクトル (C4, 第3帯域, 平坦)

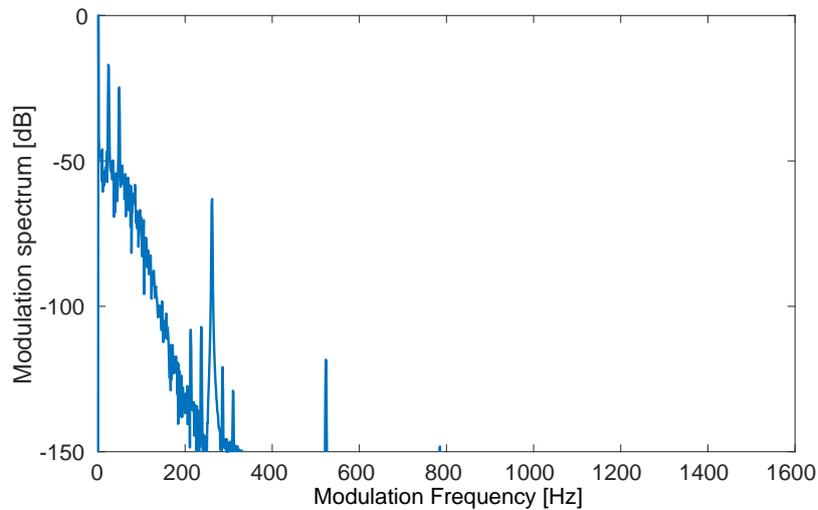


図 5.21: 変調スペクトル (C4, 第3帯域, 右上り)

### 5.7.2 エキサイテーションパターンと実験結果

エキサイテーションパターンに見られるピークの明確さがスペクトル傾斜の条件によって異なった。右下りの条件におけるエキサイテーションパターンのピークが最も明確であること、サーストンの一対比較法による音階の配置が右下りで最も音階の並びの誤りが少ないことから、雑音駆動合成音のピッチ弁別はエキサイテーションパターンにみられるピークの明確さに影響を受けていると考えられる。図 3.1 に示すフィルタバンクにおいて、フィルタの中心周波数が 1.5 kHz 以下に該当する帯域であり、原音において倍音成分を含む帯域のエキサイテーションパターンによる音階の配置 (図 5.12) では、右下りの条件において低音域で F0 が 0 Hz となった。そのため、これらをピッチ弁別の手がかりであると考えれば、低音域のピッチ弁別は困難であることが予想される。しかし、右下りの条件におけるサーストンの一対比較法による音階の配置 (図 5.8(a)) では、低音域の並びの誤りが少なく、高音域の並びの誤りが多く生じていることから、これらがピッチ弁別の手がかりであるとは考え難い。そのため、図 3.1 に示すフィルタバンクにおいて、1.5 kHz 以下に該当する帯域のエキサイテーションパターン (場所説) を利用して雑音駆動合成音のピッチ弁別をしている可能性は低いと考えられる。原音の F0 が含まれる帯域における雑音駆動合成音のエキサイテーションの時間変化から算出された F0 の配置 (図 5.13) では、右下りおよび平坦の条件において C5 が D4 の隣に配置されたことから、1 オクターブ下の C4 とオクターブエラーを起こしている様に見られた。しかし、サーストンの一対比較法による音階の配置 (図 5.8) では C5 および C4 のオクターブエラーが生じていないことから、今回の結果とは一致していない。また、右上りの条件では低音域の算出ができていないほか、C5 および C4 のオクターブエラーは生じていない。しかし、サーストンの一対比較法によ

る音階の配置（図 5.8）では，低音域の弁別ができていないことから，今回の結果とは一致していない．そのため，原音の F0 が含まれる帯域におけるエキサイテーションの時間変化を利用して雑音駆動合成音のピッチ弁別している可能性は低いと考えられる．

### 5.7.3 変調スペクトルと実験結果

図 3.1 に示すフィルタバンクにおいて，1.5 kHz 以下に該当する帯域で得られた変調スペクトルでは，C3-B3 の低音域においてのみ F0 算出が可能であった（図 5.17）．しかし，サーストンの一対比較法による音階の配置（図 5.8）では，平坦および右上りの条件において C3-B3 の低音域の音階の配置が音楽音階と大きく異なっていた．また，右下りの条件の配置（図 5.17(a)）もサーストンの音階の配置とは一致していない．これらのことから，変調スペクトルの周期性（時間説）を利用して雑音駆動合成音のピッチ弁別をした可能性は低いと考えられる．

Shamma らは，Unresolved な倍音成分に Resolved な倍音成分を組み合わせることで振幅包絡線からピッチ算出を行った [19] が，本章では Resolved の帯域では調波性を検出できなかったため，音階の配置を正しく求めることができなかった．さらに，原音の F0 が含まれる帯域の変調スペクトルでは，自己相関法によって調波性を検出できなかった．これは，図 7 に示すように変調スペクトルで F0 が含まれる第 3 帯域の調波性が見られなかったことによるものと考えられる．

全ての刺激において Unresolved な倍音成分が含まれる第 16 帯域の変調スペクトルでは，自己相関法により算出された F0 による音階の配置が音楽音階に似ていた（図 5.18）．これは，聴覚フィルタの該当する帯域に倍音成分が多数含まれており，変調周波数軸上に調波性が見られるためであると考えられる．右下りの条件では変調スペクトルによる音階の配置とサーストンの一対比較法による音階の配置が似ているが，A4 の算出された F0 が 1 オクターブ下の A3 と同じ値でありオクターブエラーが生じている．このことから，変調スペクトルの特定の帯域における調波性（時間説）を利用して雑音駆動合成音のピッチ弁別をしている可能性は低いと考えられる．図 3.1 のフィルタバンクの低い帯域では，図 5.19-図 5.21 に示すように，変調スペクトルに見られるピーク値が低い．そのため，雑音駆動合成音のピッチを弁別できていない可能性が考えられる．したがって，変調スペクトルに見られるピークの数を増加させた場合，雑音駆動合成音のピッチを弁別できる可能性が考えられる．

第 16 帯域の変調スペクトルでは，いずれの条件においても音楽音階に似た音階の配置となった．このことから，Shamma ら [19] の報告と同様に，図 3.1 に示すフィルタバンクでフィルタの中心周波数が 1.5 kHz 以下に該当する帯域と Unresolved な帯域を組み合わせることで，雑音駆動合成音のピッチ弁別の手がかりを検討できる可能性がある．

# 第6章 雑音駆動合成音のピッチ知覚の手がかりの検討

## 6.1 実験目的

雑音駆動合成音のピッチ弁別の手がかりを検討することを目的とする。特に、第5章におけるピッチ弁別の成績が悪かったC3 – C4の1オクターブ目に着目してピッチ弁別の手がかりを検討する。また、ピッチ弁別結果と実験刺激のエキサイテーションパターンおよび変調スペクトルとの関連についても検討する。

## 6.2 実験刺激

第5章で用いた45種の調波複合音のうち、C3 – C4 (130.8 Hz – 261.6 Hz) の調波複合音24種を使用した。これらの調波複合音を原音とする振幅包絡線抽出時のカットオフ周波数  $F_c=64, 128, 256, 1024$  Hz の雑音駆動合成音45種を作成し、雑音駆動合成音のみを実験刺激とした。ここでは、雑音駆動合成音作成時に用いる低域通過フィルタについて、第4章および第5章とは異なる急峻な低域通過フィルタを用いた。実験刺激の例を図6.1 – 図6.3に示す。

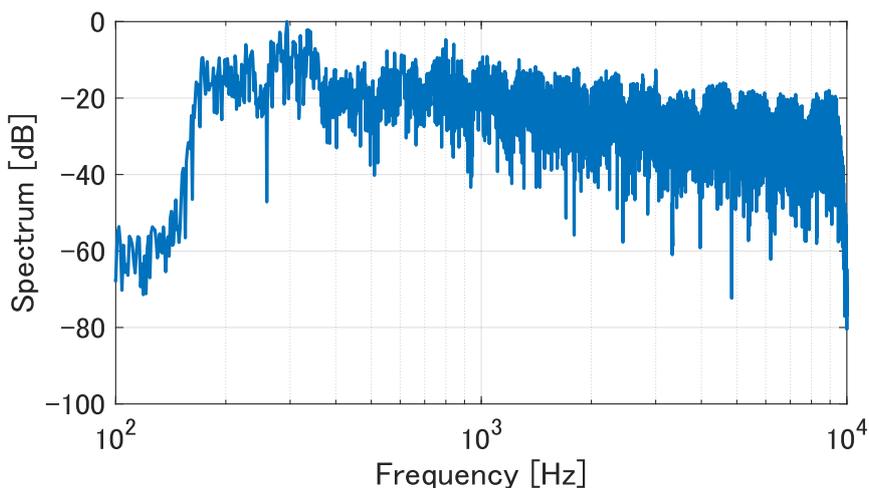


図 6.1: 実験刺激 (C4, 右下り,  $F_c=64$ Hz)

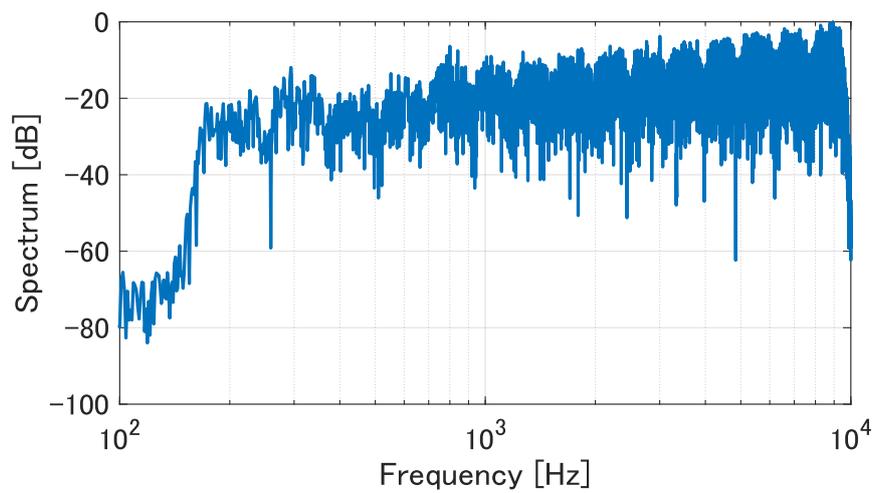


図 6.2: 実験刺激 (C4, 平坦,  $F_c=64\text{Hz}$ )

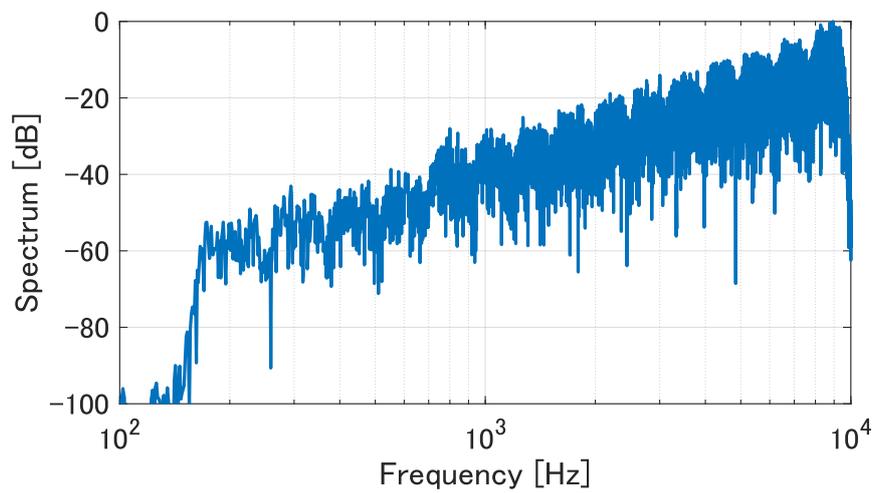


図 6.3: 実験刺激 (C4, 右上り,  $F_c=64\text{Hz}$ )

## 6.3 実験手続き

第4章および第5章と同様に，実験にはサーストンの一対比較法を用いた．同音同士を除いた刺激対を実験参加者にランダムに呈示し，音が高いと感じる方を強制的に判断させた．各刺激の呈示は1 sとし，刺激と刺激の間隔は0.5 sであった．ここでは，実験刺激のスペクトル条件を問わず総当たりで呈示した．実験は1セクション252対とし，雑音駆動合成音4セクションの計1008対行った．実験参加者は正常聴力を有する成人10名（23–28歳，うち女性2名）であった．

実験は無響室で行われ，刺激の呈示にはPC（Windows10, MATLAB），オーディオインターフェース（Fireface UCX），ヘッドフォンアンプ（AudioTechnica AT-HA21），ヘッドフォン（SENNHEISER HDA-200）を使用した．実験刺激のサンプリング周波数は44.1 kHzとした．刺激は，人工耳（B&K Artificial Ear Type 4153），マイク（B&K Type 4291），騒音計（B&K Sound Level Meter Type 2250）を用いて，ヘッドフォンからの出力がA特性音圧レベルが70 dBとなるように設定された．

## 6.4 実験結果

### 6.4.1 正答率

第4章および第5章と同様にして，振幅包絡線抽出時のカットオフ周波数毎に全刺激対の平均正答率を算出した．算出した結果を表6.1に示す．いずれのカットオフ周波数の刺激においても50%程度と低かった．

さらに，第4章および第5章と同様にして，各刺激対に対する実験参加者の平均正答率を算出した．ピッチ弁別が可能と判断された刺激対数と，全刺激対に対するピッチ弁別が可能と判断された刺激対の割合を表6.2に示す．また，本来のピッチ弁別と逆の弁別が可能であると判断できる平均正答率25%以下の刺激対数と，全刺激対数に対するその割合を表6.3に示す．ピッチ弁別が可能であると判断された刺激対数は $F_c=64$  Hzで14， $F_c=128$  Hzで16， $F_c=256$  Hzで9， $F_c=1024$  Hzで18であり，第4章および第5章と比較して大幅に少なかった．平均正答率25%以下の刺激対数は $F_c=64$  Hzで16， $F_c=128$  Hzで18， $F_c=256$  Hzで12， $F_c=1024$  Hzで36であり，ピッチ弁別が可能であると判断された刺激対数よりもわずかに多かった．特に， $F_c=1024$  Hzでは，平均正答率25%以下の刺激対数は正しくピッチ弁別が可能な刺激対数の2倍であった．

表 6.1: 全刺激対に対する全実験参加者のピッチ弁別平均正答率

カットオフ周波数 (Hz)	全刺激対に対する 全実験参加者の平均正答率 (%)
64	50.0
128	44.6
256	45.6
1024	50.2

表 6.2: ピッチ弁別が可能であると判断した刺激対数 (全刺激)

カットオフ 周波数 (Hz)	ピッチ弁別可能 な刺激対数	ピッチ弁別可能な 刺激対の割合 (%)
64	14	5.56
128	16	6.35
256	9	3.57
1024	18	7.14

表 6.3: ピッチを逆に弁別可能であると判断した刺激対数 (全刺激)

カットオフ 周波数 (Hz)	平均正答率 25% 以下の刺激対数	平均正答率 25% 以下の刺激対の割合 (%)
64	16	6.35
128	18	7.14
256	12	4.76
1024	36	14.3

表 6.4: ピッチ弁別が可能であると判断した刺激対数 (同傾斜同士)

カットオフ 宗派酩 (H)	刺激グループ	ピッチ弁別可能な 刺激対数	ピッチ弁別可能な 刺激対の割合 (%)
64	右下り	2	7.14
	平坦	4	14.3
	右上り	1	3.57
128	右下り	4	14.3
	平坦	0	0
	右上り	0	0
256	右下り	2	7.14
	平坦	1	3.57
	右上り	2	7.14
1024	右下り	4	14.3
	平坦	0	0
	右上り	4	14.3

#### 6.4.2 サーストンの一対比較法による音階の配置

第4章および第5章と同様に、サーストンの一対比較法によって各刺激に対して実験参加者が知覚した音階の配置を求めた。ここでは、同じスペクトル傾斜を持つ刺激同士の尺度を算出した。雑音駆動合成音の音階の配置を図6.4-6.7に示す。

$F_c=64$  Hzの音階の配置では、いずれのスペクトル条件においても音階の並びが音楽音階とは異なっていた。右下りの条件においては、C4は最も高く知覚されているが、1オクターブ下のC3が隣に並んでいた。間隔変動については、A4 - C4は等間隔に近い配置で並んでいるが、D3 - B3については間隔が狭く、B3 - A3の間隔が広がっていた。平坦の条件においては、D3が特に低く知覚されており、隣り合ったC3との間隔がかなり広がっていた。C3 - F3、A3 - E3の間隔も広がっており、その他の間隔は狭くなっていた。右上りの条件では、C4が最も低く知覚されており、間隔を空けて1オクターブ下のC3が隣に配置されていた。D3 - F3の間隔も広がっており、その他の間隔は比較的等間隔に並んでいた。

$F_c=128$  Hzの音階の配置では、いずれのスペクトル条件においても音階の並びが音楽音階とは大きく異なっていた。右上りの条件および右下りの条件では、G3が最も低く、F3が最も高く知覚されていた。これらは、本来であれば隣り合う刺激であるが、1オクターブ離れて知覚されていた。平坦の条件ではD3が低く知覚されており、その他の刺激は狭い間隔で並んでいた。右上りの条件では、並びは音楽音階と異なるが、間隔はG3 - A3、D3 - F3で比較的等間隔に並んでいた。

$F_c=256$  Hzの音階の配置では、いずれのスペクトル条件においても音階の並びが音楽音階とは大きく異なっていた。右下りおよび右上りの条件において音階の

配置が二分化し、平坦の条件においては各刺激同士が比較的離れて知覚されていた。右下りの条件では A3 - D3, C4 - B3 の間隔が広く、D3 - C4 の間隔が広くなっていた。平坦の条件では F3 - D3 の間隔が狭くなっていた。また、最も低く知覚された A3 と B3 の間隔が広くなっていた。右上りの条件では、F3 - A3, E3 - C4 の間隔が狭く、A3 - B3 の間隔が広くなっていた。

$F_c=1024$  Hz の音階の配置では、いずれのスペクトル条件においても音階の並びが音楽音階とは大きく異なっていた。間隔変動については、他のカットオフ周波数の刺激と比較して音楽音階に近い間隔で並んでいた。右下りの条件では、E3 と C4 が同程度の高さとして知覚された。D3 - E3 については、各々の間隔は広いが、D3 - E3 の中では等間隔に配置された。G3 - D3 については、音楽音階に似たほぼ等間隔で並んでいた。平坦の条件では、F3 - C3 の間隔が狭く、C3 - G3 の間隔が広くなっていた。右上りの条件では、E3 - C4, G3 - C3 の間隔が広くなっていた。

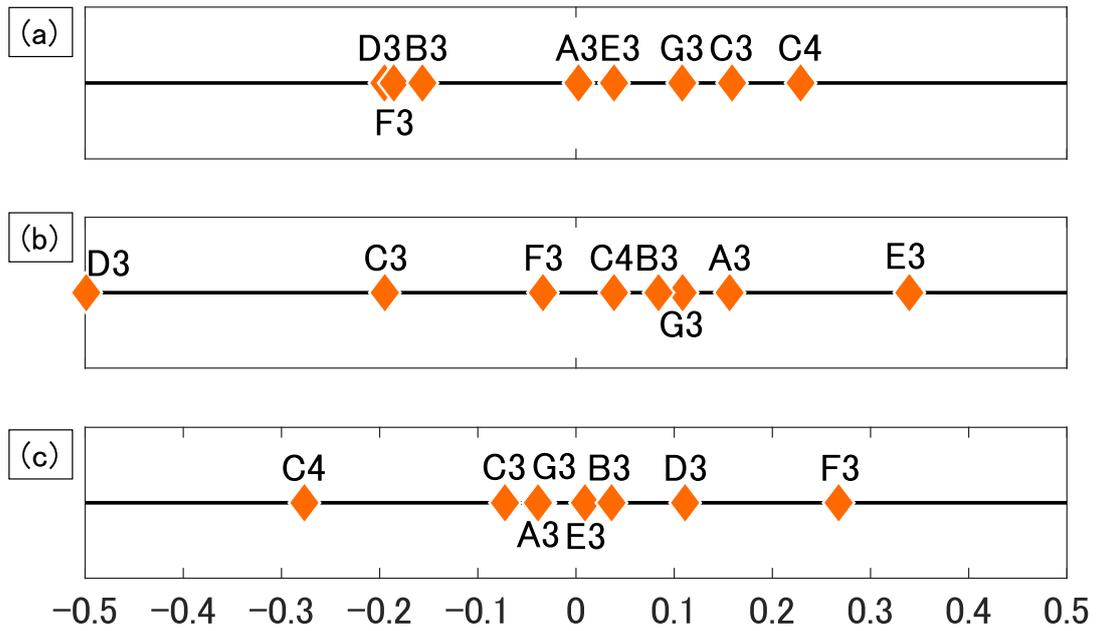


図 6.4: サーストンの一対比較法による音階の配置 ( $F_c=64$  Hz)

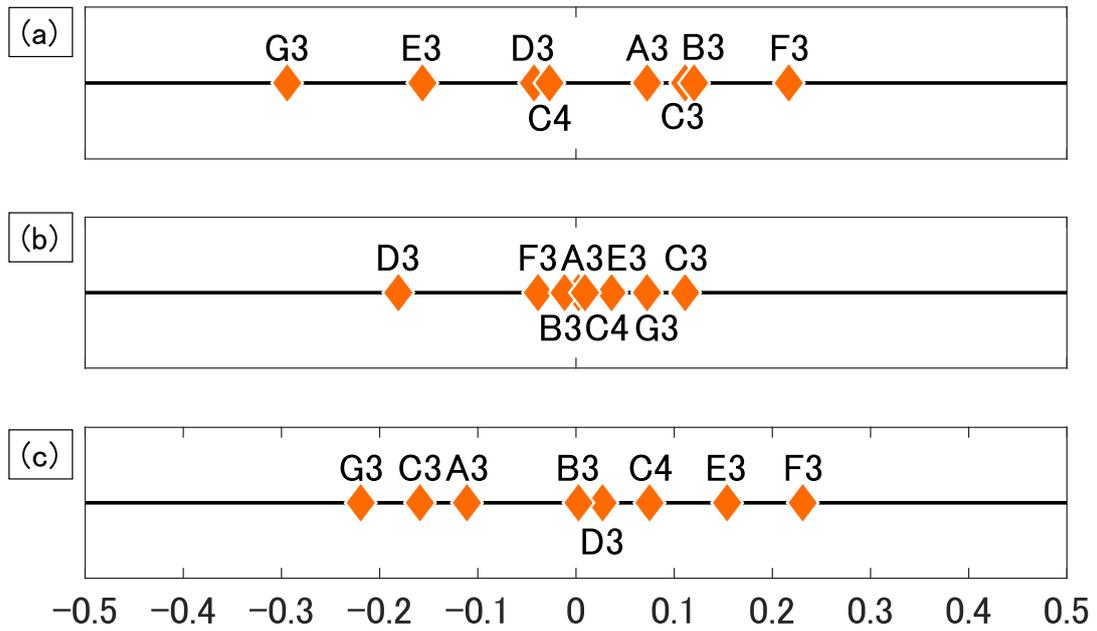


図 6.5: サーストンの一対比較法による音階の配置 ( $F_c=128$  Hz)

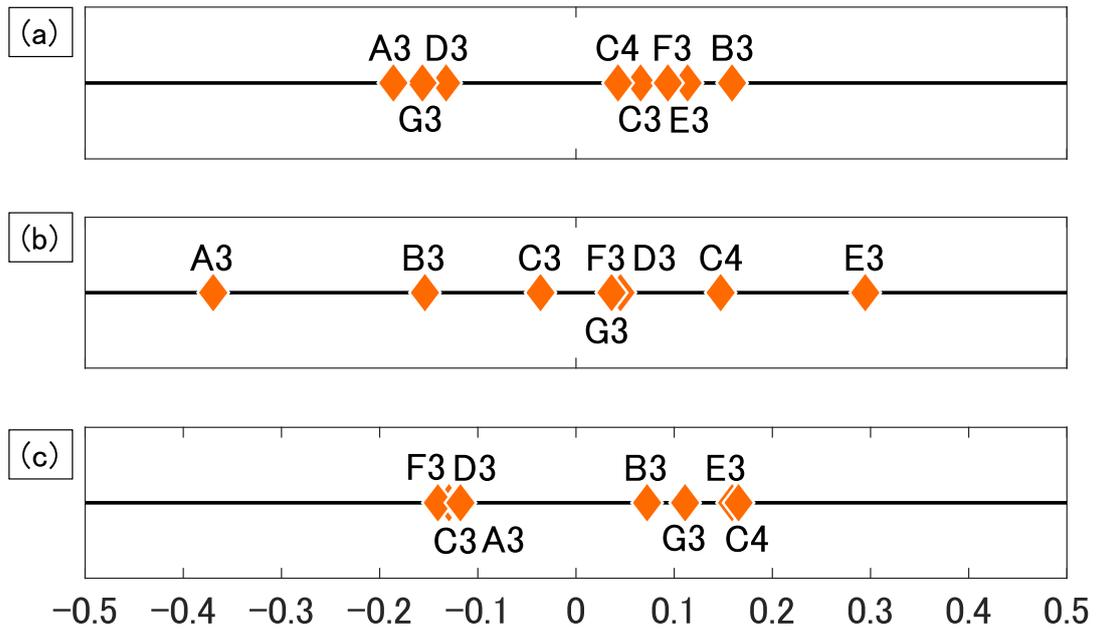


図 6.6: サーストンの一対比較法による音階の配置 ( $F_c=256$  Hz)

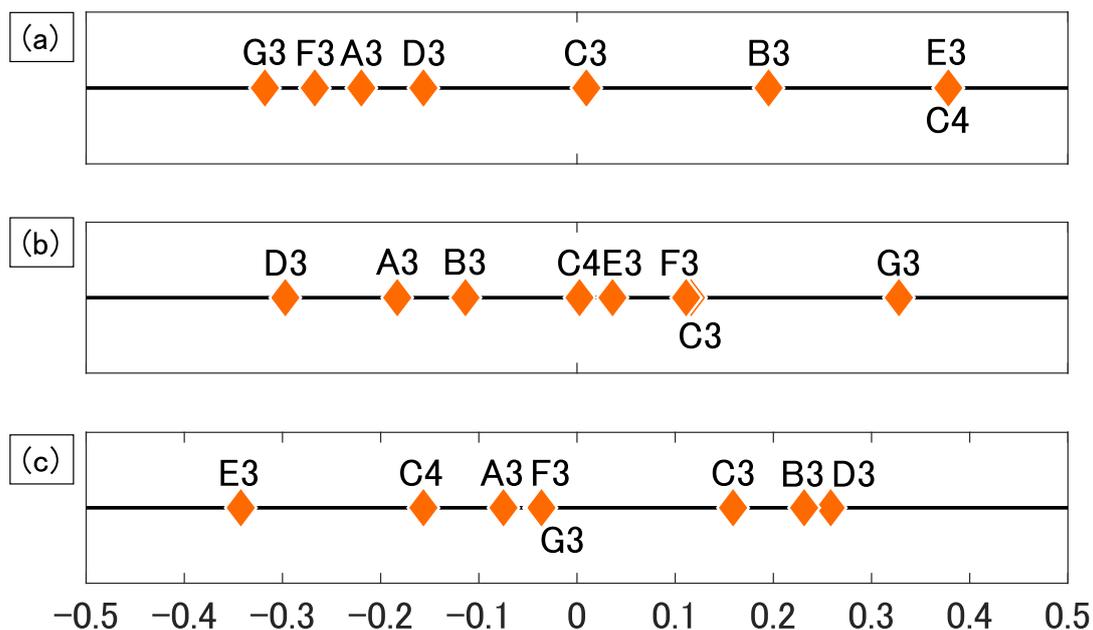


図 6.7: サーストンの一対比較法による音階の配置 ( $F_c=1024$  Hz)

## 6.5 エキサイテーションパターンの分析

### 6.5.1 エキサイテーションパターンの形状

第5章と同様にして，実験刺激のエキサイテーションパターンの算出した．算出したエキサイテーションパターンの例を図6.8－図6.10に示す．図中の点線は原音のエキサイテーションパターンを，実線は雑音駆動合成音のエキサイテーションパターンを示す．第5章の刺激と同様に，原音では1.5 kHzより低域側のエキサイテーションパターンで倍音構造が見られたのに対し，雑音駆動合成音のエキサイテーションパターンでは顕著な倍音構造が見られなかった．エキサイテーションパターンの倍音構造は右下りの条件において最も顕著であり，右上りの条件では顕著な倍音構造がほぼ見られなかった．雑音駆動合成音の一部に倍音構造のようなものが見られる．これらは第5章と同様に，倍音成分が含まれない聴覚フィルタが存在することでその帯域にエキサイテーションパターン上の谷が形成され，見かけ上のピークが生じたことによるものである．

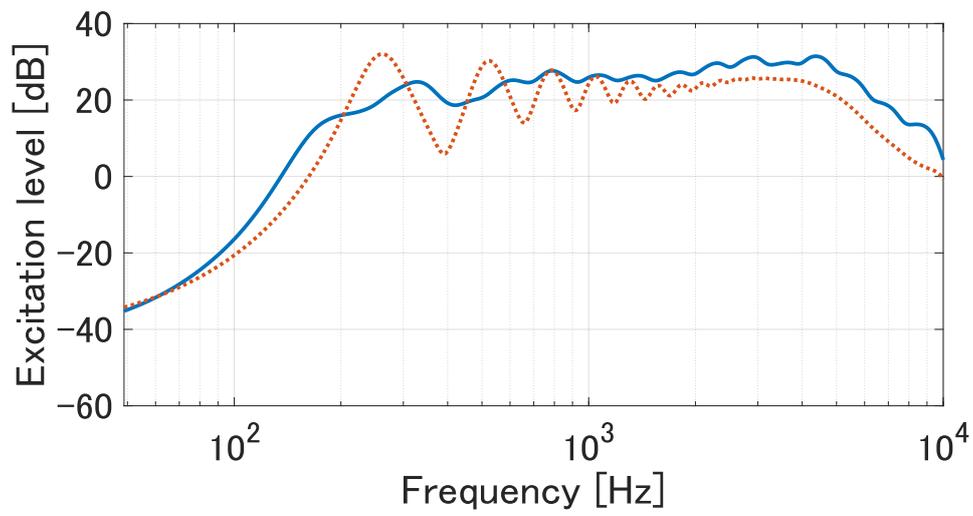


図 6.8: エキサイテーションパターン ( $F_c=64$  Hz, 右下り)

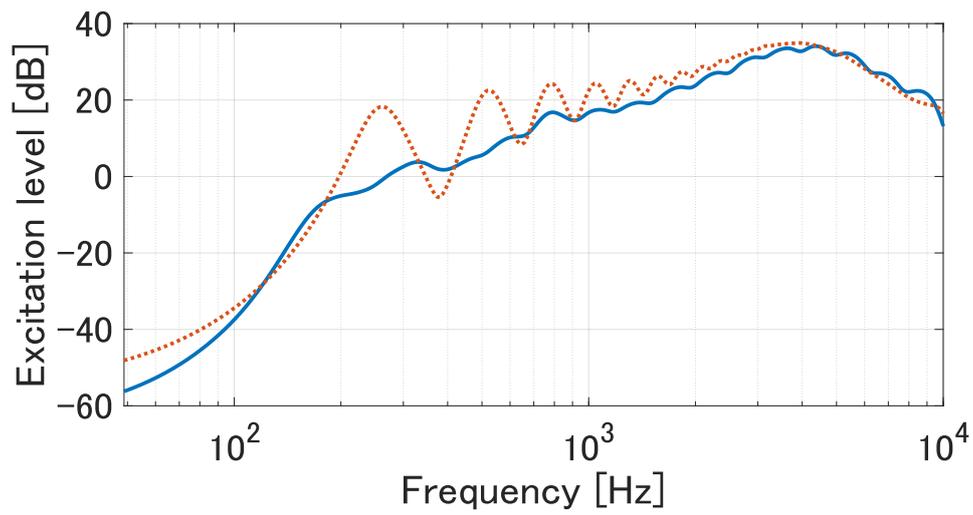


図 6.9: エキサイテーションパターン ( $F_c=64$  Hz, 平坦)

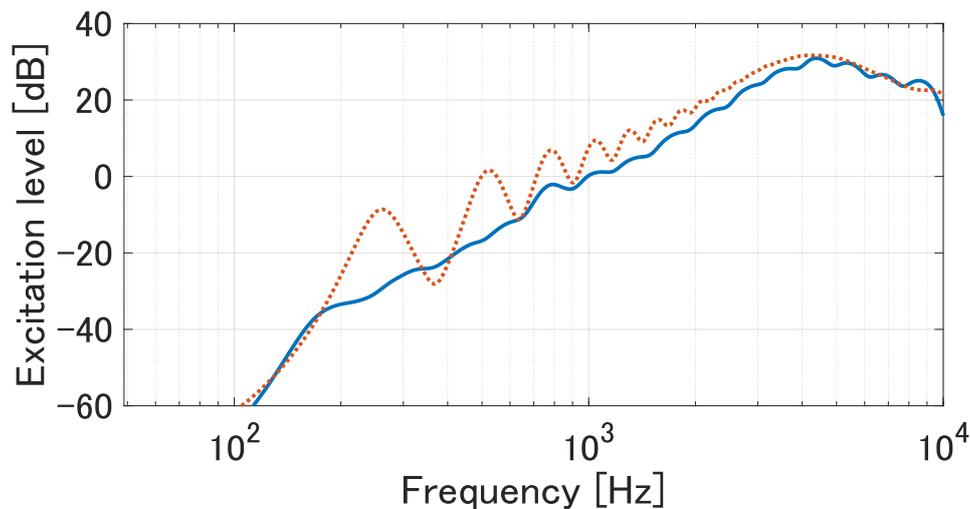


図 6.10: エキサイテーションパターン ( $F_c=64$  Hz, 右上り)

### 6.5.2 エキサイテーションパターンから算出した $F_0$ による音階の配置

図 3.1 に示すフィルタバンクにおいて中心周波数 1.5 kHz 以下に該当する帯域のエキサイテーションパターンから算出された  $F_0$  による音階の配置

図 3.1 に示すフィルタバンクにおいて中心周波数 1.5 kHz 以下に該当するエキサイテーションパターンから算出された  $F_0$  による音階の配置を求めた。その結果を図 6.11 に示す。なお、カットオフ周波数によらず音階の配置および算出された  $F_0$  が同値であった。いずれのスペクトル傾斜の条件においても 0 Hz に配置された刺激が多く見られ、右下りの条件での B3 のみ  $F_0$  を算出することができた。算出した B3 の  $F_0$  は、実際の  $F_0$  と比較して僅かに高い値であった。平坦および右上りの条件では、どの刺激においても調波性を検出できなかったため、音階の配置として並べることができなかった。

原音で  $F_0$  が含まれる帯域のエキサイテーションパターンから算出された  $F_0$  による音階の配置

原音で  $F_0$  が含まれる帯域におけるエキサイテーションの時間変化から算出された  $F_0$  を音階として配置したものを図 6.12-図 6.15 に示す。

$F_c=64$  Hz の音階の配置において、右下りの条件では音楽音階に似た配置となっているが、C3 の配置が C3 - E3 で入れ替わっており、F3 - G3, B3 - C4 の間隔が広がっていた。平坦の条件では  $F_0$  を算出できなかった刺激が多く見られた。C3 および D3, E3 の  $F_0$  は、実際の  $F_0$  とほぼ同値であった。

$F_c=128$  Hz では、右下りおよび右上りの条件の音階の配置が  $F_c=64$  Hz と同様であった。平坦の条件では、C3 および D3, E3 の F0 が実際の F0 とほぼ同値として配置された。また、G3 が実際の F0 よりも大幅に低い値で配置された。

$F_c=256$  Hz では、右下がりの条件の音階の配置が  $F_c=64, 128$  Hz と同様であった。平坦の条件では D3 および E3 のみ F0 が算出されたが、並びが入れ替わっていた。D3 は実際の F0 とほぼ同値で配置されているが、E3 は実際の F0 よりも大幅に低い値で配置された。右上りの条件では、D3 - C4 は実際の F0 とほぼ同値で配置されたが、C3 および F3, D3 は実際の F0 より低い値で配置された。 $F_c=64, 128$  Hz と比較して C3 - F3 の間隔が広がっていた。

$F_c=1024$  Hz では、右下りおよび右上りの条件の音階の配置が  $F_c=64, 128$  Hz と同様であった。平坦の条件では、D3 が実際の F0 とほぼ同値で配置された。B3 および G3, E3 は実際の F0 よりも大幅に低い値で配置された。これらの結果は、図 6.4 - 図 6.7 サーストンの一対比較法による音階の配置とは一致しなかった。

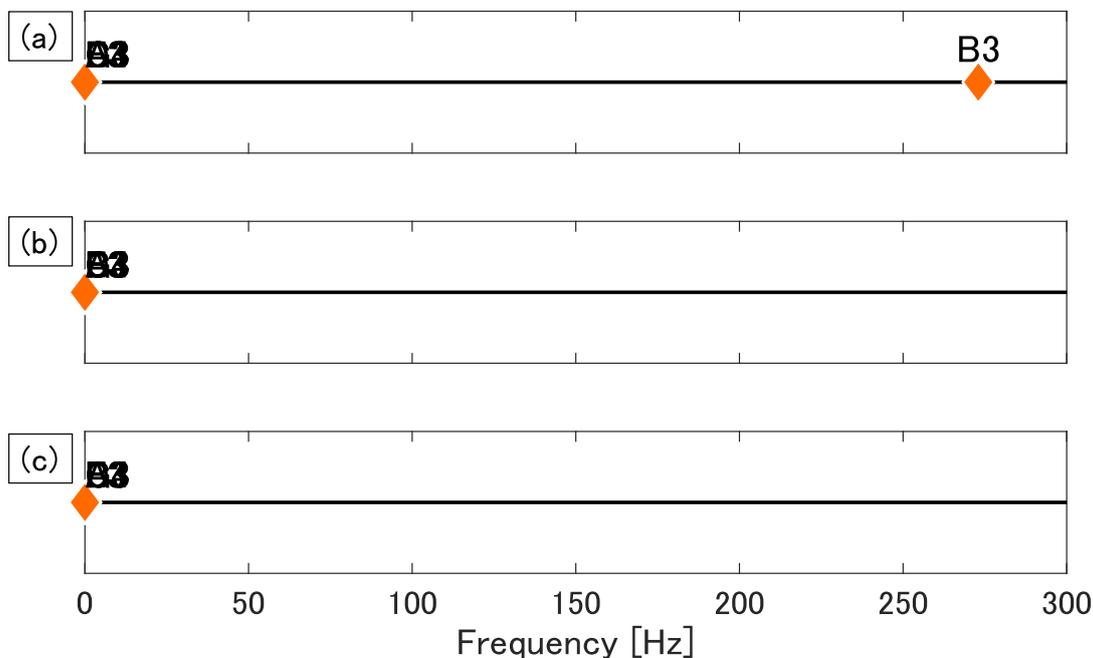


図 6.11: 図 3.1 に示すフィルタバンクにおいて 1.5 kHz 以下の帯域のエキサイテーションパターンから求めたサーストンの一対比較法による音階の配置 ( $F_c=64$  Hz)

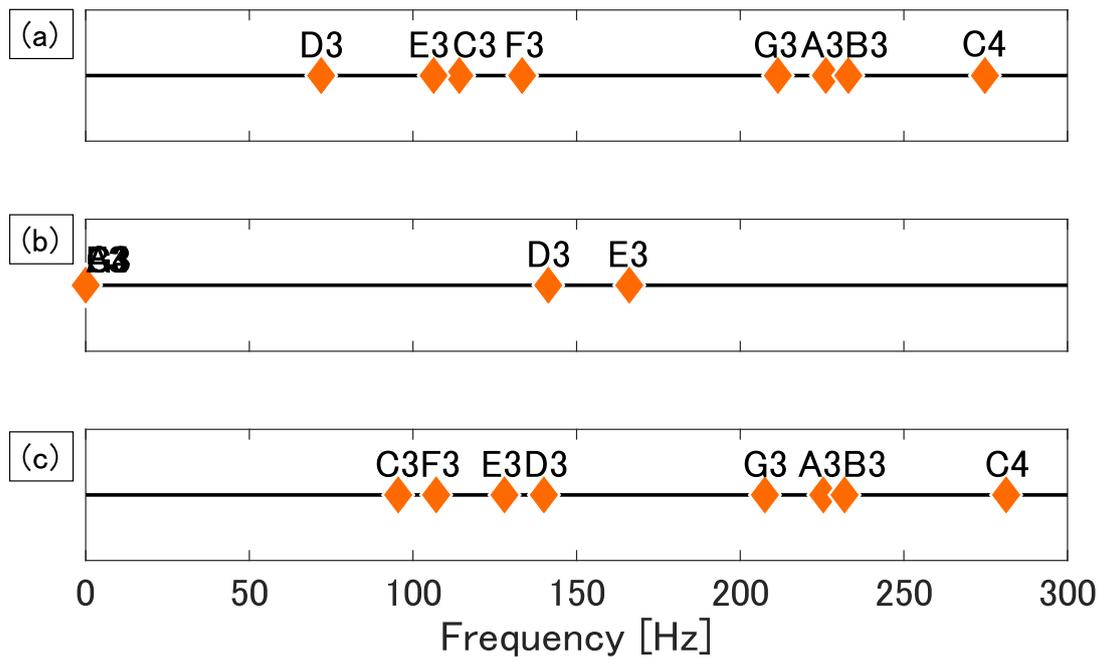


図 6.12: 原音で  $F_0$  を含む帯域のエキサイテーションパターンから求めたサーストンの一対比較法による音階の配置 ( $F_c=64$  Hz)

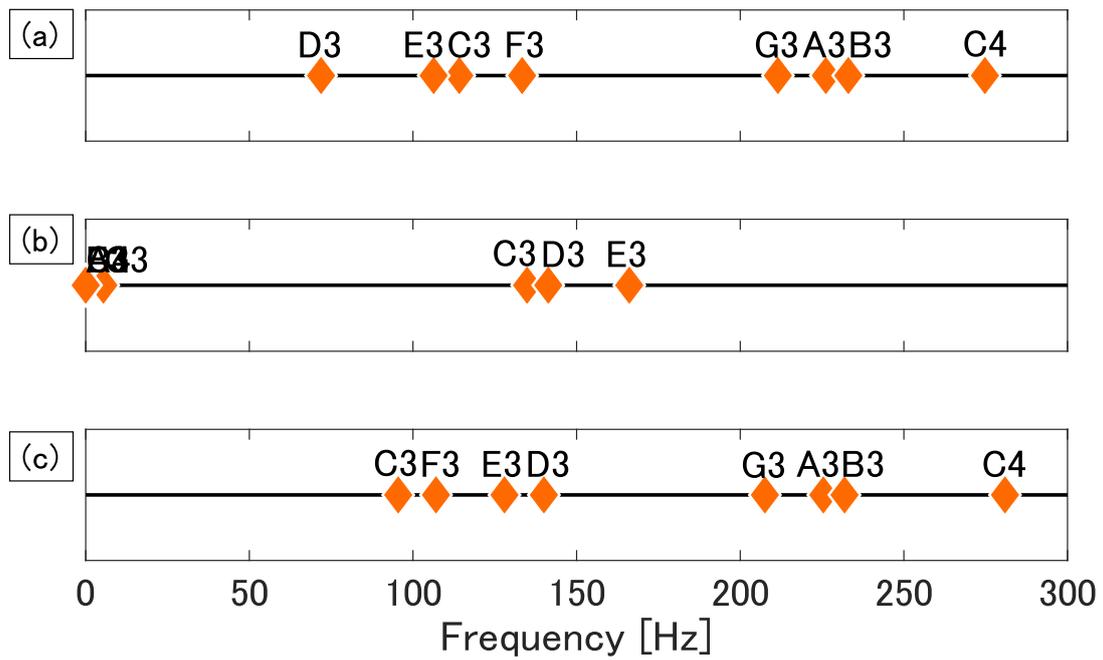


図 6.13: 原音で  $F_0$  を含む帯域のエキサイテーションパターンから求めたサーストンの一対比較法による音階の配置 ( $F_c=128$  Hz)

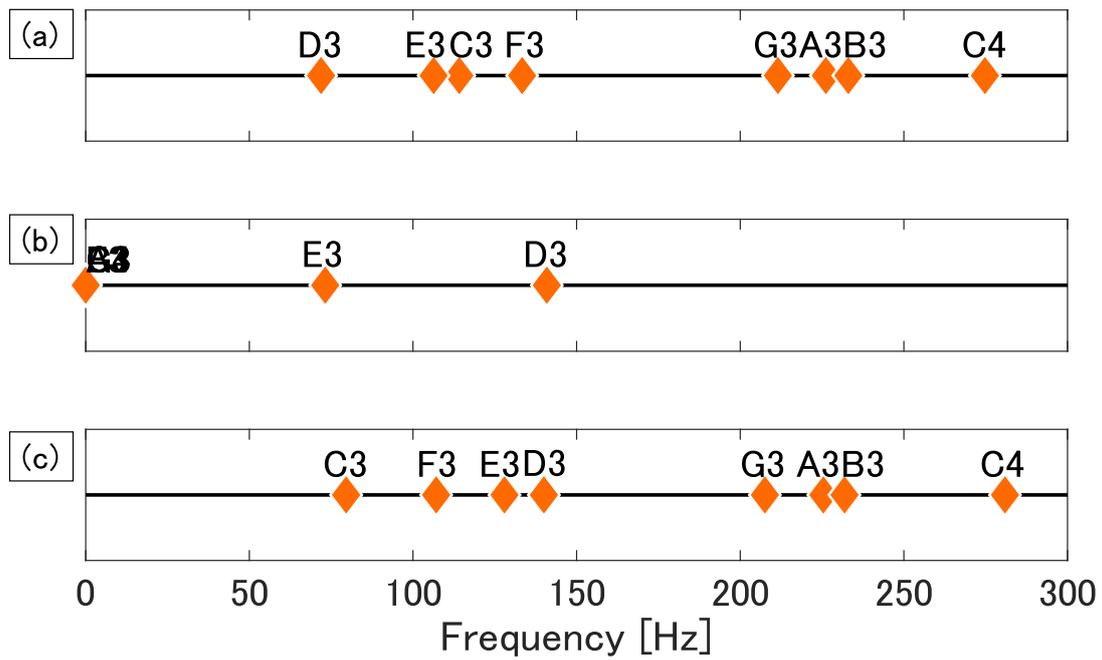


図 6.14: 原音で F0 を含む帯域のエキサイテーションパターンから求めたサーストンの一対比較法による音階の配置 ( $F_c=256$  Hz)

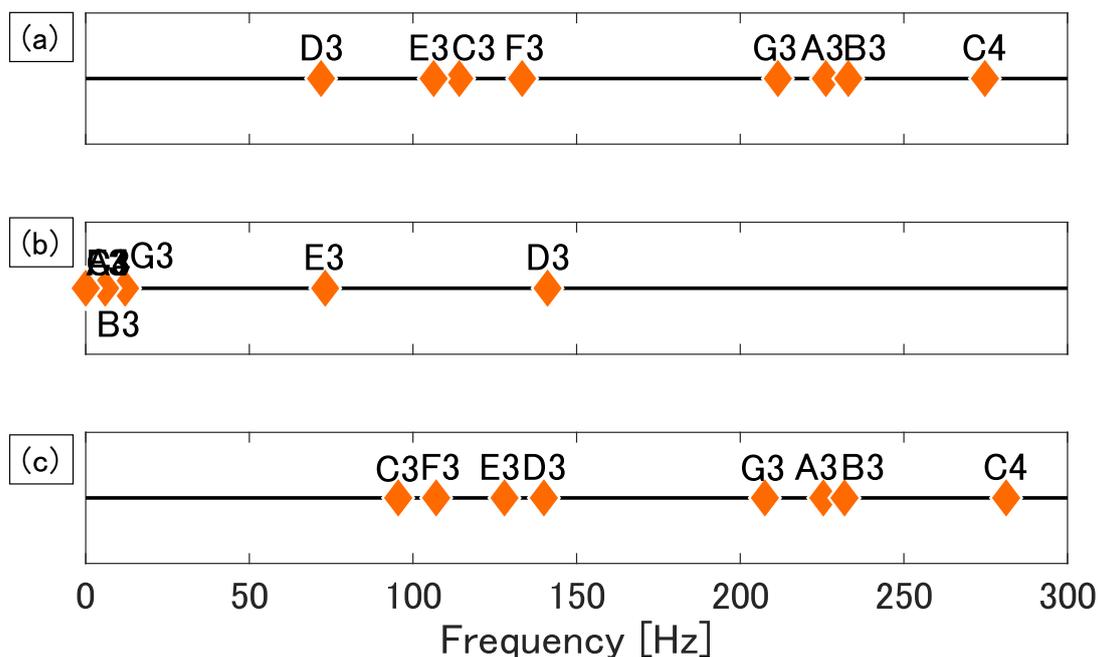


図 6.15: 原音で F0 を含む帯域のエキサイテーションパターンから求めたサーストンの一対比較法による音階の配置 ( $F_c=1024$  Hz)

## 6.6 変調スペクトルの分析

### 6.6.1 変調スペクトル

第5章と同様にして、実験刺激の変調スペクトルの算出を行った。算出した変調スペクトルの例として、各カットオフ周波数における C4 の変調スペクトルを図 6.16 – 6.19 に示す。いずれのカットオフ周波数においても、変調スペクトル上に F0 に対応する調波性が見られた。  $F_c=64$  Hz の変調スペクトルでは、原音で C4 の F0 を含む帯域である第3帯域では調波性が見られず、ピーク値も低くなっていた。一方、第9帯域以上の帯域では、基本周波数に対応するピークを除くピークの方が低くなっていたが、調波性が見られた。

$F_c=128, 256, 1024$  Hz の変調スペクトルでは、原音で C4 の F0 を含む帯域である第3帯域においても調波性が見られ、各ピーク値も  $F_c=64$  Hz と比較して高くなっていた。  $F_c=256, 1024$  Hz では、基本周波数に対応するピークよりも値の高いピークが見られる帯域もあった。

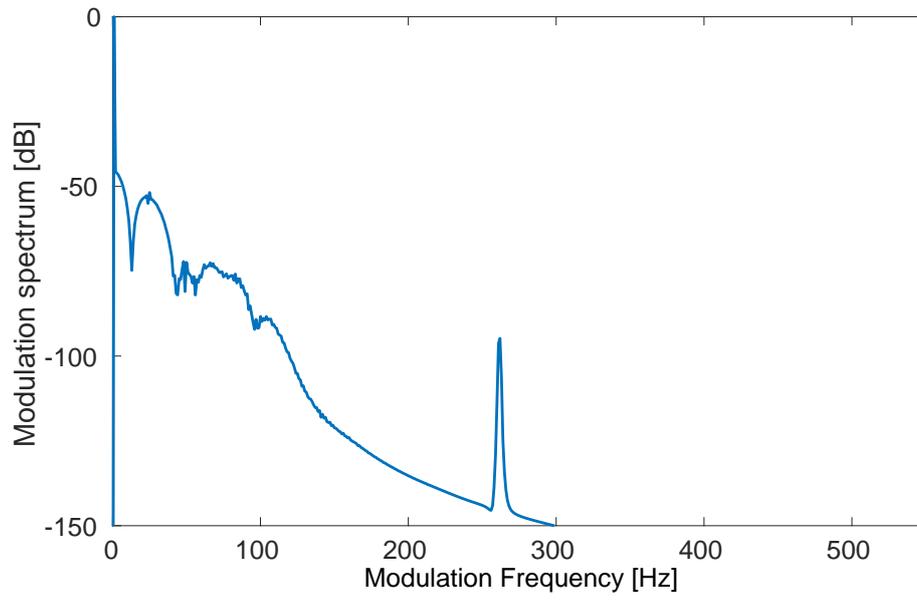


図 6.16: 変調スペクトル ( $F_c=64$  Hz)

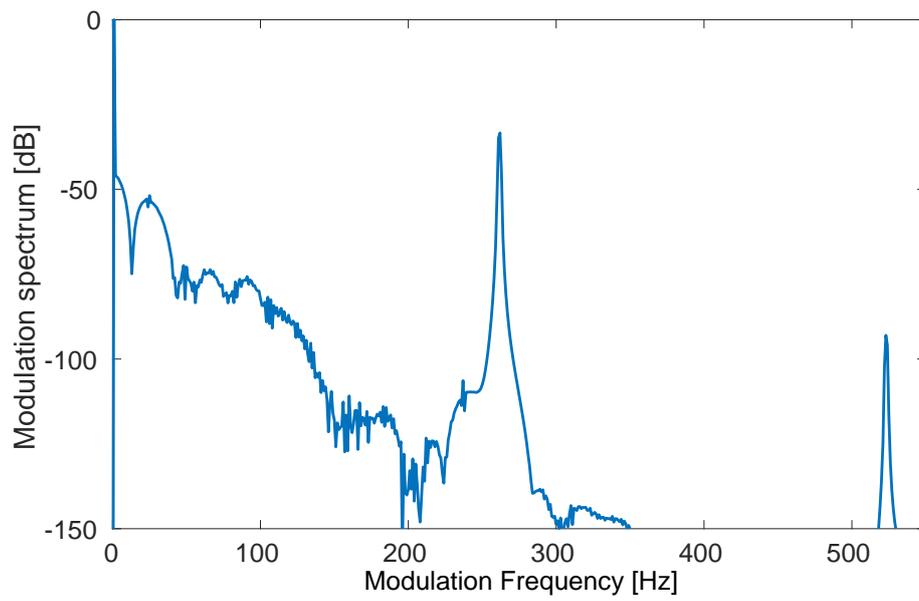


図 6.17: 変調スペクトル ( $F_c=128$  Hz)

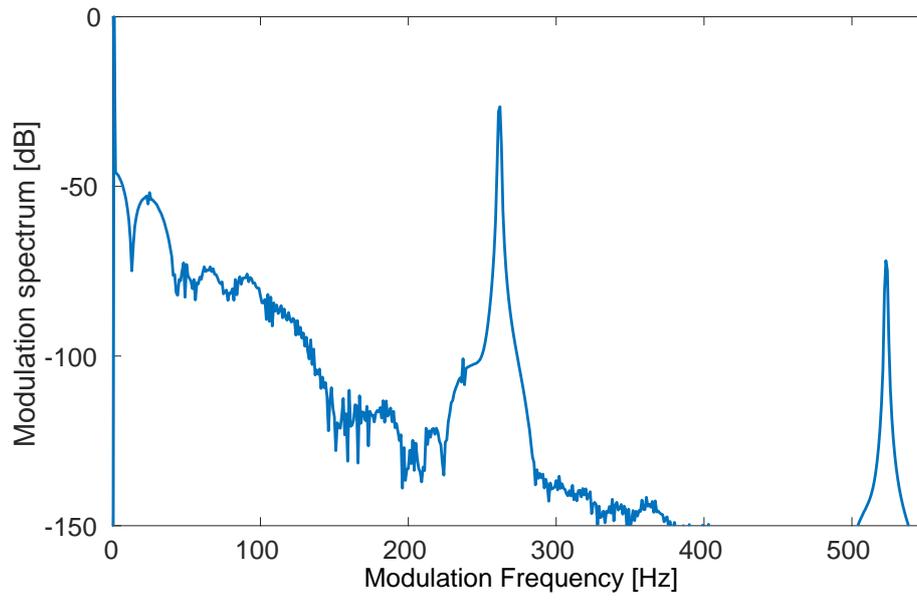


図 6.18: 変調スペクトル ( $F_c=256$  Hz)

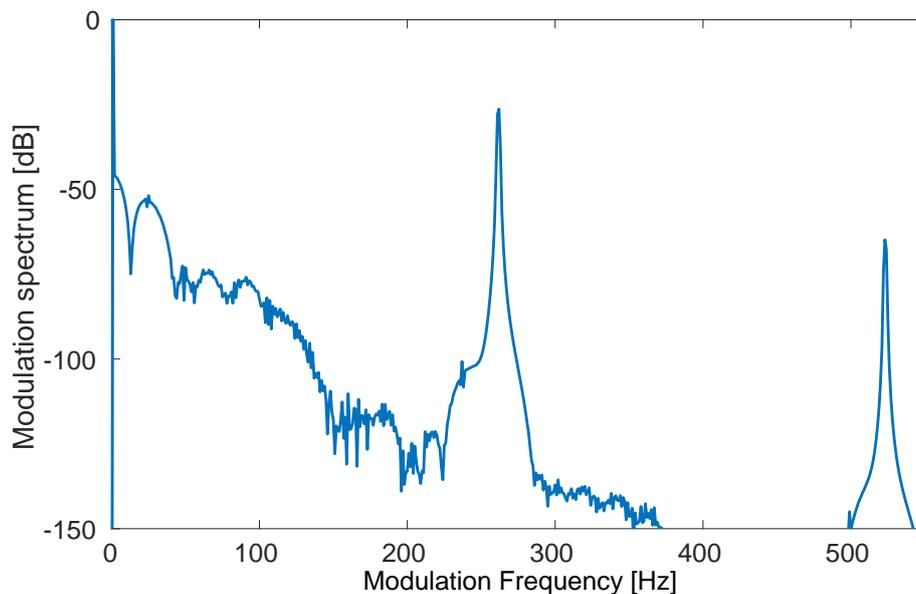


図 6.19: 変調スペクトル ( $F_c=1024$  Hz)

### 6.6.2 変調スペクトルから算出した $F_0$ による音階の配置

図 3.1 に示すフィルタバンクで中心周波数 1.5 kHz 以下に該当する帯域の変調スペクトルから算出された  $F_0$  による音階の配置

第 5 章と同様にして、変調スペクトルから  $F_0$  算出を行った。図 3.1 に示すフィルタバンクで 1.5 kHz 以下に該当する帯域の変調スペクトルから算出された  $F_0$  を音階として配置したものを図 6.20-図 6.23 に示す。いずれのカットオフ周波数およびスペクトル条件においても、音階の配置が音楽音階と大きく異なった。また、 $A_3$  が最も低く配置される傾向にあった。

$F_c=64$  Hz の音階の配置では、刺激全体が 0 Hz - 50 Hz 付近に配置され、間隔が大幅に狭くなっていた。 $F_0$  を算出することができなかった刺激数は、右下りの条件で 4、平坦の条件で 5、右上りの条件で 1 であった。図 6.4 に示すサーストンの一対比較法による音階の配置とは、音階の並び、間隔共に大きく異なった。

$F_c=128$  Hz の音階の配置では、どのスペクトル条件においても  $E_3$  が高く配置された。右下りの条件では  $C_4$  が最も低く配置された。 $C_4 - B_3$  の間隔が広がっており、その他の刺激の間隔は狭くなっていた。平坦の条件では、全体の間隔が右下りの条件と比較して広がっていた。 $A_3 - G_3$ ,  $C_4 - B_3$ ,  $C_3 - D_3$  の間隔が広がっていた。右上りの条件では、 $E_3$  および  $F_3$  が実際の  $F_0$  とほぼ同値として配置された。 $C_4 - C_3$  が隣り合って配置され、間隔は広がっていた。 $D_3 - E_3$  の間隔も広がっていた。図 6.5 に示すサーストンの一対比較法による音階の配置

とは、右上りの条件の E3, F3 が最も高く配置された部分は共通したが、その他の音階の並び、間隔は大きく異なった。

$F_c=256$  Hz の音階の配置においても、 $F_c=128$  Hz と同様に、E3 が高く配置された。いずれの条件においても、B3 と D3 が隣り合い、間隔が狭く配置された。平坦の条件における G3 – F3, 右上りの条件における A3 – C3 において間隔が特に広がっていた。平坦の条件における E3, 右上りの条件における E3 および F3 が実際の F0 とほぼ同値として配置された。図 6.6 に示すサーストンの一対比較法による音階の配置とは並び方が大幅に異なった。

$F_c=1024$  Hz の音階の配置では、右下りおよび右上りの条件において刺激全体の間隔が狭く配置され、F3 – E3 の間隔が広がっていた。平坦の条件では、A3 – C3 の間隔が広がっていたが、A3 – C3 間では等間隔に配置された。図 6.7 に示すサーストンの一対比較法による音階の配置とは並び方が大幅に異なった。

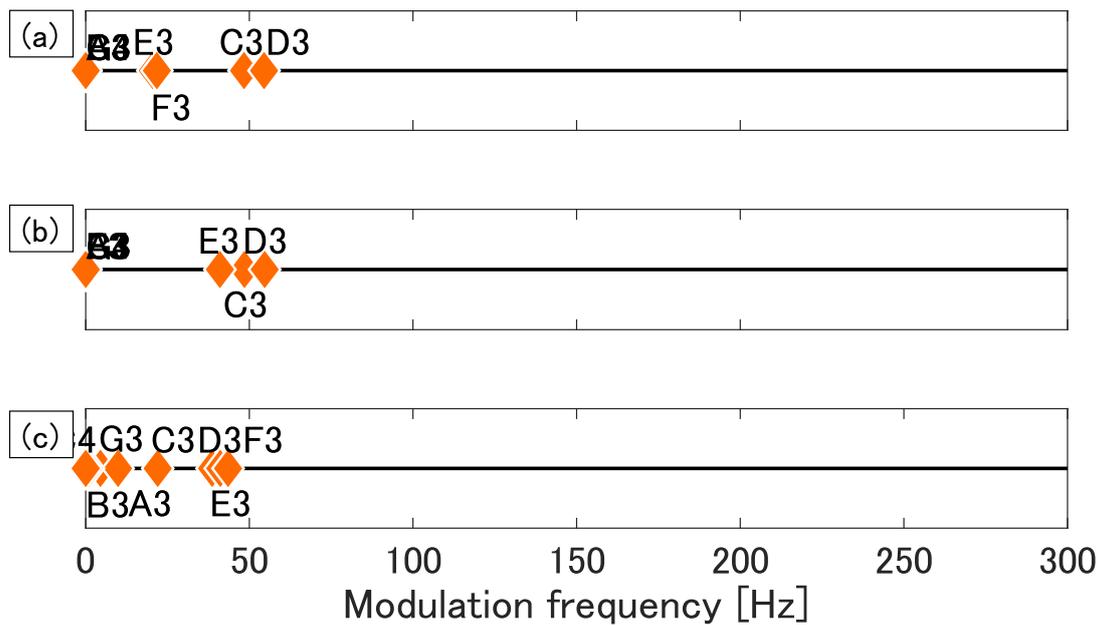


図 6.20: 図 3.1 に示すフィルタバンクにおいて 1.5 kHz 以下の帯域の変調スペクトルから求めたサーストンの一対比較法による音階の配置 ( $F_c=64$  Hz)

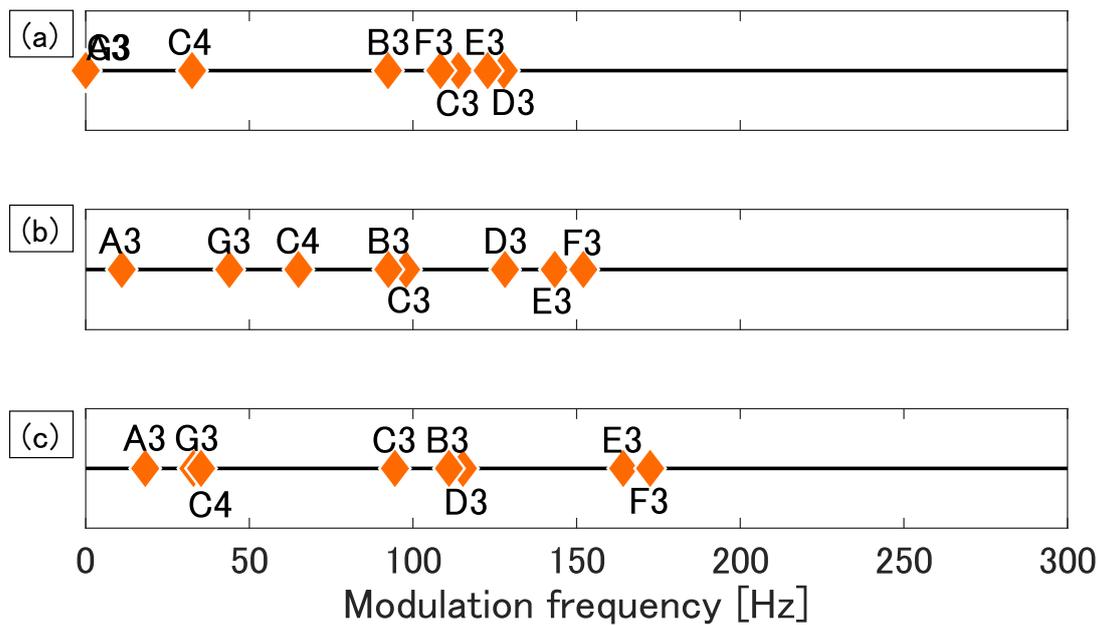


図 6.21: 図 3.1 に示すフィルタバンクにおいて 1.5 kHz 以下の帯域の変調スペクトルから求めたサーストンの一対比較法による音階の配置 ( $F_c=128$  Hz)

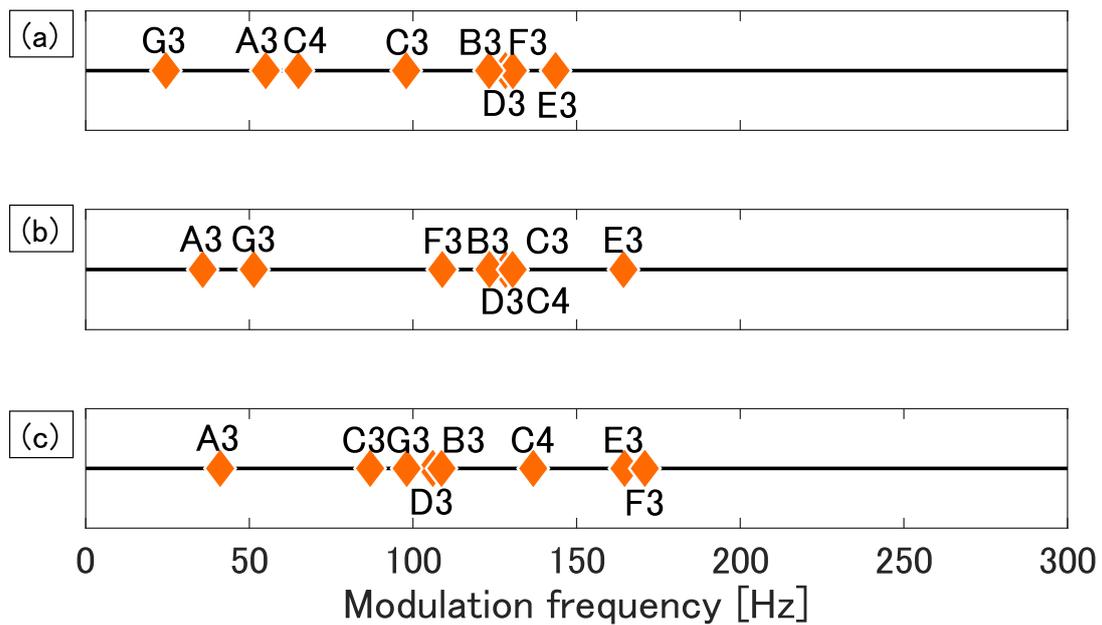


図 6.22: 図 3.1 に示すフィルタバンクにおいて 1.5 kHz 以下の帯域の変調スペクトルから求めたサーストンの一対比較法による音階の配置 ( $F_c=256$  Hz)

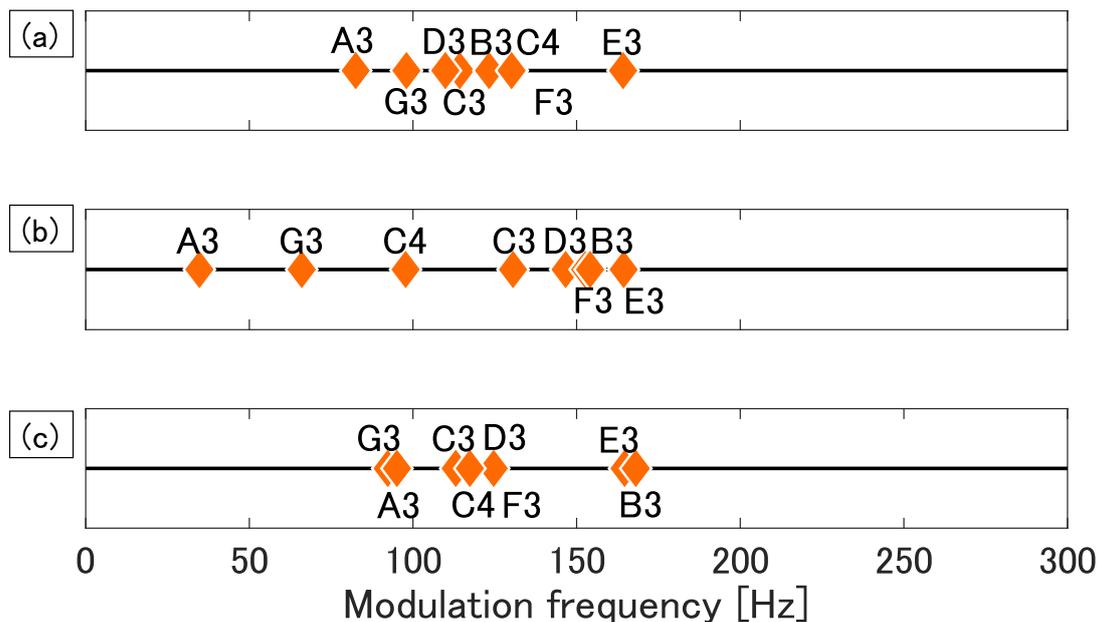


図 6.23: 図 3.1 に示すフィルタバンクにおいて 1.5 kHz 以下の帯域の変調スペクトルから求めたサーストンの一対比較法による音階の配置 ( $F_c=1024$  Hz)

#### 原音で $F_0$ が含まれる帯域の変調スペクトルから算出された $F_0$ による音階の配置

図 3.1 に示すフィルタバンクで 1.5 kHz 以下に該当する帯域と同様にして，原音で  $F_0$  が含まれる帯域の変調スペクトルを算出した．算出された  $F_0$  を音階として配置したものを図 6.24 – 図 6.27 に示す．

$F_c=64$  Hz では，いずれの条件においても自己相関法によって  $F_0$  の算出ができなかった． $F_c=128, 256, 1024$  Hz では，音階の配置が似ていた．

$F_c=128$  Hz の音階の配置では， $F_0$  の算出ができなかった刺激が多く見られた．右下りおよび平坦の条件では， $E_3$  および  $F_3$  の  $F_0$  が実際の  $F_0$  とほぼ同値として配置された．右上りの条件では， $D_3, E_3, F_3$  の  $F_0$  が実際の  $F_0$  とほぼ同値として配置された． $A_3, G_3, C_3$  は実際の  $F_0$  よりも大幅に低く算出された．

$F_c=256$  Hz においては，右下りの条件の音階の配置が  $F_c=128$  Hz の音階の配置と同様であった．平坦の条件では， $C_3 - F_3$  の  $F_0$  が実際の  $F_0$  とほぼ同値として配置された．右上りの条件では， $E_3$  および  $F_3$  の  $F_0$  が実際の  $F_0$  とほぼ同値として配置された． $C_3$  および  $D_3$  については，正しい  $F_0$  の  $1/2$  の値で算出された． $G_3$  および  $A_3$  は同値で算出されたが，実際の  $F_0$  よりも大幅に低い値であった．

$F_c=1024$  Hz においては，右下りの条件の音階の配置が  $F_c=128, 256$  Hz の音階の配置と同様であった．平坦の条件では， $F_c=256$  Hz と同様に  $C_3 - F_3$  の  $F_0$  が実際の  $F_0$  とほぼ同値として配置された． $G_3$  の  $F_0$  も算出されたが，本来の  $F_0$  よりも大幅に低い値であった．右上りの条件では， $E_3$  および  $F_3$  の  $F_0$  が実際の  $F_0$

とほぼ同値として配置された。A3, D3, C3, G3 は本来の F0 よりも大幅に低い値で F0 が算出された。C3 と G3 は同値で算出された。

いずれのカットオフ周波数においても、図 6.4 – 図 6.7 に示すサーストンの一対比較法による音階の配置とは並び方が大幅に異なった。

F0=64 Hz の原音で F0 が含まれる帯域の変調スペクトルでは、自己相関法によって F0 の算出ができなかった。そのため、第 16 帯域の変調スペクトルから算出された F0 を音階として配置したものを図 6.28 に示す。その結果、いずれの条件においても音楽音階に似た音階の配置となった。算出された F0 は、実際の F0 とほぼ同値であった。右下りおよび右上りのスペクトル条件では、A3 は F0 算出ができなかった。図 6.4 に示すサーストンの一対比較法による音階の配置とは大きく異なった。

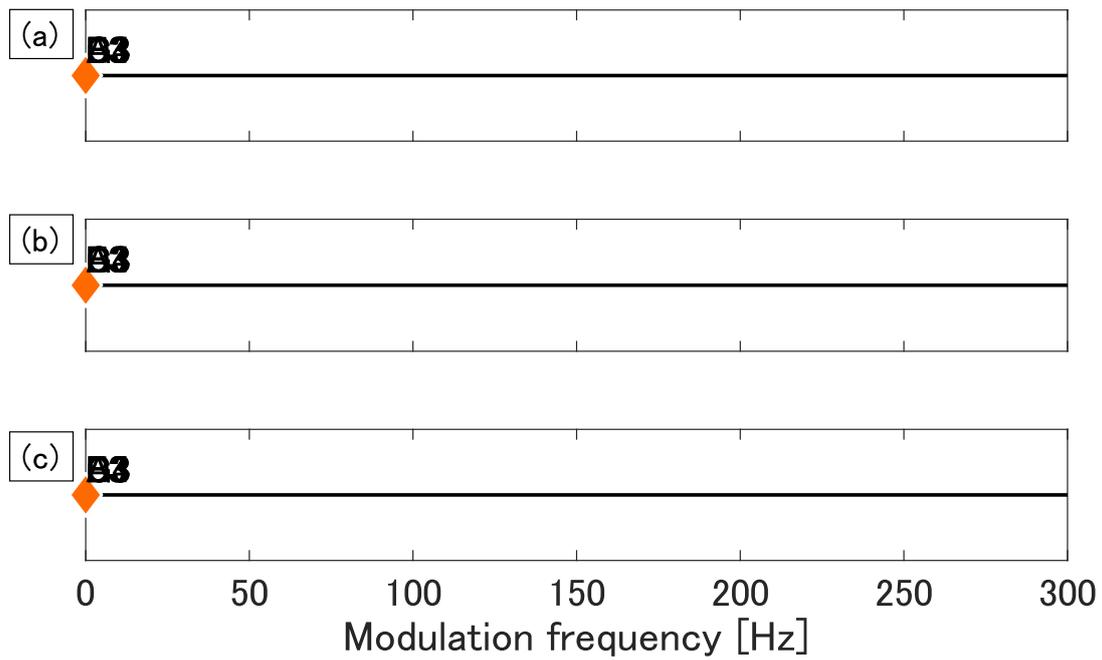


図 6.24: 原音で  $F_0$  を含む帯域の変調スペクトルから求めたサーストンの一対比較法による音階の配置 ( $F_c=64$  Hz)

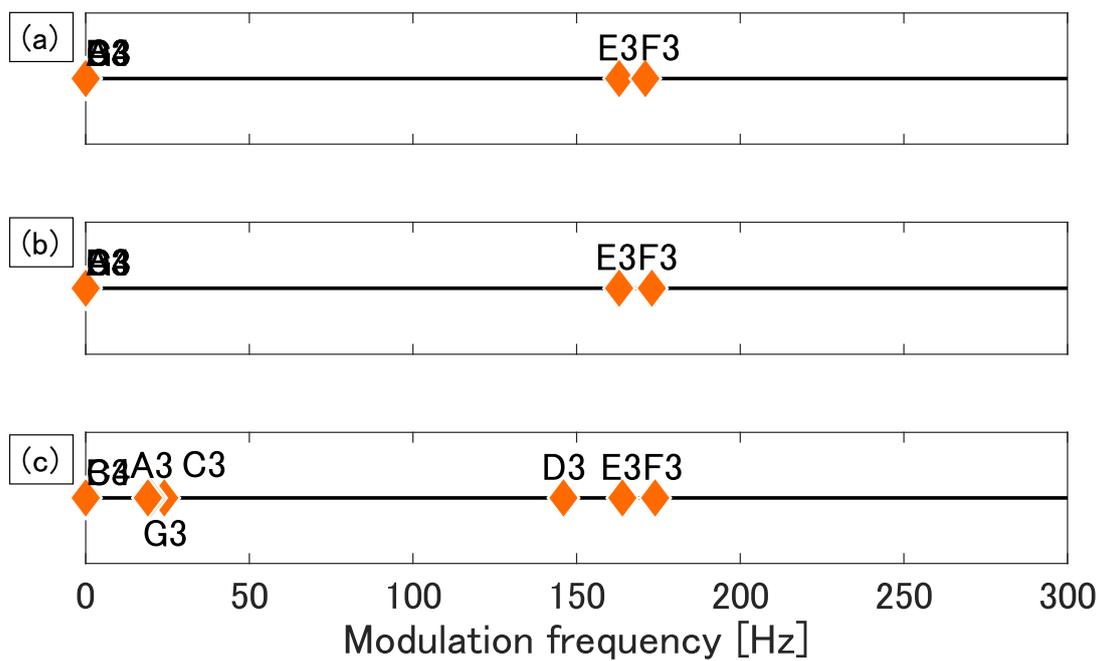


図 6.25: 原音で  $F_0$  を含む帯域の変調スペクトルから求めたサーストンの一対比較法による音階の配置 ( $F_c=128$  Hz)

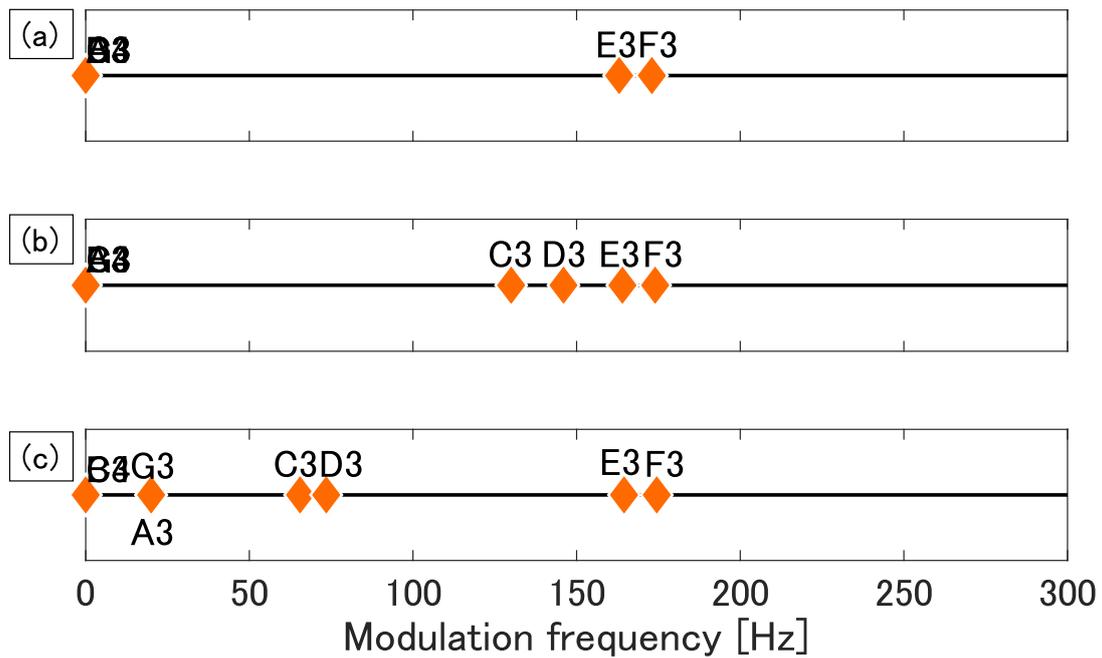


図 6.26: 原音で  $F_0$  を含む帯域の変調スペクトルから求めたサーストンの一対比較法による音階の配置 ( $F_c=256$  Hz)

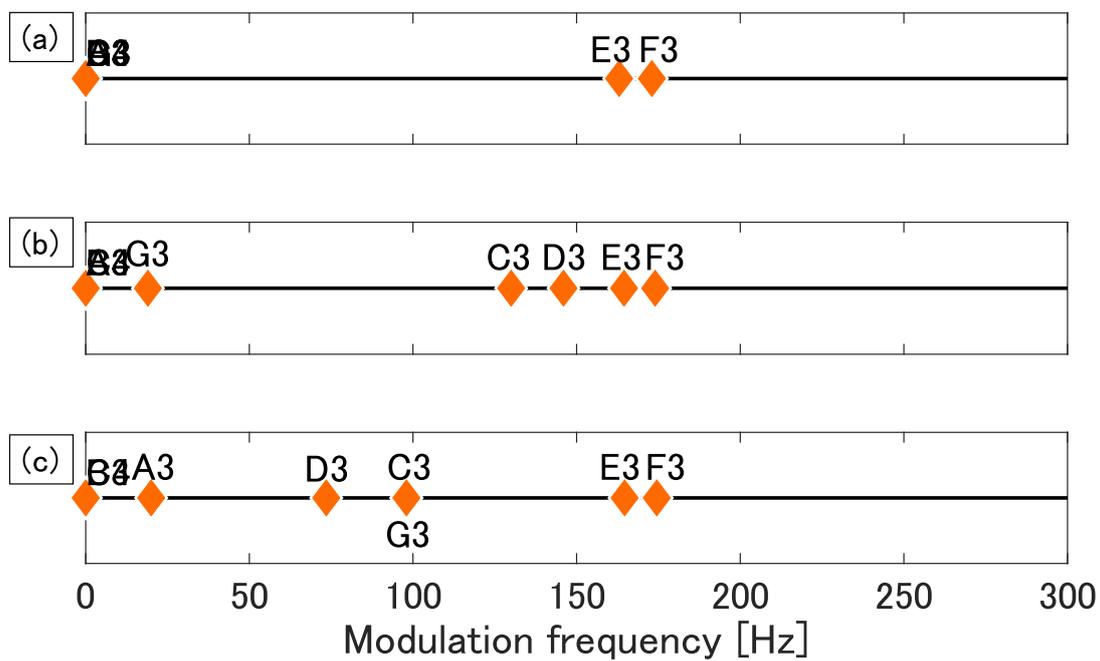


図 6.27: 原音で  $F_0$  を含む帯域の変調スペクトルから求めたサーストンの一対比較法による音階の配置 ( $F_c=1024$  Hz)

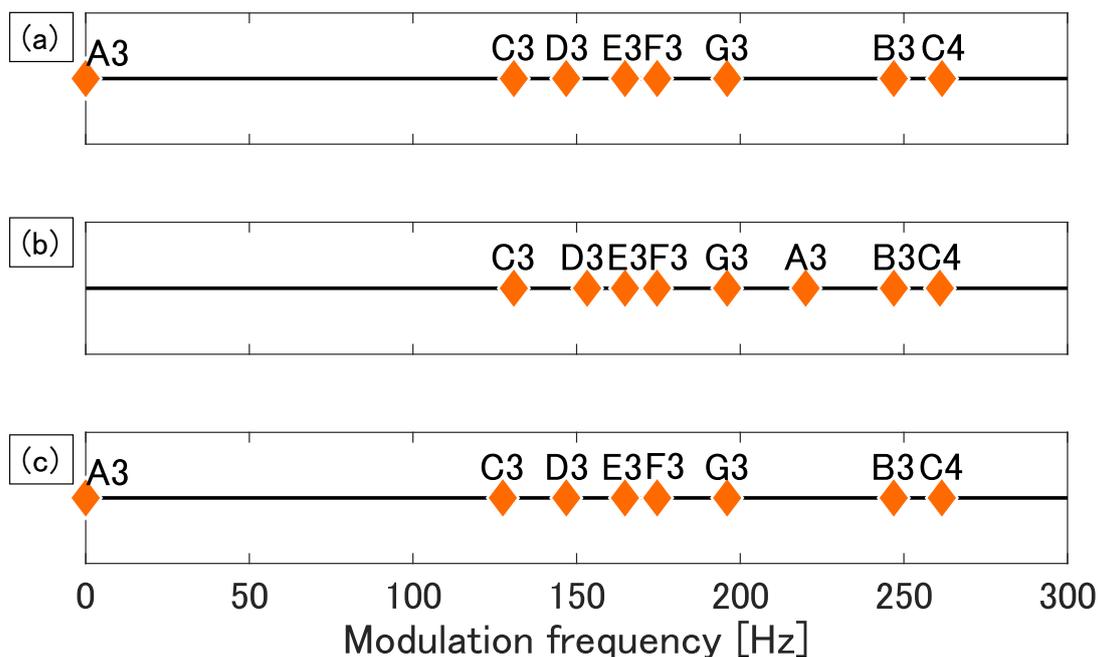


図 6.28: 第 16 帯域の変調スペクトルから求めたサーストンの一対比較法による音階の配置 ( $F_c=64$  Hz)

## 6.7 考察

### 6.7.1 実験結果の考察

いずれのカットオフ周波数の刺激においてもピッチ弁別の平均正答率が 50%程度であったこと，図 6.2 に示すようにピッチ弁別が可能と判断された刺激対が少ないため，本章で用いた雑音駆動合成音のピッチ知覚は困難であると考えられる．また，ピッチ弁別が可能と判断された刺激対と，誤った（本来とは逆の）ピッチ弁別が可能と判断された刺激対を合わせても，全刺激対に対する割合が 8.3 – 21.4%と低かった．そのため，多くの刺激対のピッチ弁別が困難であったと考えられる．

サーストンの一対比較法による音階の配置は，いずれの条件においても音楽音階とは異なっていた． $F_c=64$  Hz では，C4 と C3 が隣り合って知覚されていたことから，オクターブエラーを起こしていると考えられる． $F_c=128$  Hz の平坦の条件では，刺激同士の間隔が狭かったことから，これらの刺激が似て聴こえていると考えられる． $F_c=64, 256, 1024$  Hz の平坦の条件では，刺激同士の間隔が広い箇所が多く，各刺激の音の高さが異なるものとして聴こえていると考えられる． $F_c=1024$  Hz の音階の配置では，音階の並びは音楽音階と異なるが，間隔では他のカットオフ周波数と比較していずれの条件においても音楽音階に最も似ていた．

いずれのカットオフ周波数においても、スペクトル条件によって音階の配置が異なることから、雑音駆動合成音のピッチ弁別にはスペクトル傾斜が影響すると考えられる。

### 6.7.2 エキサイテーションパターンと実験結果

第5章と同様に、エキサイテーションパターンのピークの顕著さがスペクトル傾斜の条件によって異なった。右下りの条件でエキサイテーションパターンのピークが最も顕著であるが、サーストンの一対比較法による音階の配置の誤りがスペクトル条件によらない。これらのことから、雑音駆動合成音のピッチ弁別がエキサイテーションパターンにみられるピークの顕著さに影響を受けているとは考えにくい。

図3.1に示すフィルタバンクでフィルタの中心周波数が1.5 kHz以下に該当する帯域のエキサイテーションパターンによる音階の配置（図6.11）では、多くの刺激のF0が算出できなかった。そのため、図3.1に示すフィルタバンクでフィルタの中心周波数が1.5 kHz以下に該当する帯域のエキサイテーションパターン（場所説）をピッチ弁別の手がかりであるとする、雑音駆動合成音のピッチ弁別は困難であることが予想される。

原音のF0が含まれる帯域における雑音駆動合成音のエキサイテーションの時間変化から算出されたF0の配置（図6.13－図6.15）では、いずれの条件においても、右下りおよび右上りの条件では全ての刺激でF0が算出された。G3－C4においては、F0が実際のF0とほぼ同値で算出されたため、これらを手がかりにピッチ弁別をしているとすれば、雑音駆動合成音のピッチ弁別は可能であると予想される。しかし、音階の配置は音楽音階ともサーストンの一対比較法による音階の配置（図6.5－図6.7）とも大幅に異なった。そのため、原音のF0が含まれる帯域におけるエキサイテーションの時間変化を利用して雑音駆動合成音のピッチ弁別している可能性は低いと考えられる。

### 6.7.3 変調スペクトルと実験結果

カットオフ周波数が高くなることで振幅包絡線に含まれる情報が増加したため、 $F_c=64$  Hzの変調スペクトルと比較して $F_c=128, 256, 1024$  Hzの変調スペクトルではより調波性が見られ、ピーク値も高くなったと考えられる。

図3.1に示すフィルタバンクでフィルタの中心周波数が1.5 kHz以下に該当する帯域の変調スペクトルでは、 $F_c=64$  Hzのいずれの条件においてもF0を算出できなかった刺激が存在し、算出できたF0値も実際のF0値とは大幅に異なっていた。また、サーストンの一対比較法による音階の配置（図6.4）では、音階の並びは音楽音階と大きく異なるが、間隔は広がっていた。そのため、いずれの条件においても、図3.1に示すフィルタバンクでフィルタの中心周波数が1.5 kHz以下に該当

する帯域の変調スペクトルによる音階の配置とサーストンの音階の配置とは一致していない。  $F_c=128, 256, 1024$  Hz では、実際の  $F_0$  とほぼ同値で  $F_0$  を算出できた刺激も一部見られたが、  $F_c=128$  Hz の右上りのスペクトル条件において E3 および F3 が最も高く配置された部分が共通したのみで、図 3.1 に示すフィルタバンクでフィルタの中心周波数が 1.5 kHz 以下に該当する帯域の変調スペクトルによる音階の配置とサーストンの音階の配置（図 6.5 – 図 6.7）とは一致していない。これらのことから、変調スペクトルの周期性（時間説）を利用して雑音駆動合成音のピッチ弁別をした可能性は低いと考えられる。

$F_c=64$  Hz の原音で  $F_0$  が含まれる帯域の変調スペクトルでは、 $F_0$  を検出できなかった。これは、図 6.20 に示すように、変調スペクトルで C4 の  $F_0$  が含まれる第 3 帯域の調波性が見られなかったことによるものと考えられる。しかし、サーストンの一対比較法による音階の配置（図 6.4）では、知覚された音階を配置できている。また、変調スペクトルによる音階の配置とサーストンの一対比較法による音階の配置が一致しない。  $F_c=128, 256, 1024$  Hz では、変調スペクトルによる音階の配置が似ている。実際の  $F_0$  とほぼ同値で  $F_0$  を算出できた刺激も一部見られたが、図 6.5 – 図 6.7 に示すサーストンの一対比較法による音階の配置はカットオフ周波数によって異なった。また、カットオフ周波数の増加によって変調スペクトルによる音階の配置に顕著な違いは見られなかった。これらのことから、原音で  $F_0$  が含まれる帯域の変調スペクトルにおける調波性（時間説）を利用して雑音駆動合成音のピッチ弁別をしている可能性は低いと考えられる。

全ての刺激において Unresolved な倍音成分が含まれる第 16 帯域の変調スペクトルでは、自己相関法により算出された  $F_0$  による音階の配置が音楽音階に似ていた（図 6.28）。これは、聴覚フィルタの該当する帯域に倍音成分が多数含まれており、変調周波数軸上に調波性が見られるためであると考えられる。変調スペクトルから算出された  $F_0$  は音楽音階と似て配置されているが、変調スペクトルによる音階の配置とサーストンの一対比較法による音階の配置が大幅に異なることから、変調スペクトルの特定の帯域における調波性（時間説）を利用して雑音駆動合成音のピッチ弁別をしている可能性は低いと考えられる。

以上のことから、カットオフ周波数を高くすることで変調スペクトルに見られるピークの数を増加させたが、カットオフ周波数による雑音駆動合成音のピッチ弁別成績の向上は見られなかった。そのため、カットオフ周波数の増加は雑音駆動合成音のピッチ弁別に影響しないと考えられる。しかし、本章では C3 – C4 のみの調査を行ったため、C4 – C5 の刺激への影響も検討する必要がある。また、エキサイテーションパターンおよび変調スペクトルとサーストンの一対比較法による音階の配置に関連が見られなかった。

## 第7章 全体考察

### 7.1 雑音駆動合成音のピッチ弁別の可能性

第4章において雑音駆動合成音のピッチ弁別の可能性を検討した。その結果、原音と雑音駆動合成音のピッチ弁別が同程度に可能であることが明らかにされた。また、第5章において、スペクトル傾斜を持つ雑音駆動合成音のピッチ弁別の検討を行った。その結果、雑音駆動合成音ではスペクトル条件が右下りから平坦、右上りと変わることによってピッチ弁別可能な刺激対が減少することが明らかにされた。これらのことから、既報 [8-14] とは異なり、振幅包絡線情報のみを持つ雑音駆動合成音であっても、右下りのスペクトル条件であればピッチを知覚できる可能性があると考えられる。

第6章では、第4章および第5章で用いた刺激とは異なり、振幅包絡線抽出時の低域通過フィルタを急峻なものに変更して作成した刺激を用いたピッチ弁別の検討を行った。ピッチ弁別が可能と判断された刺激対が  $F_c=64$  Hz で 5.56%,  $F_c=128$  Hz で 6.35%,  $F_c=256$  Hz で 3.57%,  $F_c=1024$  Hz で 7.14% と少なかった。そのため、低域通過フィルタを急峻なものに変更した場合には雑音駆動合成音のピッチ弁別は困難であると考えられる。第4章の検討では、原音のピッチ弁別調査と雑音駆動合成音のピッチ弁別調査が時間を空けずに行われたため、原音のピッチ弁別が雑音駆動合成音のピッチ弁別の注意になっている可能性が考えられる。そこで、第4章の雑音駆動合成音のピッチ弁別調査を再度行った。結果を図7.1に示す。実験参加者は第4章での聴取実験の参加者のうち10名であった。その結果、再度調査した音階の配置と図4.4に示す第4章での音階の配置がほぼ同様であった。そのため、第4章での実験結果に、原音による注意が影響しているとは考えにくい。したがって、低域通過フィルタを急峻にした場合、雑音駆動合成音のピッチ弁別は困難であると考えられる。

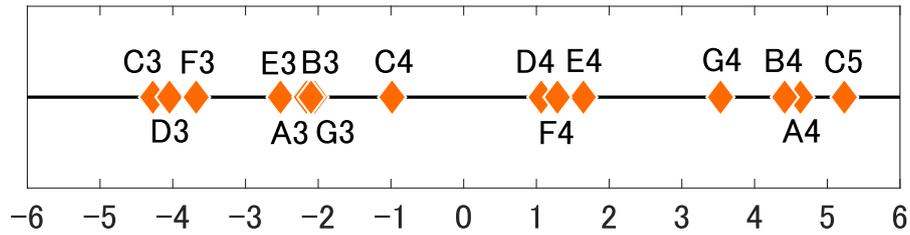


図 7.1: サーストンの一対比較法による音階の配置 (第 4 章 NVS 再実験)

## 7.2 雑音駆動合成音のピッチ弁別の手がかり

第 5 章においてスペクトル傾斜の条件によってピッチ弁別の結果が異なった。また、第 6 章の検討では、いずれのカットオフ周波数においてもスペクトル条件によって音階の配置が異なったことから、雑音駆動合成音のピッチ弁別にはスペクトル傾斜が影響すると考えられる。

エキサイテーションパターンに見られるピークが、右下りの条件において最も顕著であること、サーストンの一対比較法による音階の並びの誤りが右下りで最も少ないことから、雑音駆動合成音のピッチ弁別はエキサイテーションパターンにみられるピークの顕著さに影響を受けると考えられる。

図 3.1 に示すフィルタバンクにおいて、フィルタの中心周波数が 1.5 kHz 以下に該当する帯域のエキサイテーションパターンから求めた  $F_0$  による音階の配置 (図 5.12) は、サーストンの一対比較法による音階の配置 (図 5.8) と一致しなかった。そのため、エキサイテーションパターンを手がかりに場所説の観点からピッチ弁別をした可能性は低いと考えられる。また、原音の  $F_0$  が含まれる帯域における雑音駆動合成音のエキサイテーションの時間変化から算出された音階の配置 (図 5.13) は、サーストンの一対比較法による音階の配置とは一致しなかった。そのため、原音の  $F_0$  が含まれる帯域におけるエキサイテーションの時間変化を利用して雑音駆動合成音のピッチ弁別をした可能性は低いと考えられる。

図 3.1 に示すフィルタバンクでフィルタの中心周波数が 1.5 kHz 以下に該当する帯域の変調スペクトルによる音階の配置では、 $F_c=128$  Hz の右上りのスペクトル条件において E3 および F3 が最も高く配置された部分が共通したのみで、それ以外の刺激は図 6.5 – 図 6.7 に示すサーストンの音階の配置と一致していない。これらのことから、変調スペクトルの周期性を手がかりとして時間説の観点から雑音駆動合成音のピッチ弁別をした可能性は低いと考えられる。さらに、原音で  $F_0$  が含まれる帯域の変調スペクトルでは、音階の配置とサーストンの一対比較法による音階の配置が一致しなかった。これらのことから、原音で  $F_0$  が含まれる帯域の変調スペクトルにおける調波性を手がかりとして時間説の観点から雑音駆動合成音のピッチ弁別をした可能性は低いと考えられる。

図5.18に示すように、全ての刺激において Unresolved な倍音成分が含まれる第16帯域の変調スペクトルの音階の配置が音楽音階に似ていた。これは、聴覚フィルタの第16帯域に倍音成分が多数含まれており、変調周波数軸上で調波性が見られたためであると考えられる。このことから、Resolved harmonics から F0 を推定できれば、Shamma ら [19] と同様に Unresolved な倍音成分を用いた雑音駆動合成音のピッチ弁別の可能性が考えられる。

カットオフ周波数を高くすることで変調スペクトルに見られるピークを増加させたが、カットオフ周波数による雑音駆動合成音のピッチ弁別成績の向上は見られなかった。また、カットオフ周波数が増加してもエキサイテーションパターンおよび変調スペクトルによる音階の配置に顕著な違いは見られなかった。そのため、カットオフ周波数の増加は雑音駆動合成音のピッチ弁別に影響しないと考えられる。しかし、第4章および第5章と異なり、C3 - C4 のみ調査を行ったため、C4 - C5 が存在した場合のピッチ弁別への影響も検討する必要がある。また、雑音駆動合成音からどのようなピッチが知覚されているかを調査することで、雑音駆動合成音のピッチ弁別の手がかりをさらに検討することができると考えられる。

## 第8章 結論

### 8.1 本研究で明らかにしたこと

本研究では、ZhuらやUnokiらの検討と同様に、雑音音駆動合成音を用いたアプローチにより、音の時間的な振幅包絡線情報を用いてピッチ知覚が可能かどうかを検討した。

はじめに、楽音を模した調波複合音と雑音駆動合成音のピッチ弁別実験を行った。その結果、既報とは異なり、雑音駆動合成音の調波複合音とピッチ弁別が同程度に可能であることを明らかにした。

次に、三つのスペクトル傾斜を持つ調波複合音とその雑音駆動合成音のピッチ弁別実験を行った。その結果、ピッチの正答率およびサーストンの一対比較法による音階の配置がスペクトル条件によって異なった。そのため、雑音駆動合成音のピッチ弁別がスペクトル傾斜に影響を受けることを明らかにした。

さらに、振幅包絡線抽出時のカットオフ周波数を高くして作成した雑音駆動合成音のピッチ弁別実験を行った。その結果、カットオフ周波数によって雑音駆動合成音のピッチ弁別の成績が変わらなかった。そのため、カットオフ周波数を増加させたことによる振幅包絡線の情報の増加は雑音駆動合成音のピッチ弁別に影響しないことが示唆された。

最後に、実験刺激のエキサイテーションパターンおよび変調スペクトルとピッチ弁別結果との関連性を考察した。その結果、エキサイテーションパターンおよび変調スペクトルによる音階の配置とピッチ弁別結果との間に関連性がないこと、雑音駆動合成音のピッチ弁別はエキサイテーションパターンにみられるピークの顕著さに影響を受けることが示唆された。

以上のことから、雑音駆動合成音のピッチ弁別は、スペクトル傾斜および振幅包絡線抽出時の低域通過フィルタ、エキサイテーションパターンのピークの顕著さに影響されるが、カットオフ周波数の増加には影響を受けないことが示唆された。

これらのことから、音の時間的な振幅包絡線情報を用いたピッチ知覚が可能である刺激と困難である刺激があることを明らかにした。

## 8.2 残された課題

### カットオフ周波数を増加させた2オクターブの刺激のピッチ弁別調査

第6章では、第4章および第5章と異なり、C3 – C4のみ調査を行った。そのため、実験刺激にC4 – C5が存在する状態でカットオフ周波数を増加させた場合のピッチ弁別への影響を検討する必要がある。また、低域通過フィルタを変更せずにカットオフ周波数を増加させた場合のピッチ弁別への影響も検討することで、雑音駆動合成音のピッチ弁別の手がかりおよびメカニズムの解明に寄与できると考えられる。

### 雑音駆動合成音のピッチの調査

本研究では雑音駆動合成音のピッチ弁別調査を行ったが、そもそも雑音駆動合成音がどのようなピッチを持つのかは明らかにできなかった。そのため、調整法 [34] 等を用いて雑音駆動合成音からどのようなピッチが知覚されているかを調査することで、雑音駆動合成音のピッチ弁別の手がかりをおよびメカニズムの解明に寄与できると考えられる。

### 人工内耳装用者による雑音駆動合成音のピッチ弁別調査

雑音駆動合成音は人工内耳装用者の聴こえを模擬した音の一つである。雑音駆動合成音のピッチ弁別の手がかりが明らかになった場合、その手がかりを強調した音を用いた人工内耳装用者によるピッチ弁別調査をすることで、人工内耳装用者のピッチ知覚の獲得の方略を議論することができると考えられる。

# 謝辞

研究を進めるにあたり、数多くのご指導およびご助言をいただきました鷓木祐史教授に深謝いたします。また、熱心なご助言をいただきました赤木正人教授に感謝いたします。

研究および実験を進めるにあたり、数多くのご助言とご討論をいただきました木谷俊介助教、小林まおり博士に感謝いたします。

京都市立芸術大学音楽学部教授 津崎実先生には、お忙しい中ご助言いただきました。深く御礼申し上げます。

お忙しい中、実験にご参加いただいた皆様に感謝いたします。

本研究は、科研費・新学術領域 (No.18H05004)、三谷研究開発支援財団・研究助成ならびに JST 未来社会創造事業 (JP-MJMI18D1) の支援を受けて行われました。

研究および大学院生活において、鷓木研究室、赤木研究室の先輩として数多くのご助言をいただきました鳥谷さん、磯山さん、高橋さん、川村さん、並河さん、畠山さん、大谷さん、松本さんに感謝いたします。また、多くの時間を共に過ごし、共に走り、共に戦ってくださった同期の水野さん、西垣さん、宮川さん、森田さん、黄さん、及川さんに感謝いたします。同期のように共に過ごしてくださった M1 の浅江くん、倉くん、木所くん、坂本さん、藤田くん、藤田くんに感謝いたします。

最後に、大学院生活を支えてくださった家族に感謝いたします。

## 参考文献

- [1] 大串健吾, “音のピッチ知覚,” コロナ社, 2014.
- [2] C. J. Plack, “The sense of hearing,” second edition, Routledge, 2016.
- [3] H. Fujisaki, “Prosody, models, and spontaneous speech, in Computing prosody,” Y. Sagisaka, N. Campbell, and N. Higuchi (Eds), Springer, pp.27–42, 1996.
- [4] C. Lorenzi, G. Gilbert, H. Carn, S. Garnier, B. C. J. Moore, “Speech perception problems of the hearing impaired reflect inability to use temporal fine structure,” Proc. Natl. Acad. Sci. USA., vol.103, no.49, pp.18866–18869, 2006.
- [5] 森尚彫, 日本における人工内耳の現状, “保健医療学雑誌, 6(1), pp.15–26, 2015.
- [6] “人工内耳について,” 一般社団法人日本耳鼻咽喉科学会  
<http://www.jibika.or.jp/citizens/hochouki/naiji.html> (2020/1/31 閲覧)
- [7] T.Green, A. Faulkner, S. Rosen, “Spectral and temporal cues to pitch in noise-excited vocoder simulations of continuous-interleaved-sampling cochlear implants,” J. Acoust. Soc. Am., vol.112, no.5, pp.2155–2164, 2002.
- [8] A. J. Oxenham, “Pitch perception and auditory stream segregation: Implications for hearing loss and cochlear implants,” Trends Amplif., vol.12, no.4, pp.316–331, 2008.
- [9] W. R. Drennan, J. T. Rubinstein. “Music perception in cochlear implant users and its relationship with psychophysical capabilities,” J. Rehabil. Res. & Dev., vol.45, no.5, pp.779–790, 2008.
- [10] 城間将江, 菊地義信, 河野淳, 鈴木衛, 加我君孝, “人工内耳装用者による音楽の知覚 (第一報),” Auditory Japan, vol.41, pp.755–764, 1998.
- [11] 大金さや香, 城間将江, 小淵千絵, “人工内耳装用者と補聴器装用者の音楽知覚の比較検討,” Auditory Japan, vol.58, pp.60–68, 2015.

- [12] 北澤茂良, 桐山伸也, 横山貴紀, 岩崎聡, 姜洪仁, “人工内耳装用者のための音楽の試作,” 信学技報, SP2009-50, pp.7–12, 2009.
- [13] V. Looi, H. McDermott, C. McKay, L. Hickson, “Music perception of cochlear implant Users Compared with that of Hearing Aid Users,” *Ear Hear.*, vol.29, no.3, pp.421–434, 2008.
- [14] J. Laneau, J. Wouters, M. Moonen, “Improved music perception with explicit pitch coding in cochlear implants,” *Audiol. Neurotol.*, vol.11, no.1, pp.38–52, 2006.
- [15] R. V. Shannon, F. G. Zeng, V. Kamath, J. Wygonski and M. Ekelid, “Speech recognition with primarily temporal cues,” *Science*, 13, vol.270, no.5234, pp.303–4, 1995.
- [16] Z. Zhu, R. Miyauchi, Y. Araki, and M. Unoki, “Contributions of temporal cue on the perception of speaker individuality and vocal emotion for noise-vocoded speech,” *Acoust. Sci. & Tech.*, vol.39, no.3, pp.234–242, 2018.
- [17] M. Unoki, M. Kawamura, M. Kobayashi, S. Kidani, M. Akagi, “How the temporal amplitude envelope of speech contributes to urgency perception,” *Proc. ICA2019*, pp.1739–1744, 2019.
- [18] 森大毅, 前川喜久雄, 粕谷英樹, “音声は何を伝えているか—感情・パラ言語情報・個人性の音声科学,” コロナ社, 2014.
- [19] S. Shamma and K. Dutta, “Spectro-temporal templates unify the pitch percepts of resolved and unresolved harmonics,” *J. Acoust. Soc. Am.*, vol.145, no.2, pp.615–629, 2019.
- [20] W. P. Shofner, J. Campbell, “Pitch strength of noise-vocoded harmonic tone complexes in normal-hearing listeners,” *J. Acoust. Soc. Am.*, vol.132, no.5, pp.398–404, 2012.
- [21] A. Preis, “An attempt to describe the parameter determining the timbre of steady-state harmonic complex tones,” *Acustica*, vol.55, no.1, pp.1–13, 1984.
- [22] 大串健吾, “複合音の音色を支配する物理的・心理的要因について,” *音響会誌*, vol.36, no.5, pp.253–259, 1980.
- [23] X. Luo, S. Soslowsky, K. R. Pulling, “Interaction between pitch and timbre perception in normal-hearing listeners and cochlear implant users,” *J. Assoc. Res. Otolaryngol*, vol.20, no.1, pp.57–72, 2019.

- [24] B. C. J. Moore, “An Introduction to the psychology of hearing, sixth edition,” Brill Academic Pub., 2013.
- [25] B. C. J. Moore, “Hearing,” , Second edition, Academic Press, 1995.
- [26] 寶坂友希菜, 木谷俊介, 鷗木祐史, “楽音を模した調波複合音の雑音駆動合成音のピッチ知覚の検討,” 聴覚研資, vol.49, no.7, pp.463–468, 2019.
- [27] 佐藤信, “統計的官能検査法,” 日科技連出版社, 1985.
- [28] 西川泰夫, 大澤光, 沼野元義, “計量心理学,” 放送大学教育振興会, 2006.
- [29] 磯山拓都, 水野滉介, 木谷俊介, 鷗木祐史, “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,” 2020年春季研究音講論(春), 1-Q-1, 2020.
- [30] B. C. J. Moore, “The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people,” J. Assoc. Res. Otolaryngol, vol.9, no.4, pp.399–406, 2008.
- [31] A. R. Palmer and I. J. Russell, “Phase-locking in the cochlear nerve of the guinea-pig and its relation to the receptor potential of inner hair-cells,” Hear. Res., vol.24, pp.1–15, 1986.
- [32] W. Hess, “Pitch and voicing determination,” in Advanced in speech signal processing, ed. S. Furui and M.M. Sondhi, pp.3–48, Marcel Dekker. Inc. New York 1992.
- [33] A. J. M.Houtsma and J. Smurzynski, “Pitch identification and discrimination for complex tones with many harmonics,” J. Acous. Soc. of Am., vol.87, no.1, pp.304–310, 1990.
- [34] T. Kuroda and E. Hasuo, “The very first step to start psychophysical experiments,” Acoust. Sci. & Tech., vol.35. no.1, pp.1–9, 2014.

# 研究業績

## 国内学会における発表

1. 寶坂友希菜, 鷗木祐史, “雑音駆動合成音のピッチ知覚の検討,” 2019年度電気・情報関係学会北陸支部連合大会, G-2, 2019.
2. 寶坂友希菜, 木谷俊介, 鷗木祐史, “楽音を模した調波複合音の雑音駆動合成音のピッチ知覚の検討,” 聴覚研資, 49(7), 463–468, 2019.
3. 寶坂友希菜, 木谷俊介, 鷗木祐史, “雑音駆動合成における調波複合音のピッチ知覚の検討,” 聴覚研資, 2020.
4. 寶坂友希菜, 木谷俊介, 鷗木祐史, “雑音駆動合成を用いた調波複合音のピッチ知覚の検討,” 2020年春季研究音講論（春）, 1-Q-5, 2020.

## 国際学会における発表

1. M. Unoki, Y. Hosaka, and S. Kidani, “Pitch perception of noise-vocoded harmonic complex tones mimicking musical instruments,” Proc. Forum Acusticum 2020, 2020 (accepted).