

ハイレゾリューションオーディオの研究

西 口 敏 行

電気通信大学 情報システム学研究科
博士（工学）の学位申請論文

2009 年 3 月

ハイレゾリューションオーディオの研究

博士論文審査委員会

主査 出澤 正徳 教授

委員 井上 誠喜 客員教授

委員 阪口 豊 准教授

委員 長岡 浩司 教授

委員 森田 啓義 教授

著作權所有者

西口敏行

2009

A Study on Human Hearing of High Resolution Audio

Toshiyuki Nishiguchi

Abstract

It is generally accepted that the frequency bandwidth of human hearing by air conduction does not far exceed 20 kHz. However, some papers discussed the influence of extending the frequency by such high resolution audio formats on the auditory impression and brain activity. Studies on the bandwidth of human hearing are important for defining the frequency bandwidth for sound recordings and for developing electro-acoustic transducers, next-generation broadcasting systems and the definition of the recording format of audio archives.

If differences of sound impression among high resolution audio formats indeed exist, it is necessary to consider two factors: influence of reproduction of very high frequency, and difference of sound quality in auditory frequency band among the sampling formats.

Regarding the former factor, we conducted subjective evaluation tests concerning the influence of very high frequency components on the human perception in musical sounds. To make a precise evaluation, the test system was designed to exclude any influence from very high frequency components in the audible frequency range. Tests showed that the subjects could discriminate between musical sounds with and without very high frequency components.

While, to study the difference of sound quality in the auditory frequency band among several high sampling digital recording formats, we also conducted subjective evaluation tests of perceptual discrimination among the following digital recording formats: 24 bit/48 kHz, 24 bit/192 kHz and DSD. The results showed no significant difference, namely, sound quality of the auditory frequency band in this experiment system does not depend on the sampling format.

According to the results of these tests, extension of the frequency range by high resolution audio affects the perception of sound and it is caused by the reproduction of very high frequency components. It could be worth recording music by high resolution audio systems with a very wide frequency range. However various issues have been pointed out regarding the very high frequency band, i.e., the non-linear distortion of the amplifier or loudspeaker, time jitter and so on. Very accurate handling is required to reproduce the very high frequency components.

ハイレゾリューションオーディオの研究

西口 敏行

概 要

人間の可聴帯域上限は 20kHz とされていたが、近年、20kHz 以上の超高域成分が楽音の知覚や生理・心理にも影響を与えるとの報告がなされ、ハイサンプリングオーディオとよばれる 100kHz 近くまで記録可能なメディアも販売されている。収録・再生帯域をどこまで高めるべきかは、マイクロホンやスピーカなどオーディオ機器の開発、放送や音源の収録・アーカイビングの方式を定めるうえで重要な要因となる。

ハイレゾリューションオーディオが聴感に影響を与えるとすると、その要因としては、可聴域を超える周波数成分（超高域成分）の影響と可聴域の音質変化が考えられる。

本研究では、まず、楽音に含まれる超高域成分の有無を弁別できるか否かを主観評価実験により確認した。超高域成分が聴こえに影響を与えるとすると、その要因は、2つに分けて考える必要がある。ひとつは、超高域成分を聴取することにより聴感へ影響が出る場合、もうひとつは、超高域成分を再生したことにより、可聴域の音に変化し、これが弁別の手がかりになっている場合である。後者の要因として、実験システムの非線形歪が指摘されている。本研究では、評価音再生時の非線形歪を測定し、歪のレベルが充分小さいことを確認したシステムを用いて実験を行った。また、過去の報告では、スーパーオーディオ CD (Super Audio CD: SACD) や DVD-Audio など市販の音楽ソフトを評価音源として用いたケースがあるが、この場合、収録条件が判らず、超高域成分の量も管理できない。そのため、本研究における評価音源としては、主に独自収録した広帯域音源から作成したものをを用いた。

本研究では、実験 1 と 2 の 2 つの超高域成分（21kHz 以上とした）の主観弁別実験を行った。実験 1 では、超高域成分による音質差異が存在する可能性があるかないか、また、その影響は、音源の種類や被験者によって異なるのかどうかを確認する目的で、様々な種類の音源を用意し、比較的多数の被験者により基礎的な評価実験を行った。実験方法は、R, A, B の 3 つの評価音を試聴し、A と B の中から R と同じ方を選択させる *duo-trio test* を用いた。被験者 36 人で 20 種の音源をそれぞれ 40 回評価した。この結果、音源毎には有意な弁別結果は得られなかったが、被験者 1 人が正答率 0.75 をマークした。そこで、この被験者に対して追加実験を行った。より信頼性を高めるため、音源を 6 種類に選別して、それぞれの音源を 20 回評価した。この結果、どの音源に対しても有意な弁別結果は得られなかった。

実験 1 の結果と、超高域の影響には、音源の提示時間が関係するという先行研究の結果を考慮して実験方法を変更、被験者、音源を精査して実験 2 を行った。実験 2 では、超高域成分が豊富な音源を作成するため、新たに音楽録音用の超広帯域マイクを開発した。このマイクロホンは、音波の回折効果を利用して超高域の感度を高めることにより、固有雑音の大きさは従来の音楽録音用マイクとほぼ同等に抑えながら、収録帯域を 100 kHz まで拡張したものである。音源は、弦楽四重奏、筑前琵琶の弾き語り、ハープシコードの 3 種を用いた。実験方法には、2 つの評価音を対にして提示し、差があるか否かを判断させる *pair test* 法を採用した。この方法は、音源の提示時間が制御でき、かつ、微妙な差違の検知に適している。13 人の被験者で超高域有無の弁別実験を行った結果、2 人の被験者が筑前琵琶の音源において、超高域の有無を有意に弁別できた。

一方、ハイレゾリューションオーディオの可聴域の音質差異については、広帯域録音・再生（ハイサンプリングフォーマット）対応装置による 3 種の標本化フォーマット 48kHz 24bit, 192kHz 24bit, 2.8MHz 1bit（Direct Stream Digital : DSD）にて収録・再生をおこない、各フォーマット間の音質差異に関して主観弁別実験を行って検討した。

本実験では、標本化フォーマット以外の条件に差が生じないように充分留意した。音源は、収録条件を管理して独自収録を行った。音質への影響が大きい A/D,

D/A 変換器については、機種や個体差による音質差を排除するため、機種・ファームウェアが同一のものをを用い、収録・再生を通してフォーマット毎のレベル偏差は、 $\pm 0.05\text{dB}$ 以内になるように調節した。評価した音源は、邦楽、ボサノバ、ピアノトリオの 3 種、被験者は、聴力健常な音楽大学学生 4 人、バイオリニスト 1 人、録音技術者 1 人である。実験方法は、実験 2 と同様に微妙な差違の検知に適した **pair test** 法を採用した。実験の結果、被験者、音源に依らず、標本化フォーマットの違いは有意に弁別されなかった。

したがって、ハイレゾリューションオーディオによる広帯域収録再生では、可聴域を超える周波数成分が聞こえに影響を与える場合があり、超高域成分を豊富に含む音源については、ハイレゾリューションオーディオによる広帯域収録再生は価値あることと考えられる。一方、広帯域収録再生においては、非線形歪や時間ゆらぎの影響など様々な課題も指摘されている。現実的なオーディオ信号の取り扱いを考えた場合、広帯域収録再生には十分な注意が必要とされる。

目次

第 1 章 序論	1
1.1 研究の背景と目的	1
1.2 研究の概要	1
1.3 本論文の構成	3
第 2 章 聴覚とデジタルオーディオ	5
2.1 聴覚器官	5
2.2 音と聴覚	8
2.2.1 音の強さと周波数	8
2.2.2 音の種類とスペクトル	11
2.2.3 音の心理物理	16
2.2.4 可聴範囲	16
2.3 デジタルオーディオ	19
2.3.1 アナログ信号のデジタル化	19
2.3.1.1 標本化と補間	19
2.3.1.2 標本化定理と保持効果	21
2.3.1.3 帯域制限フィルタ	22
2.3.1.4 オーバサンプリング	22
2.3.1.5 量子化雑音と標本化周波数	25
2.3.2 A/D, D/A 変換方式	27
2.3.3 デジタルオーディオの変遷とハイレゾリューションオーディオ	31
2.4 第 2 章のまとめ	33
第 3 章 超高域成分の主観弁別実験	35
3.1 はじめに	35
3.2 実験システム	45
3.2.4 実験システムの構成	45
3.2.5 実験システムの特徴	47

3.3 実験 1：多種音源・多数の被験者による評価	52
3.3.1 実験 1 の目的と特徴	52
3.3.2 被験者	52
3.3.3 音源	52
3.3.4 実験方法	54
3.3.5 実験結果	54
3.3.6 実験 1 の追試	61
3.3.6.1 被験者と音源	61
3.3.6.2 実験結果	61
3.3.7 実験 1 のまとめ	62
3.4 実験 2：音源提示時間と弁別の関係	63
3.4.1 実験 2 の概要	63
3.4.2 超広帯域マイクロホンによる音源収録	63
3.4.3 実験方法の検討	67
3.4.4 実験 2.1：提示時間が長い場合	69
3.4.4.1 音源	69
3.4.4.2 被験者と実験結果	72
3.4.5 実験 2.2：提示時間が短い場合	74
3.4.5.1 音源	74
3.4.5.2 被験者と実験結果	74
3.4.6 被験者の聴覚閾値測定	76
3.5 第 3 章のまとめ	77

第 4 章 標本化フォーマットと可聴帯域内の音質 …… 79

4.1 はじめに	79
4.2 実験システム	81
4.2.1 実験システムの構成	81
4.2.2 A/D, D/A 変換器の物理特性	83
4.2.3 再生条件	85
4.3 実験 3：3 種の標本化フォーマットの主観弁別実験	86
4.3.1 被験者	86
4.3.2 音源	86
4.3.3 実験方法	88
4.3.4 実験結果	88

4.4 第4章のまとめ.....	90
第5章 結論	91
謝辞.....	93
参考文献.....	94
付録 略語表.....	97

目次

図 1. 聴覚器官の概要 [1]。	6
図 2. 蝸牛の断面模式図（亀田, 1986） [3]。	7
図 3. 基底板の振動と周波数 [4]。	7
図 4. 粗密波 [1]。	10
図 5. 音圧レベルの目安 [1]。	10
図 6. 音の波形とスペクトル [5]。	11
図 7. ブラックマン窓とそのフーリエ変換の対数振幅 ($N=64$) [8]。	14
図 8. 可聴範囲 [1]。	18
図 9. 年齢別の MAP 平均値と 95%信頼区間 [9]。	18
図 10. デジタル信号処理の流れ [7]。	20
図 11. 保持効果 [7]。	22
図 12. オーバサンプリングを用いた A/D 変換の例 [7]。	24
図 13. 標本化周波数 f_s による量子化雑音分布の変化 [7]。	26
図 14. 各種の A/D, D/A 変換方式 [7]。	29
図 15. 変換方式と周波数スペクトル [7]。	30
図 16. 8~20 kHz の最小可聴値 (Fausti ら[18])。	35
図 17. Plenge らの用いたテスト信号のスペクトル [19]。	36
図 18. Muraoka らの実験システム [21]。	38
図 19. 評価音源のレベル変化 (Muraoka ら[21])。	38
図 20. 楽音を使った弁別実験の結果 (Muraoka ら[21])。	39
図 21. 純音を使った可聴周波数上限の実験結果 (Muraoka ら[21])。	39
図 22. Oohashi らが実験に用いた音源のパワースペクトル [23]。	41
図 23. 安静状態 (baseline) と各聴取条件 (LCS, HCS, FRS, 図 22 参照) における正規化した脳波 α 波 (Alpha-EEG : alpha-electroencephalogram) 電位。	43
図 24. 実験システムのダイアグラム。	46
図 25. 帯域分割デジタルフィルタの特性。	46
図 26. 被験者とスピーカの位置, 聴取状態。	48
図 27. 超高域成分 (21 kHz 以上) の再生に用いたスーパーツイータの指向特性。	49

図 28. 実験システムの総合周波数特性。	50
図 29. 実験システムの非線形歪。	51
図 30. 実験 1 の音源毎の実験結果。	56
図 31. 実験 1 の被験者毎の実験結果。	57
図 32. 実験 1 で用いた音源の分析結果。	60
図 33. 実験 1 の追試結果。	61
図 34. 超広帯域マイクロホンの帯域拡張方法。	65
図 35. 超広帯域マイクロホンの周波数特性。	65
図 36. 超広帯域マイクロホンの外観。	66
図 37. 超広帯域マイクロホンと従来の音楽録音用マイクロホンの収録帯域の比較。	66
図 38. 実験 2.1 と 2.2 で採用した pair test 法の詳細。	68
図 39. 実験 2.1 と 2.2 で用いた音源の分析結果。	71
図 40. 実験 2.1 の結果。	73
図 41. 実験 2.2 の結果。	75
図 42. 48 kHz と 96 kHz 標本化の音質比較のための試験テープ作成方法 (Yoshikawa ら [43])。	80
図 43. Blech らの実験システム[44]。	80
図 44. 実験 3 で用いた実験システム。	82
図 45. A/D 変換・D/A 変換トータルの周波数振幅特性の実測結果。	84
図 46. A/D 変換・D/A 変換トータルの周波數位相特性の実測結果。	84
図 47. 実験 3 の実験システムの総合周波数特性。	85
図 48. 実験 3 の評価音源のスペクトル。	87
図 49. 実験 3 の結果。	89

表目次

表 1. 窓関数の例。	15
表 2. デジタルオーディオの変遷。	32
表 3. FRS と HCS (図 22 参照) の音質に対する主観評価結果 (Oohashi ら[23])。	42
表 4. 実験 1 で用いた音源。	53
表 5. 新たに開発した超広帯域マイクロホンの仕様。	66
表 6. 実験 2.1 と 2.2 で用いた評価音源。	69
表 7. 実験 2.1 と 2.2 の評価結果の比較。	75
表 8. 自由音場での聴覚閾値測定結果。	76

第1章 序論

1.1 研究の背景と目的

従来、人間の可聴域上限の周波数は約 20 kHz とされてきた。このため、コンパクトディスク (Compact Disc : CD)、DAT (Digital Audio Tape) やデジタル放送の収録帯域は通常 20kHz 前後に定められている。この値は、それまでの心理実験により測定された結果に基づいて定められたものである。

ところが、近年、ハイレゾリューションオーディオと呼ばれるスーパーオーディオ CD (Super Audio CD : SACD) や DVD-Audio など、収録帯域が 100 kHz に迫る音楽ソフトが発売されている。こうした広帯域化が音の主観的印象や脳の活動に影響を与えるという報告例もある。しかしながら、広帯域化が聴感に与える影響についての科学的検証例は少なく、十分なコンセンサスが得られているとは言えない。

収録・再生帯域の検討は、音声の収録方法、マイクロホンやスピーカといった電気音響変換器、記録・伝送フォーマットの基礎検討としても重要な課題である。したがって、本研究は、楽音による主観評価実験により、収録・再生の広帯域化が聴感に与える影響を調べることを目的とする。本研究では、この目的に鑑み、気導提示された楽音を対象とする。

1.2 研究の概要

ハイレゾリューションオーディオが聴感に影響を与えるとすると、その要因としては、可聴域を超える周波数成分（超高域成分）の影響と可聴域の音質変化が考えられる。

本研究では、まず、楽音に含まれる超高域成分の有無を弁別できるか否かを主観評価実験により確認した。超高域成分が聴こえに影響を与えるとすると、その要因は、2 つに分けて考える必要がある。ひとつは、超高域成分を聴取する

ことにより聴感へ影響が出る場合、もうひとつは、超高域成分を再生したことにより、可聴域の音に変化し、これが弁別の手がかりになっている場合である。後者の要因として、実験システムの非線形歪が指摘されている。本研究では、評価音再生時の非線形歪を測定し、歪のレベルが充分小さいことを確認したシステムを用いて実験を行った。また、過去の報告では、SACD や DVD-Audio など市販の音楽ソフトを評価音源として用いた場合があるが、この場合、収録条件が判らず、超高域成分の量も管理できない。そのため、本研究における評価音源としては、主に独自収録した広帯域音源から作成したものを用いた。

本研究では、実験 1 と 2 の 2 つの超高域成分 (21kHz 以上とした) の主観弁別実験を行った。実験 1 では、超高域成分による音質差異が存在する可能性があるかないか、また、その影響は、音源の種類や被験者によって異なるのかどうかを確認する目的で、様々な種類の音源を用意し、比較的多数の被験者により基礎的な評価実験を行った。実験方法は、R, A, B の 3 つの評価音を試聴し、A と B の中から R と同じ方を選択させる duo-trio test を用いた。被験者 36 人で 20 種の音源をそれぞれ 40 回評価した。この結果、音源毎には有意な弁別結果は得られなかったが、被験者 1 人が正答率 0.75 をマークした。そこで、この被験者に対して追加実験を行った。より信頼性を高めるため、音源を 6 種類に選別して、それぞれの音源を 20 回評価した。この結果、どの音源に対しても有意な弁別結果は得られなかった。

実験 1 の結果と、超高域の影響には、音源の提示時間が関係するという先行研究を鑑み、実験方法を検討、被験者、音源を精査して実験 2 を行った。実験 2 では、超高域成分が豊富な音源を作成するため、新たに音楽録音用の超広帯域マイクを開発した。このマイクロホンは、音波の回折効果を利用して超高域の感度を高めることにより、固有雑音の大きさは従来の音楽録音用マイクとほぼ同等に抑えながら、収録帯域を 100 kHz まで拡張したものである。音源は、弦楽四重奏、筑前琵琶の弾き語り、ハープシコードの 3 種を用いた。実験方法には、2 つの評価音を対にして提示し、差があるか否かを判断させる pair test 法を採用した。この方法は、音源の提示時間が制御でき、かつ、微妙な差違の検知に適している。被験者 13 人で超高域有無の弁別実験を行った結果、2 人の被

験者が筑前琵琶の音源において、超高域の有無を有意に弁別できた。したがって、音源と、聴取時間、被験者の条件によっては、超高域成分の有無が聞こえに影響を与え得ると考えられる。

一方、ハイレゾリューションオーディオの可聴域の音質差異については、広帯域録音・再生対応装置により、3種の標本化フォーマット 48kHz 24bit, 192kHz 24bit, 2.8MHz 1bit (Direct Stream Digital : DSD) にて収録・再生をおこない、各フォーマット間の音質差異に関して主観弁別実験を行って検討した。

評価音源については、収録条件を管理して独自収録を行い、フォーマット以外の条件に差が生じないように充分留意して作成した。さらに、音質への影響が大きい A/D, D/A 変換器については、機種や個体差による音質差を排除するため、機種・ファームウェアが同一のものをを用い、収録・再生を通してフォーマット毎のレベル偏差は、 $\pm 0.05\text{dB}$ 以内になるように調節した。評価に用いた音源は、邦楽、ボサノバ、ピアノトリオの3種、被験者は、聴力健常な音楽大学学生4人、バイオリニスト1人、録音技術者1人である。実験方法は、実験2と同様、微妙な差違の検知に適す pair test 法を採用した。実験の結果、被験者、音源に依らず、フォーマットの違いは有意に弁別できなかつた。

したがって、ハイレゾリューションオーディオによる広帯域収録再生では、可聴域を超える周波数成分が聞こえに影響を与える場合があり、超高域成分を豊富に含む音源については、ハイレゾリューションオーディオによる広帯域収録再生は価値あることと考えられる。

1.3 本論文の構成

本論に入る前に第2章で、本研究に関連する聴覚器官、聴覚、デジタルオーディオの基礎的事項を概説する。

前述のように、ハイレゾリューションオーディオが聴感に影響を与えるとすると、その要因としては、可聴域を超える周波数成分（超高域成分）の直接的影響と可聴域の音質変化が考えられる。

そこで、第 3 章で、楽音に含まれる超高域成分の有無を弁別できるか否かの主観弁別実験について述べる。3.1節では、先行研究で残された問題を指摘し、本章の実験の特徴を述べ、3.2節では、実験システムについて詳述する。3.3節では、様々な種類の音源と多数の被験者により基礎的な検討を行った実験 1 について述べ、続いて、3.4節で実験 1 の結果と先行研究に鑑み、音源の提示時間など実験方法を検討し、被験者、音源を精査して行った実験 2 について詳述した。3.5節で楽音に含まれる超高域成分の主観弁別実験をまとめ、結果と問題点について考察する。

第 4 章では、ハイレゾリューションオーディオと可聴域の音質について、3 種の標本化フォーマットの主観弁別実験を行い検討した。4.1節では、先行研究で残されている問題を述べ、4.2節で本実験システムについて、4.3節で主観弁別実験の内容と結果を述べ、4.4節で本章を総括する。

第 5 章では、本論文を総括するとともに、ハイレゾリューションオーディオの問題点について議論する。

第2章 聴覚とデジタルオーディオ

本章では、本論に入る前に本研究に関連する音響聴覚とデジタルオーディオの基礎的事項を概説する。

2.1 聴覚器官

聴覚器官の概要を図 1 に示す[1]。到来した音は、耳介から耳甲介腔を経て外耳道に導かれ鼓膜を振動させる。鼓膜の奥には、空気の満たされた鼓室があり、その中に耳小骨と呼ばれる、つち（槌）骨、きぬた（砧）骨、あぶみ（鐙）骨の 3 つの小さな骨が関節状に結合されている。つち骨は鼓膜の内側に接し、あぶみ骨の一端あぶみ骨底は蝸牛の前庭窓に陥入している。これらの小骨連鎖によって、鼓膜に生じた振動は、蝸牛内のリンパ液に伝えられる。内耳は、側頭骨錐体部にある骨でつくられた器官で、前方から蝸牛、前庭、三半規管の 3 部からなっており、互いに交流する外リンパ液を満たしている。内耳の骨の開口部は前庭窓と蝸牛窓で、前庭窓はあぶみ骨底によって蓋をされ、蝸牛窓は第 2 鼓膜と呼ばれる薄い膜によって閉じられている。聴覚に大きく関与するのは蝸牛である。

蝸牛は、らせん形に約 3 回転した管で、大きさは大豆程度である[2]。蝸牛の断面を図 2 に示す。管の内部は、前庭階、蝸牛管、鼓室階と呼ばれる 3 つに分かれており、蝸牛管は内リンパ液で満たされ、前庭階と鼓室階は外リンパ液と通じている。耳小骨から前庭窓に伝わってきた音の振動は、リンパ液を介して前庭階奥の蝸牛頂に向かって上がり、折り返して鼓室階を降りて鼓室へ抜ける。鼓室階の上面には基底板（基底膜）が張っていて、この膜の上に受容細胞である有毛細胞が並んでおり、それぞれが聴神経とつながっている。

有毛細胞には 2 種類あり、らせんの中心に向かって内側 1 列が内側有毛細胞（内有毛細胞）、外側 3 列が外側有毛細胞（外有毛細胞）と呼ばれ、それぞれ異なった機能をもつ。内側有毛細胞の場合、毛が曲がると細胞が興奮状態となり、

電気信号である神経パルスが発生する。内側有毛細胞には、聴神経の 90% がつながっており、音センサの役割を果たしている。鼓膜・耳小骨・前庭窓と伝わってきた音の振動により基板も振動し、この振動の大きさや場所は、音の周波数や大きさに依存して変化するが、これが内側有毛細胞によって検出される。前庭窓が振動すると、入り口付近、蝸牛基部側の基板上に変形が生じ、この変形は大きさを変化させながら一瞬にして板上を先端部（蝸牛頂）に向かって移動する。図 3 に示すように、移動中の変形が最大となる場所は、音の高さ（周波数）に関係し、音が高いほど蝸牛基部の方で、低いほど先端部の方で膨らみが最大となる。このようにして基板は音の周波数分析器の役割も果たしている。

一方、外側有毛細胞は、入ってきた振動に応じて伸び縮みし、基板の振動に機械的なポジティブフィードバックをかけ、その結果、基板の形状変化は 100 倍程度に増幅される。これは、音の周波数分析の分解能を高めるのに役立っている。さらに、この機能は、大きな音に対しては働かず、弱い音に対して働くため、聴覚の感度を高める役割を担っている。外耳・中耳を経て、内耳で神経パルスに変換された音の情報は、聴神経を通して大脳の聴覚野に送られる。

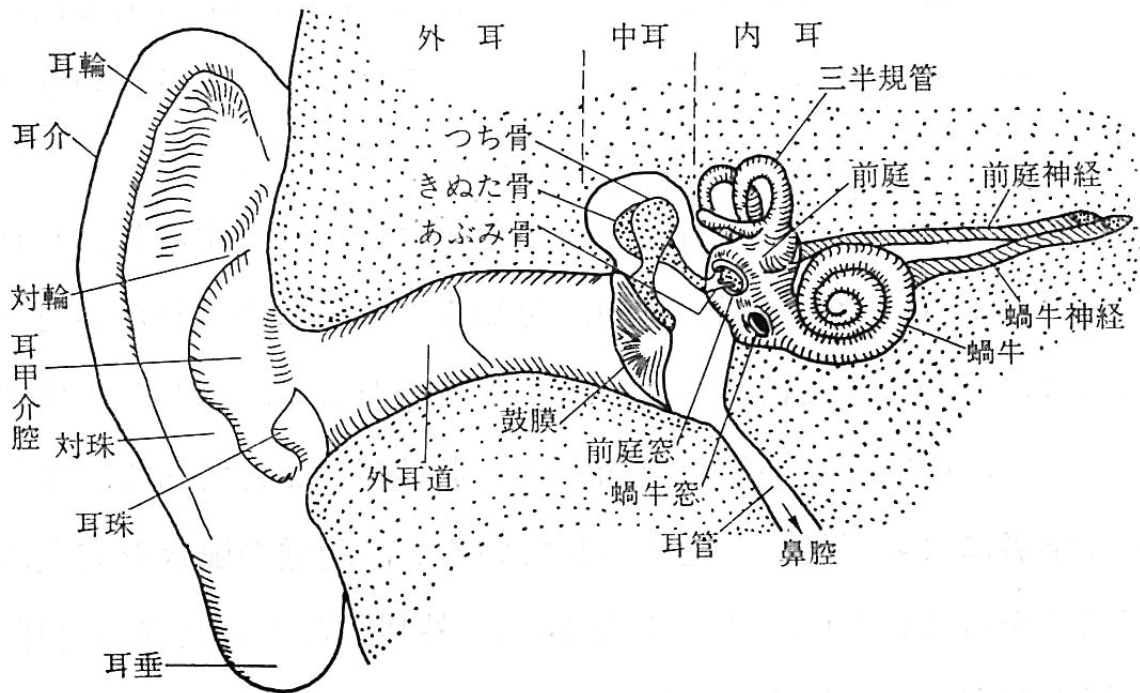


図 1. 聴覚器官の概要 [1]。

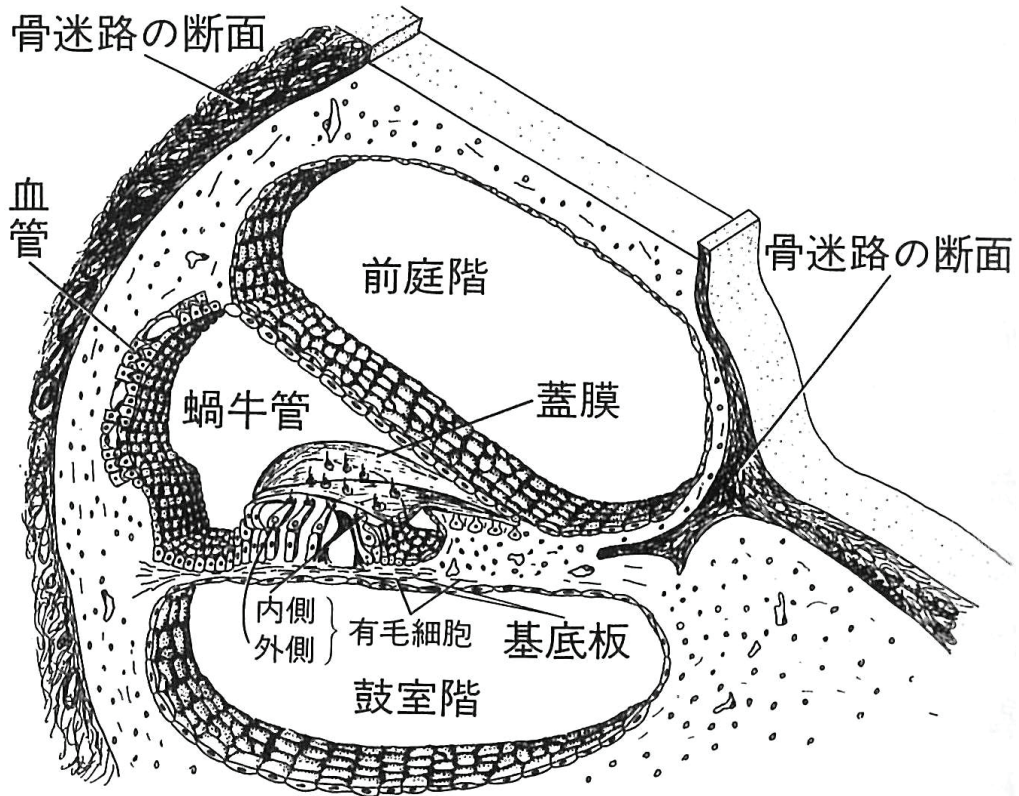


図 2. 蝸牛の断面模式図 (亀田, 1986) [3]。

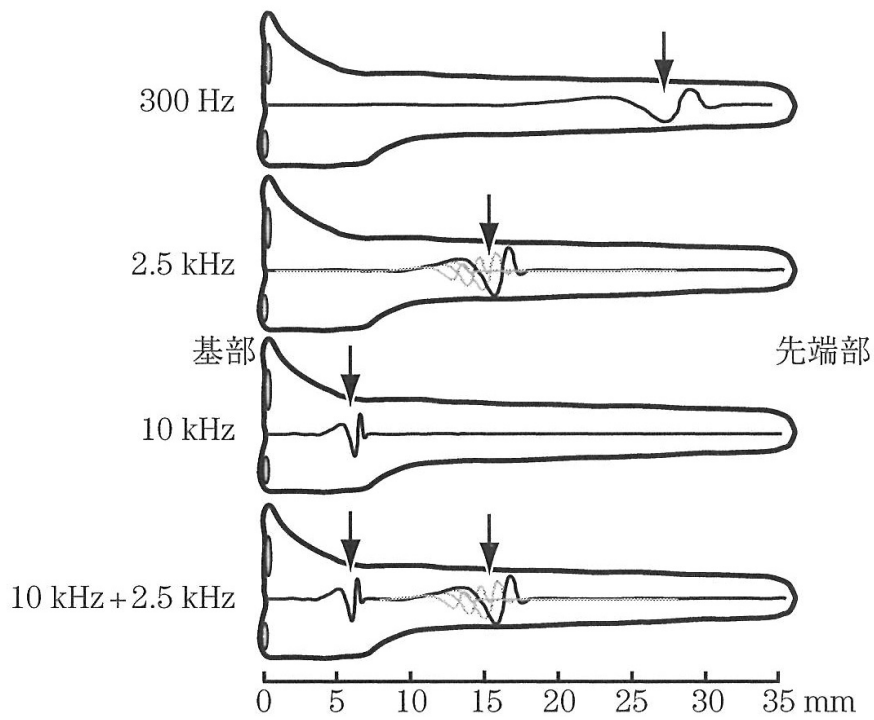


図 3. 基底板の振動と周波数 [4]。

2.2 音と聴覚

本節では、音波の物理的性質と聴覚心理の基本的な性質について概説する。

2.2.1 音の強さと周波数

音波は、図 4 (a)のように媒質の密度変化が伝搬するとも考えられるので、粗密波ともいう[1]。媒質が密なところでは、圧力が大気圧より少し上がり、粗なところでは少し下がる。ある瞬間の圧力は図 4 (b)のように大気圧を中心として微小の変化を生じていることになる。この音波が一定の速度で伝搬するから、ある 1 点で観測すると、図 4 (c)のように音は空気の圧力の時間関数として表すことができる。大気圧が音によって瞬時的に δp (Pa)変化するとき、 δp をピーク音圧として表すこともあるが、通常は圧力変化の 1 周期 T (s)を間の実効値音圧 \bar{p} で音圧を表す。

$$\bar{p} = \sqrt{\frac{1}{T} \int_0^T (\delta p)^2 dt} \quad (\text{Pa}) \quad (1)$$

また、音波の進行方向に垂直な面の単位面積を単位時間に通過するエネルギーの量で音の強さ I を表すこともある。平面波では、

$$I = \frac{\bar{p}^2}{\rho c} = \rho c \bar{v}^2 \quad (\text{W/m}^2) \quad (2)$$

で与えられる。ここで、 ρ は媒質の密度、 c は音波の伝わる速さであり、1 気圧 15°C の乾燥空気では、それぞれ 1.226 kg/m³, 340 m/s である。 ρc は固有音響インピーダンス(kg/m²s), \bar{v} は媒質の粒子速度(m/s)と呼ばれる。

我々が聴くことのできる音のエネルギーはごく微弱ではあるが、その最大と最小の比は 10¹² に及ぶため、基準量に対する比の対数をとって dB (デシベル) 表示をすることが多い。

表すべき音の強さと音圧をそれぞれ I , p としたとき、それぞれ I_0 , p_0 を基準の音の強さと音圧にすると、

$$\alpha (\text{dB}) = 10 \log_{10} \frac{I}{I_0} \quad (3)$$

$$= 20 \log_{10} \frac{p}{p_0} \quad (4)$$

として、式(3)で表したものを音の強さのレベル、式(4)で表したものを音圧レベルという。基準となる I_0 、 p_0 は、一般に空気中では、

$$I_0 = 10^{-12} \text{ W/m}^2, p_0 = 20 \text{ } \mu\text{Pa} \quad (5)$$

である。本論文では、音圧レベル (Sound Pressure Level) を式(4)、(5)で定め、単位を dB_{SPL} と表記する。音圧レベルを体験的な音で例示すると図 5 のようになる。

音の周波数は、1 秒間あたりに繰り返される圧力の周期的変化の数である。周期が T (s) であるならば、周波数 f (Hz) は、

$$f = 1/T \quad (6)$$

で、圧力の変化 δp が時刻 t に対して正弦波的に変化するとき、

$$\delta p = \delta p_m \sin(2\pi f t) \quad (7)$$

で表わされる。このような音は周波数 f の純音という。一定周期をもった音は、すべての純音、すなわち周波数 f 、 $2f$ 、 $3f \dots$ の集まりとして、次のように表すことができる。

$$\delta p = \sum_{n=1}^n \delta p_n \sin(2\pi n f t + \theta_n) \quad (8)$$

周波数 f の音を基本波音、 $2f$ 、 $3f \dots$ の音を第 2 高周波音、第 3 高周波音と呼び、 δp_n は第 n 高周波の振幅、 θ_n は位相角である。

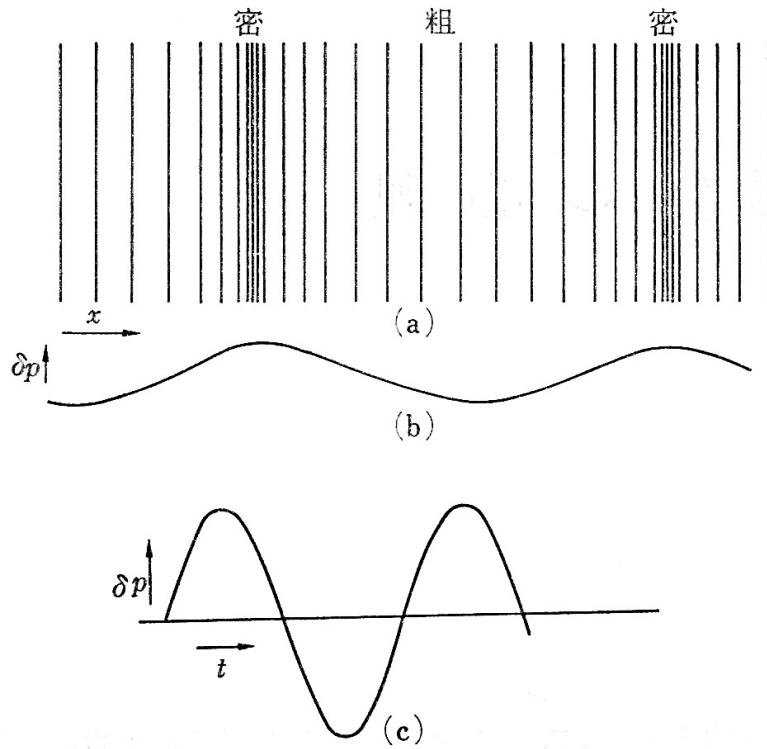


図 4. 粗密波 [1].

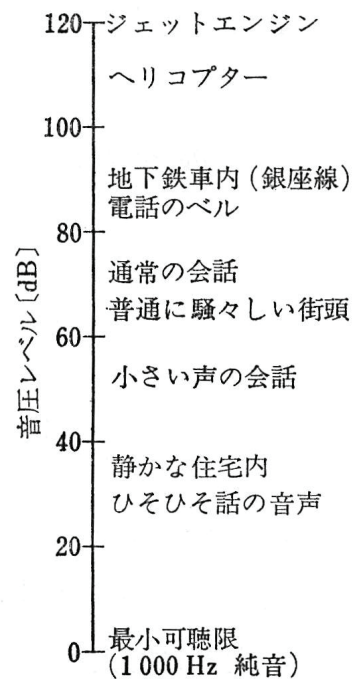


図 5. 音圧レベルの目安 [1].

2.2.2 音の種類とスペクトル

2.1節に述べたように、我々の聴覚は末梢系から周波数分析的機能を有しており、聴覚心理の面からも、音の周波数成分によって聞こえが大きく左右されるので、周波数は音の性質の重要な要素である[1]。周波数成分に関する音の構造は、フーリエ分析を用い、周波数に対して成分音の振幅をプロットした振幅スペクトルで表すことが多い[5]。または、振幅の代わりにエネルギー、あるいはパワーをプロットした、エネルギースペクトル、パワースペクトルもよく用いられる。振幅スペクトルの例を図 6 に示す。正弦波のスペクトルは、定義より単一の周波数成分からなっている。矩形波は、基本波と奇数次の高調波からなり、 n 時高調波の振幅は基本波の振幅の $1/n$ である。短いパルス列（クリック列）には、基本波および全高調波が等しい振幅で含まれている。

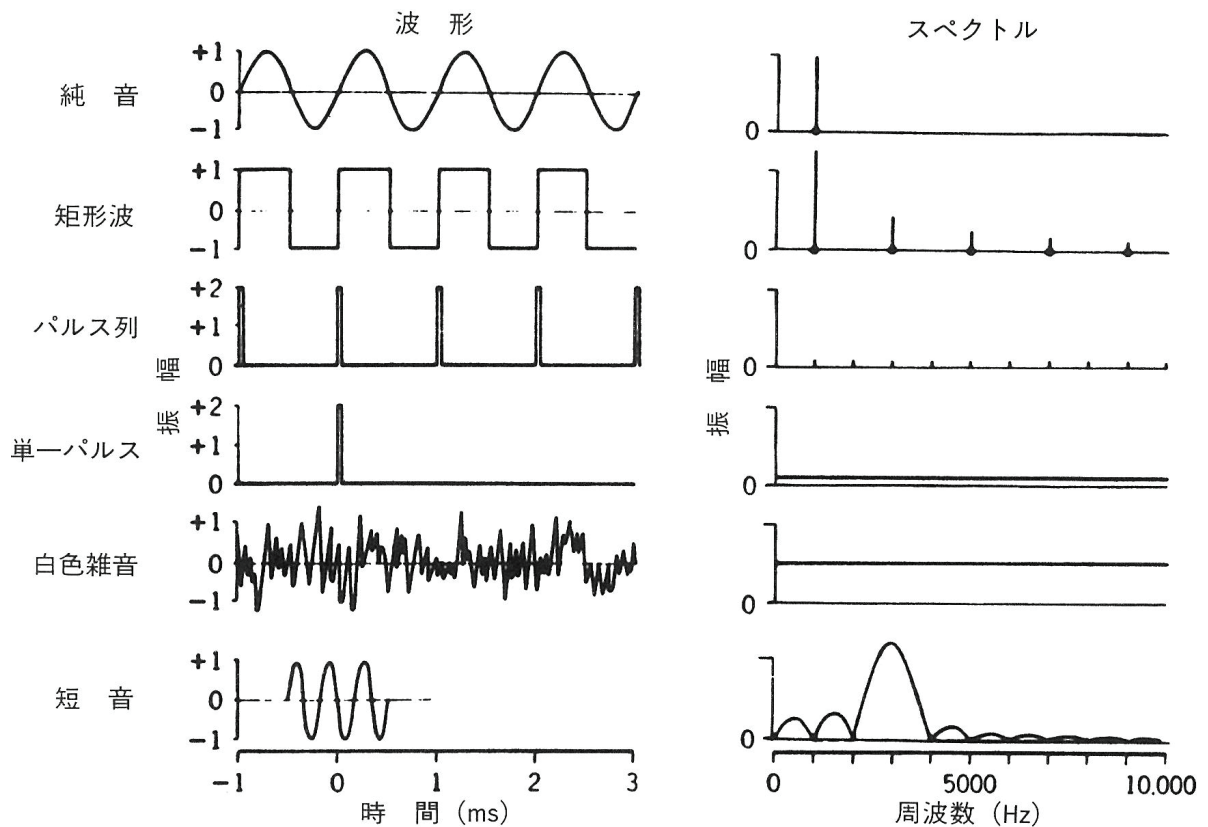


図 6. 音の波形とスペクトル [5]。

単一パルスと白色雑音 (white noise : ホワイトノイズ) は、ともに平坦で連続したスペクトルをもつ。ただし、図示していないが、各成分の位相を周波数の関数として表した位相スペクトルは異なる。白色雑音の場合、各成分の位相はランダムに分布するが、パルスでは、全成分が時刻 0 において 90 度の位相を持つ。白色雑音のスペクトルは周波数によらず一定だが、オクターブあたり 3dB 減衰するスペクトルをもつ雑音はピンクノイズと呼ばれ、単位バンド幅あたりのパワーが一定なため、測定用雑音音源として広く用いられている[6]。

短音は、正弦波を短い時間だけ出したもので、トーンバースト (tone burst)、あるいは、トーンパルス (tone pulse) とも呼ばれる。正弦波のスペクトルが線スペクトルになるのは、持続時間が極めて長いときだけである。持続時間の短いトーンバーストほど、スペクトルは広い範囲に分布する。これは、聴感上も対応していて、トーンパルスの持続時間が短いほど、高さをもった音という感じがなくなり、パルス音のように聞こえる[5]。

スペクトルを計算で求めるための数学的方法はフーリエ変換である。連続時間関数 $x(t)$ のフーリエ変換 $X(f)$ は、

$$X(f) = \int_{-\infty}^{+\infty} x(t) \exp(-j2\pi ft) dt \quad (9)$$

ここで f は周波数である[7]。関数 $x(t)$ が音響信号の波形として、 $x(t)$ の振幅スペクトルは $X(f)$ の大きさ $|X(f)|$ 、位相スペクトルは $X(f)$ の偏角、

$$\angle X(f) = \arctan\left(\frac{\text{Im}(X(f))}{\text{Re}(X(f))}\right) \quad (10)$$

で定められる。

しかし、現実の音の信号に対して、式(9)のように無限の時間にわたって波形を観測することはできないし、また、耳が無限の時間にわたり積分していないことも明白である。そこで、実際のスペクトル分析では、信号の一部を標本化 (sampling : サンプリング) して切り出し、離散時間フーリエ変換 (Discrete Fourier Transform : DFT) を適用することになる。連続時間信号を標本化周波数 (sampling frequency : サンプリング周波数) f_s で標本化し、有限長の離散

時間信号 $x(n)$, $0 \leq n < N$ を得たとする。 n は標本化周期 $1/f_s$ を単位とした時間を表す。 $x(n)$ の DFT は,

$$X(k) = \sum_{n=0}^{N-1} x(n) \exp(-j2\pi k n / N) \quad (11)$$

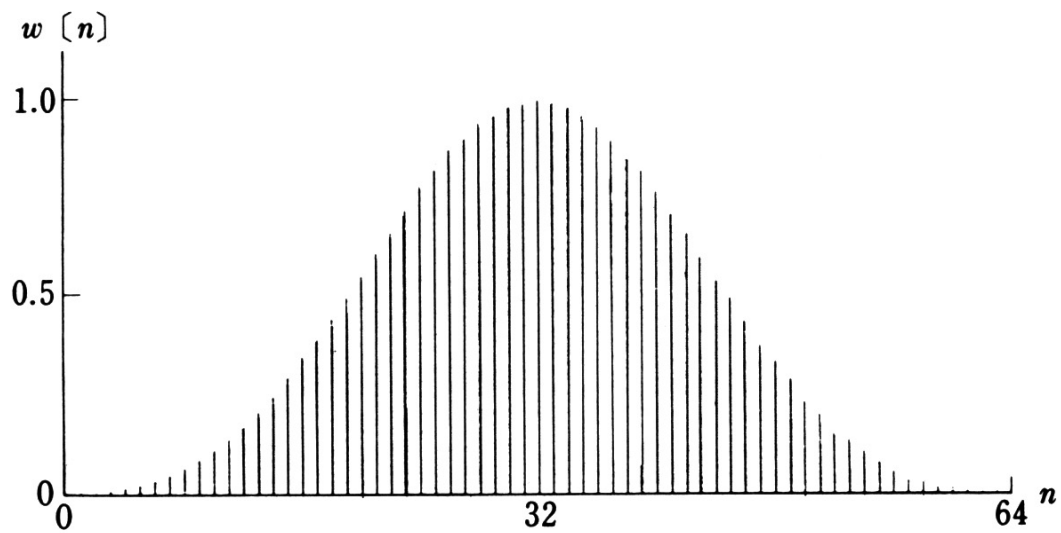
ここで k は f_s/N を単位とした周波数を表す。

このように有限長の信号を切り出してスペクトルを計算した場合、元の信号には本来存在しない周波数成分が $X(k)$ に現れてしまう。これはスペクトル漏れと呼ばれ、スペクトルを調べるうえで具合の悪いことである。そこで、スペクトル漏れをできるだけ小さくするための窓関数と呼ばれる時間の重み関数を、 $x(n)$ に乗じてから DFT を行う方法がしばしばとられる。窓関数を $w(n)$ とすると、この窓関数を施したスペクトル $X_w(k)$ 次のようになる。

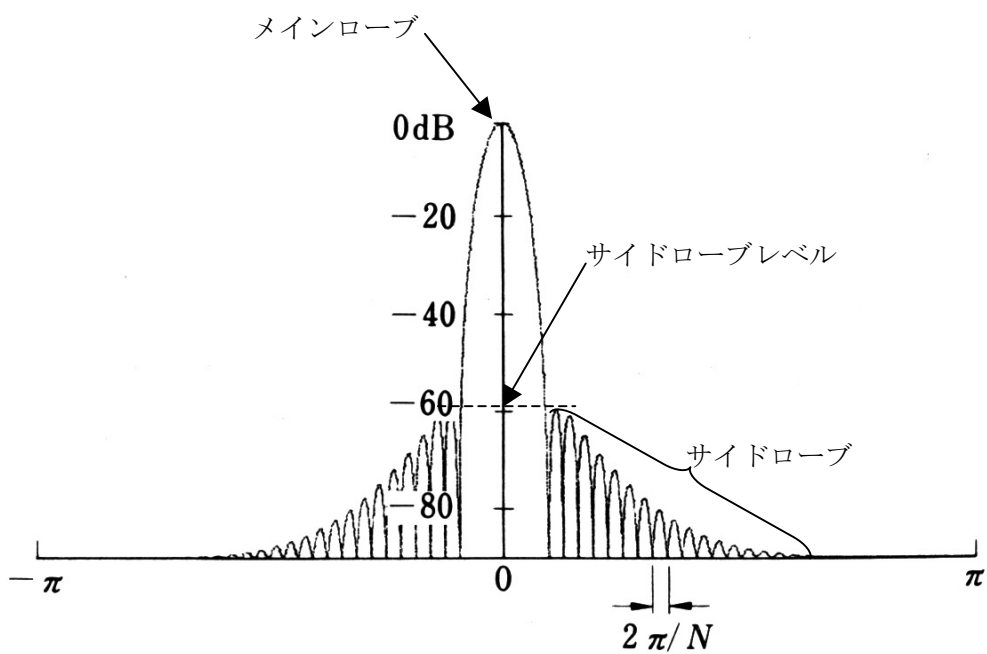
$$X_w(k) = \sum_{n=0}^{N-1} w(n)x(n) \exp(-j2\pi k n / N) \quad (12)$$

音声や楽音のように時々刻々と変化する複雑なスペクトルの時間変化を分析したい場合には、比較的長時間幅の短い窓をオーバーラップしてスペクトルを求めることもある。

窓関数の特性は、そのフーリエ変換のメインローブの帯域幅とサイドローブのレベルおよび減衰などにより特徴付けられている。窓関数とそのフーリエ変換の対数振幅の例を図 7 に示す。一般にメインローブの帯域幅が狭いほど周波数分解能が高く、サイドローブのレベルが低く減衰が急峻なほど低レベルの周波数成分の分析が正確に行える。しかし、メインローブの帯域幅とサイドローブの減衰特性はトレードオフの関係があり、表 1 のように目的に応じて様々な窓関数が提案されている [6], [8]。



(a) ブラックマン窓



(b) フーリエ変換

図 7. ブラックマン窓とそのフーリエ変換の対数振幅 ($N=64$) [8]。

表 1. 窓関数の例。

窓関数 $(0 \leq n < N)$	特 徴	メイン ローブ 帯域幅 $(2\pi/N)$	サイドローブ	
			レベル (dB)	減衰 (dB/oct)
方形窓 $w(n) = 1$	有限長区間をそのまま切り出すことに相当する。周波数分解能は最も高いが、サイドローブレベルが高く減衰も緩やか。	0.89	-13	-6
ハニング窓 $w(n) = 0.5 - 0.5 \cos\left(\frac{2\pi n}{N}\right)$	サイドローブレベルが低く、減衰も大きいので実用的に用いられる。	1.44	-32	-18
ハミング窓 $w(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N}\right)$	ハニング窓に比べ、メインローブ幅はほぼ同じで、サイドローブレベルが-43 dB と低いので、広く用いられる。	1.30	-43	-6
ブラックマン窓 $w(n) = 0.42 - 0.50 \cos\left(\frac{2\pi n}{N}\right) + 0.08 \cos\left(\frac{4\pi n}{N}\right)$	周波数分解能は上記の窓に劣るが、サイドローブレベルが低く、減衰も急峻である。レベルの低い高調波成分の検出などの用途に適している。	1.68	-58	-18
3 シグマガウス窓 $w(n) = \exp\left(-4.5 \left(\frac{2n}{N-1}\right)^2\right)$	ガウス波形は、あらゆる波形のうち実効時間幅と帯域幅の積が最小となる特徴をもつ。	1.55	-55	-6

2.2.3 音の心理物理

古くから音の感覚は 3 つの要素に分類されるといわれてきた。これを音の 3 属性という。

第 1 は、音の大きさ (loudness) で、周波数成分にも時間にも依存するが、主に音の強さに対応し、強さの増加に対応して音の大きさも増加する。

第 2 は、音の高さ (pitch) であり、周波数に対応する感覚要素といわれている。ピアノの鍵盤を次々に弾いていったような場合に、1 オクターブごとに周期的に音の類似性が戻ってくるのが知られているが、これは調性 (tonality) と呼ばれ、高さとは別の属性として扱う場合がある。

第 3 の属性として、音色^{ねいろ} (timbre) があげられている。たとえば、大きさも高さも等しい 2 つの音が異なった感じを与えるとき、その相違に対する性質が音色であると説明されている。音の心理的性質から、大きさと高さの要素を取り去った残りであるから、音のスペクトル、波形、音圧およびそれらの時間的変化など多くの物理的性質に関連している。大きさや高さは、主として一つの対応する物理的性質があげられるのに対し、音色では、そのような性質は見いだされていない。音色の心理物理的関係は整理されておらず、聴覚心理の大きな課題である。

2.2.4 可聴範囲

音が外耳道から入り、鼓膜や耳小骨連鎖を介して内耳に伝わることによって得られる聴力を気導聴力、音が頭蓋骨の振動を介して蝸牛に伝達され、これによって得られる聴力を骨導聴力という [6]。通常、骨導聴力は、気導聴力と比較することによって、難聴の鑑別診断に用いられるもので、ここでは気導聴力の聴覚範囲について述べる。

気導聴力の周波数範囲は通常 20 Hz~20 kHz とされているが、年齢や性別によって異なるほか個人差も大きい。極めて低い周波数になると皮膚が振動として感じたり、耳が大振幅の音に対して非線形特性を示して、その高調波を聴いてしまったりして、はっきりとした値を決めにくい。最低周波数は 15~20 Hz

とするのが妥当と考えられている[1], [3]。最高周波数については、次章で詳述する。

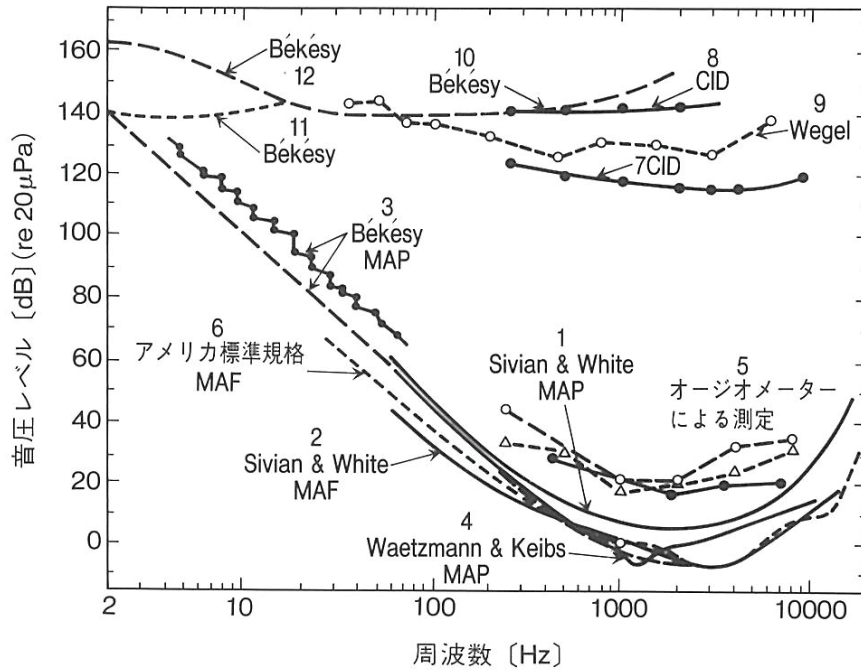
音として聞こえる音波の強さ、あるいは音圧レベルの下限は最小可聴値、最小可聴域（閾）値、聴覚域値などと呼ばれる。上限は、音の強さを増していったときに、耳に音の感覚以外の触覚や痛覚が生じ始める音の強さであり、最大可聴値や痛覚域（閾）値と呼ばれる。最小可聴値と最大可聴値は、音の周波数によって異なる値をとるため、これら両可聴値の周波数特性曲線によって囲まれた範囲を可聴範囲という。

多くの研究者により測定された結果を図 8 に示す[1]。図中の曲線 7~12 は、音の感覚以外に生じるいろいろな最大可聴値を示しているが、いずれの周波数でもおよそ 120~130 dB_{SPL} で異常を感じるといえる。さらに強い音も聴取できないわけではないが、130 dB_{SPL} を越すと、短時間でも聴覚を損なう危険がある。

最小可聴値は、通常、音の強さを小さくしてゆき、50%の確率で聴こえると判断できる音圧レベルをもって定めている。図 8 の曲線群 5 が他に比して感度が悪いのは、聴こえる限界というよりも、多くの人々の最小可聴値の平均という臨床的な基準であるからである。最小可聴値には、自由音場で測定した MAF (Minimum Audible Field) と、片耳受話器を用いて測定した MAP (Minimum Audible Pressure) の 2 種類がある。MAF では、外耳道による共振の影響が 3 kHz 付近に現れ、耳の感度を良くしている。また、MAP では、受話器を耳に圧迫するので、呼吸や血流による耳内雑音が増加し、試験音がマスクされて MAF の場合より多少域値が上昇する。

最小可聴限の周波数特性は、年齢によって大きく変化する。年齢が増すにつれて、鼓膜や耳小骨の関節、筋など伝音系の硬化や蝸牛基部の細胞損失などによって最小可聴値が上昇する。

図 9 は、高齢者 (60~95 歳) 46 人と勤労世代 (60 歳未満) 86 人の MAP を測定した例である[9]。40 代から MAP の上昇が目立ちはじめ、60 代以降はは全域の MAP が上昇し、特に 10 kHz 以上の上昇は著しい。



1, 3, 4, 5 は受話器による最小可聴限, 2 と 6 は音場での最小可聴限, 7 は不快が生じ始める最大可聴限, 8 と 10 はむずかゆい感じはし始める, 9 は痛み始める, 11 は触覚的な感じ, 12 は中耳がしくりと痛い感じのし始めるレベル。MAF (Minimum Audible Field) : 自由音場における最小可聴値, MAP (Minimum Audible Pressure) : 片耳受話による最小可聴値。

図 8. 可聴範囲 [1]。

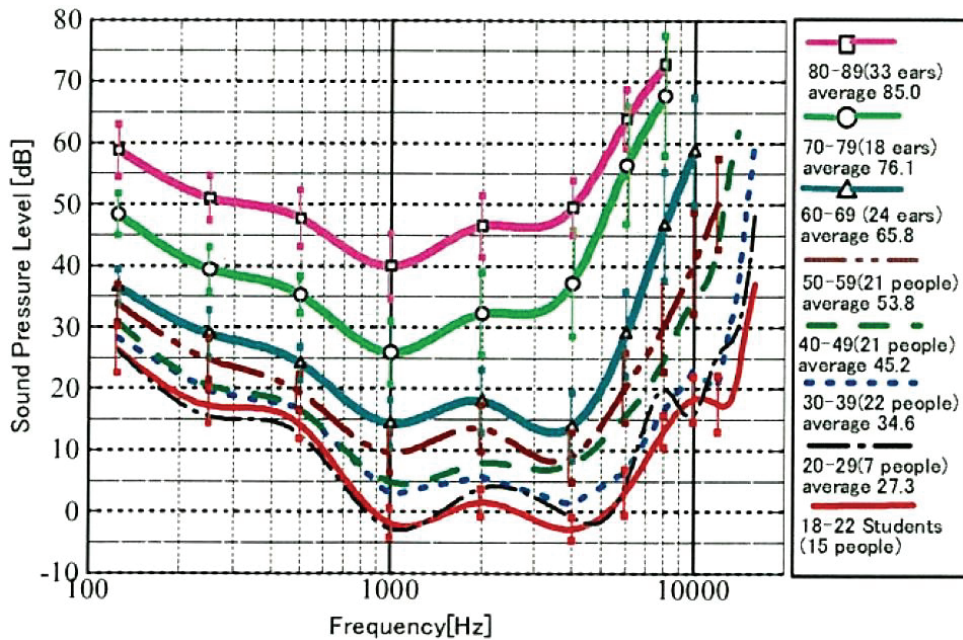


図 9. 年齢別の MAP 平均値と 95%信頼区間 [9]。

2.3 デジタルオーディオ

本節では、デジタルオーディオ技術の基礎、デジタルオーディオフォーマットの変遷、今日のハイレゾリューションオーディオを概説する。

2.3.1 アナログ信号のデジタル化

デジタル信号とは時間、振幅とも離散 (discrete) 的な信号のことである。アナログ信号をデジタル化すると、その情報量 (エントロピー) は有限な値となる。デジタル処理系の質は、基本的には、標本化周波数と量子化特性、演算語長によって決定される。デジタル伝送では伝送容量が有限な現実の機器を使用しても、劣化のない伝送・処理が期待できる[7]。

2.3.1.1 標本化と補間

図 10 にデジタル信号処理の流れを示す。アナログ信号をデジタル化 (A/D 変換) するには、時間の離散化である標本化 (sampling : サンプルング) と振幅の離散化である量子化 (quantization) の操作が必要である。通常はサンプル (アンド) ホールド回路 (sample and hold circuit : S/H) で標本化を行ったうえ、量子化器 (quantizer) で量子化するが、標本化と量子化が同時に行うこともある。なお、単に量子化器を A/D 変換器と呼ぶこともある。

標本化とは、図 10 (a) の連続信号のある時点の値を読み取る操作である。標本化周波数が標本化する信号のもつ帯域の 2 倍以下の場合には、後述する標本化定理を満足させるべく、標本化に先立ち信号の帯域制限を行う。

標本化された各標本値を (c) のように数字 (通常 2 進数) で表現する操作が量子化である。

伝送、記録等の処理を終えたデジタル信号は、デジタル信号に対応する振幅値に変換する逆量子化器で各標本値に対応したパルス列に戻し、補間フィルタ (一般にローパスフィルタ) でアナログ信号に復元する (D/A 変換)。逆量子化器を単に D/A 変換器と呼ぶこともある。(d) に示すように、標本値を帯域 $1/(2T)$ の理想ローパスフィルタ (Low Pass Filter : LPF) に通すと、標本点

ではもとの値のままで両側に減衰振動が現れるが、減衰振動の波形は、他の標本点ではすべて 0 である。すべての標本値に理想ローパスフィルタを施すと、(e) に示すように標本点では標本値に等しく、標本点の間は各標本点に対するフィルタ出力の和で連続的に埋ってアナログ信号が得られる。この操作がフィルタによる標本の補間 (interpolation) である。

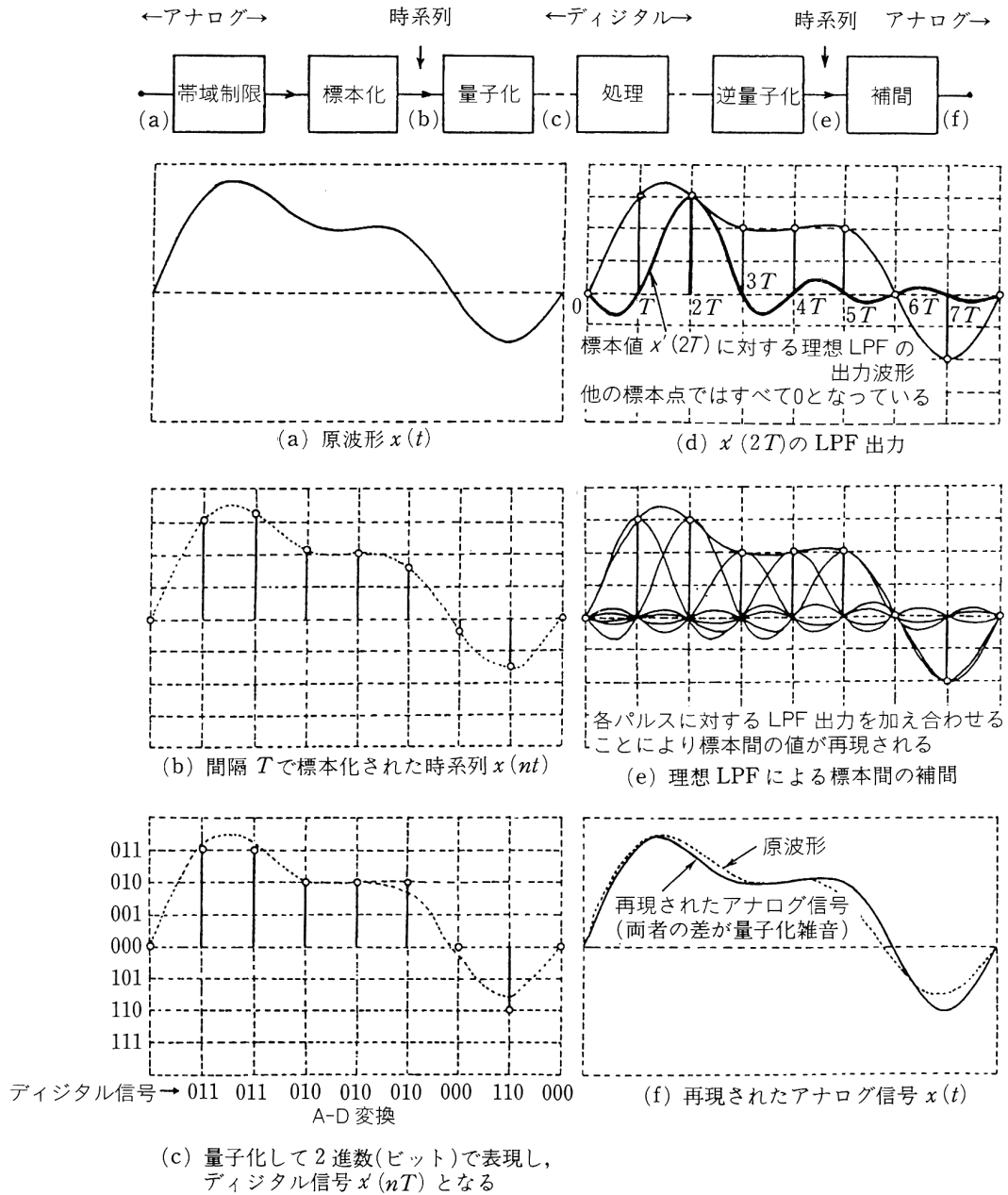


図 10. デジタル信号処理の流れ [7].

実際には、完全な帯域制限、理想的なインパルスによる標本化、幅のないパルス列、理想フィルタによる補間は実現不可能であり、現実のハードウェアでは保持効果（aperture effect）が生じる。

また、(f) の再現された信号と原波形との差が量子化により生じた量子化雑音（quantization noise）であり、デジタル化に際して原理的に避けることのできない信号劣化である。保持効果と量子化雑音については後述する。

2.3.1.2 標本化定理と保持効果

帯域 $1/(2T)$ (Hz)の信号 $x(t)$ は、時系列 $x(nT)$ を使って

$$x(t) = \sum x(nT) \cdot \frac{\sin(\pi(t-nT)/T)}{\pi(t-nT)/T} \quad (13)$$

と書くことができる。式(13)は $1/(2T)$ (Hz)に帯域制限された原信号 $x(t)$ が標本化された時系列 $x(nT)$ を帯域 $1/(2T)$ (Hz)の理想ローパスフィルタに通すことにより実現されることを示している。これが標本化定理である。

時系列からアナログ信号を再現するには、理論的には標本値に比例した幅のないパルス列を理想ローパスフィルタで補完しなければならない。現実にはこれは不可能なので、幅 τ (s)のパルスを使うと、例えば f (Hz)の純音 $x(t) = \cos(2\pi ft)$ に対し、復元出力 $y(t)$ は、

$$y(t) = \frac{\tau}{T} \cdot \frac{\sin(\tau\pi f)}{\tau\pi f} \cos\left(2\pi f\left(t - \frac{\tau}{2}\right)\right) \quad (14)$$

となり、 $\tau/2$ (s)の時間遅れと図 11 に示す高域の減衰が生ずる。この現象が保持効果と呼ばれる。通常時間遅れは問題にならないが、高域の減衰はフィルタで補正する等の対策を講ずる。

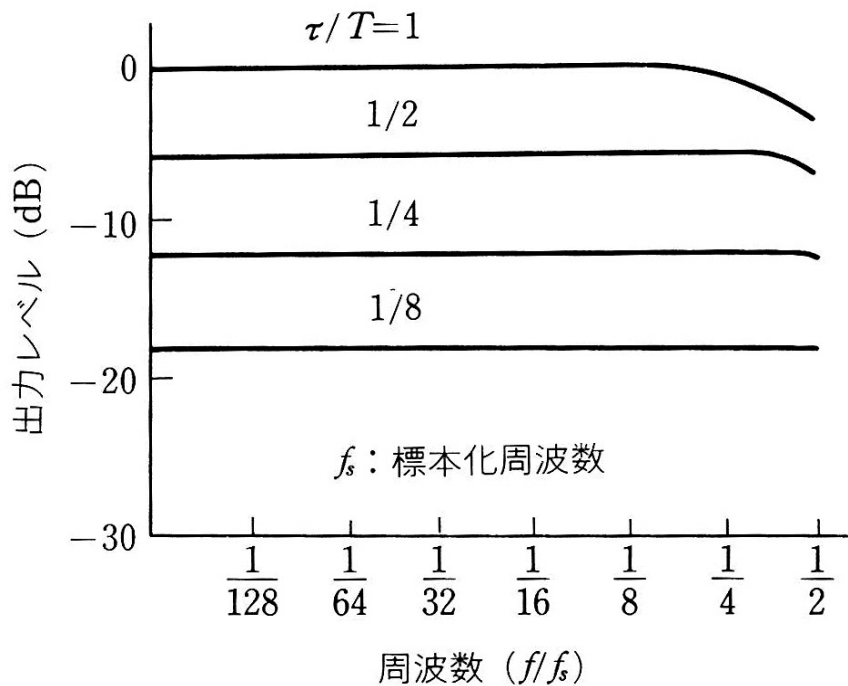


図 11. 保持効果 [7]。

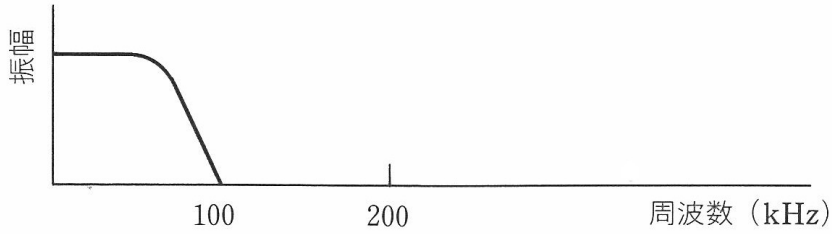
2.3.1.3 帯域制限フィルタ

原信号の帯域が予め制限されている場合や標本化周波数を充分高く設定できる場合は必要ないが、一般には、標本化に先立ち標本化定理を満足するように標本化周波数の 1/2 以内の帯域制限を行う。帯域制限が不完全だと標本化周波数の 1/2 以上の信号成分が折り返された形で帯域内に現れる。この現象は折り返し (aliasing) と呼ばれ、一度生じた折り返し雑音を後処理で取り除くことは不可能である。

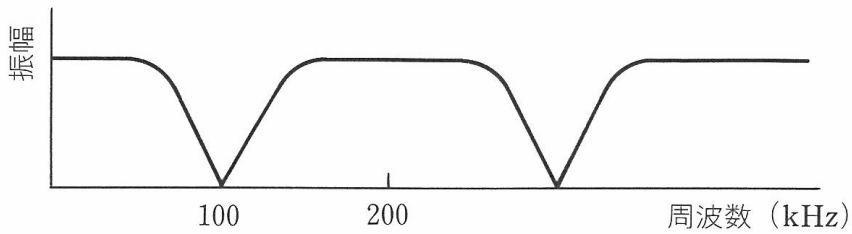
2.3.1.4 オーバサンプリング

標本化に先立つ帯域制限や補間用のフィルタには、高次の受動型あるいは能動型アナログフィルタが使われてきた。A/D, D/A 変換器やデジタル演算素子の高速化にともない、アナログフィルタの負担を軽減するため、高速標本化とデジタルフィルタを組み合わせたオーバサンプリングが用いられている。図 12 の例では、所望の信号帯域を 20 kHz、標本化周波数は 50 kHz に対し、4

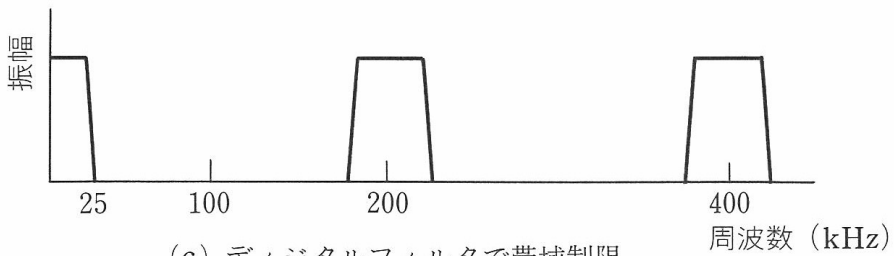
倍オーバーサンプリング（標本化周波数 200 kHz）で標本化している。標本化周波数は 200 kHz なので、アナログフィルタでは 100 kHz の帯域制限をすれば折り返し雑音は発生しない。標本化・量子化されたデジタル信号を、通過域 0~20 kHz, 阻止域 25~100 kHz のデジタルフィルタで帯域制限し、4 分の 1 に間引く操作（decimation）によって、所望の標本化周波数 50 kHz のデジタル信号が得られる。この例では、オーバーサンプリングを用いなければ、遷移域が 20 kHz~25 kHz という急峻なアナログフィルタが必要になるところ、オーバーサンプリングを用いた場合、アナログフィルタの遷移域は 20 kHz~100 kHz と緩やかなフィルタを用いることができる。また、デジタルフィルタは、FIR（Finite Impulse Response）型で構成すると、アナログフィルタでは実現不可能な、位相特性を直線とすることができ、緩やかなアナログフィルタと相まって群遅延歪みを小さくすることができる。標本化周波数が必要帯域の数倍以上であれば、特殊な場合を除きアナログフィルタを省略することもできる。同様の手法で、D/A 変換の補間フィルタの負担軽減も可能である。



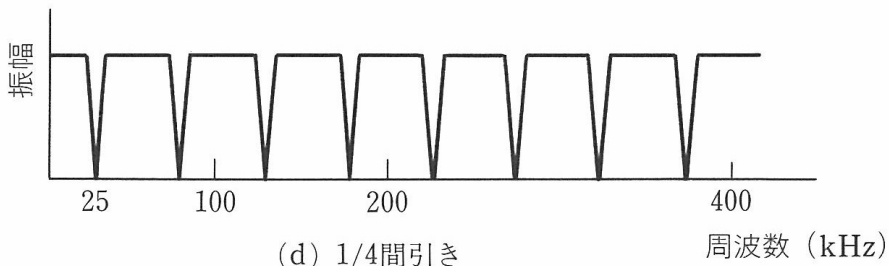
(a) アナログフィルタで 100 kHz に帯域制限



(b) 200 kHz で標本化, 量子化



(c) デジタルフィルタで帯域制限



(d) 1/4間引き

図 12. オーバサンプリングを用いた A/D 変換の例 [7].

2.3.1.5 量子化雑音と標本化周波数

量子化のステップ幅を Δ とすると、標本化周波数の1/2の帯域の量子化雑音電力は $\Delta^2/12$ である。これは、量子化ビット数を M としたとき、最大振幅で正規化すると $2^{-2(M-1)}/12$ となる。ここで、最大振幅で正規化した信号電力を σ^2 とすると、信号対量子化雑音電力比 S/N_q は、

$$S/N_q = \sigma^2 / (2^{-2(M-1)}/12) = 12\sigma^2 / 2^{-2(M-1)} \quad (15)$$

となり、dB表示すると

$$S/N_q = 6.02M + 4.77 + 20\log(\sigma) \quad (\text{dB}) \quad (16)$$

となる。振幅1の正弦波の実行値は $\sqrt{1/2}$ であるから、最大振幅の正弦波に対しては、

$$S/N_q = 6.02M + 1.76 \quad (\text{dB}) \quad (17)$$

となり、量子化ビット数が大きいほど S/N_q は大きくなり、ダイナミックレンジも広がる。

ところで、一般に標本化周波数が帯域を決定し、量子化特性がダイナミックレンジをそれぞれ独立に決定するように受け取られているが、実は、両者には密接な関係がある。信号振幅が大きく変化に富んでいる場合には、量子化雑音は、周波数領域で見ると一様に分布し、総電力は標本化周波数によらず $\Delta^2/12$ である。信号帯域を一定とした場合、図13に示すように標本化周波数が高いほど信号帯域内に分布する量子化雑音電力は小さくなる。しがたって、量子化ビット数がたとえ1 bitであっても、計算上は標本化周波数を高くすることにより、帯域内のダイナミックレンジはいくらでも広くとることが可能である。

標本化周波数を信号帯域の $2k$ 倍とすると量子化雑音の総電力は $\Delta^2/12$ 、信号帯域では $\Delta^2/12k$ となる。前節のオーバサンプリングを利用し、必要な標本化周波数の k 倍の高い標本化周波数でデジタル化し、デジタルフィルタで必要な標本化周波数のための帯域制限をしたうえ、標本を間引くことにより量子化雑音が $1/k$ になったデジタル信号が得られる。

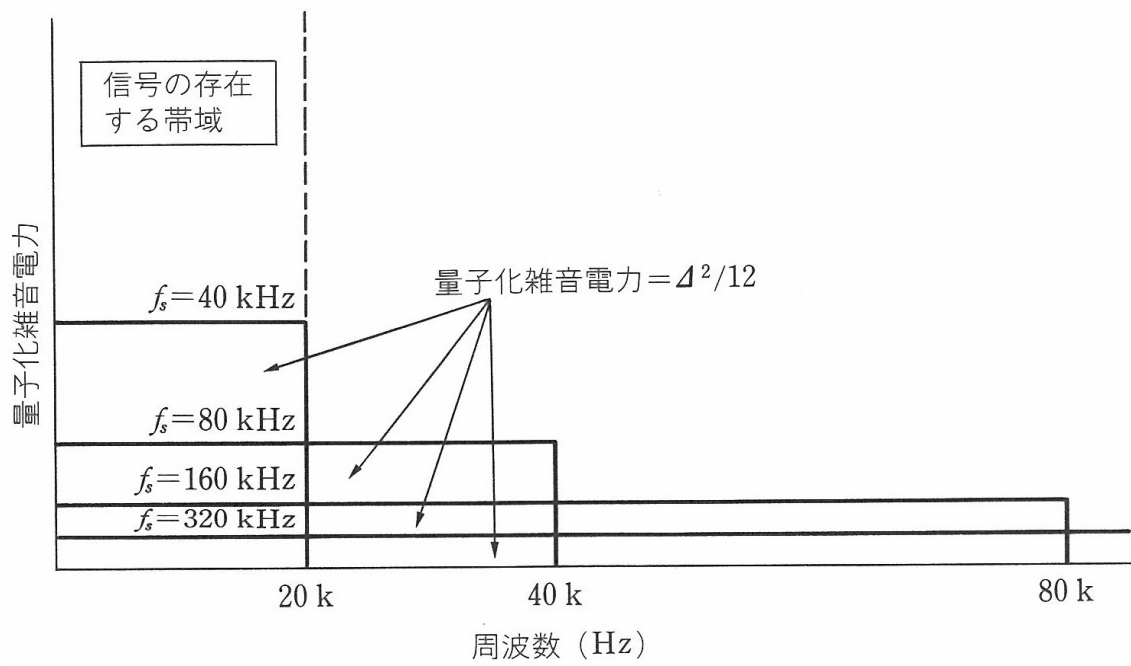


図 13. 標本化周波数 f_s による量子化雑音分布の変化 [7]。

2.3.2 A/D, D/A 変換方式

代表的な A/D, D/A 変換器の構成図を図 14 に、そのスペクトル分布を図 15 に示す。DAT や衛星放送で使われている 48 kHz 標本化、16 bit 量子化系を例に説明すると、(a) は基本的な方法で、アナログフィルタで帯域制限をしたうえで、サンプルホールド回路で 48 kHz で標本化し、16 bit の A/D 変換器で量子化する。復調時には 48 kHz, 16 bit で D/A 変換して、アナログゲート等でパルス列を作り、アナログローパスフィルタで補間を行う。折り返し雑音の影響を避ける為に、非常に急峻な遮断特性をもつアナログフィルタが要求され、一般に 10 次以上のチェビシェフ型ローパスフィルタが使われる。

(b) は、2.3.1.4 節で説明したオーバサンプリングを利用した方式である。この方式では、帯域制限がデジタルフィルタにより正確に処理されるので、アナログフィルタは低次の緩やかな遮断特性で充分である。また、2.3.1.5 節で述べた原理で、標本化周波数に応じて量子化器のビット数を減らすことができる。逆に、所定の精度の量子化器を用いた場合には、標本化周波数に応じた分だけ量子化精度が向上する。CD プレーヤ等に標本化周波数の 2~16 倍としたこの方式が広く用いられた。

(c) は現在広く使われている Σ/Δ 変調と呼ばれる変換方式で、量子化器を帰還ループの中に設けることにより、量子化雑音に高域上がりの特性を与えている。この動作をノイズシェーピング (noise shaping) と呼ぶ。標本化周波数を高く設定することにより、少ないビット数で広いダイナミックレンジが得られる。(b) と同様に帯域制限、補間はデジタルフィルタで行われるが、一般的に標本化周波数は (b) の方式より高く設定されるので、アナログフィルタの負担はさらに軽くなる。多くの場合、A/D 変換前のアナログフィルタやサンプルホールド回路は省略できる。(a) や (b) の変換には、逐次比較型や積分型の A/D 変換器、ラダー抵抗型や積分型の D/A 変換器が使われることが多く、これらの変換精度は、抵抗やコンデンサの充放電精度に頼っているのに対し、 Σ/Δ 変調方式では、時間軸の精度を利用したビット数の少ない量子化器が使われることが多い。

(d) は高速標本化 1 bit 変換方式で, 1 bit の出力をそのままデジタル信号として扱う処理系である。

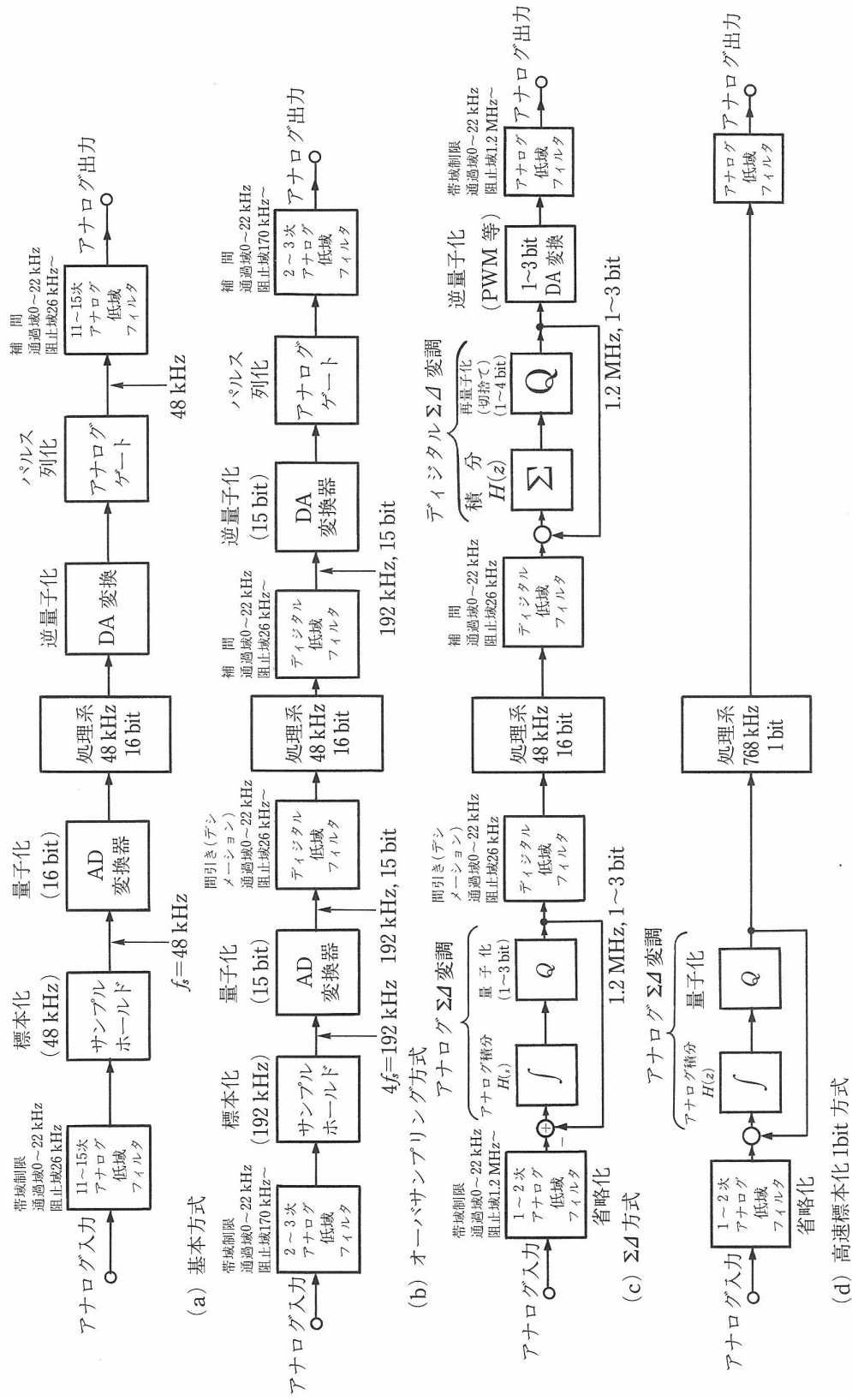
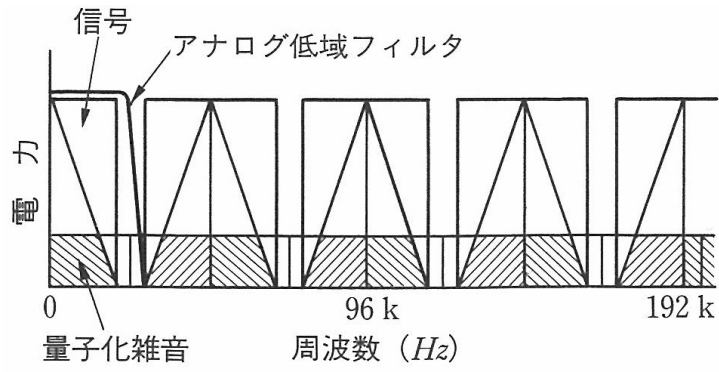
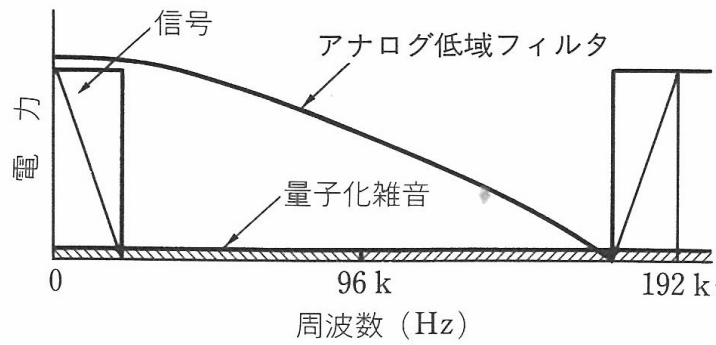


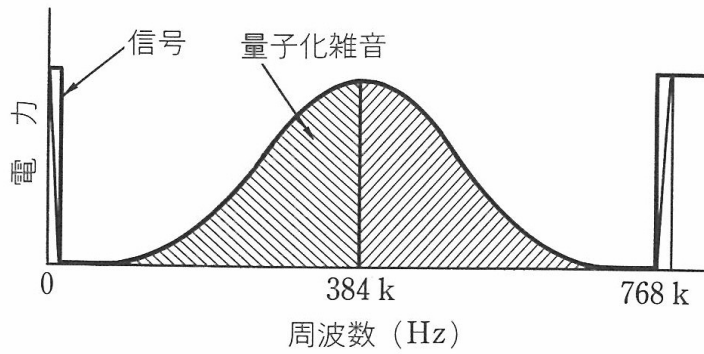
図 14. 各種の AD, D/A 変換方式 [7].



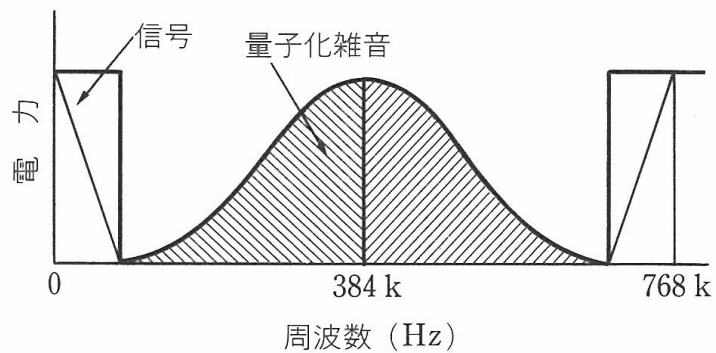
(a) 基本方式



(b) オーバサンプリング方式



(c) $\Sigma\Delta$ 方式



(d) 高速標本化 1 bit 方式

図 15. 変換方式と周波数スペクトル [7].

2.3.3 デジタルオーディオの変遷とハイレゾリューションオーディオ

表 2 にデジタルオーディオの変遷を示す[10] - [14]。デジタル信号処理の基礎となる PCM (Pulse Code Modulation) は 1937 年にフランスの A. H. Reeves が発明し、Shannon に代表される情報理論の発展とその後の半導体技術の飛躍的進歩によって、今日のデジタル信号処理の隆盛に至っている。デジタルオーディオとしては、1968 年、NHK 技研が PCM 録音機を公開、1972 年には、PCM 録音による初の LP レコードが発売されている。1981 年には、コンパクトディスク (Compact Disc : CD) が発売され、1984 年には、衛星放送でデジタル音声の放送が開始された。これまでのデジタルフォーマットは、一般的な可聴帯域と聴取環境を考慮し、標本化周波数 44~48kHz、量子化ビット数 16 bit 程度が採用されていた。

しかし、生のオーケストラや聴覚のダイナミックレンジは 100 dB を超え、楽器によっては 100 kHz の帯域をもつものもある。また、技術的には、先に述べたように、アナログローパスフィルタや量子化雑音の悪影響を軽減しようと考えられ、半導体技術の進歩もあいまって、より高い標本化周波数、量子化ビットを使ったハイレゾリューションオーディオが開発されてきた。

1997 年には、24bit/192kHz の A/D, D/A 変換器が発売。翌年、ソニー、フィリップスが提案した DSD (Direct Stream Digital) と呼ばれる 1bit/2.8MHz に対応する A/D, D/A 変換器が実用化した。この方式は、量子化ビット数を 1bit としながらも、標本化周波数を CD の 44.1 kHz の 64 倍にあたる 2.8224 MHz と高くして、2.3.1節で述べたオーバーサンプリングとノイズシェーピングを利用することで、可聴域のダイナミックレンジを 120 dB 以上確保している。24bit/192kHz は DVD-Audio, DSD は SACD (Super Audio CD) に用いられている。現在、プロフェッショナル用途に限られるが、DXD (Digital eXtreme Definition) と呼ばれる 24bit/352.8kHz の標本化フォーマットも実用化している。こうしたハイレゾリューションオーディオ対応変換器のほとんどすべては、2.3.2節で述べた $\Sigma \Delta$ 変調方式により実現されている。

表 2. デジタルオーディオの変遷。

1937	PCM (pulse code modulation) 発明 (仏 A. H. Reeves)
1948	“通信理論” (C. E. Shannon)
1968	PCM 録音機公開 (12bit/30kHz, NHK 技研)
1972	PCM 録音実用機 (13bit/47.25kHz), LP 発売 (日本コロムビア)
1977	民生用 PCM プロセッサ (13bit/44.056kHz) 発売 (ソニー)
1981	CD (16bit/44.1kHz) 発売 (フィリップス, ソニー 他)
1984	デジタル音声記録の LD 発売 (パイオニア)
1984	衛星放送 B モード音声 (16bit/48kHz) 放送開始 (NHK)
1988	24bit 対応 AD, DA 変換器発売 (英 DCS)
1997	24bit/192kHz 対応 AD, DA 変換器発売 (英 DCS)
1998	DSD (1bit/2.8MHz) 対応 AD, DA 変換器発売 (英 DCS)
1999	SACD 規格化, デジタルフォーマットは DSD (フィリップス, ソニー)
1999	DVD-Audio 規格化 (DVD フォーラム)
2004	DXD (24bit/352.8kHz) 対応変換器発売 (デンマーク DAD)

2.4 第 2 章のまとめ

本章では最初に、聴覚の根本である聴覚器官について述べ、外耳から入った音が神経パルスに変換され聴神経に送られる仕組みを概説した。2.2節では、音の強さやスペクトルなど音響物理と、聴覚の 3 要素や可聴範囲について述べ、本研究の中心となっている主観評価実験の音源や結果の分析の基礎を示した。2.3節では、デジタルオーディオの理論と技術を概説し、第 4 章で取り上げる様々なデジタルフォーマットが用いられるようになった技術的背景を示した。

これまでの CD やデジタル放送などは、周波数帯域とダイナミックレンジの点で、2.2節で述べた可聴範囲をほぼ満足するよう、標本化周波数や量子化ビット数など技術的仕様が定められたが、その後、アナログローパスフィルタや量子化雑音の悪影響を軽減して、実際的な性能の向上させるという技術的な要求から、より高い標本化周波数、量子化ビットを使ったハイレゾリューションオーディオが実現され、結果として広帯域収録再生が可能になった。しかし、広帯域収録再生の音質に対する心理音響的な観点からの科学的な検討は多くない。こうした検討は、音声の収録方法、電気音響変換器、記録・伝送フォーマットの基礎検討としても重要な課題である。そこで、本研究では、ハイレゾリューションオーディオ対応機器による広帯域収録再生に関して、可聴域を超える周波数成分（超高域成分）の影響と、可聴域の音質変化という二つの視点から検討する。

第3章 超高域成分の主観弁別実験

3.1 はじめに

気導聴力の可聴範囲については2.2.4節で概説したが、ここでは特に可聴周波数の上限について詳説する。気導提示される音の可聴周波数帯域については、古くから様々な検討がなされてきた[15] - [17]。

Fausti らは、8~20 kHz の最小可聴値の測定を行い、過去の閾値測定の結果と比較を行っているが、図 16 に示したようにいずれの結果においても、最小可聴値は 16 kHz から急激に上昇することが示されている[18]。

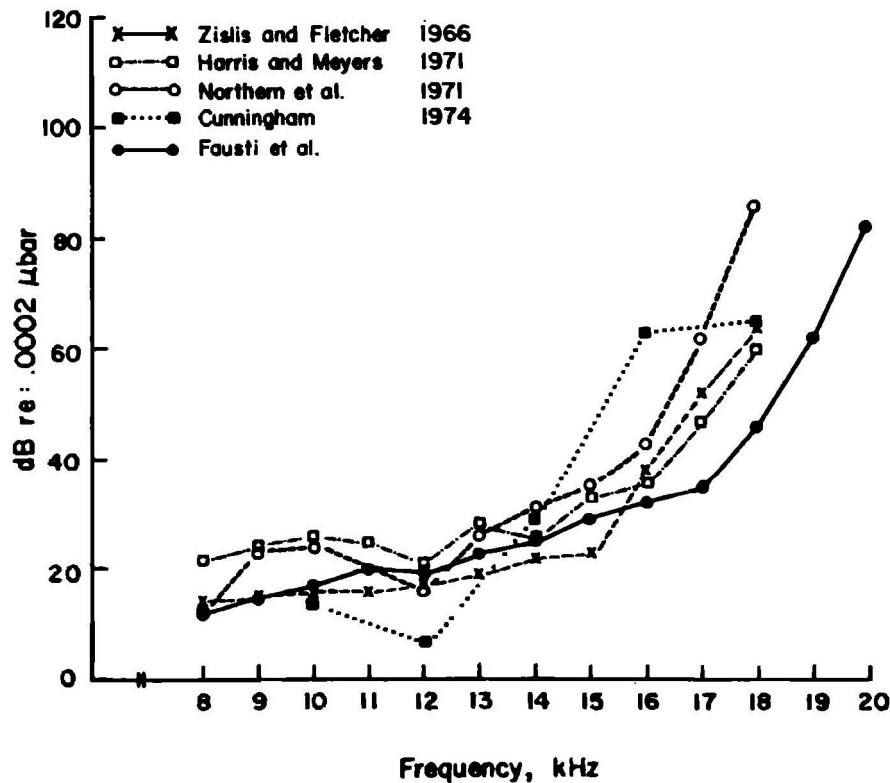


図 16. 8~20 kHz の最小可聴値 (Fausti ら[18])。

Plenge らは、デジタルオーディオにおける必要帯域の検討を目的に、インパルス性の広帯域テスト信号（図 17）の帯域をフィルタで変え、43 人の被験者で実験を行い、音の伝送帯域幅としては、15 kHz で充分と結論づけている[19]。

同様に田辺らも、番組音声伝送に必要な帯域を検討するため、広帯域合成音とローパスフィルタを用い、14 人の被験者で可聴上限の測定と主観弁別実験を行っている[20]。その結果、稀な条件下で 20 kHz を超える帯域の有無を弁別できる場合があるが、ほとんどの被験者は 19kHz の帯域制限の有無を弁別できず、高品質音声番組伝送の為の上限周波数としては、余裕をみて 22 kHz 程度に設定すればよいとしている。

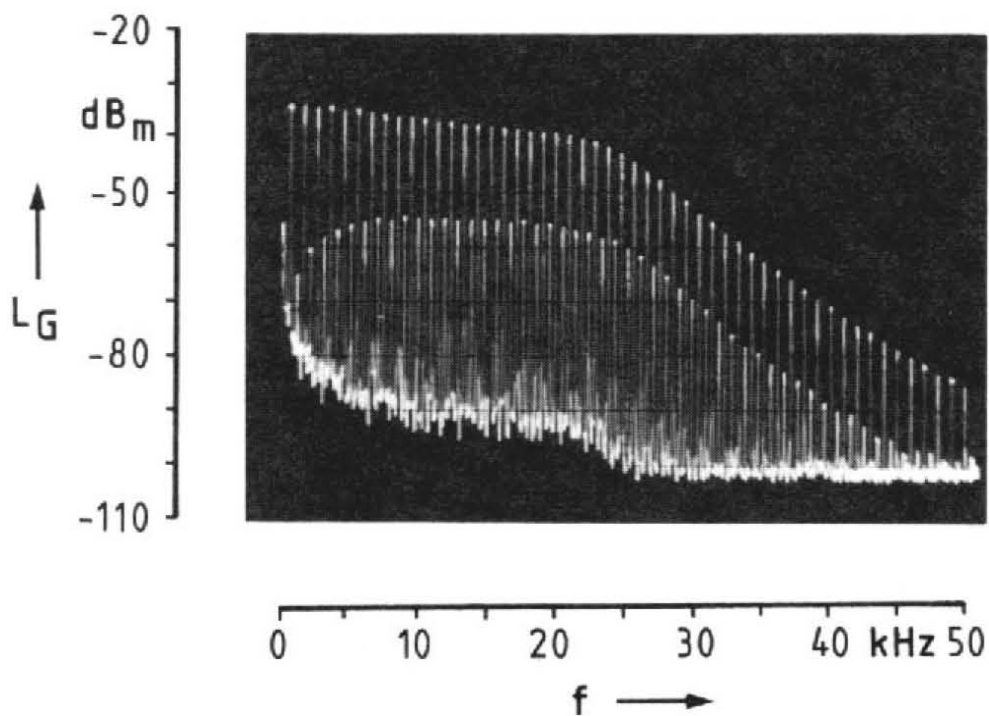
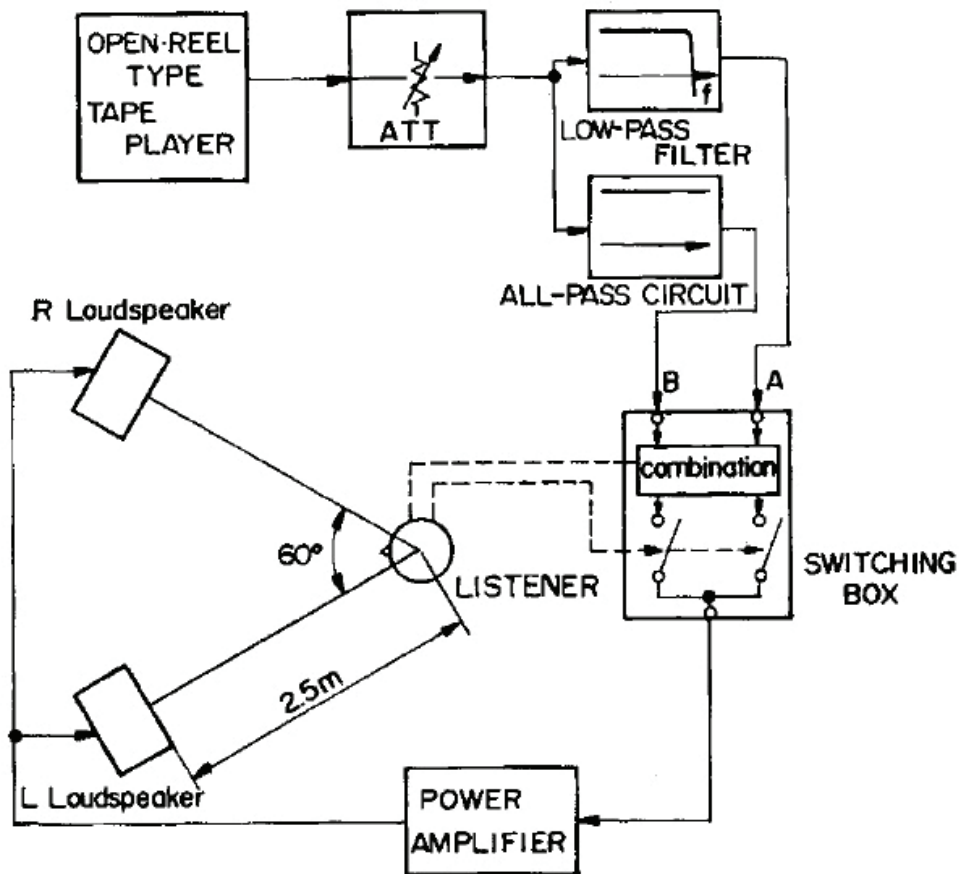


図 17. Plenge らの用いたテスト信号のスペクトル [19]。

Muraoka らも、楽音とスイープ信号を用いて実験を行っている。被験者は、音響技術者を中心とした 176 人（年齢は 15～56 歳）と極めて多い。楽音を使った実験では、図 18の実験装置で図 19のようなレベル変化をもつ楽音の再生帯域をローパスフィルタによって変えたとき、被験者が弁別できたか否かを評価している。この結果を図 20に示す。ローパスフィルタのカットオフ周波数を 16 kHz 以上とした場合、正答の平均回数が試行回数の半分の 10 となることから、カットオフ周波数の閾値は 15 kHz 周辺であると述べている。18 kHz と 20 kHz で正しく弁別できた被験者が僅かに見られるが、この結果は突出していて保障できないとしている。また、純音を使った実験では、試験音の音圧レベルが 80 dB と 90 dB の場合、圧迫感や頭痛、耳鳴りをもとに判断した被験者があり、70 dB での測定結果が正確であるとし、可聴周波数上限は 16 kHz 周辺としている（図 21）。以上のことから、Muraoka らは、20 kHz の帯域は、プロ用オーディオシステムの帯域として充分と結論している[21]。

こうした検討結果から、一般的に人間の可聴帯域は 20 kHz を大きく超えることはないと考えられ、CD や DAT、デジタル放送など通常のデジタルオーディオの音声帯域幅は 20 kHz とされてきた。



NOTE: Listener operates switching by himself.

図 18. Muraoka らの実験システム [21]。

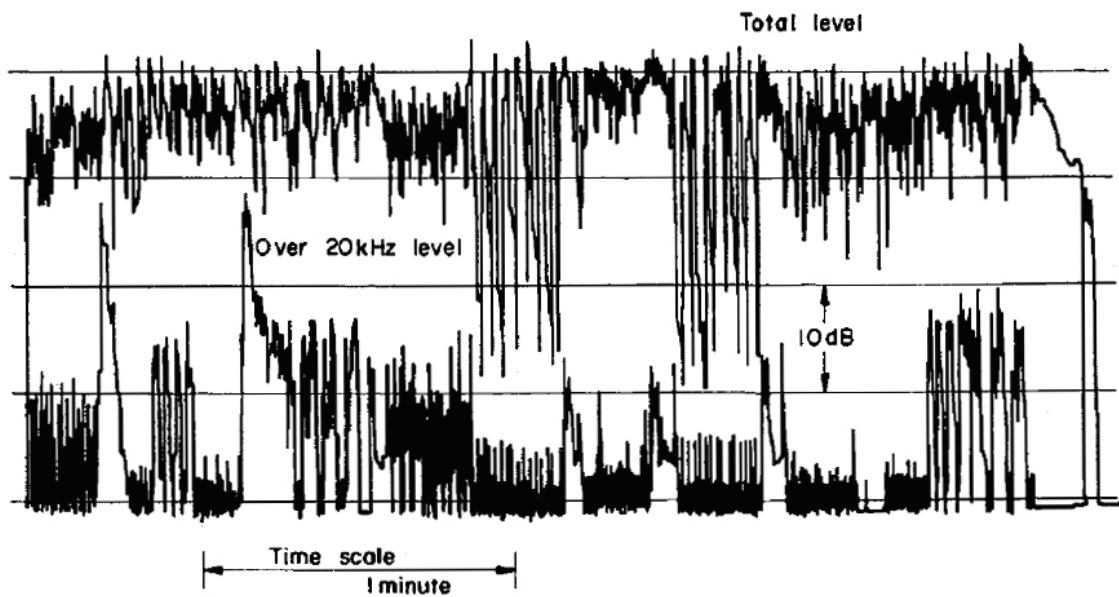


図 19. 評価音源のレベル変化 (Muraoka ら [21])。

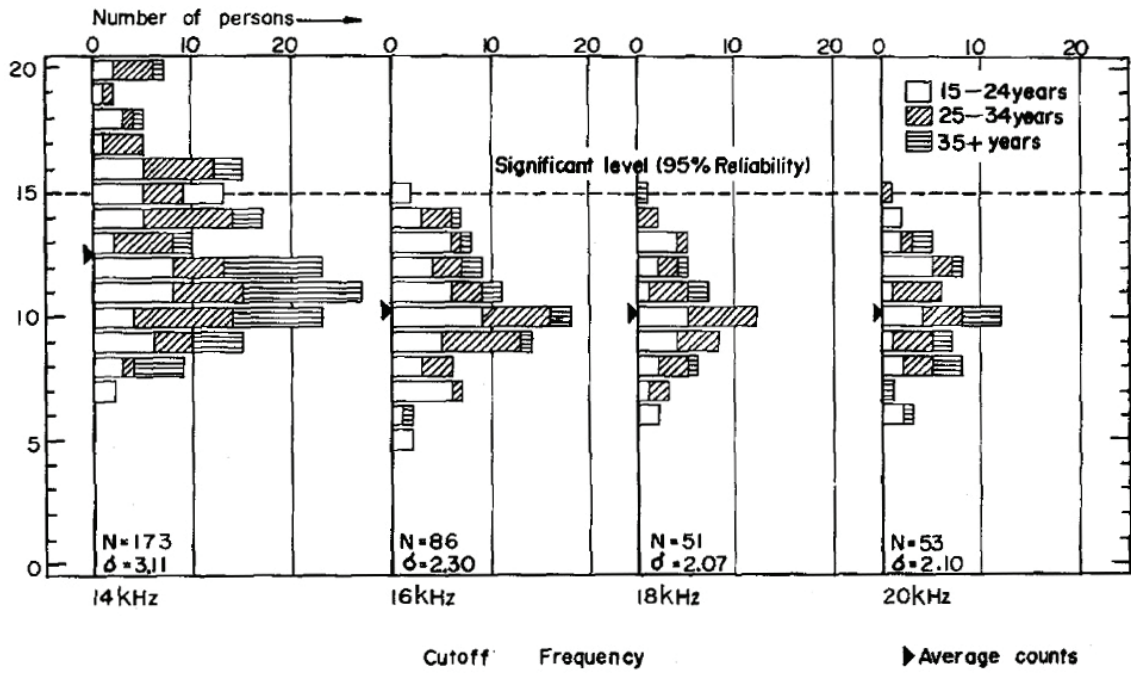


図 20. 楽音を使った弁別実験の結果 (Muraoka ら[21])。

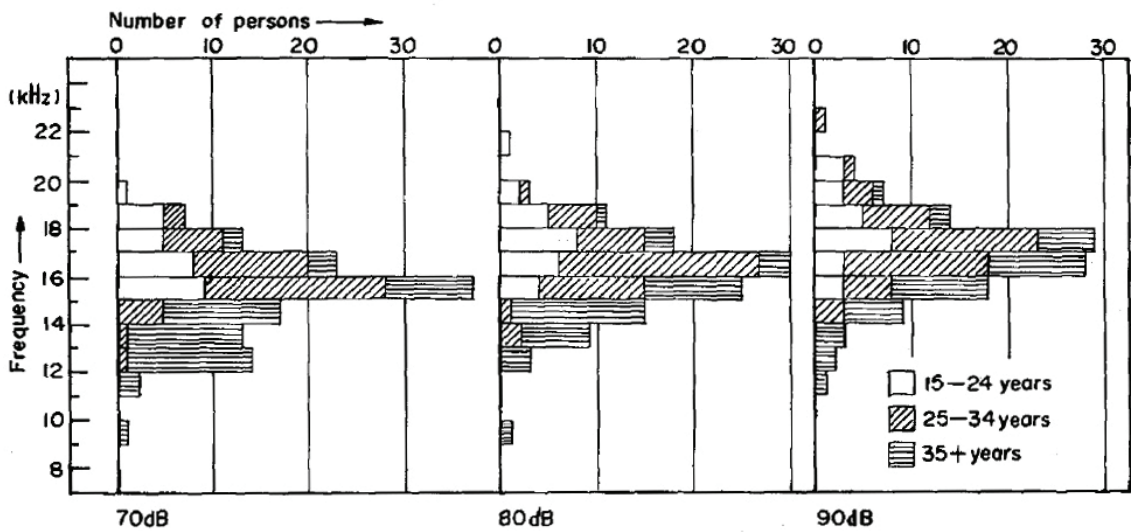


図 21. 純音を使った可聴周波数上限の実験結果 (Muraoka ら[21])。

ところが、マイクロホンやレコーダの帯域が広がり、技術的に超高域が収録可能になってくると、楽音や非定常音の 20 kHz を超える超高域成分が影響を与えるという報告が出てきた。Oohashi らは、広帯域なガムラン音楽 (図 22) を用い、22 kHz ないし 26 kHz を超える帯域の有無が、聴感 (表 3), 脳波 (図 23), 脳血流に影響を与えることを報告している [22], [23]。

松嶋らも、20 kHz 以上の超高域を豊富に含む森林の環境音を用いて実験を行い、超高域の有無が脳波に影響を与えると報告している [24]。Yoshikawa らは、帯域が 20 kHz と 40 kHz のパルス列の間隔を変えて実験を行い、帯域が広い方が聴覚上の時間分解能が高いことを示した [25]。

しかし、こうした超高域の影響は観測されなかったという報告も依然として成されている [26] - [28]。

超高域成分が聴こえに影響を与えるとすると、その要因は、2 つに分けて考える必要がある。ひとつは、超高域成分を聴取することにより聴感へ影響が出る場合、もうひとつは、超高域成分を再生したことにより、可聴域の音が変化し、これが弁別の手がかりになっている場合である。後者の要因として、実験システムの非線形歪が指摘されている [29]。非線形歪が発生する可能性があるのは、

- a. 録音系 (マイクロホンから記録装置まで)
- b. 再生系 (記録装置からスピーカまで)
- c. 音波の伝搬系 (スピーカから耳の道入り口まで)

である。このうち、a. は、録音時に確定した歪であり、超高域再生の有無に関係しないので問題ない。c. では再生音圧が極めて大きな場合、発生する可能性があるが [30], [31], 通常、楽音を再生する音圧では無視できる。問題になるのは b. であり、蘆原や Griesinger らは、スピーカやアンプで発生した非線形歪の可聴域成分が超高域再生の有無を弁別する手がかりになることを示唆している [32], [33]。したがって、こうした超高域成分の聴感への影響を検討する場合、評価音再生時の非線形歪を定量的に評価することが重要である。しかし、これまで、楽音の超高域成分が聴感に影響を与えたとした報告は、この非線形歪の大きさを定量的に明らかにしていない。

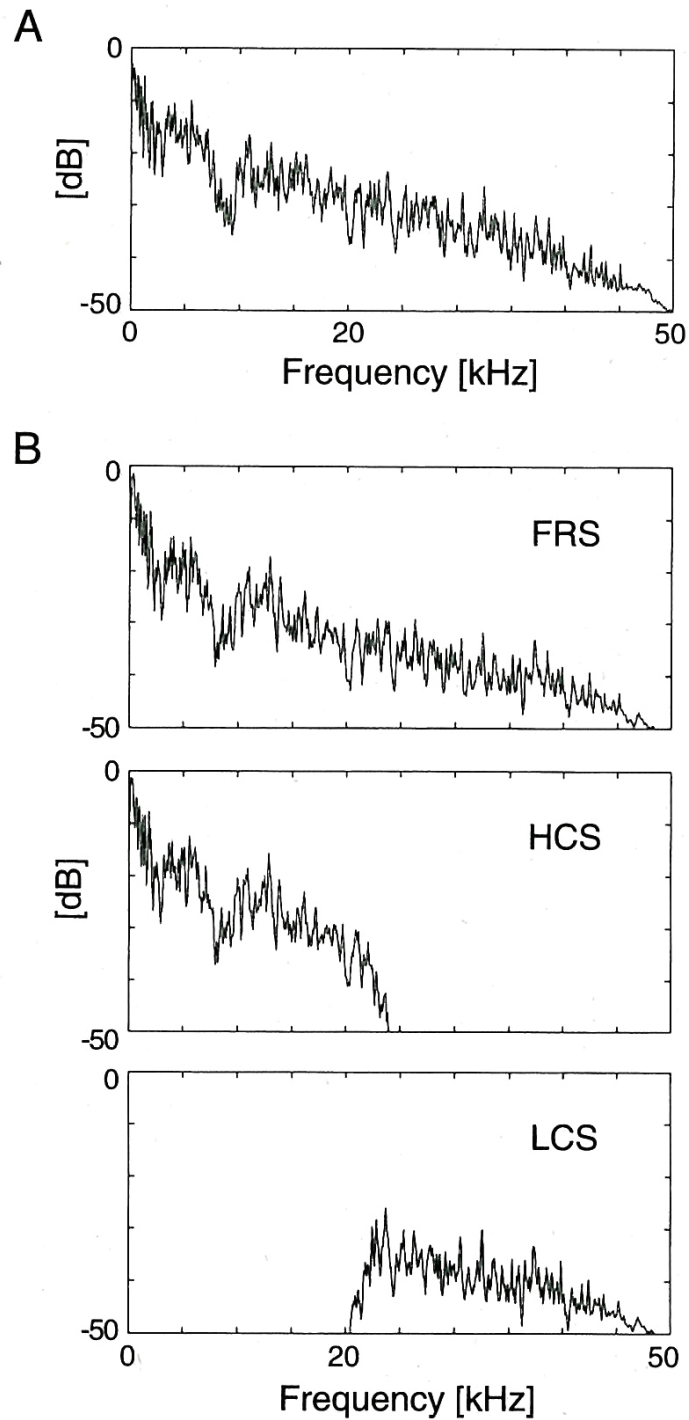


図 22. Oohashi らが実験に用いた音源のパワースペクトル [23]。音源全体 200 s の平均。元の音源 (図 A) を、アナログフィルタでローカットサウンド (図 B の LCS) とハイカットサウンド (図 B の HCS) に分割。この両帯域を合わせたのがフルレンジサウンド (図 B の FRS)。

表 3. FRS と HCS (図 22 参照) の音質に対する主観評価結果 (Oohashi ら [23])。 “Soft vs. hard” から “Rich in nuance vs. lacking in nuance” までの 5 項目で、 FRS と HCS に有意差がある。

Element for Evaluation*	Significance Level (P)†	q ‡
Soft vs. hard	<0.01	5.33
Reverberant type vs. percussive type	<0.01	5.01
Instruments in balance vs. instruments in imbalance	<0.01	4.57
Comfortable to ears vs. uncomfortable to ears	<0.01	4.44
Rich in nuance vs. lacking in nuance	<0.05	3.63
Lower tone dominant vs. higher tone dominant	—	2.25
Thick vs. thin	—	1.70
Light vs. heavy	—	1.13
Like vs. dislike	—	1.12
Finely textured vs. roughly textured	—	0.14

* Approximate English equivalents for pairs of Japanese words used for evaluation of sound quality. The subjects rated sound quality on a scale of 5 (the former) to 1 (the latter). † P indicates the significance level by which FRS showed a higher score (more favorable) than HCS for each element. ‡ q indicates each Student's interval.

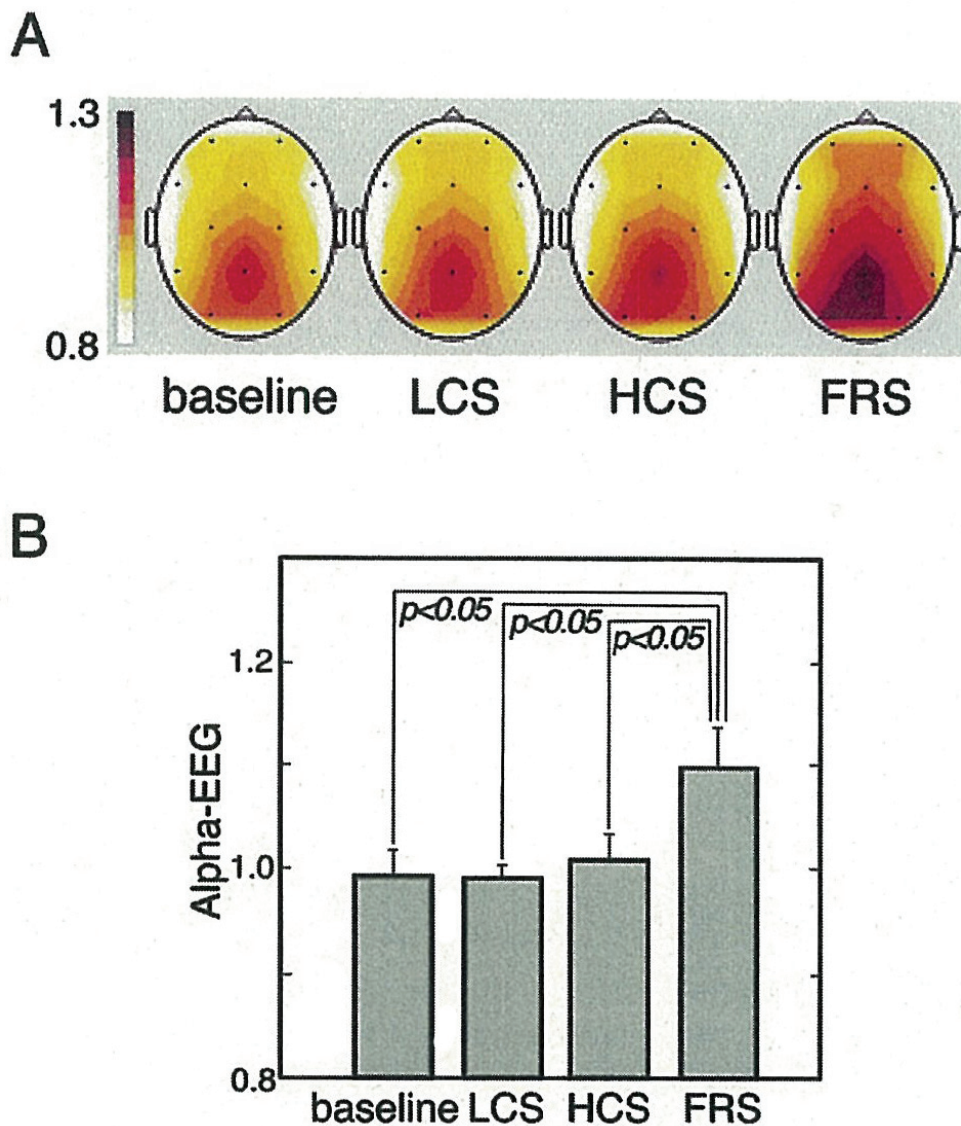


図 23. 安静状態 (baseline) と各聴取条件 (LCS, HCS, FRS, 図 22参照) における正規化した脳波 α 波 (Alpha-EEG : alpha-electroencephalogram) 電位。A は、音源の後半 100~200 s を聴取時の脳電気活動図。17 人の被験者の平均。B は 17 人の被験者の後頭部 α 波の平均と標準偏差。FRS は他の条件と比べ有意に α 波が強められている (Oohashi ら[23])。

本研究では、実験装置の発生する非線形歪を聴取位置において測定することにより明確にした。また、過去の報告では、SACD や DVD-Audio など市販の音楽ソフトを評価音源として用いたケースがあるが、この場合、収録条件が判らず、超高域成分の量も管理できない。そのため、本研究では、独自に広帯域録音を行って評価音源を作成し、実験に用いた。

本章では、3.2節で実験システムの詳細、3.3節で多種音源・多被験者による実験 1 について[28]、3.4節では、音源の提示時間を考慮した実験 2 について述べ[34]、3.5節で本章をまとめる。

3.2 実験システム

3.2.4 実験システムの構成

本実験システムを図 24 に示す。本実験では、可聴域を 21 kHz 以下、超高域を 21 kHz 以上の帯域として、広帯域音源をこの 2 帯域に完全に分離し、この 2 帯域それぞれに対応して 2 系統の D/A 変換器、アンプ、スピーカ、電源を用い、可聴域側と超高域側を完全に分離した系で再生した。これにより、同一の系で再生するよりも、それぞれの系に入力される音響エネルギーが小さくなり、歪みを小さくすることができる。

帯域の分割は、図 25 に示した 1024 タップ直線位相 FIR ローパス (LP) / ハイパス (HP) デジタルフィルタを用いた。それぞれのフィルタは、21 kHz を中心に 1 kHz のガードバンドをもたせ、阻止量は 100 dB 以上とした。このように、各々の帯域から漏洩を完全に排除している。このフィルタを、演算処理により元の広帯域音源の標本化データに施し、分割された 2 帯域の音源データは、デジタルオーディオワークステーション (Digital Audio Workstation : DAW) から同時に再生される。

超高域をカットして提示する場合は、超高域側 D/A 変換器の出力に挿入したミュートユニット (Mute Unit) により超高域側のみ信号を遮断する。このミュートユニットは、CdS 光導電セルと LED のアナログ・フォトカップラ¹を用いて、信号の通過・遮断の切り替えノイズを発生しないよう特別に配慮して製作されている。以上により、超高域再生の有無が可聴域に与える影響を排除した。システムのデジタル系は、標本化周波数 192 kHz、量子化ビット数 24 bit で動作させた。

¹ CdS 光導電セルは、光によって電気抵抗が変化するデバイス。メカニカルリレーでは、切り替え時に信号が不連続になる為、切り替えノイズが発生するが、光導電セルの抵抗変化に 100 ms～数 100 ms 程度の比較的大きな時定数を持ち、滑らかに信号が切り替わるため、ノイズの発生を避けることができる。

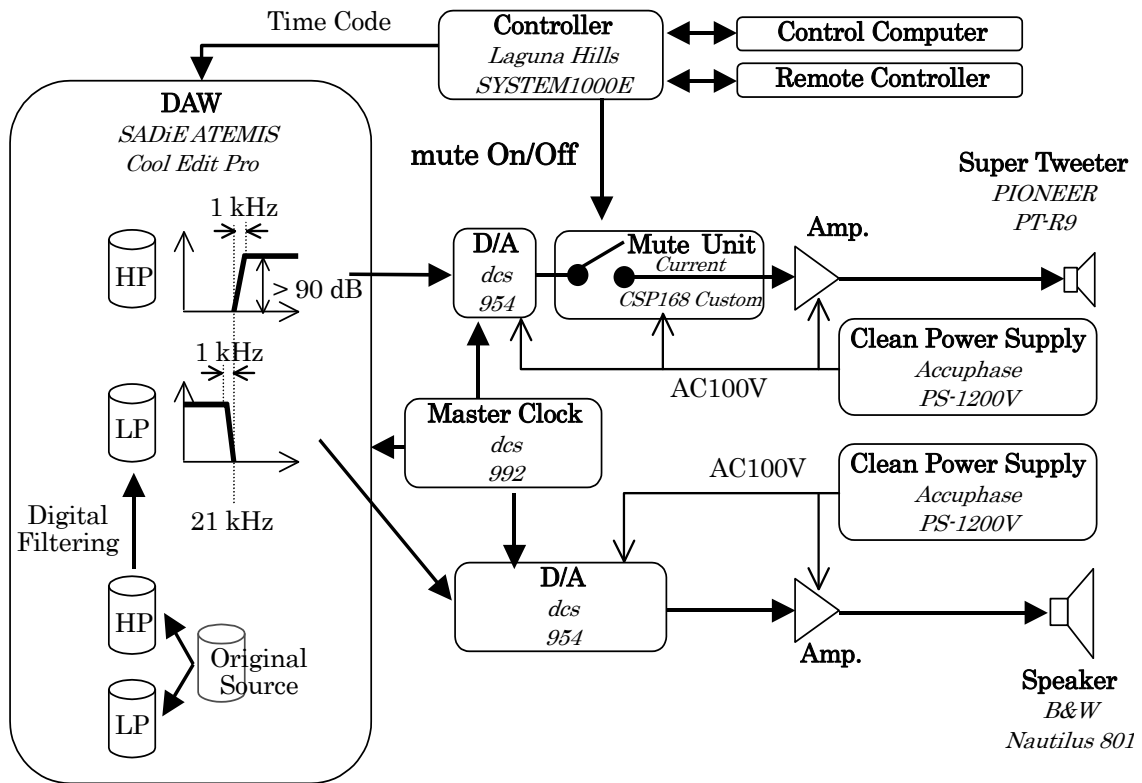


図 24. 実験システムのダイアグラム。実験システムは、超高域再生が可聴域に影響を与えないように、可聴域（21 kHz 以下）と超高域（21 kHz 以上）が独立したシステムを持つ。

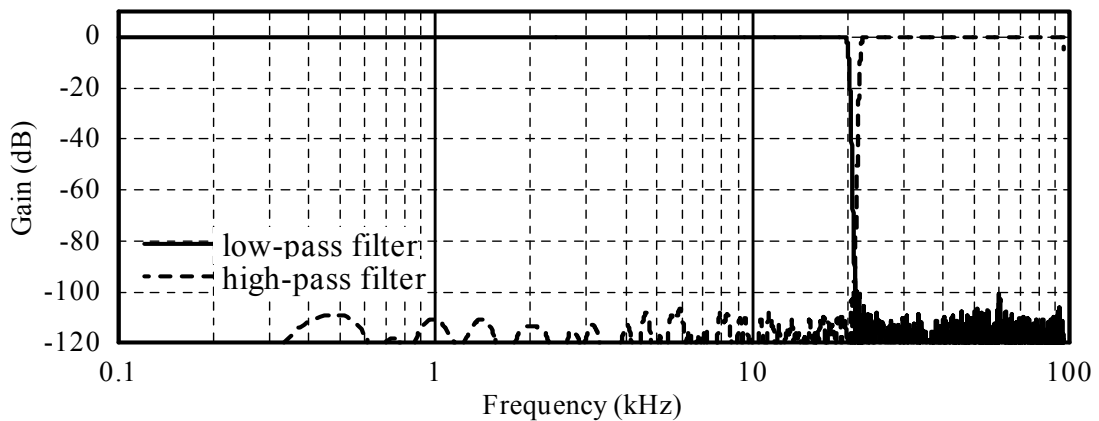


図 25. 帯域分割デジタルフィルタの特性。広帯域音源は、このフィルタによって可聴域（21kHz 以下）と超高域（21 kHz 以上）に完全に分離される。

3.2.5 実験システムの特徴

実験は ITU-R BS1116 [35]に規定された Reference listening room に完全準拠した音響評価室で行った。被験者とスピーカの位置，聴取状態を図 26 に示す。スピーカは，2チャンネルステレオ再生のセッティングとして，被験者との距離は 3 m とした。

超高域再生用のスーパーツイータ (Super tweeter) は，可聴域再生用のスピーカの振動が直に伝わらないように専用のスタンドを設けて，可聴域再生用スピーカのすぐ脇に設置した。スーパーツイータは，複数の機種 of 非線形歪を測定し，歪の小さい機種を選択した。無響室で実測したスーパーツイータの指向特性を図 27 に示す。80 kHz における水平方向の半値幅は 30 度であるのに対し，垂直方向の半値幅は非常に小さい。このため，図 26 に示したように，座面の高さが調節可能な椅子を用い，被験者の耳の高さをスーパーツイータの高さに合わせて実験を行った。

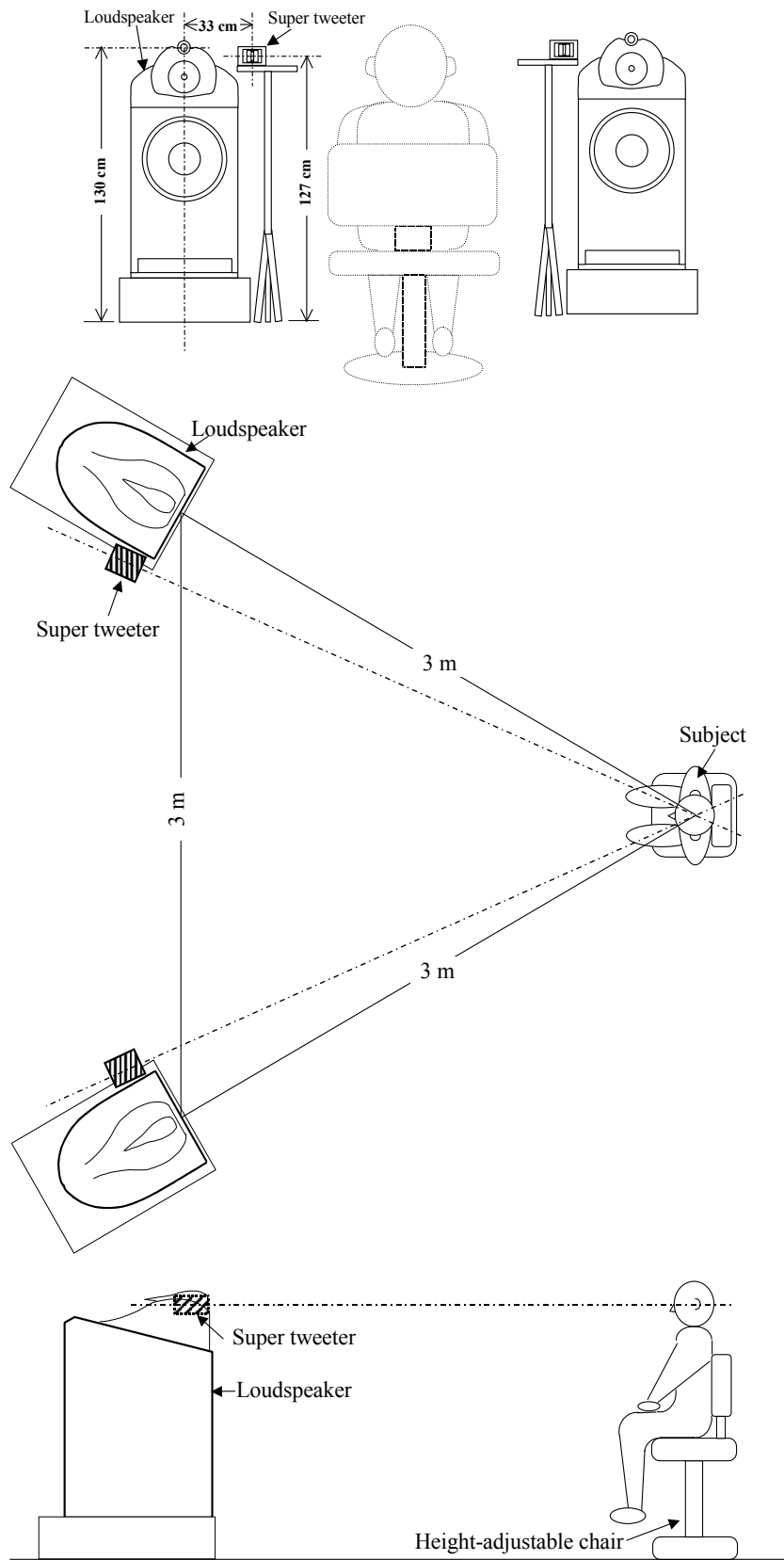


図 26. 被験者とスピーカの位置，聴取状態。

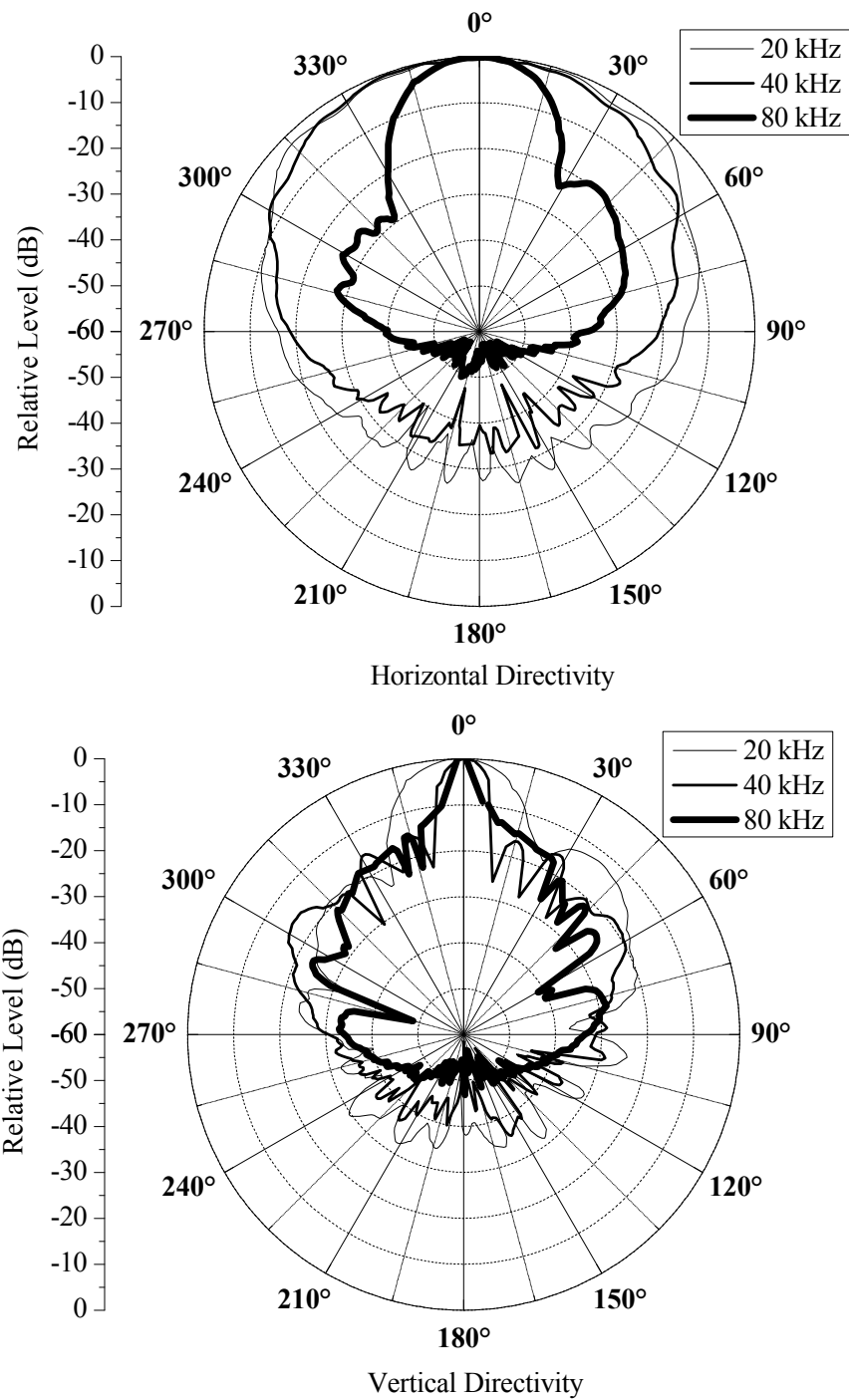


図 27. 超高域成分（21 kHz 以上）の再生に用いたスーパーツイータの指向特性。無響室で実測した水平方向の指向特性（上）と垂直方向の指向特性（下）を示す。

実験システムの聴取位置における総合周波数特性を図 28 に示す。これは、ホワイトノイズのデータを DAW から再生し、聴取位置においた測定用マイクで收音し、これを FFT アナライザに入力して測定したものである。実験システムの再生周波数上限は 70 kHz (−10 dB, 1 kHz 基準) である。この特性は前述の帯域分割フィルタ、DAW、D/A 変換器、スピーカ、スピーカから聴取位置までの音波伝搬の超過減衰を含むトータルの特性である。

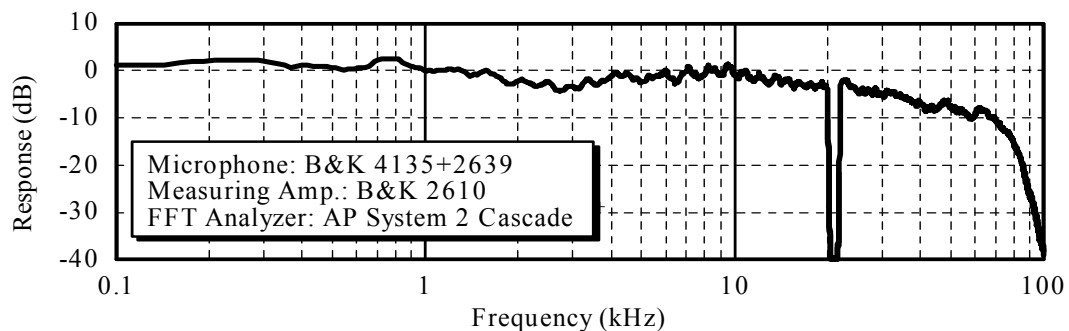


図 28. 実験システムの総合周波数特性。

さらに、広帯域評価音源を再生したときの、可聴帯域内に生じる非線形歪も測定した。被験者が超高域成分再生の有無を弁別できた場合、超高域成分の有無を弁別したのか、或いは、超高域再生によって生じた非線形歪の可聴域成分を弁別したのかを検討する上で重要な測定である。測定では、超高域側の再生システムを無響室内に設置し、実際に実験に用いる音源を再生する。このとき、スーパーツイータから再生される非線形歪の可聴域成分を測定する。測定には、

マイクロホン：B&K 社 4149+2669

マイクロホンアンプ：B&K 社 2636 と Current 社 CSP-353

A/D 変換器：DCS 社 904

を用いた。測定位置は、聴取位置が実際の状況に即しているが、もともと、非線形歪の小さいスーパーツイータを選別しているため、発生する歪は極めて小さく、主に測定用マイクロホンの固有雑音にマスクされてしまい、測定が困難である。そこで、マイクロホンをスーパーツイータから 19 cm の位置に近づけ

て測定を行い，距離減衰を考慮して聴取位置での音圧レベルに換算した。スーパーツイータから評価音源の超高域側のみ再生し，これをマイクロホンで收音，DAWに記録，このデータに図 25 の LP フィルタを施して非線形歪の可聴域成分のみを取り出し，聴取位置での音圧レベルに換算した。評価実験に用いた音源のうち，超高域のエネルギーが最も大きな音源について，測定結果を図 29 に示す。この図では，非線形歪の可聴域成分の音圧レベル ($0 \text{ dB}_{\text{SPL}} = 20 \times 10^{-6} \text{ Pa}$) と，音源の可聴域と超高域それぞれについて，聴取位置での音圧レベルの計算値を時間とともにプロットしてある。超高域再生によって可聴帯域内に生じた非線形歪の音圧レベルが最大 $20 \sim 25 \text{ dB}_{\text{SPL}}$ であるのに対し，可聴帯域の音源の音圧レベルが $70 \text{ dB}_{\text{SPL}}$ 前後あり，マスキング効果を考慮すれば，非線形歪が弁別の手がかりになるとは考えられない。

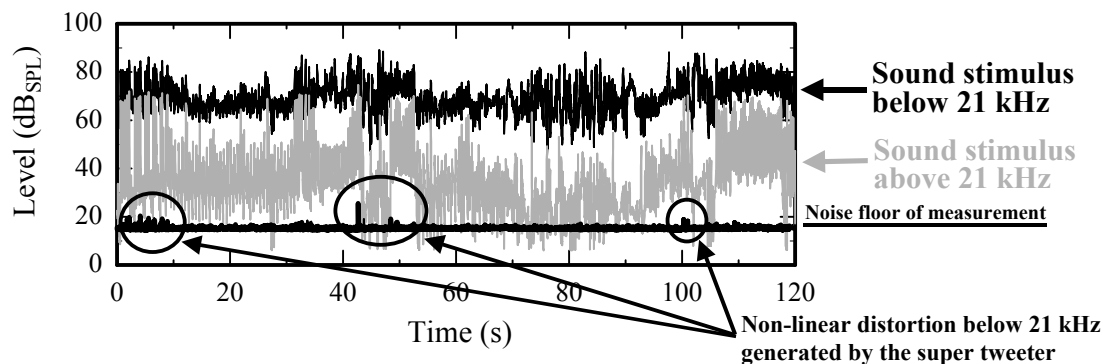


図 29. 実験システムの非線形歪。実験に使用した音源の超高域(21 kHz)と可聴域 (21 kHz 以下) の音圧レベルと，可聴域内の非線形歪の音圧レベルを示す。この非線形歪は，超高域成分を再生時に，スーパーツイータから発生した歪の可聴域成分である。

3.3 実験 1: 多種音源・多数の被験者による評価

3.3.1 実験 1 の目的と特徴

前述のように、楽音の超高域の有無が聴感に与える影響については、これまでも検討されてきているが、有意差の有無は報告によりまちまちであり、音源や被験者についての実験データも多いとは言えない。そこで、本研究では、超高域成分による音質差異が存在する可能性があるかないか、また、その影響は、音源の種類や被験者によって異なるのかどうかを確認する目的で、様々な種類の音源を用意し、比較的多数の被験者により基礎的な主観評価実験を行った。

3.3.2 被験者

聴力健常な男性 30 人、女性 6 人の計 36 人とした。専門・職業は、音響関係の研究者と音声制作技術者 33 人、学生 2 人、評価音源の演奏家 1 人である。年齢構成は、10 代が 3 人、20 代が 12 人、30 代が 16 人、40 代が 3 人、50 代が 2 人である。超高域の影響を報告している先行研究[22], [23]では、可聴域をカットして超高域のみを提示しても、聞こえや生体反応 (図 23) に影響が発現しないとしているため、予め聴力検査によって被験者を超高域が聴取できる者に限定しなかった。

3.3.3 音源

音源の一覧を表 4 に示す。本実験の目的から、様々な音楽ジャンル、楽器からなる音源を用意した。全 20 種のうち、音源 11~13 以外は、独自収録した音源を用いた。收音のためのマイクロホンは、周波数帯域の広さだけならば、帯域が 100 kHz におよぶ測定用マイクロホンも存在するものの、固有雑音レベルが大きく、楽音の評価実験に用いる音源収録には不向きと考えた。そこで、40~50kHz までフラットな周波数特性を有する音楽録音用の全指向性マイクを用い、必要に応じて単一指向性の補助マイクも加え、FFT アナライザを用いて可聴域と超高域のバランスを管理し 2 チャンネルステレオ録音を行った。A/D 変

換は、できる限り広帯域かつ、デジタルフィルタによる帯域分割など後処理を考慮し、標本化周波数 192kHz, 24 bit 量子化を採用した。

音源 11 と 12 は、参考のため市販の SACD から選択した。音源 13 は、音源 3 の 21 kHz 以上の帯域をホワイトノイズ（聴取位置で 70 dB_{SPL}）に置き換えて作成した。

再生音圧レベルは、音源の主要なピークが 80dB_{SPL} (A 特性, Fast) 前後となるレベルを標準レベルとして、-3 から+5 dB の範囲で被験者に調節を許した。これは、被験者が、音響研究者、音声制作技術者、演奏家、学生といった様々の職種で構成され、被験者の希望する聴取レベルに差があり、これに対応したためである。結果として、標準レベル+5 dB で聴取した被験者は 1 人、標準レベル±3 dB の範囲で聴取が 9 人、標準レベルで聴取が 26 人であった。

表 4. 実験 1 で用いた音源。様々な楽器、音楽ジャンルから 20 種の音源を用意した。No. 11, 12, 13 以外はすべて独自収録した音源である。

No.	Title	Musical Instrument
1	"Satsuma-Biwa"	"Satsuma-Biwa" (Japanese traditional music)
2	Litha	Drums, Bass, Pf (Jazz piano trio)
3	Meditation	Vn, Pf
4	Romanian Folk Dances	Vn, Pf
5	Intermezzo de "Carmen"	Fl, Pf
6	Beethoven: Sym. No.9 4th Mov.	Picc
7	Bach: Suite for Vc No.2 - Prelude	Sax
8	Bach: Suite for Vc No.6 - Prelude	Sax
9	Piece en forme de Habanera	Sax, Pf
10	Partie	Sax, Pf, Perc
11	Sednalo	Bulgarian Chorus (SACD ARHS-1002)
12	TihViatar	Bulgarian Chorus (SACD ARHS-1002)
13	Meditation+White Noise	Vn, Pf, High frequency band consists of only white noise.
14	Airs Valagues	Fl, Pf
15	Tchaikovski: Sym. No.6 3rd Mov.	Full Orchestra
16	Doralice	Vo, Gt (Bossa Nova)
17	Chega de Sauadade	Vo, Gt, Pf, Perc (Bossa Nova)
18	tiny rose	Vo, Pf, Gt, Fl, Perc (Pop music)
19	butterfly	Vo, Pf, Gt, Perc (Pop music)
20	Autumn Leaves	Drums, Bass, Pf (Jazz piano trio)

(Gt: Guitar, Fl: Flute, Perc: Percussion, Picc: Piccolo, Pf: Piano, Sax: Saxophone, Vc: Cello, Vn: Violin, Vo: Vocal)

3.3.4 実験方法

R, A, B の 3 刺激を試聴し, A と B を比較して R と同じ方を選択させる duo-trio test を用いた。R は常に可聴域+超高域の全帯域が再生され, A と B のどちらか一方は超高域をミュートして可聴域のみ再生される。A と B のどちらが R と同じ全帯域再生かは, 音源毎にランダムであるが, A が R と同じ確率, および, B が R と同じ確率は等しく 0.5 になるように提示した。被験者には, リモートコントローラにより R, A, B を自由に切り替え, 任意の場所を納得がいくまで繰り返し比較試聴させた。音源の提示順序は試行毎にランダムとした。実験は, 図 26 のように被験者 1 人ずつ行った。実験の 1 セッションを 30 分間として, 各の被験者は, セッション間に 10~15 分の休憩をはさみ, 1 日に 3~4 セッションですべての実験を終了した。

3.3.5 実験結果

音源毎の正答率を図 30 に, 被験者毎の正答率を図 31 に示す。この実験では, 20 種の音源を 32 人の被験者が 1 回評価し, 4 人の被験者 (被験者 No. 1~4) が 2 回評価した。よって, 図 30 の各音源の評価は, 40 回繰り返されたことになる。また, 図 31 の被験者は, No. 1~4 は 40 試行, これ以外の被験者は 20 試行の評価を行ったことになる。

被験者は各々の評価で, A か B の 2 者択一で, 聞こえが R と同じと感じられ方を選択しなければならない。もし, 被験者にとって A と B に聞こえの差がなければ, 各々の試行は, チャンスレベル 0.5 のベルヌーイの試行と考えられ, 正答率は 2 項分布に従う。正答率の確率密度関数は, 次のうように 2 項分布から与えられる[36]。

$$P(x) = \frac{N!}{x!(N-x)!} p_x^x (1-p_x)^{N-x} \quad (18)$$

ここで, x は正答数, N は試行数, p_x は正答となる確率(=0.5)である。もし, 超高域の有無が聞こえに対して影響を与えなければ, すなわち, A と B の聞こえの差がなければ, 正答率は 0.5 に収斂する。そこで, 帰無仮説 H_0 : 正答率 =

0.5, 対立仮説 H_1 : 正答率 $\neq 0.5$ として, 2 項検定を行う。2 項検定の危険率 (P 値) は, 次のように求められる。

$$p = \begin{cases} 2 \sum_{n=0}^x P(n) & (x < \frac{N}{2}) \\ 1 & (x = \frac{N}{2}) \\ 2 \sum_{n=x}^N P(n) & (x > \frac{N}{2}) \end{cases} \quad (19)$$

音源毎の正答率 (図 30) の P 値を見ると, 特に有意な音源はないが, No. 1, 10, 15 の正答率が高い。一方, No. 2, 3, 13 は, ほぼチャンスレベルである。これらの音源について, 聴取位置での帯域毎の音圧レベルと, 代表的な周波数スペクトルを計算により求めた結果を図 32 に示す。この分析には, 図 28 に示した実験装置の総合周波数特性が加味されている。音圧レベルは 50 ms 区間毎の実効値を算出し, 聴取位置での音圧に換算した。周波数スペクトルは, 500 ms の区間を 4096 サンプル (21.3 ms) のブラックマン窓 (表 1) を用いて分割し, 得られたスペクトル群を平均化して示した。

音源 13 は, 音源 3 の超広域をホワイトノイズに置き換えて作成した音源で, 超高域の音圧が 70 dB_{SPL} もあるにもかかわらず, 弁別されなかった。正答率が比較的高かった音源は, 超高域のレベルが高く, かつ, 超高域成分が非定常という特徴がみられる。

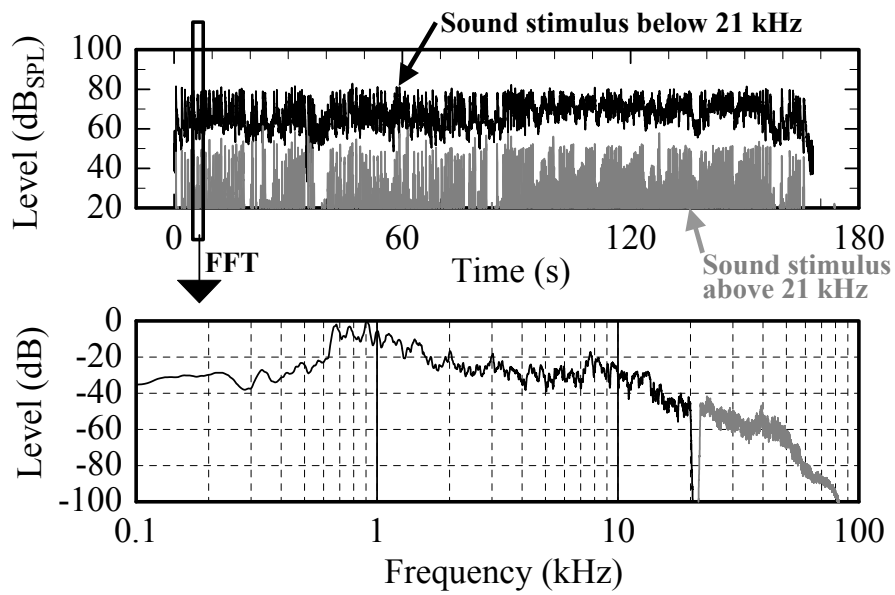
被験者毎の正答率 (図 31) から, 被験者 No. 30 (17 歳女性) は, 正答率 0.75 を達成し, 危険率 $p=0.0417$ は, 有意水準 5% の検定で有意となる値である。そこで, より信頼性の高い結論を得るため, この被験者に対して追試を行った。

Stimulus number	Correct response rate											<i>p</i>
	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	
1												0.0807
2												0.8746
3												1.0000
4												0.6358
5												0.8746
6												0.6358
7												0.8746
8												0.2682
9												0.4296
10												0.1539
11												0.6358
12												0.8746
13												0.6358
14												0.6358
15												0.0807
16												0.8746
17												0.6358
18												0.8746
19												0.8746
20												1.0000

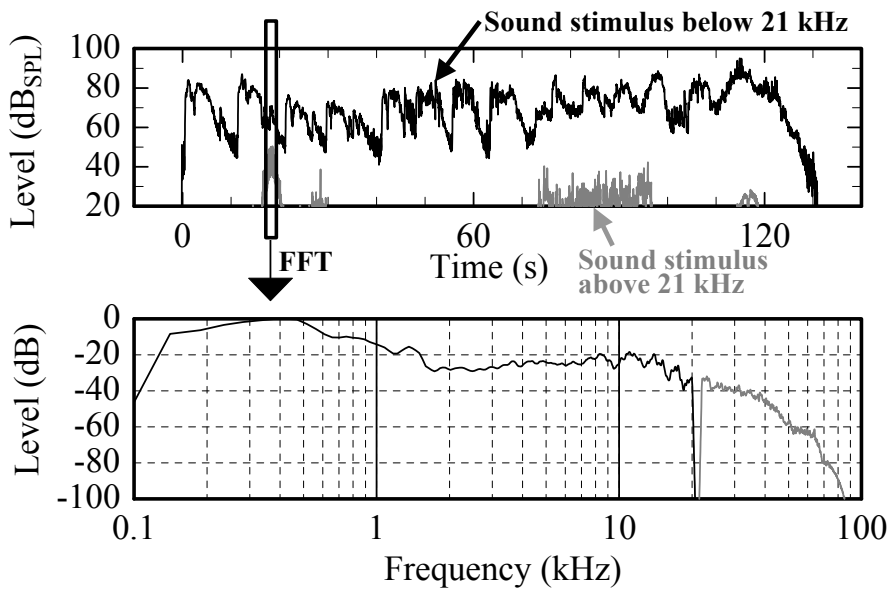
図 30. 実験 1 の音源毎の実験結果。音源毎の正答率と P 値を示す。

Subject No.	Correct response rate											<i>p</i>
	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	
1												0.6358
2												0.6358
3												0.8746
4												0.0807
5												0.8238
6												0.5034
7												0.8238
8												0.8238
9												1.0000
10												0.8238
11												1.0000
12												0.8238
13												0.2632
14												0.5034
15												0.8238
16												0.1153
17												0.8238
18												0.8238
19												0.5034
20												0.5034
21												0.5034
22												0.2632
23												0.2632
24												0.5034
25												0.8238
26												0.8238
27												0.2632
28												1.0000
29												1.0000
30												0.0414
31												0.8238
32												1.0000
33												0.5034
34												0.8238
35												0.5034
36												0.2632

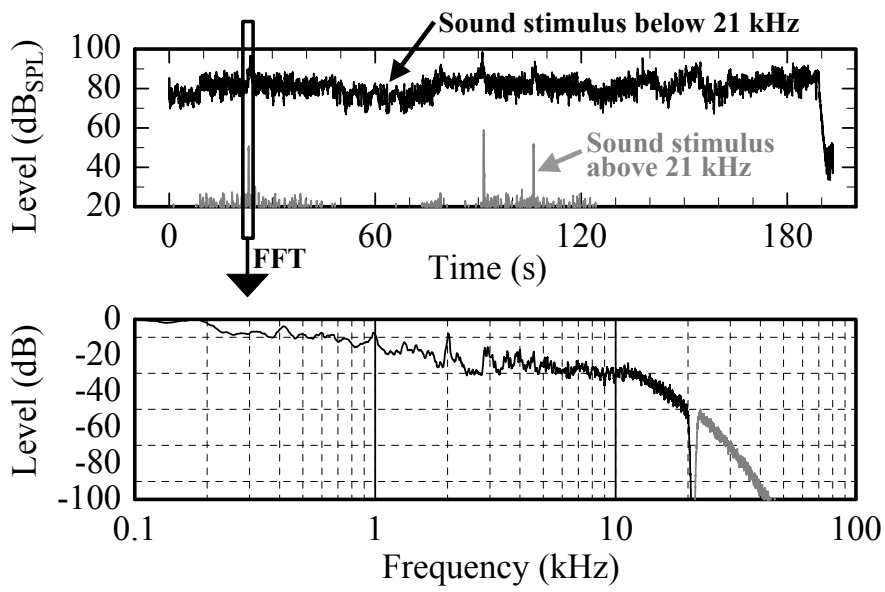
図 31. 実験1の被験者毎の実験結果。被験者毎の正答率とP値を示している。



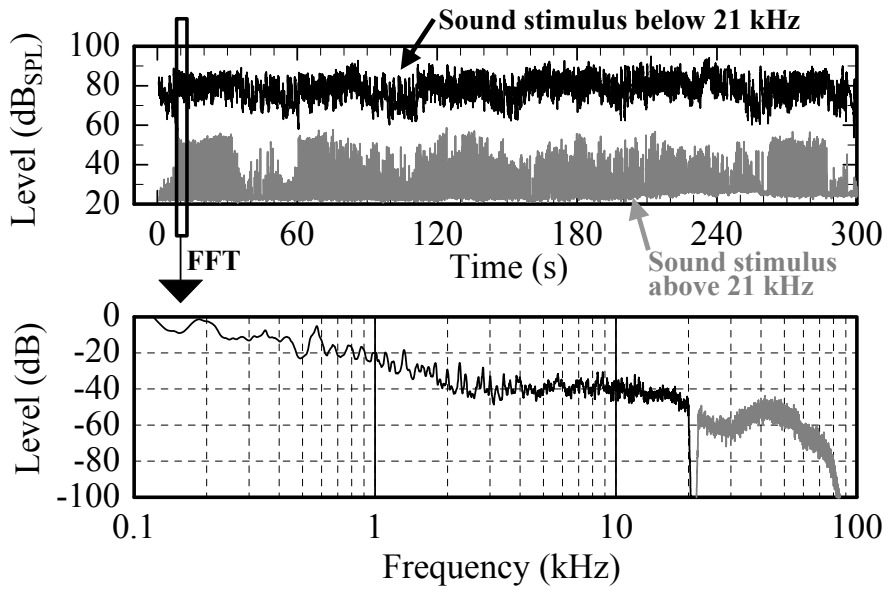
☒ 32 (a) No. 1 (Satsuma-Biwa)



☒ 32 (b) No. 10 (Sax, Pf, Perc)



☒ 32 (c) No. 15 (Full orchestra)



☒ 32 (d) No. 2 (Jazz piano trio)

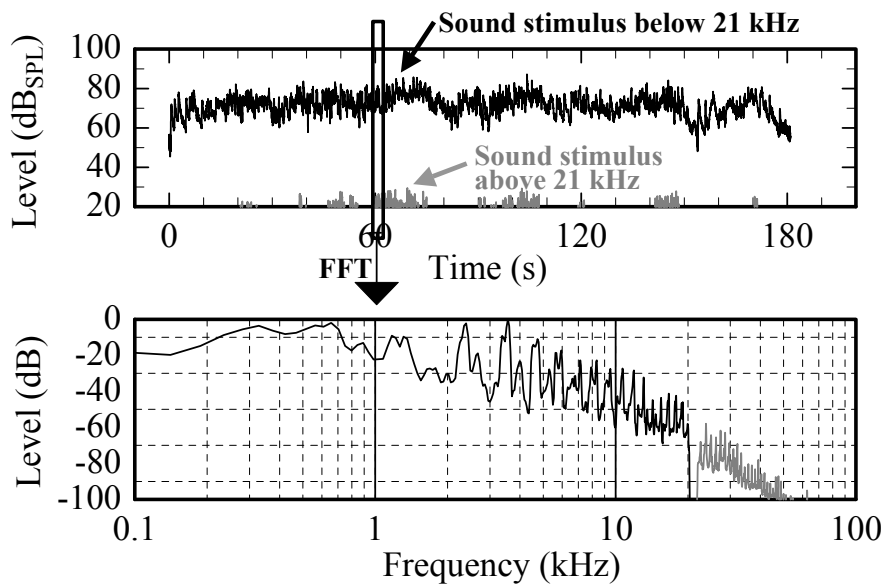


図 32 (e) No. 3 (Vn, Pf)

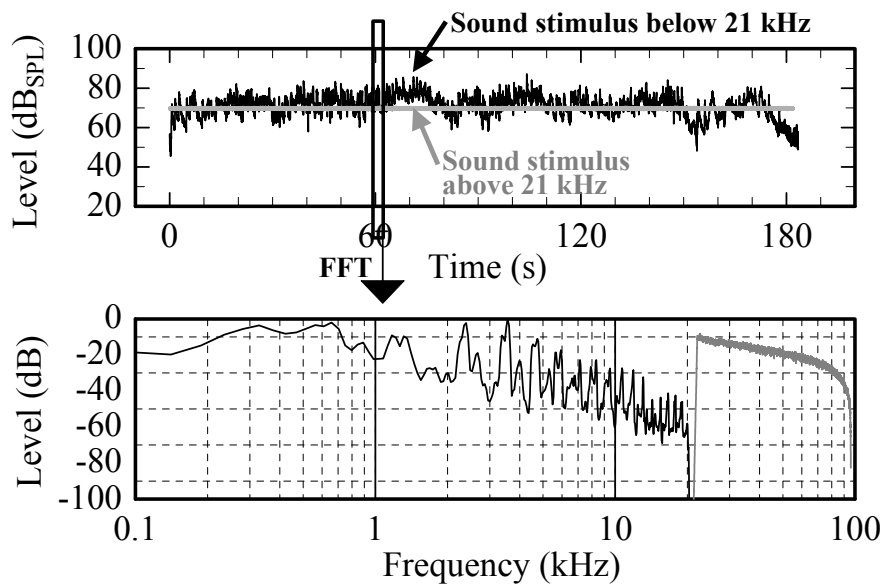


図 32 (f) No.13 (Vn, Pf, High frequency band consists of only white noise)

図 32. 実験 1 で用いた音源の分析結果。各音源の上のグラフは、可聴域 (21 kHz 以下) と超高域 (21 kHz 以上) の音圧レベルの時間変化を示す。下のグラフは、上のグラフの方形で囲んだ部分のスペクトルを示す。

3.3.6 実験1の追試

3.3.6.1 被験者と音源

被験者は、前述の実験で最も高い正答率を得た No. 30 である。

音源は、表 4 から No. 1, 2, 10, 15, 16 の 5 種を選択した。音源 No. 1, 10, 15 は、前述の実験 (図 30) で高い正答率を得た音源である。音源 No. 2 と 16 は、高い正答率は得ていないが、No. 2 は、図 32 (d) のとおり非定常かつ高いレベルの超高域成分を有していること、また、No. 16 は、弁別しやすいという被験者内観を得たことから選択した。また、参考として高い正答率を得た音源 No. 1 の超高域を、レベル一定 (聴取位置で 70 dB SPL) のホワイトノイズに置き換えた音源 No. 21 を新たに作成して用いた。

3.3.6.2 実験結果

すべての音源を、それぞれ 20 回評価した結果を図 33 に示す。いずれの音源に対しても、P 値が高く、有意な結果は得られていない。したがって、被験者 No. 30 は、この 6 種の音源に関して、超高域の有無を弁別できていないという結果を得た。

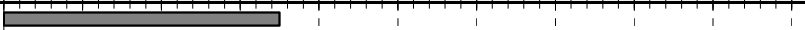
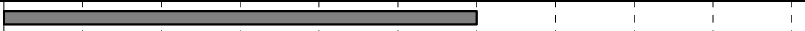
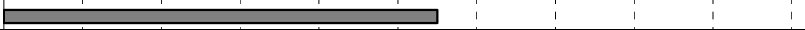



Stimulus number	Correct response rate											<i>p</i>
	0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	
1												0.2632
2												0.5034
10												0.8238
15												0.8238
16												0.5034
21												0.2632

図 33. 実験 1 の追試結果。被験者 30 の正答率と P 値を音源毎に示している。

3.3.7 実験1のまとめ

20種の音源を用い、36人の被験者で超高域成分の有無について主観弁別実験を行った。音源別には、有意な結果は得られなかったが、被験者別では、1人の被験者が正答率0.75を得た。そこで、音源を選別し、この被験者の追試を行った。追試では、6種の音源を20回繰り返して評価した結果、どの音源に対しても有意な弁別結果は得られなかった。音源に関しては、超高域のレベルが高く、かつ、非定常な音源で正答率が高い傾向がみられたが、信頼性の高い結果を得るためには、被験者、音源を精査し、同一被験者で同一音源を繰り返し評価する必要があると考えられる。

3.4 実験 2: 音源提示時間と弁別の関係

3.4.1 実験 2 の概要

実験 1 では、超高域の有無について有意な結果は得られなかった。しかし、この実験から、

- ・ 超高域のレベルが高く、かつ、非定常な音源が、より聴感への影響を与えると推測される。
- ・ 被験者、音源を精査し、同一被験者で同一音源を繰り返し評価する必要がある。

また、超高域の影響には、音源の提示時間が関係するという報告がある[22]、[23]。そこで本実験では、下記のように、新たに開発した音楽用超広帯域マイクロホンを用いて、超高域を豊富に含む評価音源を作成するとともに、超高域の影響を報告した先行研究を考慮して実験方法を変更した。

3.4.2 超広帯域マイクロホンによる音源収録

実験 1 では、可聴域の S/N や音質を考慮して、従来の音楽録音用のマイクロホンを用いた。これらのマイクロホンの周波数帯域は 40~50 kHz 程度であるが、この値はマイクロホンの主軸方法から音波が到来する場合の特性であり、これ以外の方向から到来する音に対しての収録帯域はさらに狭くなる。一方、評価音源の収録に用いたデジタル録音システムは、標本化周波数 192 kHz で動作するため、理論的な帯域上限は 96 kHz である。これに比べ、従来の音楽録音用マイクロホンの収録帯域は不十分である。

一般に全指向性コンデンサ型マイクロホンの帯域上限は、音波の回折効果による感度上昇のピークが観測される周波数 f_d と、振動膜の共振周波数 f_0 とで決まり、通常はこれ以下の平坦な帯域が使用可能とされる。原理的に f_d は、マイクロホンの振動膜の直径と同じ波長の音波の周波数となる。100 kHz の帯域上限を得ようと単純に f_d を 100 kHz とすると、振動膜の直径は 3.4 mm 程度と極めて小さくしなければならない。マイクロホンの感度は、ほぼ振動膜の有効面

積に比例するので、このように振動膜を小さくすると著しく感度が低下する。計測用マイクロホンには 100 kHz の收音帯域をもつものもあるが、振動膜を小さくし、感度を犠牲にして広帯域化しているため、音楽録音用マイクロホンと比べて著しく固有雑音が大きく、可聴域の S/N が低くなる。こうしたマイクロホンは、本実験のように微妙な音質の違いを評価するための音源収録には不適當と考えられる。

そこで我々は、音楽録音用の広帯域マイクロホンを新たに開発した[37], [38]。図 34 のように、マイクロホンの回折効果による感度上昇は f_d の奇数倍で生じるため、本マイクロホンでは、 $3f_d$ が帯域上限の 100 kHz となるように振動膜の直径を決めた。これにより、振動膜の直径として 10 mm を確保できる。一方、 $2f_d$ では回折効果による感度上昇が生じない。そこで、振動膜の張力を調節して共振周波数 f_0 を $2f_d$ とし、回折効果による感度上昇が生じない $2f_d$ 付近の感度を振動膜の共振で補うようにした。このように、音波の回折効果と振動膜の共振を積極的に利用して超高域の感度を高めることにより、固有雑音の大きさは従来の音楽録音用マイクとほぼ同等に抑えながら、収録帯域は図 35 のように 100 kHz まで拡張することができた。開発したマイクロホンの仕様と外観をそれぞれ図 36 と表 5 に示す。

この超広帯域マイクロホンと収録帯域が 40 kHz 程度の従来の音楽録音用マイクロホンで収録したハープシコードの周波数スペクトルを図 37 に示した。この図からハープシコードは、明らかに 40 kHz 以上の帯域を有しているが、従来のマイクロホンでは、その収録帯域に制限されていることが判る。

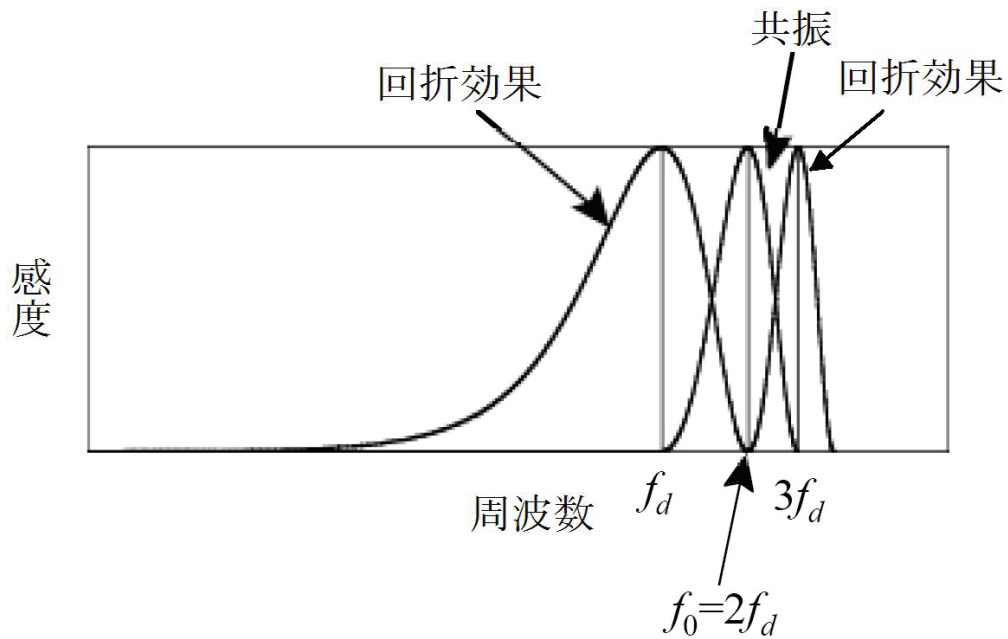


図 34. 超広帯域マイクロホンの帯域拡張方法。回折効果による f_d 、 $3f_d$ での感度上昇と、周波数 $f_0 = 2f_d$ での振動膜の共振を利用して、感度を損なうことなく帯域を拡張した。

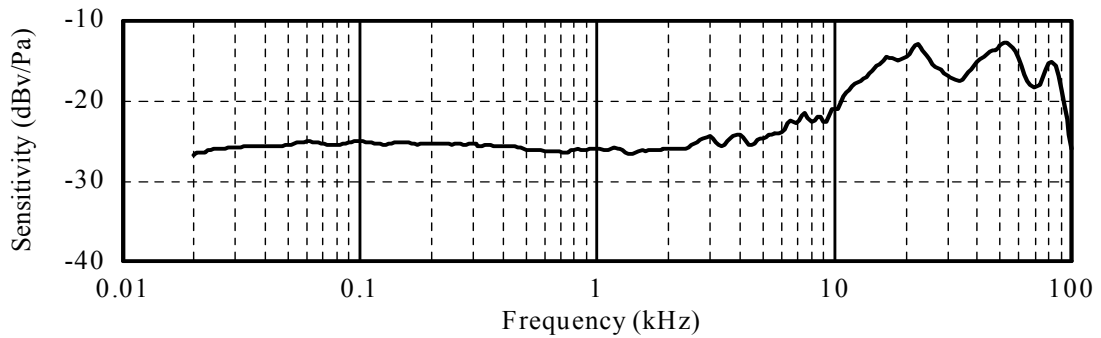


図 35. 超広帯域マイクロホンの周波数特性。



図 36. 超広帯域マイクロホンの外観。

表 5. 新たに開発した超広帯域マイクロホンの仕様。

方式	直流バイアスコンデンサー
指向性	全指向性
定格感度レベル	-26 dB (0 dB=1 V/Pa)
周波数帯域	20 Hz~100 kHz
固有雑音レベル	20 dB _{SPL} (A 特性) 以下
出力インピーダンス	150 Ω
最大入力音圧レベル	125 dB _{SPL}
ファンタム電源	直流 48 V
消費電流	6 mA 以下
重さ	150 g

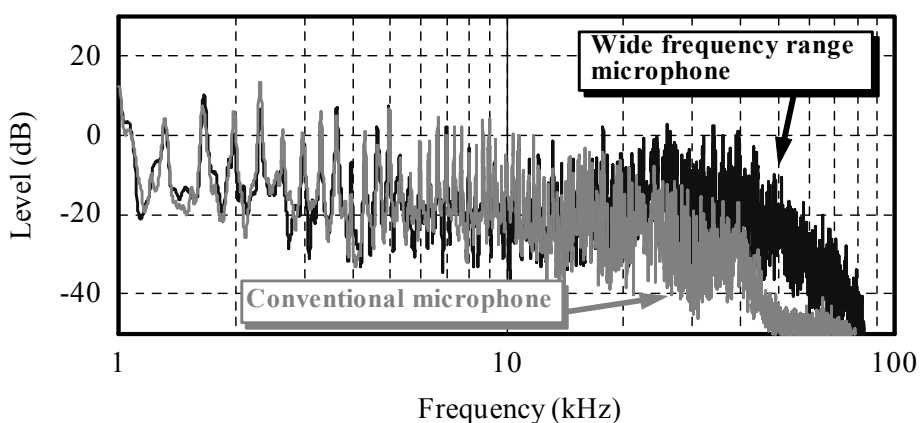


図 37. 超広帯域マイクロホンと従来の音楽録音用マイクロホンの収録帯域の比較。音源はハーブシコード。超広帯域マイクロホンでは、40 kHz 以上の帯域も収録できている。

3.4.3 実験方法の検討

楽音の超高域成分が聴こえや脳波などに影響を与えるとした先行研究によると、広帯域音源を聴取してから脳波に影響が現れるまで、数十秒の時間がかかり、この影響は、超高域をカットしても 60~200 秒程度残留し、聴こえの主観評価にも影響すると報告されている[22], [23]。そこで、本実験では評価音源の聴取時間を制御するため、実験 1 とは方法を変え、pair test 法を採用した(図 38)。被験者に、A と B のラベルをつけた評価音を対にして連続的に提示する。A と B は同一の曲だが、次のように再生帯域の違いにより 2 種類の対がある。

Type(1) : A と B のどちらか一方は、超高域をカットして可聴域のみ再生。
残る一方は、全帯域が再生される。

Type(2) : A と B 両方ともに、超高域をカットして可聴域のみ再生。

Type(1)と**(2)**のどちらの対が出現するかは一様乱数で決められ、トータルの出現回数は等しい。被験者は、AB の対を聴き終った後、A と B の音の印象が同じか違うかを判定する。**Type(1)**が提示され、被験者の判定が「違う」ならその判定は正解。同様に、**Type(2)**が提示され、判定が「同じ」なら正解。これ以外のケースは誤答である。

先行研究の実験方法を参考に、音源の提示時間を 85 秒~120 秒と長くするとともに、**Type(2)**は、超高域カットの対として、前述の残留効果が打ち消されるようにした。一方、評価音源の提示時間が長いと、前の音の印象を保持するのが困難であり、ITR BS-1116 [35]では、標準的な提示時間は 15~25 秒としている。そこで、提示時間を約 20 秒にした実験も行った。

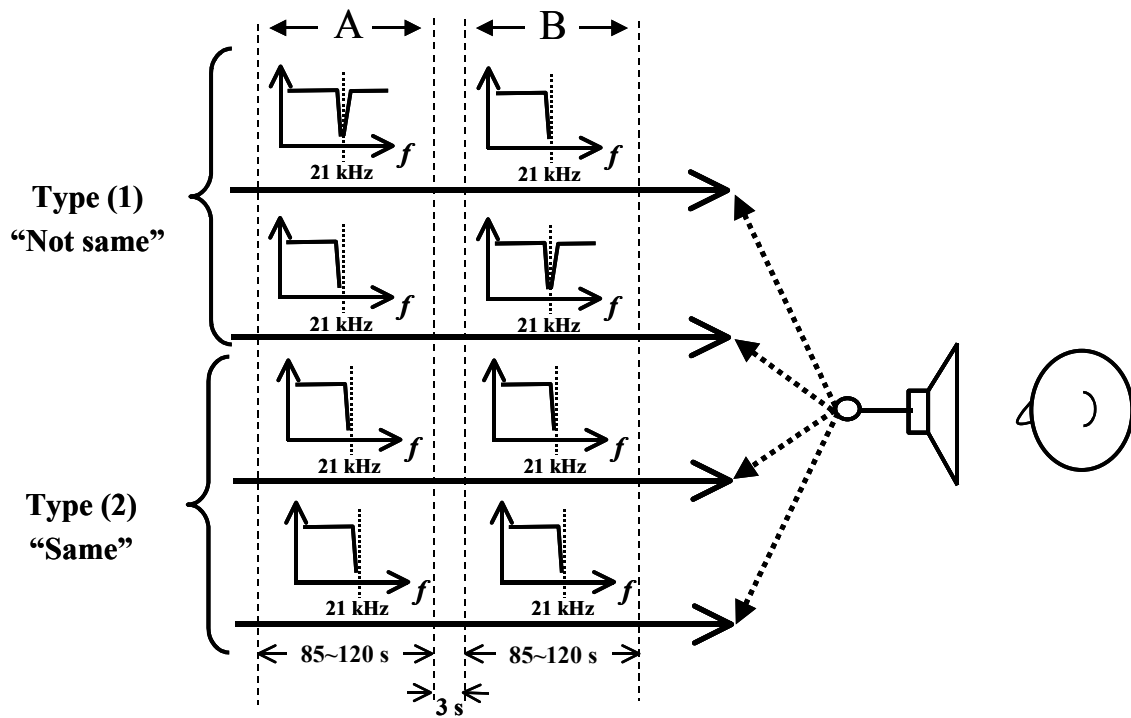


図 38. 実験 2.1 と 2.2 で採用した pair test 法の詳細。被験者に、A と B のラベルをつけた評価音を対にして連続的に提示する。A と B の再生帯域が同じ場合と違う場合があり、これを被験者が判定する。

3.4.4 実験 2.1: 提示時間が長い場合

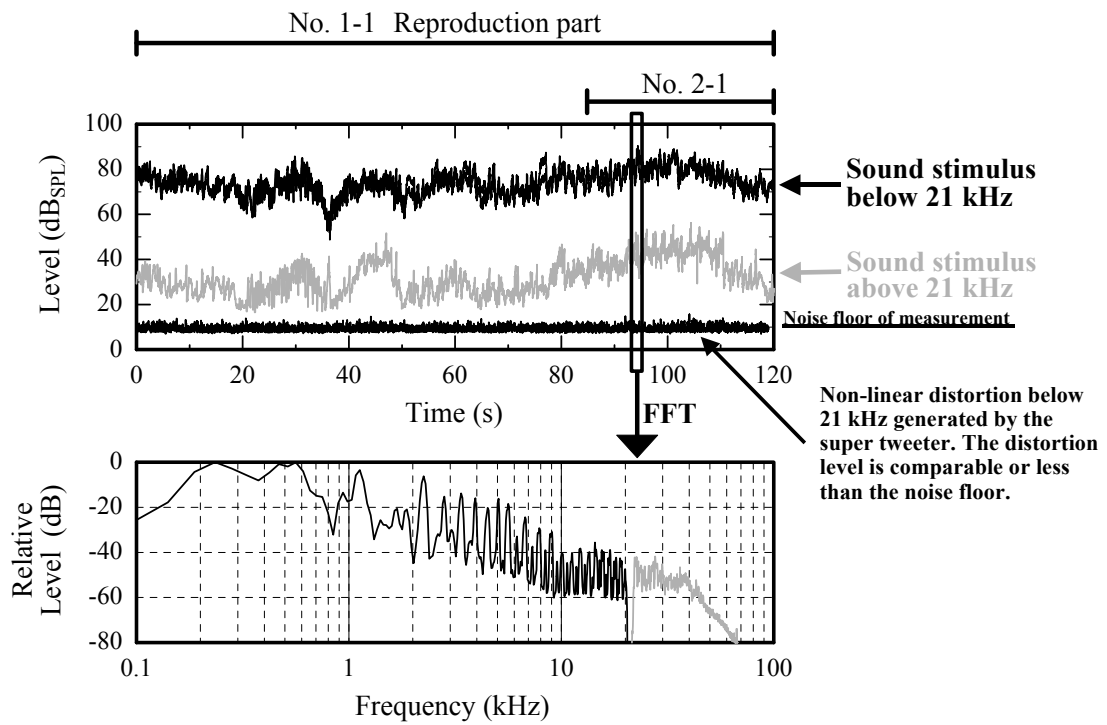
3.4.4.1 音源

弦楽四重奏, 筑前琵琶の弾き語り, ハープシコードの 3 種の音源を用意した (表 6 の No. 1-1, 1-2, 1-3)。音源の提示時間は 85~120 秒である。各音源の分析結果を図 39 に示す。この図は, 聴取位置における帯域別の音圧レベルと周波数スペクトルの計算値を示している。この分析には, 図 28 に示した実験装置の総合周波数特性が加味されている。音圧レベルは 50 ms 区間毎の実効値を算出し, 聴取位置での音圧に換算した。周波数スペクトルは, 500 ms の区間を 4096 サンプル (21.3 ms) のブラックマン窓 (表 1) を用いて分割し, 得られたスペクトル群を平均化して示した。

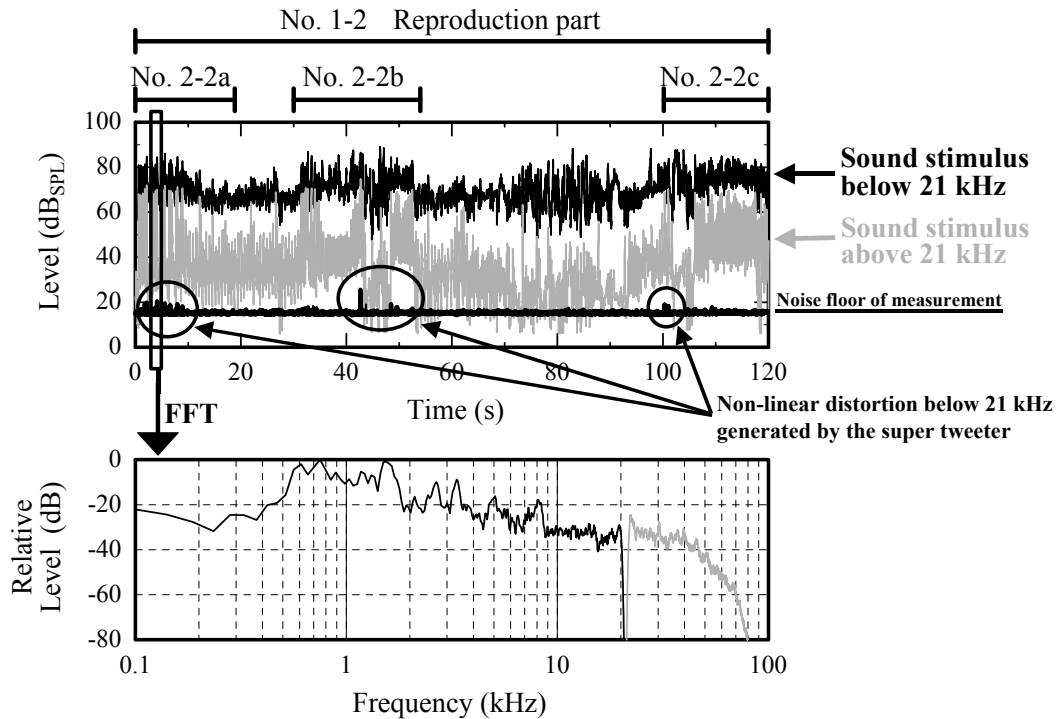
特に筑前琵琶音源は, 21 kHz 以上の超高域が, 常時 40~50 dB_{SPL} 程度出しており, 70dB_{SPL} に達する箇所もある。周波数帯域についても 70 kHz と広いことがわかる。

表 6. 実験 2.1 と 2.2 で用いた評価音源。音源は 3 種であるが, 実験 2.1 と 2.2 では音源の提示時間と提示部分が異なる。

No.		Title	Musical Instrument
Experiment 2.1	Experiment 2.2		
1-1	2-1	Faure: String Quartet	Violins, Viola, ViolinCello
1-2	2-2a, b, c	"Heike Monogatari"	"Chikuzen-Biwa", Female narration
1-3	2-3	J.S. Bach: France Suite No. 2 - Air	Harpsichord



☒ 39 (a) Nos. 1-1, 2-1 (Faure: String Quartet)



☒ 39 (b) Nos. 1-2, 2-2a, b, c (Chikuzen-Biwa)

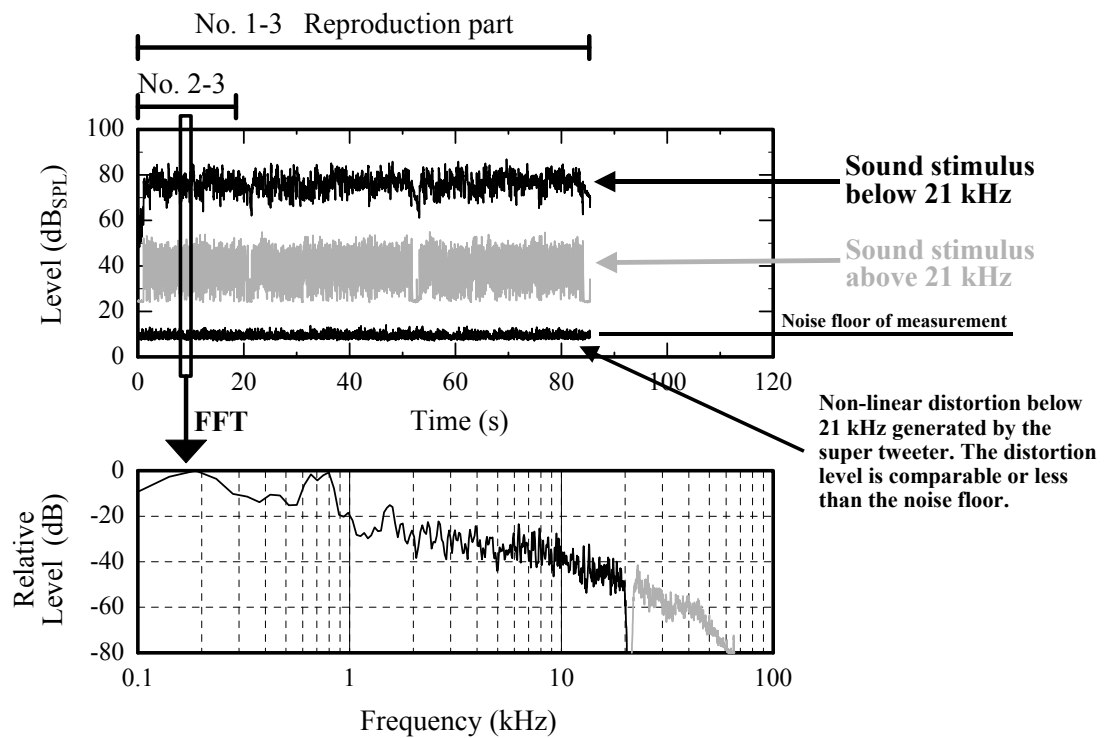


図 39 (c) Nos. 1-3, 2-3 (Harpisichord)

図 39. 実験 2.1 と 2.2 で用いた音源の分析結果。各音源の最上部の横線は、それぞれの実験で再生した部分 (No. 1-x は実験 2.1, No. 2-x は実験 2.2 での再生部分) を示す。上のグラフは、可聴域 (21 kHz 以下) と超高域 (21 kHz 以上) の音圧レベル、および、超高域再生時に発生した非線形歪の可聴域成分の音圧レベルを示す。下のグラフは、上のグラフの方形で囲んだ部分のスペクトルを示す。

3.4.4.2 被験者と実験結果

被験者と実験結果の詳細を図 40 に示す。被験者は、男性 5 人、女性 8 人の計 13 人である。職業・専門の内訳は、音楽大学学生 7 人、音楽大学教員 3 人、作曲家 2 人、演奏家 1 人、年齢は 19～51 歳であった。

被験者は、各音源について 32 回の判定（試行）を行った。実験は、約 30 分を 1 セッションとして、1 セッション内に 8 試行、セッション間に 10～15 分の休憩を挟んで、1 日に 3 セッション実施した。各被験者は 4 日で計 12 セッションの実験を行った。

実験結果（図 40）の P 値は、実験 1 と同様、式(18), (19)から求めた。有意水準 5%の二項検定では、被験者 No. 2 ($p=0.0201$)と 9 ($p=0.0070$)が音源 No.1-2 (筑前琵琶)の超高域有無を有意に弁別できたことになる。被験者 2 と 9 はそれぞれ、20 歳女性、30 歳男性で、ともに音学大学の学生であった。

Subject			Stimulus No.	Correct response rate											<i>p</i>
No.	Age	Gender		0.0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	
1	19	male	1-1												0.8601
			1-2												1.0000
			1-3												0.5966
2	20	female	1-1												1.0000
			1-2												0.0201
			1-3												1.0000
3	20	female	1-1												0.5966
			1-2												0.2153
			1-3												0.5966
4	21	female	1-1												0.8601
			1-2												0.8601
			1-3												1.0000
5	22	female	1-1												0.3771
			1-2												0.5966
			1-3												0.3771
6	24	female	1-1												0.2153
			1-2												0.5966
			1-3												0.2153
7	25	female	1-1												0.1102
			1-2												0.8601
			1-3												0.5966
8	29	male	1-1												0.3771
			1-2												0.3771
			1-3												0.8601
9	30	male	1-1												0.8601
			1-2												0.0070
			1-3												0.5966
10	31	male	1-1												0.2153
			1-2												0.2153
			1-3												0.8601
11	33	female	1-1												0.5966
			1-2												1.0000
			1-3												0.8601
12	36	female	1-1												0.8601
			1-2												0.5966
			1-3												0.8601
13	51	male	1-1												0.2153
			1-2												0.5966
			1-3												0.0501

図 40. 実験2.1の結果。各被験者、音源毎の正答率とP値を示している。

3.4.5 実験 2.2: 提示時間が短い場合

3.4.5.1 音源

実験 2.1 の 3 種の音源から、超高域のレベルが高い部分を選んで 5 種類の音源を用意した。音源の内容と分析結果を表 6, および, 図 39 の No. 2-1, 2-2a, 2-2b, 2-2c, 2-3 に示した。音源の提示時間は約 20 秒である。特に, 筑前琵琶の音源 (実験 2.1 の No. 1-2) は, 前述の実験で有意差を得たので, この実験では, 冒頭, 中間, 終わりの 3 箇所 (No. 2-2a, 2-2b, 2-2c) を選択して用いた。

3.4.5.2 被験者と実験結果

被験者は, 実験 2.1 の被験者から 3 人を選んだ。3 人の内 2 人は, 実験 2.1 で, 筑前琵琶の音源で有意差を得た者である。

実験は, 約 25 分を 1 セッションとして, 1 セッション内に 20 試行, セッション間に 10~15 分の休憩を挟んで, 1 日に 2 セッション実施した。各被験者は 5 日で計 10 セッションの実験を行い, トータルで, 各被験者は各音源について 40 回の判定 (試行) を行った。

被験者と実験結果の詳細を図 41 に示す。被験者 No. 6 の音源 No. 2-2a で正答率が極めて小さいため, P 値も小さい値を示している。3.4.3 節で述べた先行研究[22]では, 音源の提示時間が短い場合, 広帯域音源聴取の脳波への影響が時間的に残留するため, 順序効果が発生し, 評価結果が矛盾することがあると推論しているが, 少なくとも同異について正しい弁別とは言えない。また, 被験者 No. 2 と No. 9 は, 音源の提示時間を長くとした前述の実験において, 筑前琵琶の音源を有意に弁別していたにもかかわらず, 本実験の筑前琵琶音源 (No. 2-2a, 2-2b, 2-2c) は弁別できなかった。

そこで, 音源の提示時間が, 弁別結果に影響を与えているか否かを統計的に検討するため, 実験 2.1 と本実験の筑前琵琶音源に対する評価結果について, 独立性検定を行った。表 7 は, 実験 2.1 と 2.2 における共通の被験者 No. 2, 6, 9 の筑前琵琶音源 (No. 1-2, 2-2a, 2-2b, 2-2c) での判定結果をまとめたク

ロス集計表である。これに対して、 χ^2 検定を施す[36]。帰無仮説は、「実験 2.1 と実験 2.2 の正答数は独立である」である。検定統計量（連続性補正あり）は、

$$\chi_0^2 = \frac{648 \times \left(\left| 166 \times 192 - 122 \times 168 \right| - \frac{648}{2} \right)^2}{288 \times 360 \times 334 \times 314} = 7.729,$$

自由度 1, 有意水準 5%の両側検定の棄却限界値は 3.841 ($< \chi_0^2$) であるから、帰無仮説は棄却される。したがって、音源の提示時間と、筑前琵琶音源での正答数には関連があると考えられる。

Subject			Stimulus No.	Correct response rate	p
No.	Age	Gender			
2	20	female	2-1		0.8746
			2-2a		0.6358
			2-2b		0.6358
			2-2c		0.4296
			2-3		0.1539
6	24	female	2-1		0.1539
			2-2a		0.0166
			2-2b		0.6358
			2-2c		0.6358
			2-3		0.4296
9	30	male	2-1		0.8746
			2-2a		0.8746
			2-2b		0.4296
			2-2c		0.8746
			2-3		0.6358

図 41. 実験 2.2 の結果。各被験者、音源毎の正答率と P 値を示している。

表 7. 実験 2.1 と 2.2 の評価結果の比較。音源の提示時間と評価結果の関係を検定するため、両実験に共通の被験者の筑前琵琶音源に対する実験結果を集計。

	Number of correct responses	Number of incorrect responses	Total
Experiment 2.1 (Long duration)	166	122	288
Experiment 2.2 (Short duration)	168	192	360
Total	334	314	648

3.4.6 被験者の聴覚閾値測定

実験 2.1 では、2 人の被験者の結果が有意差を得た。この被験者は、超高域成分を聴取することが出来たのかどうかを検討するため、超高域の聴覚閾値測定を試みた。一般的なオーディオメータ (audiometer) では、20 kHz を超える帯域での測定ができないため、無響室内でスピーカを発音源とし、2 区間 2 肢強制選択と 3 down 1 up 変形上下法[39]を組み合わせた方法で、純音の絶対閾値を測定した[40]。

被験者には、2 つの区間 A, B が提示される。それぞれの区間は 2 秒、区間と区間の間は 0.5 秒で、どちらか一方の区間にのみ検査音が提示される。検査音は、純音のトーンバースト波で、250 ms の無音を挟んで 250 ms の純音が 4 回で構成される。被験者は、A と B どちらの区間で音が聞こえたかを答え、誤答なら検査音の音圧レベルを上げ、3 回連続で正答なら下げる。レベルの変化幅は、測定開始から最初に検査音レベルの下降と上昇が反転するまでは 4 dB、これ以降は 2 dB とした。測定は 5 回目の反転が生じたときに中止し、4 回目と 5 回目の反転時のレベルを平均して閾値とした。実験 2 の評価音源の分析結果 (図 39) から、超高域の最大音圧レベルは 90 dB_{SPL} を超えていないことと被験者の安全性を考慮し、検査音の提示レベルが 90 dB_{SPL} を超えた時点で測定を打ち切った。

測定結果を表 8 に示す。16 kHz の閾値にはかなり差があるが、20 kHz の閾値は 80 dB_{SPL} を超えており、これまでの聴覚閾値の報告[18], [40]と同様、16 kHz 付近から急激に閾値が上昇する傾向がある。22 kHz の閾値は 90 dB_{SPL} を超えていることから、被験者は、超高域の純音は聴取できないと考えられる。

表 8. 自由音場での聴覚閾値測定結果。実験 2.1 で超高域有無を有意に弁別した被験者 2 人の聴覚閾値を、無響室内で測定した。

Subject No.	Frequency (kHz)		
	16	20	22
2	29	83	>90
9	83	>90	>90

(dB SPL)

3.5 第3章のまとめ

本章では、楽音の超高域の有無が聴感に与える影響について検討した。

超高域成分の聴感への影響を主観評価実験で調べる場合、スピーカやアンプで発生した非線形歪の可聴域成分が弁別の手がかりとなる可能性がある。本実験では、評価音再生時の非線形歪を測定し、歪のレベルが充分小さいことを確認したシステムを用いて実験を行った。

実験1では、超高域成分による音質差異が存在する可能性があるかないか、また、その影響は、音源の種類や被験者によって異なるのかどうかを確認する目的で、様々な種類の音源を用意し、比較的多数の被験者により基礎的な主観評価実験を行った。被験者36人で20種の音源を、それぞれ40回評価した。この結果、音源毎には有意な弁別結果は得られなかったが、被験者1人が正答率0.75 ($p=0.0417$) をマークした。そこで、この被験者に対して追加実験を行った。より信頼性を高めるため、音源を6種類に選別して、それぞれの音源を20回評価した。この結果、どの音源に対しても有意な弁別結果は得られなかった。

実験1の結果と、超高域の影響には、音源の提示時間が関係するという先行研究[22], [23]から、実験方法を検討、被験者、音源を精査して実験2を行った。この実験2は、音源の提示時間が長い(85~120秒)実験2.1と20秒程度の短い実験2.2で構成した。提示時間が長い場合、2人の被験者が筑前琵琶の音源において、超高域の有無を有意に弁別できた。一方、提示時間が短い場合は、有意に正しい弁別結果は得られなかった。この結果は、先行研究[22], [23]の結果と整合する。

筑前琵琶音源の超高域を再生したときに発生する非線形歪の可聴域成分は、最大20~25 dB_{SPL}であり、可聴域の音源のレベル70 dB_{SPL}に対して極めて小さい。さらに、音源の提示時間を短くした実験2.2で、同じひずみが発生しているにも関わらず有意に弁別できなかつたことから、非線形歪の可聴域成分で弁別したとは考え難い。

以上の検討結果から、音源と、聴取時間、被験者の条件によっては、超高域成分の有無が聞こえに影響を与え得ると考えられる。

実験 2.1 で有意な弁別を行った 2 人の被験者が、超高域成分を聴取することが出来たのかどうかを検討するため、超高域の聴覚閾値測定を試みた。一般的なオーディオメータでは、20 kHz を超える帯域での測定ができないため、無響室内で 16, 20, 22 kHz の純音の閾値測定を行った。この結果、2 人の被験者の 22 kHz 純音の閾値は 90 dB_{SPL} を超えており、直接的に筑前琵琶音源の超高域を聴取できたとは考え難い。本実験の範囲では、超高域成分の有無をどのように聞き分けているかは説明できない。これまで、骨導で超音波が聞こえることは判明しているが、そのメカニズムについての定説はない[41]。気導聴覚系以外に、未知の情報チャンネルの存在を示唆する報告[42]もあるが、実験例が少なく更なる検証が必要と考えられる。

第4章 標本化フォーマットと可聴帯域内の音質

4.1 はじめに

ハイレゾリレーションオーディオ機器による広帯域収録再生が聴感に影響を与えるとする、その要因としては、可聴域を超える周波数成分（超高域成分）そのものの影響と、それ以外、すなわち可聴域の音質変化が考えられる。前者の超高域成分の影響については前章で述べた。本章では、後者の要因として、192kHz 24bit PCM や DSD (2.8MHz 1bit) といった広帯域な標本化フォーマットが、可聴域の音質に影響を与えるかどうか検討する。

これまで、標本化フォーマットと音質について学術的に検証した例は多くない。Yoshikawa らは、標本化周波数が 48 kHz と 96 kHz の DAT の音質について、楽音を用いて評価実験を行い、この 2 つの標本化周波数間の音質差を弁別できると結論している[43]。この実験では、図 42 に示すように、超高域を含む広帯域音源を標本化周波数 96 kHz のレコーダー（図中の X-86S）で収録し、標本化周波数 48 kHz の音源は、標本化周波数変換器（fs-convertor）を用いて 96 kHz の音源から作成している。この方法では、音質差が超高域の有無から生じたものか、或いは、標本化周波数変換の過程で可聴域の音質が変化したのか判然としないという問題がある。Blech ら[44]は、図 43 のように 2 つのフォーマット（176.4 kHz 24 bit PCM と DSD）ごとに同一の A/D, D/A 変換器の組み合わせを用いたシステムを使って音質比較を行っている。結果として、176.4 kHz 24 bit PCM と DSD の間には有意な音質差がないことを報告しているが、従来の標本化周波数 44.1～48 kHz との比較は行っていない。

本研究では、標本化フォーマット以外の条件を可能な限り同一にして、3 種のフォーマット 48kHz 24bit, 192kHz 24bit, DSD にて収録・再生をおこない、各フォーマット間の音質差異に関して主観評価実験を行った[45]。

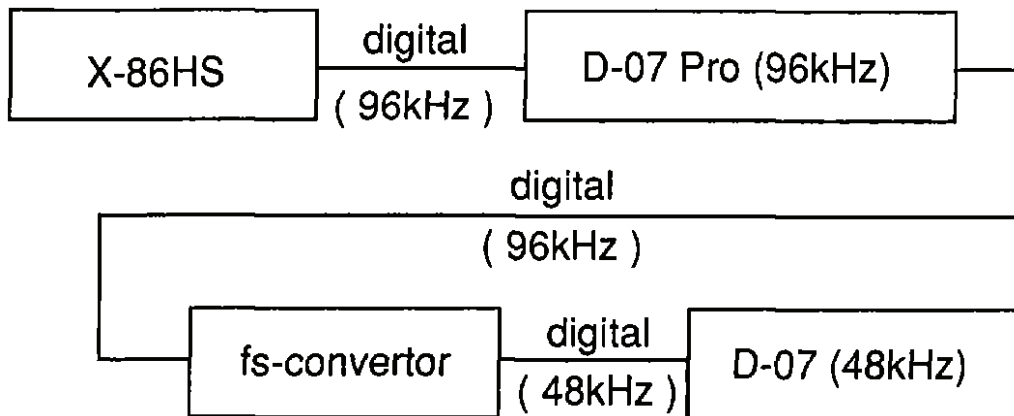


図 42. 48 kHz と 96 kHz 標本化の音質比較のための試験テープ作成方法 (Yoshikawa ら[43])。

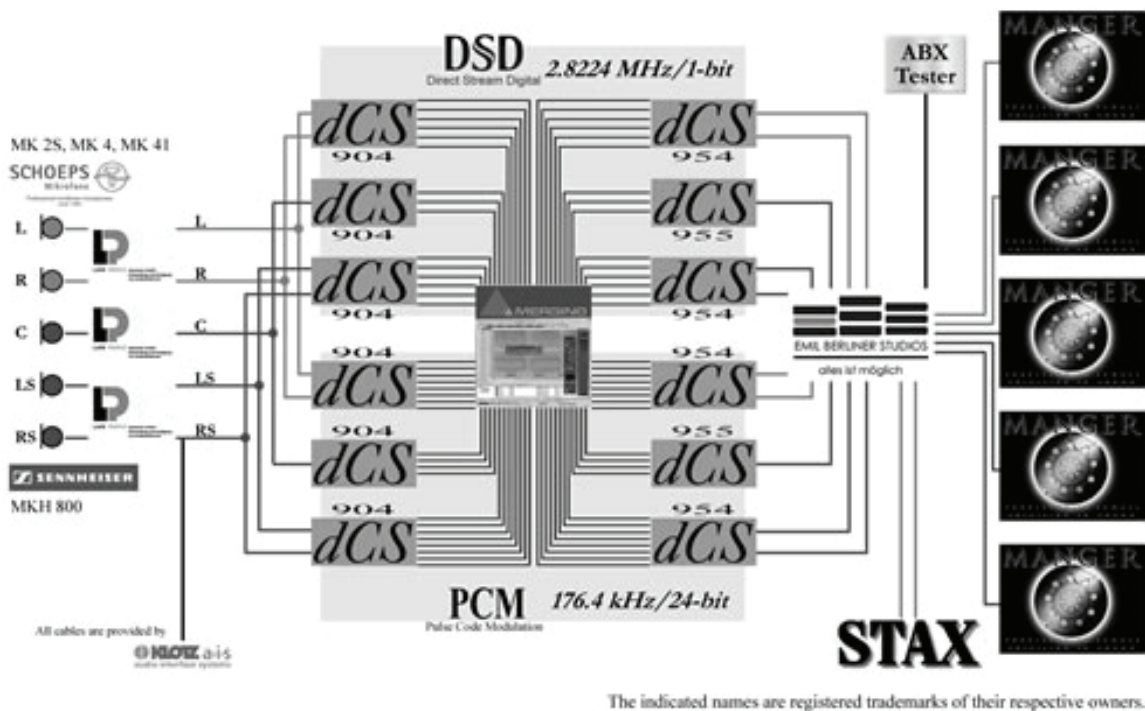


図 43. Blech らの実験システム[44]。同一機種のア/D 変換器 (dCS 社 904), D/A 変換器 (dCS 社 954 と 955) を用い、2 つのフォーマット (176.4 kHz 24 bit PCM と DSD) の音質比較を行っている。再生は、ヘッドホン (STAX) とスピーカー (MANAGER)。

4.2 実験システム

4.2.1 実験システムの構成

音源収録から評価実験までのシステムを図 44 に示す。本実験では、標準化フォーマット以外の条件に差が生じないようにする必要がある。音源収録では、アナログ音声卓の出力を 3 並列にして、それぞれ 48kHz 24bit, 192kHz 24bit, DSD の 3 種フォーマットで動作する A/D 変換器 3 台に接続した。収録プロセスで特に音質への影響が大きい A/D 変換器については、機種や個体差による音質差を排除するため、機種・ファームウェアが同一のものをを用い、変換器の入力感度差は $\pm 0.03\text{dB}$ 以内になるよう調節した。音源は、MO (Magneto-Optical) ディスクに記録された後、1 台のデジタルオーディオワークステーション (DAW) にデジタルダビングして、評価実験で再生する箇所の切り出しを行った。DAW の信号処理精度による影響を避けるため、評価音の冒頭と終了に 1 秒以下のフェード処理を施した以外、一切の信号処理を行っていない。こうして作成した評価音を、フォーマット毎に MO ディスクレコーダにデジタルダビングし、機種・ファームウェアが同一の D/A 変換器で再生した。再生フォーマットの切り替えは、機械接点のリレー、CdS 光導電セル/LED のアナログ・フォトカプラを用いたアナログスイッチャーを製作して用いた。このアナログスイッチャーは、入力信号系統の切り替えは機械接点リレーで行い、切り替え時のノイズは、出力段に設けたアナログ・フォトカプラでカットする構成となっている。この構成では、信号系統毎に異なる増幅素子が入らず、原理的に信号系統間でノイズ、歪、レベルなどの特性差が発生しない。さらに、収録・再生を通してフォーマット毎のレベル偏差は、 $\pm 0.05\text{dB}$ 以内になるように D/A 変換器を調節した。

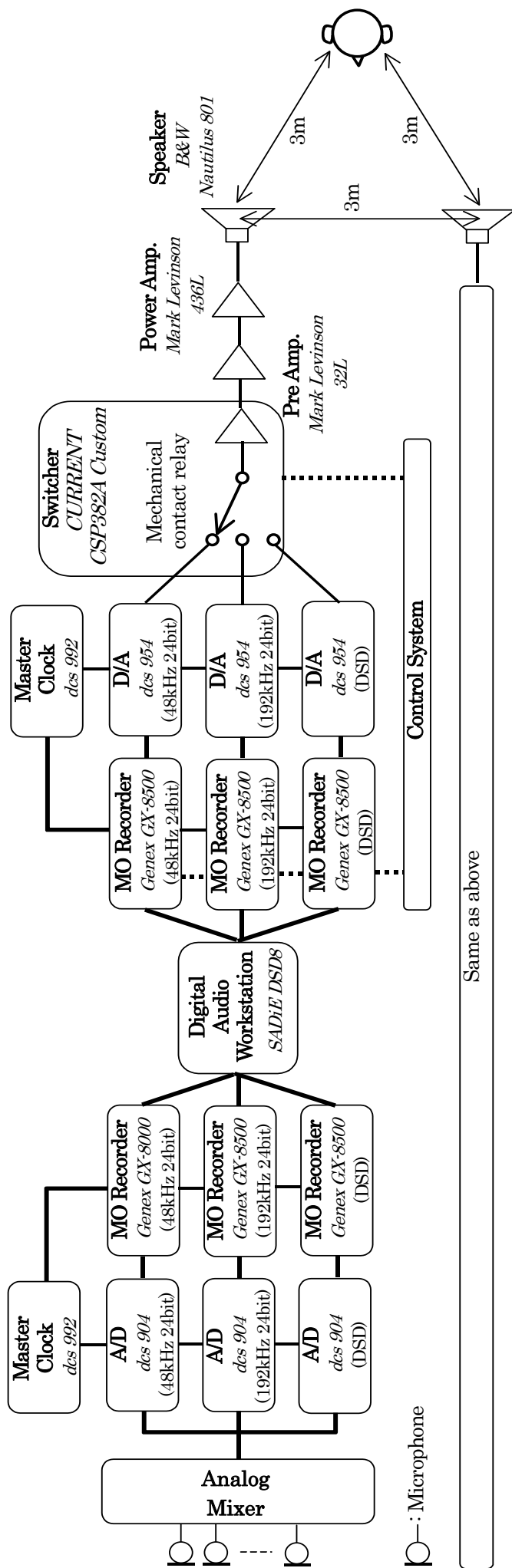


図 44. 実験 3 で用いた実験システム。評価音源の収録から再生にかけてのプロセスで、標準化フォーマット以外の条件を完全にそろえるため、同一機種、同一フォーマットの A/D, D/A 変換器を用いている。

4.2.2 A/D, D/A 変換器の物理特性

実験システムの中で、フォーマットごとの音質を左右する A/D 変換、D/A 変換について、周波数振幅特性と周波数位相特性を測定した。A/D 変換器のデジタル出力を D/A 変換器のデジタル入力に直接接続し、A/D 変換器にアナログの試験信号を入力し、D/A 変換器のアナログ信号出力を測定した。測定は、Audio Precision 社 System Two Cascade Plus (SYS-2722)を用いた。

周波数振幅特性の測定結果を図 45 に示す。本実験に用いた A/D 変換器(dCS 社 904)と D/A 変換器(dCS 社 954)は、幾つか特性の異なるアンチエイリアシングフィルタを選ぶことが出来るが、本実験では、折り返し歪の影響が出ないよう、最も急峻なカットオフ特性のフィルタを選択した。測定結果から、48kHz 24bit, 192kHz 24bit, DSD の 3 種のフォーマット間で、帯域の広さは標準化周波数に応じて異なっているが、20 kHz 以下の可聴帯域の周波数振幅特性には全く差が見られない。

周波数位相特性の測定結果を図 46 に示す。位相は 360 度でラップしてプロットされている。3 種のフォーマットとも、位相が周波数に対して直線的に変化していることから、A/D, D/A 変換系は単純遅れ系であり、フォーマットに依らず位相歪みの影響はないと考えられる。

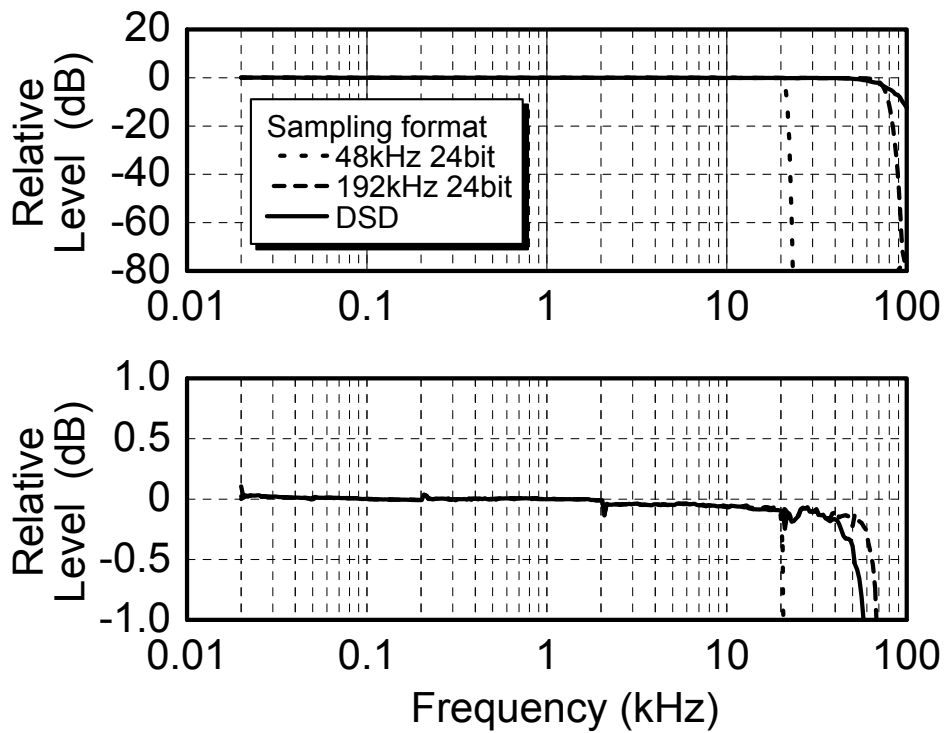


図 45. A/D 変換・D/A 変換トータルでの周波数振幅特性の実測結果。下のグラフは、上のグラフの縦軸を拡大したもの。

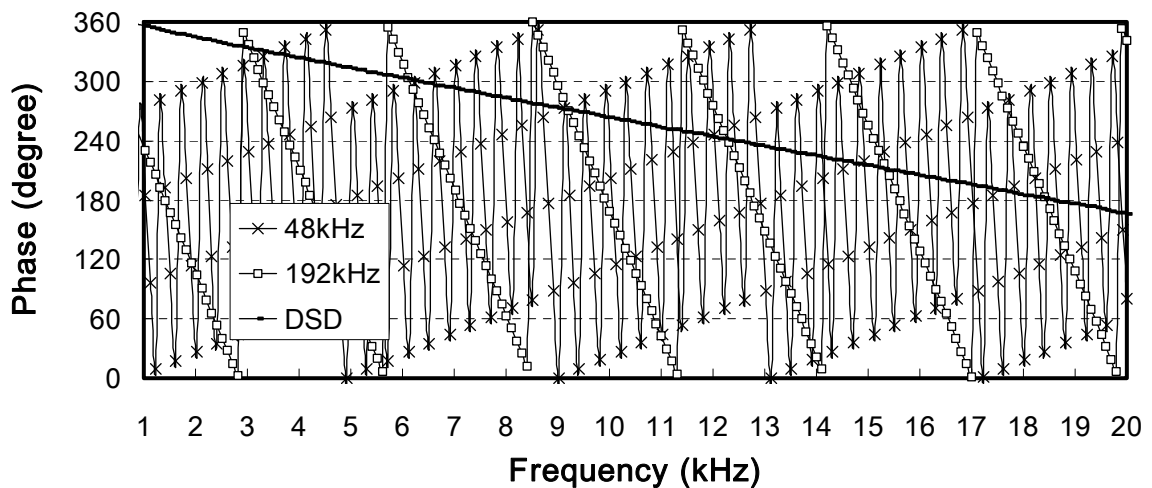


図 46. A/D 変換・D/A 変換トータルでの周波數位相特性の実測結果。360 度でラップしてプロットした。

4.2.3 再生条件

実験は ITU-R BS1116 [35] に規定された Reference listening room に完全準拠した音響評価室で行った。再生方法は、前述の超高域弁別実験と同様 2 チャンネルステレオ再生であるが (図 26), 今回の実験は可聴域の音質評価が目的であるため, スーパーツイータは使用しなかった。聴取位置における実験システムの総合周波数特性を図 47 に示す。

評価実験時の再生音圧は, 評価音源の主要なピークが約 75~80dB_{SPL} (A 特性, Fast) となるよう調節した。

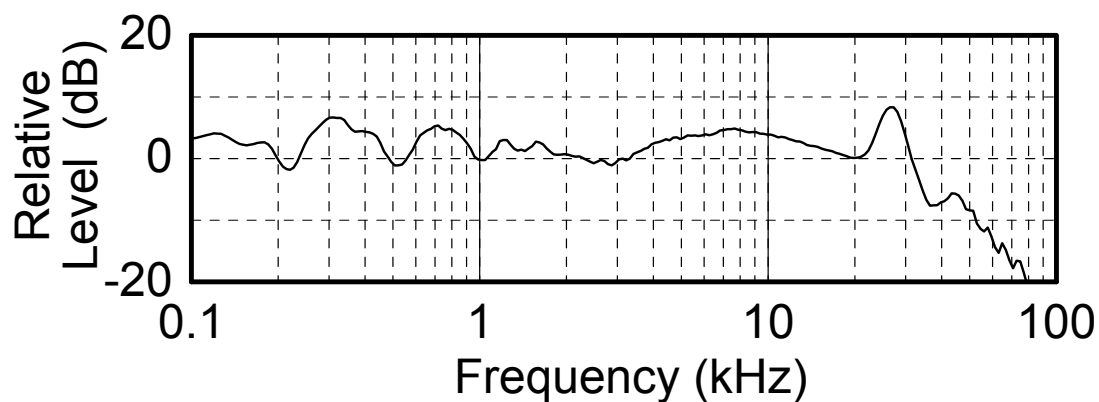


図 47. 実験 3 の実験システムの総合周波数特性。

4.3 実験 3:3 種の標本化フォーマットの主観弁別実験

4.3.1 被験者

聴力健常な音楽大学学生 4 人 (30 代男性 1 人, 20 代女性 3 人), 音源「ピアノトリオ」の演奏に加わったバイオリニスト 1 人 (30 代女性), 録音技術者 (30 代男性) 1 人である。

4.3.2 音源

「邦楽」(箏, 尺八), 「ボサノバ」(ギター, 女性ボーカル), 「ピアノトリオ」(ピアノ, チェロ, バイオリン) の 3 種である。それぞれ, 図 44 のシステムにより 48kHz 24bit, 192kHz 24bit, DSD の 3 種のフォーマットで収録を行った。本実験では, デジタルフォーマットの違いによる可聴域の音質の差を評価することを目的としたため, 意図的な収録帯域の拡大は避け, 音楽録音用マイクを用いた通常の方法によって音源収録を行った。評価音の代表的な部分の周波数分析結果を図 48 に示す。各音源の周波数帯域は, ほぼ可聴帯域内 (20 kHz 以下) に収まっており, これらの音源では, 超高域の影響を受けず, 可聴域の音質のみ評価が可能であることが確認できる。音源の提示時間は, フレーズの切れ目のよい部分で切り出し, 約 20~30 秒とした。

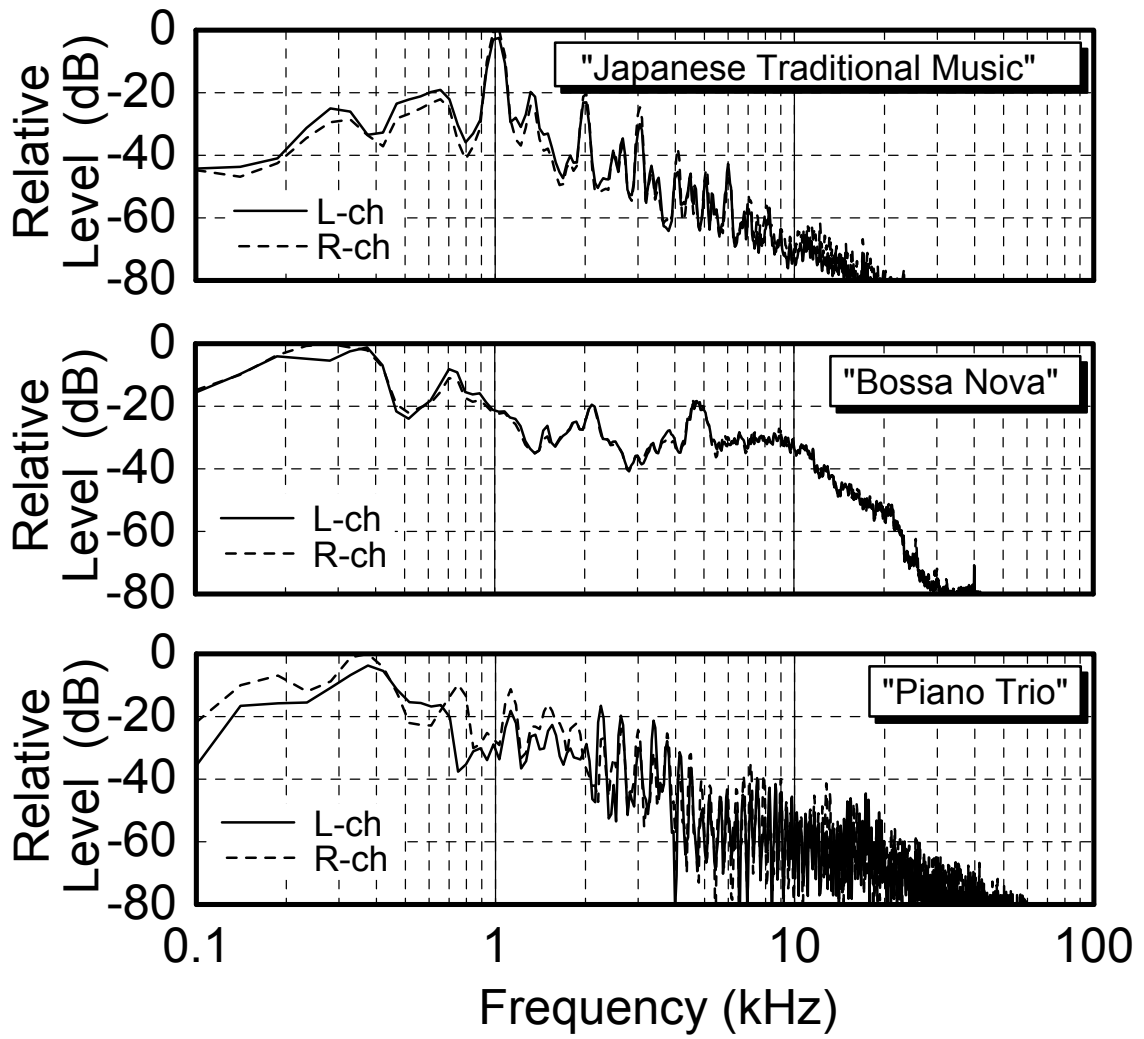


図 48. 実験 3 の評価音源のスペクトル。

4.3.3 実験方法

本実験では、微小な差の検出に適した pair test 法を用いた。被験者には、曲は同じでフォーマットが異なる評価音 A, B から AA, BB, AB, BA の 4 種の対をつくり、ランダムに提示した。評価音は、3 秒の間隙を挟んで連続提示する。被験者は、評価音の対を聴き終わったあと、音が同じか違うかを判定した。提示時間は A と B とともに約 20~30 秒である。比較するフォーマットは 3 種あるため、次の 3 種の実験を、休憩を挟んで同じ日に行った。

実験 3.1 : 48kHz 24bit と 192kHz 24bit の比較

実験 3.2 : 192kHz 24bit と DSD の比較

実験 3.3 : DSD と 48kHz 24bit の比較

4.3.4 実験結果

実験は、約 30 分を 1 セッションとして、1 セッション内に 24 試行、セッション間に 10~15 分の休憩を挟み、1 日に 3 セッション実施した。各被験者は 3 日で計 9 セッションの実験を行った。結果として、各被験者はすべての比較条件（フォーマット 3 種×音源 3 種）について 24 回繰り返し評価を行った。実験 3.1 から 3.3 の結果を図 49 に示す。この中で、もっとも高い正答率は 66.7% である。被験者は、同じ/違うの 2 者択一で答えるため、正答率は 2 項分布に従い、危険率 (P 値) は式(19)から算出できる。有意水準 5% で有意となるためには、24 試行の場合、少なくとも 75% の正答率を得る必要がある。したがって、結果からは、被験者、音源に依らず、フォーマットの違いは有意に弁別できなかった。

Subject		Stimulus	Experiment 3.1. 48kHz/192kHz			Experiment 3.2. 192kHz/DSD			Experiment 3.3. DSD/48kHz		
No.	Occupation		Age	Gender	Correct response rate (%)	Correct response rate (%)	Correct response rate (%)	Correct response rate (%)	Correct response rate (%)	Correct response rate (%)	Correct response rate (%)
1	student of univ. of music	21	female	Japanese Traditional music	50	48	55	55	30	50	55
				Bossa Nova	50	35	50	50	50	35	50
				Piano Trio	65	60	65	60	60	65	65
2	student of univ. of music	21	female	Japanese Traditional music	50	35	50	50	35	50	50
				Bossa Nova	50	68	50	68	68	50	50
				Piano Trio	65	60	65	60	65	65	65
3	student of univ. of music	21	female	Japanese Traditional music	48	42	48	42	42	48	42
				Bossa Nova	35	55	35	55	55	35	35
				Piano Trio	50	50	50	50	50	50	50
4	student of univ. of music	31	male	Japanese Traditional music	60	48	60	48	48	60	48
				Bossa Nova	48	42	48	42	42	48	42
				Piano Trio	35	50	35	50	50	35	35
5	violinist	35	female	Japanese Traditional music	68	35	68	35	35	68	35
				Bossa Nova	55	60	55	60	60	55	55
				Piano Trio	55	50	55	50	50	55	50
6	recording engineer	33	male	Japanese Traditional music	55	48	55	48	48	55	48
				Bossa Nova	40	60	40	60	60	40	40
				Piano Trio	30	50	30	50	50	30	30

図 49. 実験 3 の結果。各被験者・音源の正答率を示す。

4.4 第4章のまとめ

ハイレゾリューションオーディオが聴感に影響を与えるとすると、そのひとつの要因としては、可聴域の音質変化が考えられる。本章では、これを確かめるため、3種の標本化フォーマット、48kHz 24bit, 192kHz 24bit, DSD (2.8MHz 1bit) 間の可聴域の音質差異について、主観弁別実験を行って検討した。

本実験では、標本化フォーマット以外の条件に差が生じないように実験装置や音源の作成方法に充分注意を払った。

音源は、収録条件を管理するため、すべて独自収録を行って作成した。音質への影響が大きい A/D, D/A 変換器については、機種や個体差による音質差を排除するため、機種・ファームウェアが同一のものをを用い、収録・再生を通してフォーマット毎のレベル偏差は、 $\pm 0.05\text{dB}$ 以内になるように調節した。

評価音源は、邦楽、ボサノバ、ピアノトリオの3種、被験者は、聴力健常な音楽大学学生4人、バイオリニスト1人、録音技術者1人である。実験方法は、微小な差の検出に適した pair test 法を採用した。実験の結果、被験者、音源に依らず、フォーマットの違いは有意に弁別されなかった。

さらに、実験システムの中で、フォーマットごとの音質を左右する A/D 変換、D/A 変換について、周波数振幅特性と周波數位相特性を測定した。測定の結果、3種のフォーマット間で、20 kHz 以下の可聴帯域の周波数振幅特性は全く差が見られなかった。また、可聴域の周波數位相特性は、どのフォーマットも直線位相を示し、フォーマットに依らず位相歪みの影響はないと考えられる。主観弁別実験の結果は、この測定結果を支持している。したがって、本実験システムの可聴域の音質に関して、フォーマット間に聴感的差異は生じていないと考えられる。

第5章 結論

ハイレゾリューションオーディオが聴感に与える影響について検討を行った。ハイレゾリューションオーディオによる広帯域収録再生が聞こえ影響を与えるとすると、その要因としては、可聴域を超える周波数成分の影響と、可聴域の音質変化が考えられる。本論文の第3章では、前者の要因について検討するため、楽音に含まれる超高域成分の有無を弁別できるか否か主観評価実験を行った。2人の被験者が、帯域が広く(60 kHz)超高域を常時豊富に含む(聴取位置で40~70 dB_{SPL})筑前琵琶音源を2分間聴取した実験で、超高域の有無を有意に弁別できた。筑前琵琶音源の超高域を再生したときに発生する非線形歪の可聴域成分は、最大20~25 dB_{SPL}であり、可聴域の音源のレベル70 dB_{SPL}に対して極めて小さい。さらに、音源の提示時間を短くした実験で、同じひずみを発生しているにも関わらず有意に弁別できなかったことから、非線形歪の可聴域成分で弁別したとは考え難い。

以上の検討結果から、音源、聴取時間、被験者の条件によっては、超高域成分の有無が聞こえに影響を与え得ると考えられる。

有意な弁別を行った2人の被験者が、超高域成分を聴取することが出来たのかどうかを検討するため、超高域の聴覚閾値測定を試みたが、2人の被験者の22 kHz純音の閾値は90 dB_{SPL}を超えており、直接的に筑前琵琶音源の超高域を聴取できたとは考え難い。本研究の超高域成分の有無をどのように聞き分けられているかは、本実験の範囲で説明できない。これまで、骨導で超音波が聞こえることは判明しているが、そのメカニズムについての定説はない[41]。気導聴覚系以外に、未知の情報チャンネルの存在を示唆する報告[42]もあるが、実験例が少なく更なる検証が必要と考えられる。

一方の可聴域の音質変化については、第4章にて、3種の標本化フォーマット(48kHz 24bit, 192kHz 24bit, DSD)にて収録・再生をおこない、フォーマット間の可聴域の音質差異に関して主観評価実験により検討した。この結果、本

実験のシステムでは、フォーマット間で可聴域の音質には差が見られなかった。現在、変換精度が十分配慮された A/D, D/A 変換システムでは、2.3.2節で述べたように Σ/Δ 変調方式を採用し、オーバサンプリングによってアナログフィルタの負担を軽減するとともに、量子化精度を時間精度に負わせているため、一般的な 48kHz 標本化と、より高い標本化周波数を用いたフォーマット間の可聴域の音質差異は充分小さいと考えられる。

以上のことから、ハイレゾリューションオーディオによる広帯域収録再生では、可聴域を超える周波数成分が聞こえに影響を与える場合があり、超高域成分を豊富に含む音源については、広帯域収録再生は価値あることと考えられる。

しかしながら、広帯域収録再生の様々な課題も指摘されている[46], [47]。帯域を広げるため標本化周波数を高くすると、非線形歪や時間ゆらぎの影響が増大する。また本研究でも明らかのように、超高域成分は容易に聴取することができないため、超高域にノイズが混入している音楽ソフトが散見され、品質管理の難しさも指摘されている。現実的なオーディオ信号の取り扱いを考えた場合、広帯域収録再生には、十分な注意が必要とされる。

謝辞

本研究を進めるにあたり、マイクロホンの試作にご協力いただいた三研マイクロホン株式会社殿、スイッチャー等の製作にご協力いただいた株式会社カレント殿に感謝いたします。

評価実験の被験者として実験に参加いただいた、東京藝術大学、桐朋学園大学の皆様、NHK 放送センターの音声技術者の皆様、音源の録音にご協力いただいた演奏家の皆様に厚く御礼申し上げます。

本論文をまとめるにあたり、ご指導を賜りました電気通信大学の出澤正徳教授、井上誠喜客員教授、阪口豊准教授、長岡浩司教授、森田啓義教授に心から感謝申し上げます。

研究を進めるにあたり、ご指導、ご助言をいただいた、NHK 放送技術研究所の安藤彰男主任研究員、濱崎公男主任研究員、NHK の岩城正和チーフエンジニア、NHK 放送技術研究所の小野一穂主任研究員に深く感謝いたします。さらに、研究遂行と論文執筆あたり、さまざまなご協力をいただいた NHK 放送技術研究所 高臨場感音響研究グループの皆様に厚く御礼申し上げます。

参考文献

- [1] 境久雄, 中山剛, 「聴覚と音響心理」 (コロナ社, 東京, 1978) .
- [2] 日本音響学会 編, 「音のなんでも小事典」 (講談社, 東京, 1997) .
- [3] 大山正, 今井省吾, 和気典二 編, 「新編 感覚・知覚心理学ハンドブック」 (誠信書房, 東京, 1994) .
- [4] 内川恵二 編, 「聴覚・触覚・前庭感覚」 (朝倉書店, 東京, 2008) .
- [5] B. C. J. Moore 著, 大串憲吾 監訳, 「聴覚心理学概論」 (誠信書房, 東京, 1994) .
- [6] 日本音響学会 編, 「新版 音響用語辞典」 (コロナ社, 東京, 2003) .
- [7] 大賀寿郎, 山崎芳男, 金田豊, 「音響システムとデジタル処理」 (電子情報通信学会, 東京, 2001)
- [8] 今井聖, 「デジタル信号処理」 (秋葉出版, 東京, 1988) .
- [9] 小森智康, 清山信正, 都木徹, “高齢者の聴力を考慮した放送音声サービスの基礎検討”, 映像情報メディア学会2006年冬季大会講演予稿集, 4-5 (2006).
- [10] 原田益水, 「CD技術のすべて」 (電波新聞社, 東京, 1996) .
- [11] TEDDIGITAL, 「世界初の実用 P C M 録音機 DENON DN-023R」 , <http://www.teddigital.com/DN023RADC.htm>
- [12] 岡俊雄, 「マイクログループからデジタルへ(下)」 (ラジオ技術社, 東京, 1981)
- [13] オーディオ回顧録, 「SONY PCM-1」 , <http://page.freett.com/knisi/pcm-1.html>
- [14] Digital Audio Denmark, “A state-of-the-art Multi-channel PCM/DSD/DXD Converter”, http://www.digitalaudio.dk/Press_Axion_0405.htm
- [15] W. B. Snow, “Audible Frequency Ranges of Music, Speech, and Noise”, *J. Acoust. Soc. Am.*, **3**, pp. 155-166 (1931).
- [16] L. J. Sivian and S. D. White, “On Minimum Audible Sound Fields”, *J. Acoust. Soc. Am.*, **4**, pp. 288-321 (1933).
- [17] D. K. Gannett and I. Kerney, “The discriminability of Changes in Program Bandwidth ”, *Bell Sys. Tech. J.*, **23**, pp. 1-10 (1944).
- [18] S. A. Fausti, R. H. Frey, D. A. Erickson, B. Z. Rappaport, E. J. Cleary and R. E. Brummet, “A system for evaluating auditory function from 8000-20000 Hz”, *J. Acoust. Soc. Am.*, **66**, pp. 1713-1718 (1979).

- [19] G. H. Plenge, H. Jakubowski and P. Schöne, “Which Bandwidth is Necessary for Optimal Sound Transmission”, presented at the 62nd Convention of the Audio Engineering Society, convention paper 1449 G-6 (1979).
- [20] 田辺逸雄, 藤田尚, “番組音伝送における上限周波数の検討”, 日本音響学会講演論文集, pp. 231-232 (1979.6).
- [21] T. Muraoka, M. Iwahara and Y. Yamada, “Examination of Audio-Bandwidth Requirements for Optimum Sound Signal Transmission”, *J. Audio Eng. Soc.*, **29**, pp. 2-9 (1981).
- [22] T. Oohashi, E. Nishina, N. Kawai, Y. Fuwamoto and H. Imai, “High-frequency Sound above the Audible Range Affects Brain Electric Activity and Sound Perception”, presented at the AES 91st Convention of the Audio Engineering Society, convention paper 3207 W-1 (1991).
- [23] T. Oohashi, E. Nishina, M. Honda, Y. Yonekura, Y. Fuwamoto, N. Kawai, T. Maekawa, S. Nakamura, H. Fukuyama and H. Shibasaki, “Inaudible High-Frequency Sounds Affect Brain Activity: Hypersonic Effect”, *J. Neurophysiology*, pp. 3548-3558 (2000).
- [24] 松島博行, 崔鍾仁, 堀田健治, 山崎憲, “超音波領域の音が人間に与える生理・心理的影響に関する検討”, 日本音響学会講演論文集, pp. 539-540 (2002.3).
- [25] S. Yoshikawa, S. Noge, T. Yamamoto and K. Saito, “Does High Sampling Frequency Improve Perceptual Time-Axis Resolution of Digital Audio Signal?”, presented at the 103rd Convention of the Audio Engineering Society, convention paper 4562 I-3 (1997).
- [26] K. Ashihara, S. Kiryu, K. Kurakata, N. Nakamura, A. Shibasaki and T. Iwaki, “Perceptual effects caused by high- and ultrasonic-frequency components in musical sounds”, *Proc. AES 9th Regional Convention*, 18-21 (2001).
- [27] K. Kurakata, N. Nakamura, A. Shibasaki, T. Iwaki, K. Ashihara and S. Kiryu, “Perceptual effects of high- and ultrasonic-frequency components in musical sounds”, *Proc. AES 9th Regional Convention*, pp. 22-25 (2001).
- [28] T. Nishiguchi, K. Hamasaki, M. Iwaki and A. Ando, “Perceptual discrimination between musical sounds with and without very high frequency components”, presented at the 115th Convention of the Audio Engineering Society, convention paper 5876 (2003).
- [29] 宮坂栄一, “高周波音の知覚について”, 日本音響学会誌, **55**(8), pp. 569-572 (1999).
- [30] D. A. Webster and D. T. Blackstock, “Finite amplitude saturation of plane sound wave in air”, *J. Acoust. Soc. Am.*, **62**, pp. 518-523 (1977).
- [31] M. O. J. Hawksford, “High-Definition Digital Audio in 3-Dimensional Sound Reproduction”, presented at the AES 103rd Convention of the Audio Engineering Society, convention paper 4560 (1997).
- [32] 蘆原郁, 桐生昭吾, “複合音における可聴周波数帯域外成分の聴取実験”, 日本音響学会講演論文集, pp. 467-468 (1998.9).

- [33] D. Griesinger, “Perception of mid frequency and high frequency intermodulation distortion in loudspeakers, and its relationship to high-definition audio”, AES 24th International Conference, Presentation Slides, retrieved from <http://world.std.com/~griesngr/intermod.ppt> (2003).
- [34] K. Hamasaki, T. Nishiguchi, K. Ono and A. Ando, “Perceptual Discrimination of Very High Frequency Components in Musical Sound Recorded with a Newly Developed Wide Frequency Range Microphone”, presented at the 117th Convention of the Audio Engineering Society, convention paper 6298 (2004).
- [35] ITU-R Recommendation BS.1116-1, “Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems” (1997).
- [36] 佐藤信, 「統計的官能検査法」 (日科技連, 東京, 2003) .
- [37] 岩城正和, 安藤彰男, 西口敏行, 濱崎公男, 管野靖彦, 今永敬嗣, “回折効果を利用したマイクロホンの超広帯域化の検討”, 電子情報通信学会技術研究報告, EA2003-50, pp. 17-21 (2003).
- [38] A. Ando, M. Iwaki, K. Ono, Y. Kanno and K. Imanaga, “A Wide Frequency Range Microphone using Diffraction Effect”, *Proc. 18th International Congress on Acoustics*, Mo4.E.3, pp. 345-348 (2004).
- [39] H. Levitt, “Transformed Up-down Methods in Psychoacoustics”, *J. Acoust. Soc. Am.*, **49**, pp. 467-477 (1971).
- [40] 蘆原郁, 倉方憲治, 水浪田鶴, 松下一馬, “20 kHzを超える純音の聴覚閾値測定”, 日本音響学会聴覚研究会資料, **33**(5), pp. 335-340 (2003).
- [41] H. Hidaka and S. Kakehata, “Clinical Use of Auditory Function Tests: Nowadays and in the Future”, *Tech. Rep. IEICE*, HDA01-18, pp. 1-7 (2001).
- [42] M. Honda, E. Nishina, N. Kawai, R. Yagi, M. Morimoto, T. Maekawa and T. Oohashi, “Physiological mechanism of the emergence of the hypersonic effect”, *Proc. Spring Meet. Acoust. Soc. Jpn.*, pp. 449-450 (2006).
- [43] S. Yoshikawa, S. Noge, M. Ohsu, S. Toyama, H. Yanagawa and T. Yamamoto, “Sound Quality Evaluation of 96kHz Sampling Digital Audio”, presented at the 99th Convention of the Audio Engineering Society, convention paper 4112 O-3 (1995).
- [44] D. Blech and Min-Chi Yang, “DVD-Audio versus SACD”, presented at the 116th Convention of the Audio Engineering Society, convention paper 6086 (2004).
- [45] T. Nishiguchi, K. Hamasaki, “Differences of Hearing Impressions among Several High Sampling Digital Recording Formats”, presented at the 118th Convention of the Audio Engineering Society, convention paper 6469 (2005).
- [46] 蘆原郁, 桐生昭吾, “周波数帯域の拡張に伴うスピーカの非線形歪の増加”, 日本音響学会誌, **56**(8), pp. 549-555 (2000).
- [47] 蘆原郁, 桐生昭吾, “ハイサンプリングオーディオの現状と課題”, 日本音響学会誌, **62**(12), pp. 871-876 (2006).

付録 略語表

CD

Compact Disc : コンパクトディスク

DAT

Digital Audio Tape : デジタルオーディオテープ

DAW

Digital Audio Workstation : デジタルオーディオワークステーション

DSD

Direct Stream Digital : ダイレクトストリームデジタル

DXD

Digital eXtreme Definition : デジタルエクストリームデフィニション

FIR

Finite Impulse Response : 有限インパルス応答

HP

High Pass : ハイパス

LP

Low Pass : ローパス

LPF

Low Pass Filter : ローパスフィルター

PCM

Pulse Code Modulation : パルス符号変調

SACD

Super Audio CD : スーパーオーディオ CD

関連論文の印刷公表の方法および時期

- (1) 全著者名 : Toshiyuki Nishiguchi, Kimio Hamasaki Kazuho Ono,
Masakazu Iwaki, Akio Ando

論文題目 : Perceptual Discrimination of Very High Frequency Components in
Wide Frequency Range Musical Sound

印刷公表の方法および時期 : Applied Acoustics, Elsevier,
<http://dx.doi.org/10.1016/j.apacoust.2009.01.002>, 2009年2月 Web 上掲載,
Volume 70, Issue 7, pp. 921-934, 2009年7月 誌上掲載

(第3章の内容)

- (2) 全著者名 : Toshiyuki Nishiguchi, Kimio Hamasaki

論文題目 : Differences of Hearing Impressions among Several High Sampling
Digital Recording Formats

印刷公表の方法および時期 : The 118th Convention of the Audio Engineering
Society, convention paper 6469, 2005年5月

(第4章の内容)

参考論文の印刷公表の方法および時期

(第1 著者分のみ)

- (1) 全著者名：西口敏行, 岩城正和, 濱崎公男
論文題目：楽音に含まれる超高域成分の弁別に関する一検討
印刷公表の方法および時期：日本音響学会講演論文集, pp. 719-720, 2003 年 3 月
- (2) 全著者名：Toshiyuki Nishiguchi, Kimio Hamasaki, Masakazu Iwaki, Akio Ando
論文題目：Perceptual discrimination between musical sounds with and without very high frequency components
印刷公表の方法および時期：The 115th Convention of the Audio Engineering Society, convention paper 5876, 2003 年 10 月
- (3) 全著者名：西口敏行, 濱崎公男, 岩城正和, 安藤彰男
論文題目：広帯域収録音源による超高域成分の主観弁別実験
印刷公表の方法および時期：日本音響学会講演論文集, pp. 581-582, 2004 年 3 月
- (4) 全著者名：西口敏行, 濱崎公男
論文題目：ハイサンプリングデジタルオーディオと聴感についての一検討
印刷公表の方法および時期：日本音響学会講演論文集, pp. 483-484, 2005 年 3 月

著者略歴

西口 敏行

- 1970年10月 東京都に生まれる。
- 1994年3月 電気通信大学 電気通信学部 電子情報学科 卒業。
- 1994年4月 電気通信大学大学院 情報システム学研究科
情報ネットワーク学専攻 博士前期課程 入学。
- 1996年3月 電気通信大学大学院 情報システム学研究科
情報ネットワーク学専攻 博士前期課程 修了。
- 1996年4月 日本放送協会 入局。
- 1998年7月 日本放送協会 放送技術研究所 着任。次世代音響デバイス、
ハイレゾリューションオーディオ、高臨場感音響の研究に従事。
- 2005年4月 電気通信大学 情報システム学研究科
情報ネットワーク学専攻 博士後期課程 入学。
- 2008年3月 電気通信大学 情報システム学研究科
情報ネットワーク学専攻 博士後期課程 単位取得済退学。
- 2009年3月 博士（工学）（電気通信大学）。