

(410) 情報その他

音楽経験に依存しないハイレゾリューション音源の識別

Auditory Discrimination for High-Resolution Audio without Musical Experience

福田 祐樹[†]

石光 俊介[†]

Yuki Fukuda[†]

Shunsuke Ishimitsu[†]

[†]広島市立大学 大学院情報科学研究科

1 緒言

近年、コンパクトディスク(CD)よりもサンプリング周波数や量子化ビット数の大きい「ハイレゾリューション音源」(以下、ハイレゾ音源)が普及を始めている。それに伴い、ハイレゾ音源と既存の音源(以下、非ハイレゾ音源)の識別可能性について、様々な検討が行われている。

西口ら^[1]は、21 kHz以上の高域を含む音源を独自に収録し、その音源と非ハイレゾ音源を被験者に比較させた。実験の結果、複数人の被験者が全ての刺激を識別できたものの、4名の音楽大学生、1名のバイオリニストと1名の録音エンジニアの合計6名に対し行われた音源のフォーマット(WAV, FLAC, DSDなど)の識別実験では有意差が得られなかったとした。

山本ら^[2]は、27名の被験者に対し、ハイレゾ音源とCD音源を比較させたところ、57%の割合で識別できたと報告した。また、被験者のうち、楽器演奏や日常的なハイレゾ音源の鑑賞の経験のある被験者を含む9名の被験者が100%の割合で全ての刺激を識別し、さらに量子化ビット数の違いに関する比較でも高確率で識別する傾向が得られたと報告した。

勝呂ら^[3]は、量子化ビット数の異なる2つの音源を被験者のそれぞれの耳にヘッドホンを通して呈示したところ、量子化ビット数の大きい音源の方向に音像定位が生じて聴こえたと報告した。

本研究では、先行研究の多くが楽器を用いた音源(以下、楽音)のみを刺激に用いて実験を行っていることに着目し、楽音を用いずにハイレゾ音源の識別の可能性について検討した。

2 ハイレゾ音源の定義

ハイレゾ音源の定義については、様々な組織が公式に発表している^[4,5]。ここでは、本研究におけるハイレゾ音源の定義を明確にするためにその定義の一部を紹介する。

2.1 電子情報技術産業協会 (JEITA)

JEITAはハイレゾ音源を定義する前に、「CDフォーマット」を「44.1 kHzまたは48 kHzのサンプリング

リング周波数を持ち、16 bitの量子化ビット数を持つデジタル音源」と定義した^[4]。さらに、「ハイレゾ音源を「CDフォーマットよりもサンプリング周波数あるいは量子化ビット数が大きく、なおかつもう一方の数値もCDフォーマット以上の音源」と定義した^[4]。表1にJEITAによるハイレゾ音源の定義の例を示す。なお、Fsはサンプリング周波数、bitsは量子化ビット数を示す

表1. JEITAによるハイレゾ音源の詳細例^[4]

Fs (kHz)	bits (bit)	ハイレゾ音源
44.1	16	×
44.1	24	○
48	16	×
96	16	○
192	12	×

2.2 アメリカレコード協会 (RIAA)

RIAAはハイレゾ音源を「アーティストやプロデューサー、エンジニアが独自に作成した、CD品質(サンプリング周波数48 kHz、量子化ビット数20 bit)を超えるマスター音源から作成された、全ての周波数帯域を収録できる、損失のない音源」と定義した^[5]。

本研究では、JEITAの定義に則ってハイレゾ音源を定義した。

3 ハイレゾ音源の評価における問題点

ここでは、本研究の動機を明らかにするために、ハイレゾ音源の識別に楽音を用いることによって生じる問題点を4つ列挙する。

- 問題1: もし研究者が限定的な種類の楽音を用いていた場合、異なる種類の音源を用いた際に実験結果が異なる可能性がある。
- 問題2: 結果が被験者の音楽的経験に左右される可能性がある。つまり、研究者が被験者の音楽の聴取や楽器演奏の経験を踏まえて、注意して刺激を選定する必要がある^[6]。
- 問題3: 先行研究により、楽器演奏経験者と未経験者の間に音楽を聴いた際の身体反応や情

動が優位に異なるとする報告がある⁶⁾。したがって、被験者の音楽的な経験の有無が主観評価の結果に影響を及ぼす可能性がある。

- ・ 問題 4: 複数の種類の音楽において、非ハイレゾ音源とハイレゾ音源のマスターが異なる場合があり、同じタイトルでも周波数特性が異なる可能性がある⁷⁾。

これらの問題点に基づき、一般性のある考察を得るために楽音を用いずにハイレゾ音源の識別を検討する必要があると考える。

4 評価用信号の選定

主観評価実験の準備として、刺激にホワイトノイズを選定し、リニア PCM 信号となるように量子化⁸⁾を行った。

楽音の代わりに、実験刺激としてメルセンヌ・ツイスタにより生成されたホワイトガウシアンノイズ（以下、ホワイトノイズ）を以下の条件に基づいて選定した。

- ・ 条件 1: 信号が平坦な周波数特性を持つこと
- ・ 条件 2: 楽音ではないため、被験者の音楽的な経験が実験結果に影響を及ぼす可能性が小さいこと
- ・ 条件 3: 限られた区間で様々な振幅値を持つため、量子化ビット数の影響を検討できること
- ・ 条件 4: 信号の再生時間の影響を検討できること

本検討では、サンプリング周波数が 48 kHz, 96 kHz, 192 kHz のホワイトノイズを作成し、CD 音源と同じ符号付き 16 bit 整数に量子化した。

5 実験環境

ホワイトノイズにおけるハイレゾ音源と非ハイレゾ音源の識別の可能性を検討するため、ABX 法と MUSHRA の 2 つの実験を行った。実験環境を表 2 に示し、実験系の概略図を図 1 に示す。

表 2. 実験環境

実験室	無響室
USB-DAC	FOSTEX HP-A4BL
ヘッドホン	Sennheiser HD-650
スピーカ	ECLIPSE TD-M1

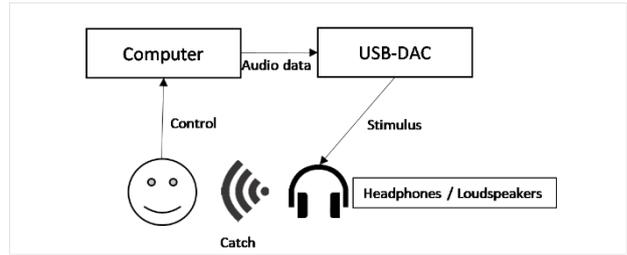


図 1. 実験系の概略図

なお、実験系に利用した USB-DAC やスピーカ、ヘッドホンは全てハイレゾ音源の再生に対応したものを利用し、各装置において 80 kHz までの出力を確認している。

6 ABX 法によるハイレゾ音源識別実験⁹⁾

最初に、サンプリング周波数の違いを識別できるかを調べるために、刺激 A, 刺激 B, 刺激 X を繰り返し聴取した後、刺激 X が刺激 A と刺激 B のどちらと同じかを判断する ABX 法⁹⁾を行った。

6.1 実験方法

まず、被験者の負担を極力軽減するために、被験者自身にホワイトノイズを聴きながら再生系の音量を調整してもらう。その後、刺激 A, 刺激 B, 刺激 X を繰り返し聴いてもらい、刺激 X が刺激 A と刺激 B のどちらと同じものかを判断してもらう。

刺激 A と刺激 B にはそれぞれサンプリング周波数の異なるホワイトノイズが格納されており、刺激 X にはランダムに刺激 A あるいは刺激 B と同じ信号が格納される。刺激 A と刺激 B の信号も解答のたびにランダムに入れ替わるようにし、二重盲検法となるように実験系を構築した。

偶然による正答率を下げるため、回答数は一つの組み合わせに対し 4 問とし、組み合わせが 3 つ、さらに呈示環境が 2 つ（ヘッドホンとスピーカ）あるため、被験者一人あたりの合計試行数は 24 とした。表 3 に実験条件を示す。

表 3. ABX 法の実験条件

被験者数	7
男女比	男:5 女:2
年齢	22.3±1.6
1 人あたり試行数	24
呈示環境	ヘッドホン/スピーカ

6.2 実験結果

ABX 法により得られたデータを有意水準 0.05 の二項検定により検定した。表 4 に各呈示環境における正答数を示す。

表 4. 正答数

サンプリング周波数	ヘッドホン	スピーカ
48 kHz vs. 96 kHz	19	22
48 kHz vs. 192 kHz	17	21
96 kHz vs. 192 kHz	20	21

表 5 と表 6 にヘッドホンとスピーカにおける正答数の二項検定の結果をそれぞれ示す。表 5 と表 6 より、多くの条件において識別に有意差が得られていることがわかる。しかし、一部の条件においてサンプル数が不十分（検出力 0.8 未満）であるため、今後も継続した実験によるデータの収集が不可欠である。

表 5. ABX 法における二項検定結果（ヘッドホン）

サンプリング周波数	p 値 (*: $p < 0.05$)
48 kHz vs. 96 kHz	0.0178*
48 kHz vs. 192 kHz	0.0925
96 kHz vs. 192 kHz	0.0063*

表 6. ABX 法における二項検定結果（スピーカ）

サンプリング周波数	p 値 (*: $p < 0.05$)
48 kHz vs. 96 kHz	0.0005*
48 kHz vs. 192 kHz	0.0019*
96 kHz vs. 192 kHz	0.0019*

7 MUSHRA^[10]によるハイレゾ音源の主観評価実験

各サンプリング周波数を持つホワイトノイズを聴取した際の聴感印象差を比較するために MUSHRA^[10]を行った。

7.1 実験方法

最初に、ABX 法と同様に被験者に再生系の音量を調整してもらい、その後、192 kHz のサンプリング周波数を持つホワイトノイズを基準音として被験者が把握するまで繰り返し呈示し続け、基準音を記憶してもらい。

基準音の呈示後、サンプリング周波数 192 kHz, 96 kHz, 48 kHz のホワイトノイズがそれぞれランダムに格納された刺激 A, 刺激 B, 刺激 C と基準音を制限なく呈示し、被験者に各刺激の点数を基準音を 100 点として採点してもらい。つまり、192kHz のホワイトノイズに近いと感じれば点数は 100 点に近く、異なると感じるほど 0 点に近い点数をつける。この採点をヘッドホンとスピーカの両方の呈示環境で行った。

実際の規格からルールを変更した点として、本来の MUSHRA では基準音に最も近い刺激音に 100 点

を採点する必要があるが、今回は 192 kHz の点数も解析に用いるべく、この規則は除外している。

採点された点数を有意水準 0.05 の二元配置分散分析で解析した。

表 7 に MUSHRA の実験条件を示す。なお、被験者 7 名のうち 6 名は ABX 法と同一の被験者である。

表 7. MUSHRA の実験条件

被験者数	7
男女比	男:6 女:1
年齢	22.1±0.98
呈示環境	ヘッドホン/スピーカ

7.2 実験結果

表 8 に二元配置分散分析の結果を示す。なお、Fs はサンプリング周波数を示す。また、交互作用はサンプリング周波数と呈示環境の交互作用である。表 8 より、サンプリング周波数のみ有意な主効果が得られているため、有意水準 0.05 の多重比較検定 (Steel-Dwass 検定) により各サンプリング周波数の組み合わせに対する得点の分布の有意差を調べた。

表 8. 二次元配置分散分析結果

要因	自由度	F 値	p 値 (*: $p < 0.05$)
呈示環境	1	0.10	0.7480
Fs	2	7.09	0.0022*
交互作用	2	2.64	0.0829

表 9 に多重比較検定の結果を示す。なお、Fs はサンプリング周波数を示す。表 9 より、最もサンプリング周波数の間隔の大きい 48 kHz と 192 kHz のサンプリング周波数を持つホワイトノイズの分布に有意差があることがわかる。

表 9. 多重比較検定結果

Fs	t 値	p 値 (*: $p < 0.05$)
48 kHz vs. 96 kHz	2.9907	0.0868
48 kHz vs. 192 kHz	3.9015	0.0160*
96 kHz vs. 192 kHz	2.2424	0.2158

8 結言

本検討では、ハイレゾ音源の識別に楽音を用いることによる一般性の欠如の可能性という問題点から、楽音以外の信号としてホワイトノイズを用いたハイレゾ音源と非ハイレゾ音源の識別可能性について検討を行った。ABX 法では多くの組み合わせにおいてサンプリング周波数の違いを有意に識別できるとする結果が得られた。また、MUSHRA では聴取環境

の影響なくサンプリング周波数の違いでのみ得点に有意な主効果が得られ、多重比較においてもその点数に有意差が得られるサンプリング周波数の組み合わせの存在が認められた。したがって、ホワイトノイズを用いてハイレゾ音源と非ハイレゾ音源を識別できる可能性があると考ええる。

しかし、7人という被験者数の少なさや、実験系が1系統のみなど、客観的に識別できると考察できるかが疑問であるため、これらの実験環境については今後被験者数を増やす、実験系の特性を詳細に計測するなどして改善する。

今後は、本検討の実験の継続によるサンプルの採集によって統計的解析の検出力を強化することに加え、量子化ビット数の違いや音源の再生チャンネル数（ステレオ、モノラル）の要因による識別の可能性について検討を行う。また、それらの結果と被験者の音楽関係との相関関係を調査し、音楽経験に依存しないハイレゾ音源の識別方法について引き続き検討する。

参考文献

- [1]. 西口, “ハイレゾリユーションオーディオの研究”, 電気通信大学博士論文 (2009).
- [2]. 山本, 金只, 水町, “ハイレゾリユーションオーディオの音質評価”, 産業応用工学会論文誌, Vol. 1, No. 2, pp.52-57 (Sept. 2013).
- [3]. A. Suguro and M. Miura, “Quality discrimination on high-resolution audio with difference of quantization accuracy by sound image localization”, AES Conference on Spatial Reproduction, e-Brief No.74 (2018).
- [4]. 照井, 西田, “ハイレゾオーディオの呼称について (周知)”, JEITA (2014, March), [online]<https://home.jeita.or.jp/page_file/20140328095728_rhsiN0Pz8x.pdf>, 2019年8月19日アクセス.
- [5]. Record Industry Association of America, “High Resolution Audio Initiative Gets Major Boost with New “Hi-Res MUSIC” Logo and Branding Materials for Digital Retailers”, RIAA (2015), [online]<<https://www.riaa.com/high-resolution-audio-initiative-gets-major-boost-with-new-hi-res-music-logo-and-branding-materials-for-digital-retailers/>>, 2019年8月19日アクセス.
- [6]. 安田, “音楽聴取による感動の心理学的研究—聴取者の音楽経験から見た感動, 身体反応, 情動—”, 日本心理学会第76回発表論文集, 3PMA13 (2012).
- [7]. ONKYO, “e-onkyo music”, ONKYO (2019), [online]<<https://www.onkyo.com/news/2514/>>, 2019年8月19日アクセス.
- [8]. 金井, “信号処理 第2回講義”, 北海道大学 (2018), [online]<<http://sdmwww.ssi.ist.hokudai.ac.jp/lecture/signal/presen2.pdf>>, 2019年8月19日アクセス.
- [9]. W. A. Munson and M. B. Gardner, “Standardizing Auditory Tests”, The Journal of the Acoustical Society of America, Vol. 22, pp. 675 (1950).
- [10]. International Telecommunication Union, “Recommendation ITU-R BS. 1534-3 (10/2015): Method for the subjective assessment of intermediate quality level of audio systems” (2015).