

聴感特性を考慮した効率的な高解像度音の提案と快適性評価 Proposal and Comfort Evaluation of Efficient High Resolution Sound in Consideration of Auditory Characteristics

精密工学専攻 15 号 奥原 裕幸
Hiroyuki Okuhara

1. はじめに

手軽に入手できる高品位なオーディオコンテンツとして高解像度音源が注目されている。高解像度音源とは、CD等の従来音源のスペックを超えた情報量を持つ音源のことを指し、パソコンを利用し音楽データを保存することが一般的になったことにより、オーディオ愛好家を中心に広まりつつある。しかしながら、高解像度音源は従来音源よりも情報量が多いにもかかわらず、聴感上の高品位性が認められ難い⁽¹⁾ことや、最近は無線通信をはじめとした手軽に音楽を楽しむ方式が一般に浸透していることが、高解像度音源のさらなる普及に向けての課題となっている。

そこで本研究では、どのような音が従来音源と高解像度音源との音質の差を知覚したり、生体へ影響を与えるかを把握するため、主観的および客観的手法により解像度の異なる音源を用いた音質評価を実施する。その際、音質の違いの影響を独自の信号指標や聴覚の等感特性に着目した考察を行うことで音の持つ特徴を把握する。次に、把握した特徴を維持したまま模擬的に情報量を減少した効率的な高解像度音として量子化ビット数変動音を提案する。さらに、主観的および客観的手法により量子化ビット数変動音の高音質性や快適性を評価し、その有用性を検討する。

なお、本研究における音質評価について、被験者に対し十分なインフォームドコンセントを行った上で実施している。

2. 様々な音楽の音質評価

本章では、様々な種類の音楽の解像度の違いを主観および客観評価による音質評価によって把握する。

2.1 SD法による音質評価

5種類の音楽について、高解像度音源と従来音源のSD法の相対評価により2音源間に生ずる聴感上の差異を把握し、物理特性との関係を考察する。

2.1.1 実験条件

評価方法として、SD法の7段階相対評価を用いる。評価する形容詞対は明瞭性、空間性、迫力性を考慮して選定した計18種類である。洋楽女声クラシック(CL1)、洋楽男声ジャズ(JZ1)、洋楽男声ロック(R01)、邦楽女声アニメソング(AN1)、自然音波音(NA1)の5種類の音楽について、基準音として従来音源であるサンプリング周波数が44.1kHz、量子化ビット数が16bit(以下、44/16と記す)、評価音として高解像度音源である96/24を用いる。基準音と評価音の比較をしやすくするため、モニタに音楽プレーヤを映し出し、基準音と評価音を順序自由に何度でも再生できるようにする。被験者は手元にあるコントローラを使いプレーヤを操作し、納得いくまで音源の比較を行った後、各形容詞対の基準音に対する評価音の印象の得点を評価用紙に記入する。

音質評価は無響室で行い、音源は2mの間隔で設置されたB&W社製スピーカCM6 S2とMarantz社製アンプHD-AMP1を用いて再生する。スピーカから被験者までの距離は2mで、各音源間に音の大きさのばらつきが生じないように、頭部中心位置のA特性ピーク音圧レベルが約90dBAとなるように音

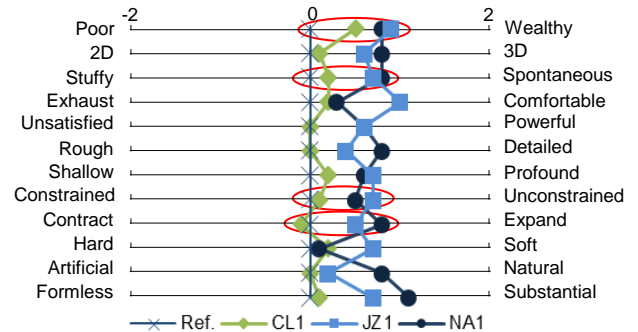


Fig. 1 Profile diagram of CL1, JZ1, NA1

量を調節する。被験者は20代10名である。

2.1.2 実験結果と考察

Fig. 1に代表的な形容詞対について、CL1, JZ1, NA1の96/24の全被験者平均のプロフィールを示す。

CL1, AN1について、この2音楽は基準音と評価音に各形容詞対における明確な差異がほとんどない。このことから、音源の解像度の違いは聴感上の差異としては非常に小さく、音楽によってはその差を聴き分けられないことがわかる。

JZ1, R01, NA1については、基準音と評価音に差がない形容詞対もあるが、いくつかの形容詞対で差異がみられた。特に、「瘦せた-豊かな」「詰まった-伸びのある」「のびやかである-のびやかでない」「広がりがない-広がりのある」は3つの音源で0.5ポイント以上の差があり、それぞれ「豊かな」「伸びのある」「のびやかである」「広がりのある」の方向に得点が振れている。つまり、音源の解像度の違いは、空間性に関連する感性に影響を与えやすいと考えられ、高解像度音源はより広々とした音空間を感じられると推察される。

2.1.3 Signal Differenceの定義と評価結果との関係⁽²⁾

2.1.2の結果より、解像度が異なることが、必ずしも聴感上の差異とならないことが考えられた。そこで、信号における解像度の差の度合いが聴感上の音質差異へ影響を与えると推測し、その度合いを可視化する手段として信号差分(Signal Difference: SD)を定義する。

Fig. 2のように、高解像度音源とCD音質程度である従来音源の音声信号を時間軸で描画し拡大すると、サンプリング周波数および量子化ビット数が異なるため、波形は完全には一致しない。そこで、この信号の差分をとり、絶対値にしたものをSDとする。

デジタル信号では、サンプリングの概念があるため、サンプリング周波数の異なる2信号間の差をそのまま取ることはできない。そこで、従来音源を横軸方向に引き伸ばし、高解像度音源のサンプリング周波数にあわせることで、データ上では高解像度音源と同じサンプリング周波数を持つが、見かけ上は従来音源と同等の信号波形となるようにする。従来音源の信号を横軸方向に引き伸ばすとき、縦軸の値は直前の値と同値で補充し、計算を行う。

Fig. 3に、CL1, JZ1, NA1の信号差分を音声信号の最大値で割ることで正規化したのち、10ms毎の移動平均によって

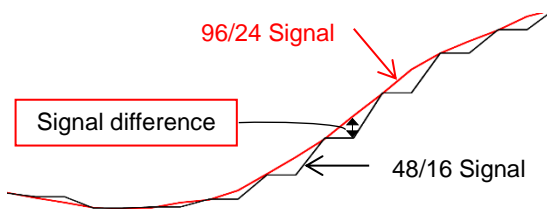


Fig. 2 Definition of SD

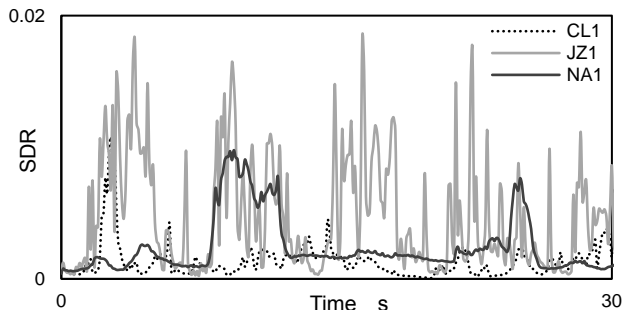


Fig. 3 SDR of CL1, JZ1, NA1

平滑化して算出した Signal Difference Ratio (以下, SDR) の時間変動を示す。

CL1 の SDR は他の音楽と比較しても値が小さく、大小変動も多くない。つまり、SDR が小さいことで、聴感上の音質差も小さいと考えられる。一方、JZ1 の SDR は全体的に値が大きく、時間に合わせて大小の変動が多い傾向がある。また、その様な傾向が全時間帯で一様に現れている。つまり、SDR が大きく、大小変動が多い音源は聴感上の音質差も感じやすくなると考えられる。以上より、高解像度音源と従来音源の間に聴感上の音質差を生じやすい音楽は、SDR が一様に大きく、大小変動が多い特徴を持つと考えられる。

NA1 は上述のような特徴を持っていないが、10 秒付近に SDR が大きくなる瞬間があり、音楽全体の中でも特徴的なシーンである。結果、被験者の評価対象区間となり、聴感上の音質差に影響を与えたと推測される。

以上より、高解像度音源と従来音源の間に聴感上の音質差が生じる場合は、物理的な特徴として、信号差分の大きさが影響を与えていると考えられる。

2.2 心電計測による音質評価

本節では、心電計測によって安静状態で高解像度音源と従来音源の音楽を聴いた際の生体への影響を把握する。

2.2.1 実験条件

実験方法は、心電計測によるストレス負荷評価を行う。計測した心電図からとり合う R 波の間隔 (RRI: R-R interval) を取り出し、RRI から RRI の中央値 (RRIm) と、横軸に n 番目の RRI、縦軸に n+1 番目の RRI をプロットしたローレンツプロットの面積⁽³⁾ (LPs) を算出する。評価音楽は波の音である NA2 とロックジャンルの RO2 を用い、音源はそれぞれ 96/24 と 48/16 を用意する。被験者は計測装置を装着した状態で着席する。生体情報への視覚的な影響を排除するために評価室は消灯する。評価の流れとして、90 秒間の無音区間の後、120 秒間音源を提示する。提示される音源は NA2, RO2 どちらかの 96/24 または 48/16 の解像度である。これを 1 セットとし、NA2, RO2 それぞれ 2 音源ずつ、計 4 セットを実施する。心電計測は無音区間の後半 60 秒間と音源提示区間の後半 60 秒間を計測し評価に用いる。実験条件は 2.1 節と同様で、被験者は 20 代 4 名とする。

2.2.2 実験結果と考察

評価で得られた心電データは、音源提示区間の RRIm, LPs の値を無音区間の RRIm, LPs の値で割り正規化することで、

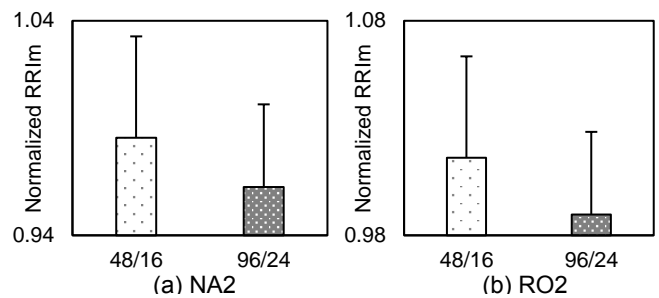


Fig. 4 Average RRIm

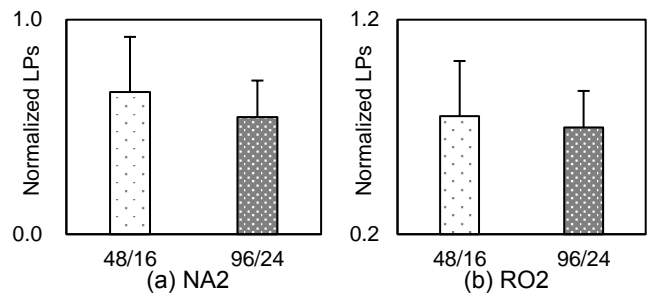


Fig. 5 Average LPs

無音区間のストレス負荷から音源提示区間のストレス負荷の変化率を算出する。RRIm は値が小さいほど心拍間隔が短くなることからストレス負荷状態、LPs は値が小さいほど心拍揺らぎが少なくなることからストレス負荷状態と考える。

Fig. 4 に各音楽の正規化した RRIm の平均値、Fig. 5 に各音楽の正規化した LPs の平均値を示す。全体的な傾向として 48/16 と比較して 96/24 の RRIm, LPs の値が小さくストレスが負荷される傾向である。しかし、音源間の値の差は大きくなく、高解像度音源を感受したことで不快感が大きく上昇するとは考えにくい。つまり、高解像度音源は従来音源と比較してわずかに活動的な生理状態となることが考えられ、緊張感や覚醒感として生理活動に影響する可能性が考えられる。

2.3 脳血流計測による音質評価

本節では、脳血流量計測によって安静状態で高解像度音源と従来音源の音楽を聴いた際の生体への影響を把握する。

2.3.1 実験条件

実験方法として、近赤外線分光法 (NIRS: Near-infrared spectroscopy) による酸化ヘモグロビン濃度 (Oxy-Hb) の計測を行う。計測には光トポグラフィ装置を用い、計測対象部位は前頭葉の 52 ch とする。評価音楽は波の音である NA3 を用い、音源は 96/24 と 48/16 を用意する。被験者は計測装置を装着した状態で着席する。生体情報への視覚的な影響を排除するために評価室は消灯する。評価の流れとして、60 秒間の無音区間の後、300 秒間 96/24 あるいは 48/16 の音源を提示する。被験者への疲労を考慮して評価は 1 日に 1 音源とし、2 日間にわけて実施する。評価場所および使用機器は 2.1 節と同様で、被験者は 20 代 4 名とする。

2.3.2 実験結果と考察

被験者毎に取得した 52 ch の Oxy-Hb 時間変動データから、相関係数が 0.6 以上のチャンネルを自動抽出し、加算平均することで、前頭用の Oxy-Hb の時間変動とする。Fig. 6 は音源が提示されている時間帯の 96/24 と 48/16 の全被験者の Oxy-Hb の時間変動平均値を示している。Fig. 6 より、前半部分はどちらの音源も Oxy-Hb が減少し、およそ 180 秒を境に 96/24 は再び Oxy-Hb が減少することが示されている。主観的に優れたシステムでは従来システムと比較して Oxy-Hb が減少することが報告されており⁽⁴⁾、本評価においても 96/24 の解像度が優れていたことが Oxy-Hb の減少に影響を与えたと考えられる。

ここで、Fig. 7 に 48/16 と、96/24 の Oxy-Hb 変動に差が

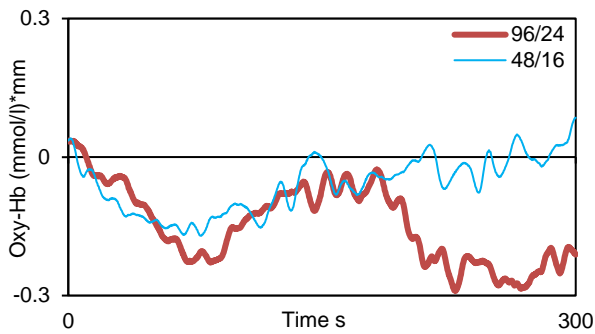


Fig. 6 The average of all subjects with time change of Oxy-Hb

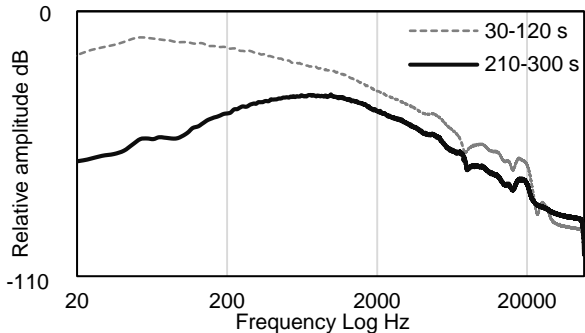


Fig. 7 Frequency characteristics by time zone

ない前半区間 (30-120 s) と差が現れた後半区間 (210-300 s) それぞれ 90 秒間ずつの周波数特性を示す。Fig. 7 より、相対的に約 1 kHz 付近を中心とした中高音域が相対的に大きい特性を持つとき、96/24 の Oxy-Hz が低下していると考えられる。これは、人の聴覚の等感曲線を考慮すると、約 1~6 kHz は最小可聴値が小さくなる帯域であることから、この帯域が目立った時間帯における解像度の違いが脳活性に影響したと考察する。

3. 量子化ビット数変動音の提案⁽⁵⁾

本章では、聴感特性を考慮した効率的な高解像度音として量子化ビット数変動音を提案する。

3.1 量子化ビット数変動音の概説

2章の実験結果から、信号差分の大きい時間帯を多く含む特徴や最小可聴値が小さくなる帯域が相対的に大きくなる周波数特性を持つ音が、従来音源と高解像度音源の解像度の差を主観的および客観的に聴き分けられることが考えられた。以上の特性を考慮し、適切に解像度を変動させることで、人の聴覚系に対してより効率的に高音質な音源を伝達できると推測される。そこで、特性を考慮した設定条件を基に微小時間単位で量子化ビット数が変動する量子化ビット数変動音 (以下、QBVP) を提案する。

QBVP は、96/24 音源を 0.1 s 毎のセグメントに分割し、それぞれのセグメントについて、周波数領域と時間領域の設定条件によって、量子化ビット数を 24 bit に維持するか、24 bit の下位 8 bit をゼロ埋めし、見かけ上 16 bit 音質に変換するかを決定する。16 bit に変換する場合は、周波数領域の条件と時間領域の条件両方を満たしている。以下に周波数領域および時間領域の条件を示す。

3.2 周波数領域の設定条件

周波数領域において、量子化ビット数を 24 bit に維持して高解像度性を保つねらいとして、聴感上感度の良い帯域の解像度の強調がある。そこで、A 特性の補正值を考慮し、1~6 kHz の帯域に着目する。この帯域の 0.1 s 毎のパワースペクトルの全時間の中央値を求め、これを閾値として各セグ

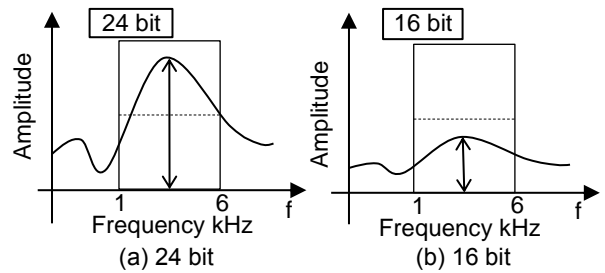


Fig. 8 Setting condition in frequency domain

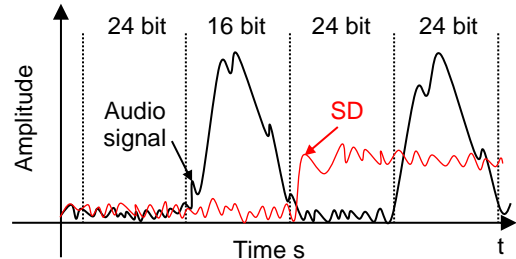


Fig. 9 Setting condition in the time domain

メントの 1~6 kHz 帯のパワースペクトルが閾値を半分以上を超えている場合に 24 bit、半分未満の場合に 16 bit に変換する。以下、処理の流れをまとめると、

- ① 96/24 に対し、1 s 幅オーバーラップ 0.9 s の短時間フーリエ変換を行う。
- ② 全時間の 1~6 kHz のスペクトル密度の中央値 (SPM_{1-6k}) を算出する。
- ③ 0.1 s セグメント毎に 1 Hz 刻みで、スペクトル密度が SPM_{1-6k} より大きい周波数の数 (LSP) を 1~6 kHz 内でカウントする。
- ④ LSP が 1~6 kHz の半分以上の場合は 24 bit を維持し、半分未満の場合は 16 bit に変換する。

周波数領域における設定条件の例を Fig. 8 に示す。

3.3 時間領域の設定条件

時間領域において、量子化ビット数を 24 bit に維持して高解像度性を保つ狙いとして、2章で定義した SD が大きい時間の解像度の強調がある。そこで、音圧のレベルが大きく、かつ SD が小さい時間の量子化ビット数を 16 bit に変換することで、SD が大きくなる時間を強調する。以下、処理の流れをまとめると、

- ① 96/24 と 48/16 の SD を算出する。
- ② 96/24 および SD の 0.001 s 毎の二乗平均平方根 ($RMS_{96/24}$, RMS_{SD}) を算出する。
- ③ $RMS_{96/24}$, RMS_{SD} それぞれの全時間の中央値 $RMSM_{96/24}$, $RMSM_{SD}$ を算出する。
- ④ $RMS_{96/24} > RMSM_{96/24}$ かつ $RMS_{SD} < RMSM_{SD}$ の場合は 16 bit に変換し、それ以外の場合は 24 bit を維持する。

時間領域における設定条件の例を Fig. 9 に示す。

4. 量子化ビット数変動音の音質評価

量子化ビット数変動音を主観評価や客観評価によって音質評価する。

4.1 量子化ビット数変動音の主観評価

様々な音楽の 96/24 と BV を主観評価し、音質への影響を把握する。

4.1.1 実験条件

音楽の種類はアニメソング (AN2)、波音 (NA4)、ジャズ (JZ2, JZ3)、J-POP (P01)、ロック (R03) である。それぞれの音楽で基準音として 48/16、評価音として 96/24 または QBVP を用

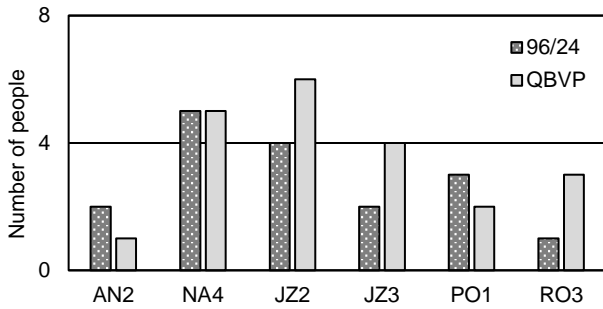


Fig. 10 Number of people who answered that sound is better than 48/16

いる。音源は QBVP の量子化の変動が多い時間帯の 10~20 秒間を切出す。

実験条件は 2.1 節と同様であり、音量は被験者の頭部中心位置における A 特性ピーク音圧レベルが約 90~100 dB となるように調節する。

評価方法は、被験者はブラインドで提示された基準音と評価音のペアを聴き、音質が良いのは「前者」「後者」「同じ」のいずれかを回答する。被験者は 20 代 8 名である。

4.1.2 実験結果と考察

Fig. 10 に、基準音に対し評価音の音質が良いと回答した人数を示す。AN2, PO1 は、96/24 の方が音質が良いと回答した人数が QBVP の方が音質が良いと回答した人数より多いことがわかる。しかし、いずれも過半数よりも人数が少なく、もとより 48/16 の音質が好まれやすい音源であったと考えられる。NA4, JZ2, JZ3, RO3 では、96/24 の方が音質が良いと回答した人数より QBVP の方が音質が良いと回答した人数の方が多いか同数である。つまり、QBVP の音質の良さが 96/24 の音質の良さと同様以上となる。特に、JZ2, JZ3, RO3 について QBVP は 96/24 より情報量が減るため、QBVP の音質が良くなることは不自然に思われるが、ノイズ的な音が含まれる場合、解像度の高さが主観的な音質の低下に影響する可能性もある⁽⁶⁾ため、音楽の内容によって高音質性に適切な解像度が異なるとすると、QBVP のような量子化ビット数の変動によって余分な情報を省いて高音質化できると考察できる。以上より、高解像度音源から情報量を減少させた音源であっても、聴感特性を考慮して適切に処理すれば、主観的な高音質性を維持できると考えられる。

4.2 量子化ビット数変動音の客観評価

本節では、量子化ビット数変動音が快適性へ与える影響を把握するため、安静状態における音源提示時の心拍間隔計測による音質評価を行う。

4.2.1 実験条件

実験方法は、心電計測によるストレス負荷評価を行う。計測した心電図から RRI_m と LPs を算出する。対象音楽は 4.1 節で BV と 96/24 の主観的な音質の良さが同等であった NA4 を用い、音源は BV, 96/24, 48/16 を用意する。被験者は計測装置を装着した状態で着席する。生体情報への視覚的な影響を排除するために評価室は消灯する。

評価の流れとして、90 秒間の無音区間の後、120 秒間音源を提示する。提示される音源 BV, 96/24, 48/16 のいずれかである。これを 1 セットとし、計 3 セットを実施する。心電計測は無音区間の後半 60 秒間と音源提示区間の後半 60 秒間を計測し評価に用いる。

4.2.2 実験結果と考察

評価で得られた心電データは、音源提示区間の RRI_m, LPs の値を無音区間の RRI_m, LPs の値で割り正規化することで、無音区間のストレス負荷から音源提示区間のストレス負荷の変化率を算出する。Fig. 11, Fig. 12 に正規化された RRI_m,

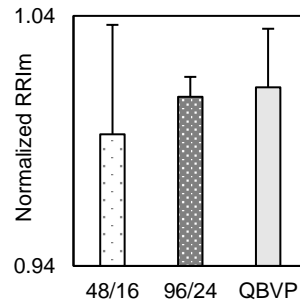


Fig. 11 Average RRI_m

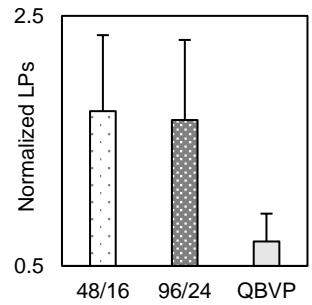


Fig. 12 Average LPs

LPs の各音源の平均値を示す。全体的には可聴域外の超高域成分が含まれる QBVP と 96/24 は RRI_m が大きくなり、比較的にリラックス傾向であると考えられる。48/16 は、LPs, RRI_m ともに被験者間のばらつきが大きく、快適性が被験者によって異なる。96/24 は RRI_m が大きく、LPs も大きくなり、3 音源の中で最も快適性が高いことが考えられる。QBVP は RRI_m が大きくなるが、LPs は他の音源と比較して小さく、ばらつきも小さい。よって大きなストレス負荷はかからないものの、完全なリラックス状態とも言えないため、活動的な生理状態であると考えられる。2 章の結果では 96/24 は比較的活動的な状態となったが、本評価ではばらつきが大きく、活動状態の維持には不安定である。そこで QBVP を用いることで、より効率的に活動状態を維持する可能性が考えられる。

5. 研究成果

- (1) 従来音源と高解像度音源の主観的評価と信号差分の結果から、信号差分が大きくなる場合に解像度の主観的な音質差異が生じることがわかった。
- (2) 生体指標を用いた客観評価により、高解像度音源は緊張感や覚醒感を生体的に付与し、聴覚の等感特性に関連して脳活動へ影響を与えると考えられる。
- (3) 信号差分および聴感特性を考慮し、模擬的に情報量を減少した効率的な高解像度音として量子化ビット数変動音 (QBVP) を提案した。
- (4) 量子化ビット数変動音を音質評価し、特に主観的な高音質性について、多くの音楽で高解像度音源と同等程度の音質を維持することがわかった。

参考文献

- (1) 山本竜太, 二矢田勝行, 水町光徳, ハイレゾ音源の知覚に対する再生環境の違いの影響, 日本音響学会講演論文集(春), (2017) pp. 525-526.
- (2) 奥原裕幸, 有光哲彦, 戸井武司, 音圧および周波数の分解能変更による聴感への影響評価, 日本音響学会講演論文集(秋), (2016) pp. 697-698.
- (3) 豊福史, 山口和彦, 荻原啓, 心電図 RR 間隔のローレンツプロットによる副交感神経活動の簡易推定法の開発, 人間工学, Vol. 43, No. 4(2007) pp. 185-192.
- (4) 浅野裕俊, 石川智治, 野澤昭雄, 棚橋大介, 首藤誠志, 井出英人, 宮原誠, 前頭葉における酸素化ヘモグロビン濃度変化による音楽再生の評価の検討, 日本感性工学学会論文誌, Vol. 9, No. 2(2010) pp. 323-328.
- (5) 奥原裕幸, 戸井武司, 高解像度音源の量子化ビット数変動による機能性評価, 日本音響学会講演論文集(春), (2018) 1-13-17.
- (6) 奥原裕幸, 有光哲彦, 戸井武司, 高解像度音源の原音再現性が聴感へ与える影響評価, 日本音響学会講演論文集(春), (2017) pp. 721-722.