

2010 年度 卒業研究論文

非可逆音声圧縮について

岡山理科大学

総合情報学部

情報科学科

澤見研究室

I07I011 岩澤哲央

I07I026 川田洸輔

I07I048 中津弘迪

目次

1	はじめに	2
2	データの圧縮形式	3
3	非可逆音声圧縮の紹介	4
3.1	MP3	4
3.2	OggVorbis	4
3.3	AAC	5
3.4	WMA	5
4	非可逆音声圧縮の仕組み	6
4.1	最小可聴限	7
4.2	マスキング	8
5	ビットレート	9
6	周波数成分による評価	10
6.1	ビットレート差による比較	10
6.2	フォーマット別の比較	12
7	$1/f$ ゆらぎ	15
7.1	ゆらぎの測定	16
8	まとめ	17
	参考文献	18

1 はじめに

インターネットによる音楽配信サービスやポータブルオーディオなどの普及により、デジタル音声データを扱うことが多くなっている。ところで CD 音質を例に挙げると、1 分当たり約 10MB と非常にサイズが大きいため、パソコンのハードディスク上では扱いにくく、より記憶容量の小さいポータブルオーディオでは保存できる曲数が限られてしまい、実用性に欠ける。

この容量の問題を解決する手段のひとつとして、データ圧縮技術がある。データの表現方法を工夫することで、データのサイズを減らすことができる。現在ではパソコンの普及により、ソフトウェアによる圧縮をしてから汎用メディアへ記録するという方法が一般的である。その中でも私たちがよく利用している方法として、非可逆音声圧縮と呼ばれる音声データの圧縮方法が用いられている。

本研究では、数多くの音声圧縮形式の中から、非可逆圧縮形式のを選び、その中からいくつかを取り上げ、データ量を同じにした場合の音質に関する比較・評価を行う。

2 データの圧縮形式

音声データの圧縮形式には、非可逆圧縮形式と可逆圧縮形式の二つがある。

可逆圧縮とは、データがまったく欠落しない圧縮方式で、可逆圧縮によって圧縮されたデータは完全にデータ内容はそのままでデータを圧縮、復元することができる。プログラムや文字などのデータは、1ビットでも欠落すると内容が変化してしまうことから圧縮するにはこのデータ圧縮の手法が用いられる。

非可逆圧縮とは、多少のデータの欠落はあるものの、劇的に圧縮率が高い手法である。非可逆圧縮方式により圧縮されたデータからは圧縮前のデータを完全に復元することができない。音声や動画などアナログからデジタルデータに変換する際に生じる程度の多少音質を落としても問題のないデータはこの圧縮方式が用いられることが多い。

3 非可逆音声圧縮の紹介

本研究で比較、評価に用いる非可逆音声ファイルフォーマット MP3, OggVorbis, AAC, WMA について紹介する.

3.1 MP3

MP3 は, MPEG Audio Layer-3 の略で, 人間の耳で聞く範囲では極端な音質の劣化がなく圧縮できるため CD など音源の媒体からパーソナルコンピュータのハードディスクドライブに取り込むときのフォーマットとして広く普及した.

MP3 が普及した原因はほかにもある. 初期の MP3 は, 音声データ以外は, 再生に必要なビットレートやサンプリング周波数といった情報しか含まれていなかったが, ID3 というタグ情報を新たに埋め込むことで, 曲名やアーティスト名, アルバム名, トラック番号といった楽曲に関する情報を記録できるようになった. このことで, 再生中の楽曲の情報を確認できるだけでなく, データをアルバムごとに管理したり, トラック番号を順に曲を並べたりと言った情報管理能力の高さを発揮できることにつながった.

MP3 が普及した原動力として, MP3 は無料で自由に利用できるという解釈がなされがちだがそれは誤解である. これは MP3 を含む MPEG フォーマットが ISO によって標準化された規格であり, 自由に利用できる個々の技術に基づくものである. MP3 においては, ドイツの”フラウンフォーファー IIS-A”とフランスの”トムソンマルチメディア”の 2 社が基本特許を取得しており, 1998 年に特許侵害を指摘してライセンス料を求めるメールを, 当時 MP3 のエンコーダー (MP3 圧縮ソフト) を公開した作者たち宛てに送ったことで問題が表面化した. 両社は, MP3 再生に必要なデコーダひとつにつき 0.5 ドル. 圧縮に必要なエンコーダーに 2.5 ドル. MP3 で音楽配信した場合, 1 ダウンロードごとに料金の 1%, 最低でも 0.01 ドルのライセンス料の支払いを求めており, 現在 MP3 を利用できるソフトを販売している会社はこの 2 社とライセンス契約を結んでいる. 例えば, Windows は標準の機能で MP3 を再生できるが, これはユーザーが気づいていないだけで, OS の料金に MP3 の使用料が含まれていることになる. 拡張子は.mp3 である.

3.2 OggVorbis

OggVorbis はクリストファー・モンゴメリ氏率いるチームの開発した, MP3 の代替を目的としたオープンソースの音声圧縮フォーマットである. MP3 はさきほど紹介したように, ライセンス料が発生しているのに対し, OggVorbis は特許技術を一切含まず, 将来にわたってライセンス料が徴収される心配のない, ライセンスフリーな規格であることが宣言されている.

”Xiph.org Foundation”の推進する”Ogg プロジェクト”のうちの, 非可逆音声圧縮を扱う”Ogg Vorbis CODEC プロジェクト”の成果物である. vorbis という名前はファンタジー小説”Small Gods”に登場するおしゃれの神にちなんだものである.

MP3 と違い, OggVorbis は可変ビットレートを基本としており, 効率のよい圧縮を行えるのが特徴である. ただし, MP3 と同様に DRM 技術を持たないため, 楽曲の有料配信にはあまり向いていない. 後発であることに加え, 特定の大企業の後押しがないこともあり, 広く普及しているとは言い難い. しかし, ライセンスフリーかつソースコードが公開されていることから, 有名オンラインソフトの多くが対応を完了しており, ポータブルプレイヤーでの採用も進みつつある. 拡張子は.ogg である.

3.3 AAC

AACはAdvanced Audio Codingの略称で、DVDなどで広く利用されている動画圧縮フォーマットの標準規格MPEG-2の音声圧縮のために作られた規格である。低速回線での映像配信を目的とし、動画・音声全般をデジタルデータとして扱うための規格MPEG-4でも利用されており、それぞれMPEG-2 AAC, MPEG-4 AACとして区別されているが、仕様が異なるだけなので通常では違いを意識する必要はない。MPEG-2 AudioではMP3を含むMPEG-1 Audioとの互換性を重視したMPEG-2 Audio BCという規格がまず用意されている。しかし性能が振るわず、互換性を捨てることで高い圧縮率と音質を実現するAACが新たに作られることとなった。ステレオ2チャンネルまでの対応だったMP3と異なり、最大で48チャンネルのオーディオ信号と15チャンネルの重低音用のサブチャンネルの再生に対応し、マルチリンガルや汎用データの伝送も可能である。

最初、AACはデジタルオーディオの分野では、限られた機器やソフトでしか利用できなかったが、デジタルオーディオの火付け役であるiPodの標準音楽形式として採用され、一気に身近な存在となった。また、AAC自体はデジタル著作権管理技術を含んでいないが、FairPlayやSD-AudioといったDRM技術と組み合わせて音楽ダウンロード販売サービスなどでも利用されている。拡張子としては.m4aなどが利用されている。

3.4 WMA

WMAはWindows Media Audioの略であり、その名前からもわかるとおり、パソコン用OSとして現在最も普及している”Windows”を提供するマイクロソフトが提唱している音声圧縮フォーマットである。インターネット上で、映像や音声と言ったデジタルメディアを配信するために開発したWindows Media Technologiesの1部となる。これは、音声や動画の圧縮フォーマットだけでなく、データを配信するためのサーバー技術や、著作権を保護するためのDRM技術など、デジタルメディアの配信に必要な各種技術を集めた、包括的なプラットフォームとなっている。かつてNet Showと呼ばれている技術を大幅に拡張したものである。Windowsに標準搭載されたマルチメディアプレーヤーソフトWindows Media Playerの標準オーディオ形式として採用されており、インターネット上でも同テクノロジーの動画圧縮フォーマットWindows Media Videoとともに広く利用されているので、知らず知らずのうちに利用していることがあるかもしれない。データ配信を目的として開発された形式ということもあり、低ビットレートでの音質の良さが特徴とされており、MP3の半分のデータサイズで同等の音質が得られるとされている。拡張子は.wmaである。

4 非可逆音声圧縮の仕組み

人間の聴覚の特性を体系化したものを心理音響モデルといい、非可逆音声圧縮では音声データを圧縮する際、心理音響モデルに基づいてデータ圧縮をする知覚符号化を用いている。知覚符号化とは、人間の知覚の特性を利用し、知覚できない、または知覚しにくい細部の情報を省いて表現することで、データ量の削減を行う手法のことである。CD 品質のオーディオの場合、可逆圧縮方式では1/2前後の圧縮率なのに対し、知覚符号化を利用した非可逆音声圧縮では音質の劣化をほとんど感じさせることなく1/10前後にまで圧縮できるため、知覚符号化による圧縮は可逆圧縮方式と比べるとはるかに効率が良いといえる。知覚符号化において省略する音は、最小可聴値とマスキングにより決定されている。

4.1 最小可聴限界

人間の耳は全ての周波数を均等に聞き分けているわけではなく，個人差はあるが，聞き取れる音の周波数範囲は 20Hz~20kHz といわれている．日常の会話などでよく使われている帯域 2kHz~4kHz 内の音は最も良く聞こえるものの，高音や低音には鈍感である．このため，同じ音の大きさであっても，周波数によって感じる音の大きさは異なってくる．この性質を利用し，周波数ごとにこれ以下の大きさの音は聞き取れないという最小可聴限界をもとにして，これよりも下回る音を省略し，データ量を削減することができる（図1）．

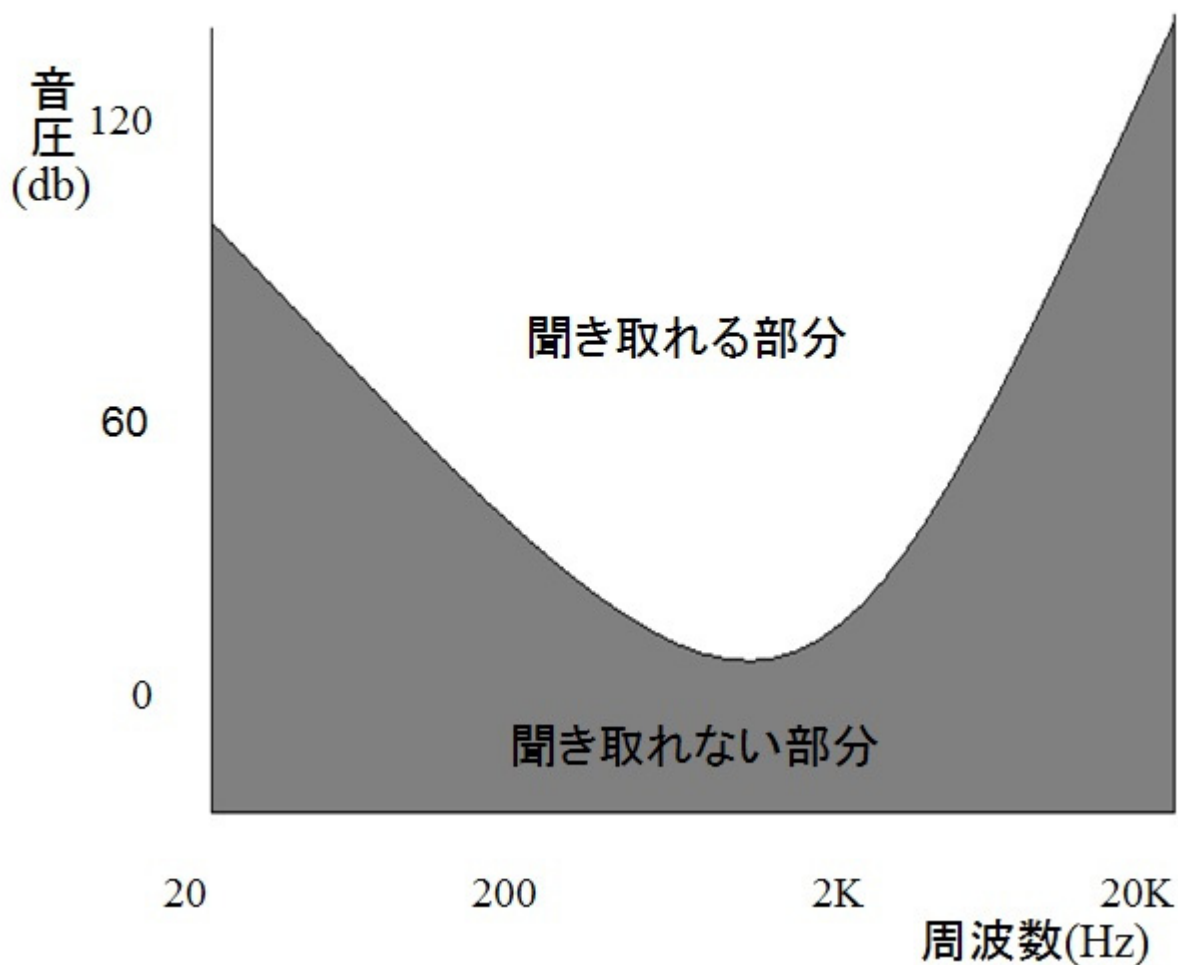


図1：最小可聴限界

4.2 マスキング

人間の耳は、周囲で大きな音が鳴っていると、それよりも小さな音はこれにかき消されてしまい、聞き取りにくくなる性質も持っている。この傾向は周波数が近いほど顕著になる。例えば、人の声が騒がしく響く場所では携帯電話の小さな声は聞き取りにくい。しかし、人の声とまったく異なる着信音であれば小さい声でも聞き分けることができる。このような効果をマスキング効果という。この効果は時間的な影響も受け、大きな音が鳴った瞬間はもちろん、音が鳴る直前にもその影響は及ぶ。この大きな音に紛れてかき消されてしまう小さな音のデータを削ることで、さらにデータを削減することができる(図2)。

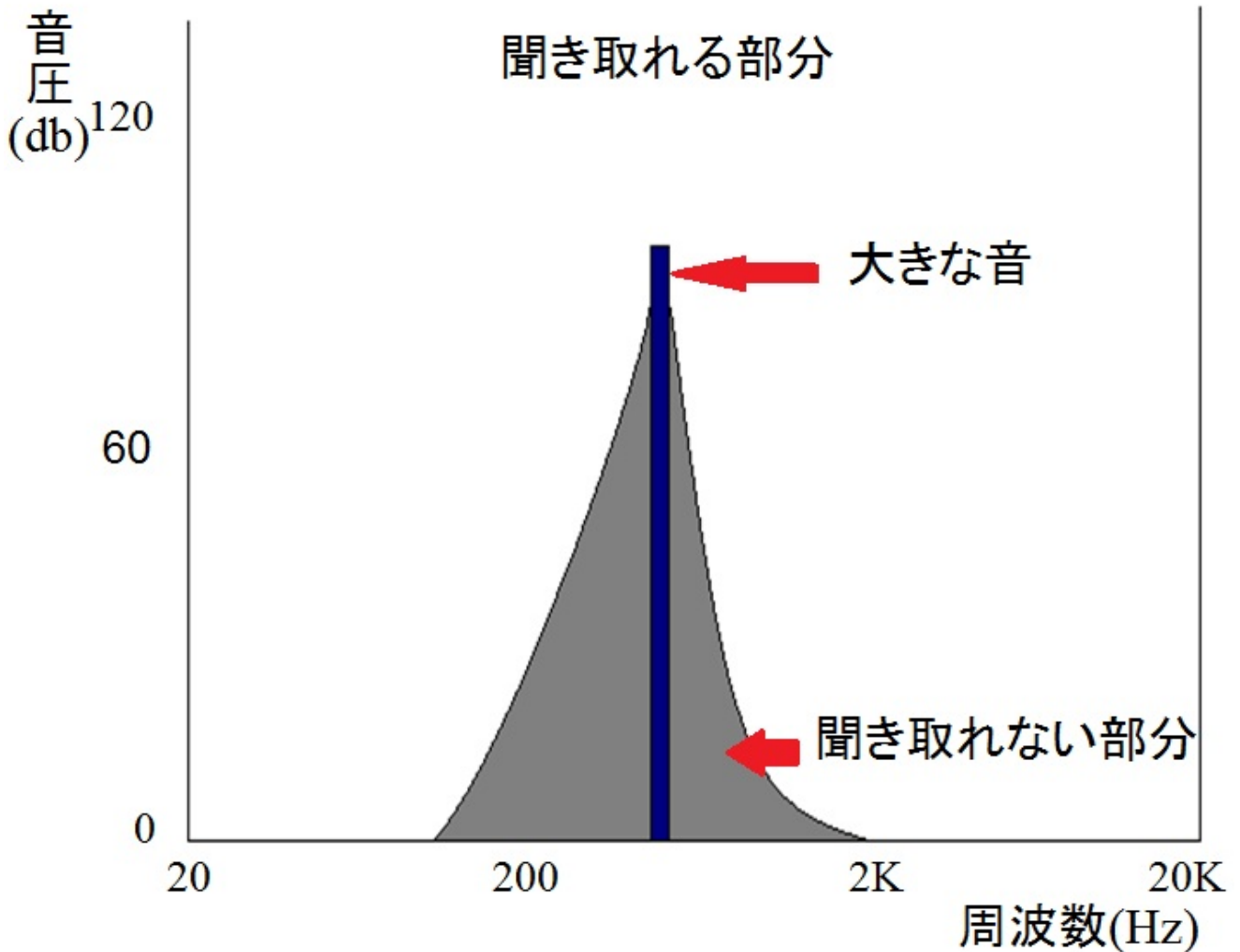


図2 : マスキング

5 ビットレート

ストリーミング再生を行う音声データに関する非可逆圧縮を行う上で欠かすことができないのが、ビットレートである。ビットレートは **bps(bit per second)** という単位で表記される。つまり 1 秒あたり何ビットの情報量になるのかを表す尺度で、以下の式で表すことができる。

$$\text{曲のサイズ (KB)} = \text{ビットレート (kbps)} \times \text{曲の秒数} \div 8$$

音声データを圧縮する場合、聞き取り難いとはいえ、データを省略することからビットレートを指定することで音質を決定することになる。ビットレートを大きくしてやれば、データのサイズは大きくなるが、その分省略されるデータは減るため音質は向上するし、ビットレートを小さくすればその逆になる。このビットレートはデータ圧縮率としても扱われている。

6 周波数成分による評価

先にも紹介したとおり，オーディオ圧縮においてビットレートが大きなカギを握ることになる．具体的には，ビットレートが違うとなにがどう変わってくるのかを考える．単純に考えて 128kbps のデータは 64kbps のデータの倍の容量になり，大雑把にいうと 64MB の容量を有する媒体に 128kbps のデータならば CD アルバム 1 枚分が収録できるのに対し，64kbps ならば 2 枚入るということである．そのことだけを考えればビットレートが小さいほうが有利だが，ビットレートを下げ過ぎるとファイルサイズは小さくなるにつれて音質は悪くなり，反対にビットレートを上げていくとファイルサイズは大きくなるが，音質は向上していくということになる．これらのことから，ファイルサイズと音質のバランスの取れたところがあり，いくらビットレートを上げててもこれ以上の音質にならないということであればその音質レベルに達する 1 番低いビットレートが最適ということになる．この場合のバランスの取れるビットレートに関する性能など AAC を例にあげて調べ比較する．

6.1 ビットレートごとの比較

音楽 CD から非圧縮のオーディオデータをリッピングし，それを圧縮して聞き比べる．素材として用意したのは 5 分 37 秒のクラシックの音楽である．単に聞き比べるだけでなく，その周波数成分がどうなっているかをも解析する．解析には WaveSpectra というソフトを利用しており，横軸が周波数，縦軸が音圧となっている．そして，曲の始めから終わりまでを再生したときの各周波数におけるピーク値を記録するように設定した．比較に用いたビットレートは 64kbps, 96kbps, 128kbps, 160kbps, 192kbps, 320kbps の 6 通りである (図 3~8)．

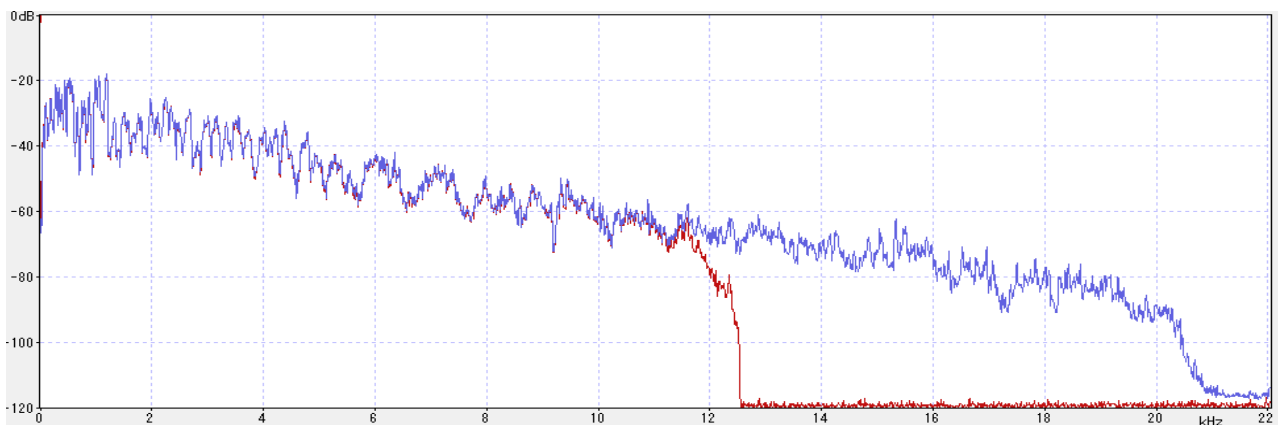


図 3：ビットレート 64kbps の周波数成分

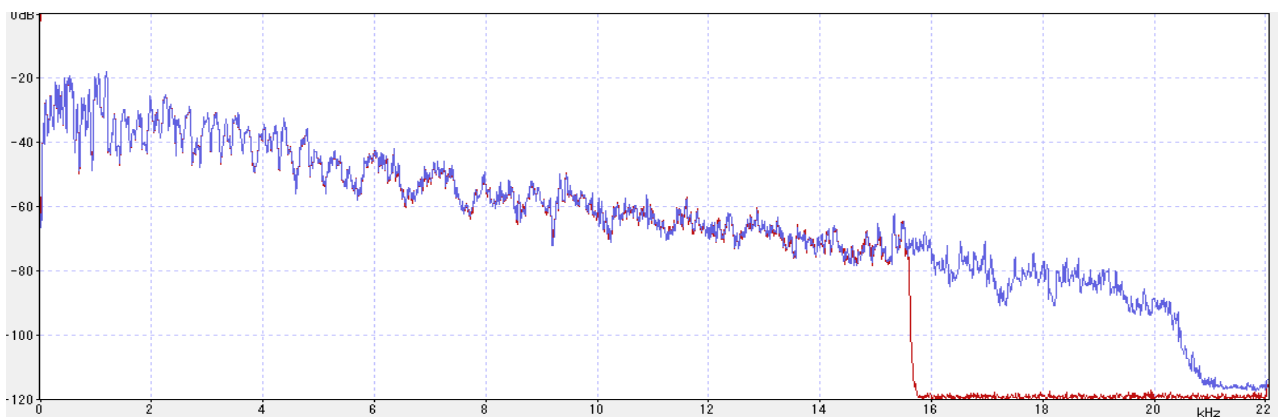


図 4 : ビットレート 96kbps の周波数成分

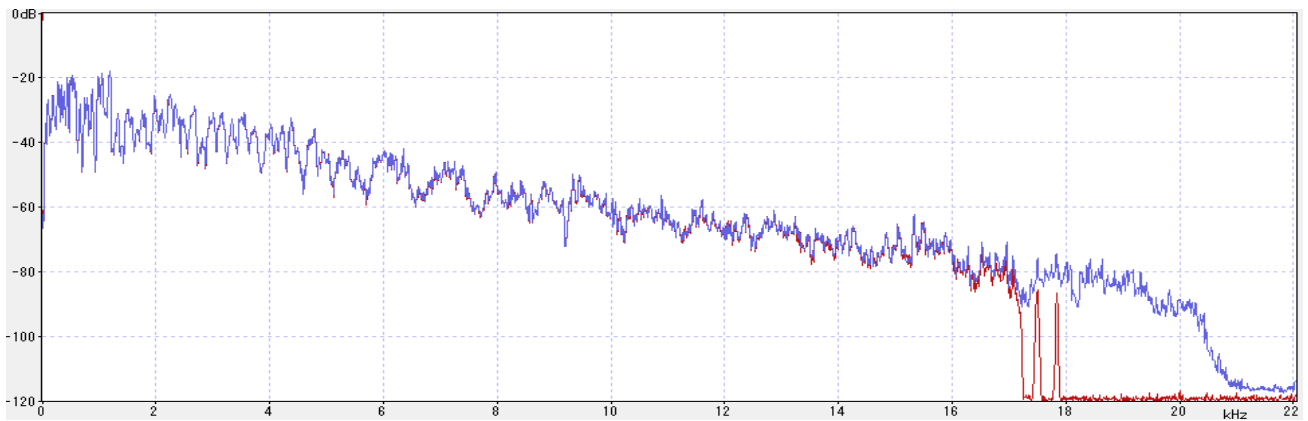


図 5 : ビットレート 128kbps の周波数成分

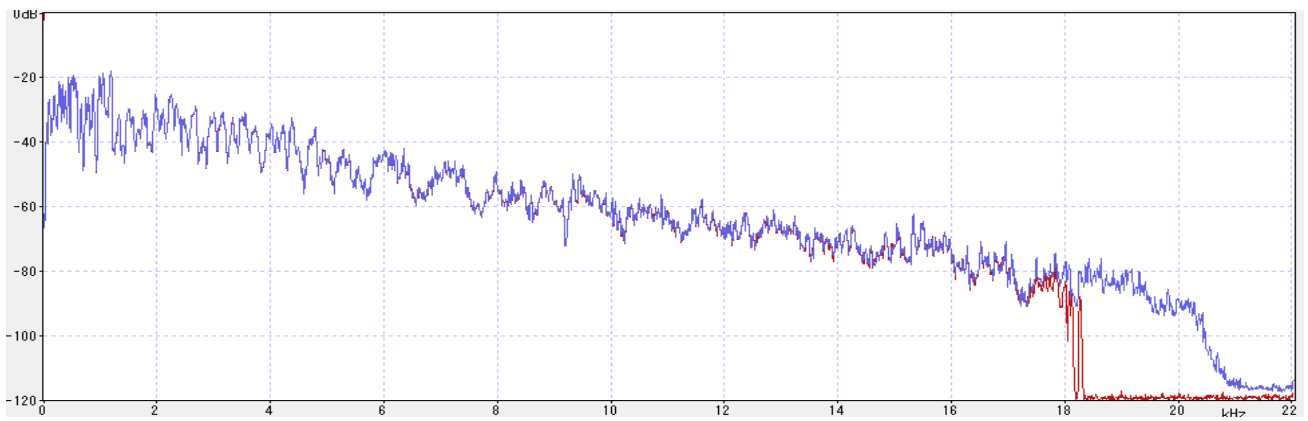


図 6 : ビットレート 160kbps の周波数成分

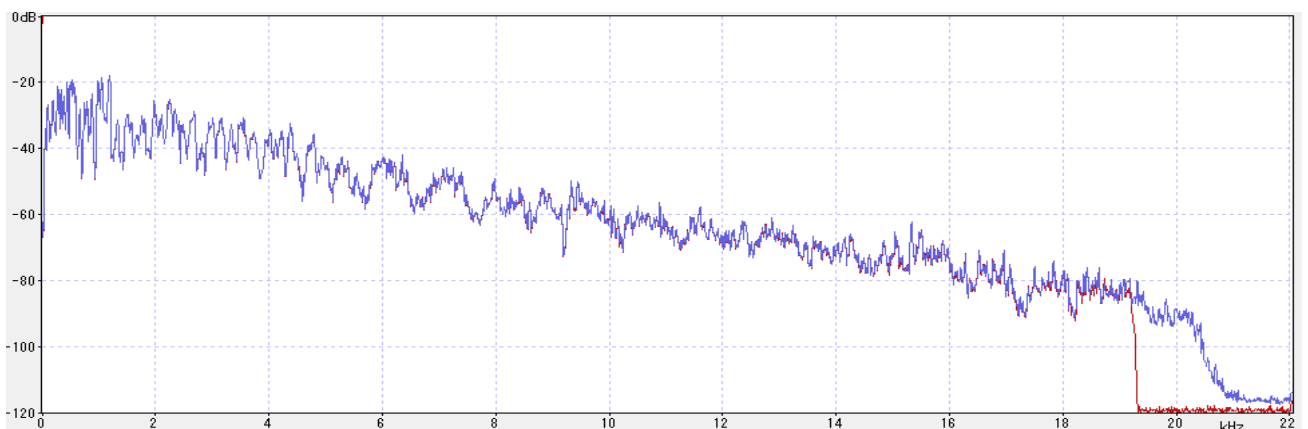


図 7 : ビットレート 192kbps の周波数成分

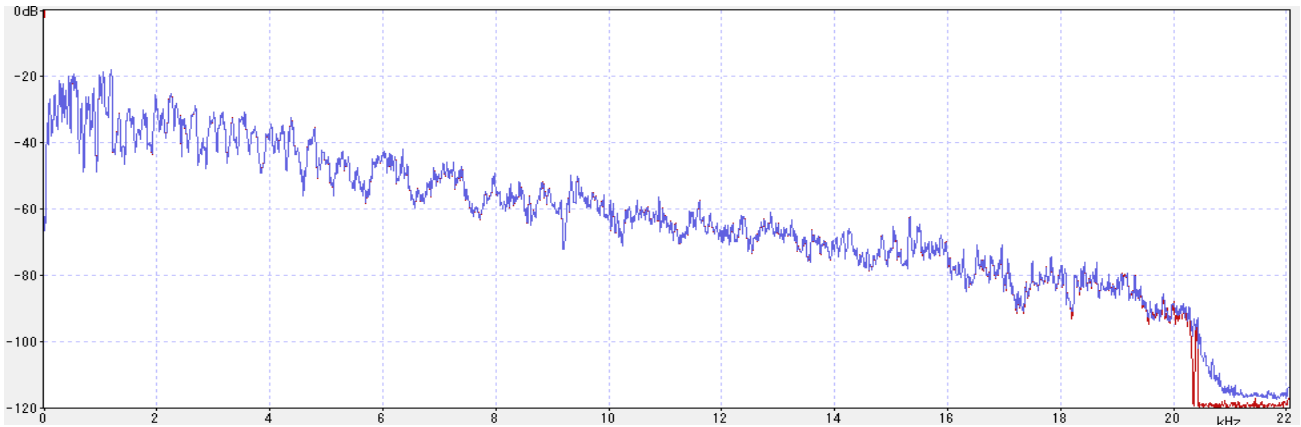


図 8 : ビットレート 320kbps の周波数成分

これらの結果より、周波数成分の差を比較する。ビットレートが 64kbps の周波数成分では約 12kHz から上の成分が切り落とされていることがわかる (図 3)。これに対し、94kbps の周波数成分では約 16kHz (図 4)、128kbps では約 18kHz (図 5)、160kbps では 18kHz 後半 (図 6)、192kbps では約 19kHz (図 7)、320kbps では約 20kHz の音域が出ている (図 8)。ビットレートが上がるにつれて成分の変化の度合いは小さくなっている。また、これらを聞き比べたとき、64kbps,96kbps,128kbps では徐々に音質が向上しているのがわかるが、それ以上のビットレートでは音質の差はほとんど感じられなかった。このことから、AAC において音質とデータサイズのバランスがとれているのは、128kbps で圧縮したときのものが最良と考えられる。

6.2 フォーマット別の比較

MP3, OggVorbis, AAC, WMA についてどのフォーマットが音質がよいのか、原音に近い音がでているのかを周波数成分を比較することにより調べた。ビットレートによって優劣があると考えられることから 96kbps,192kbps,320kbps の 3 通りを用意して比較した (図 9~14)。

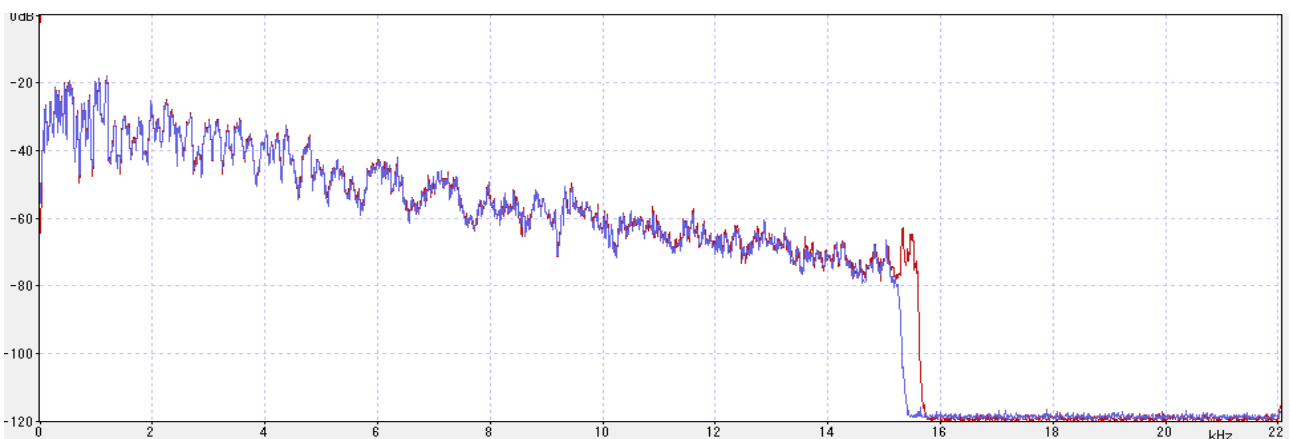


図 9 : ビットレート 96kbps(赤: AAC, 青: MP3)

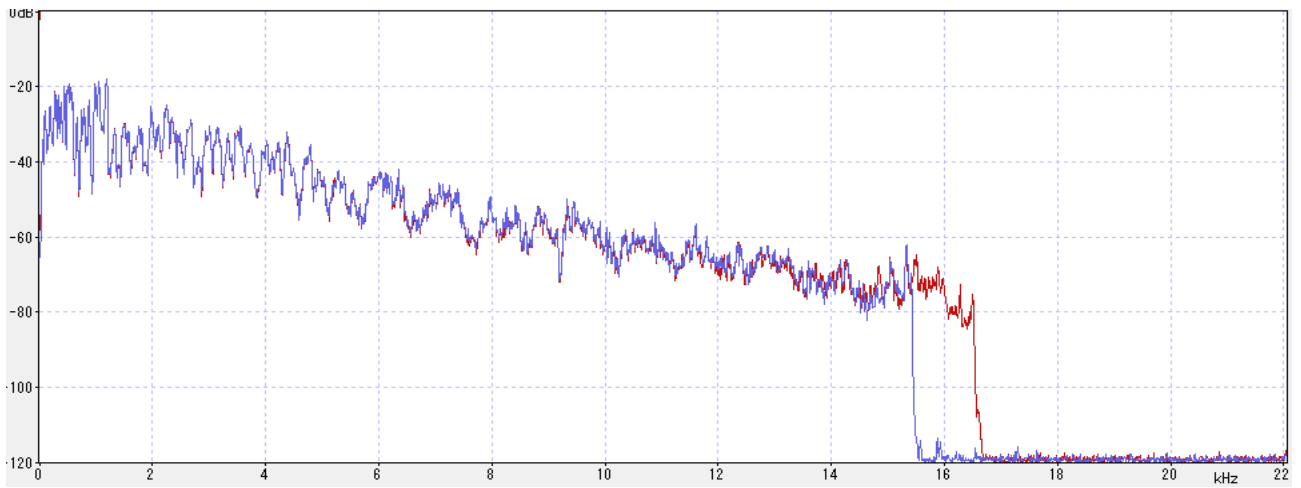


図 10 : ビットレート 96kbps(赤:OggVorbis,青:WMA)

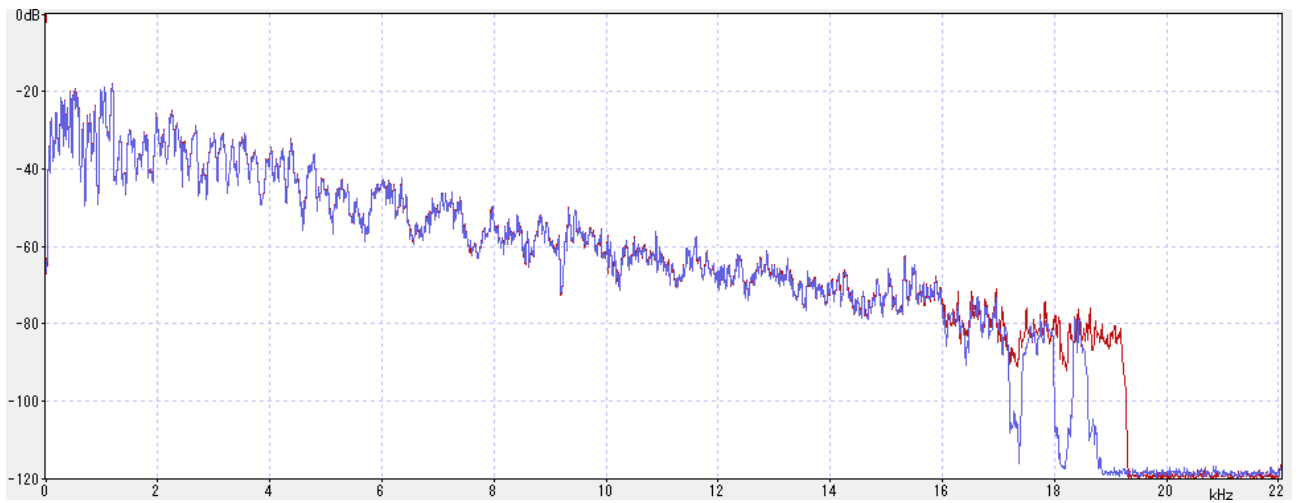


図 11 : ビットレート 192kbps(赤:AAC,青:MP3)

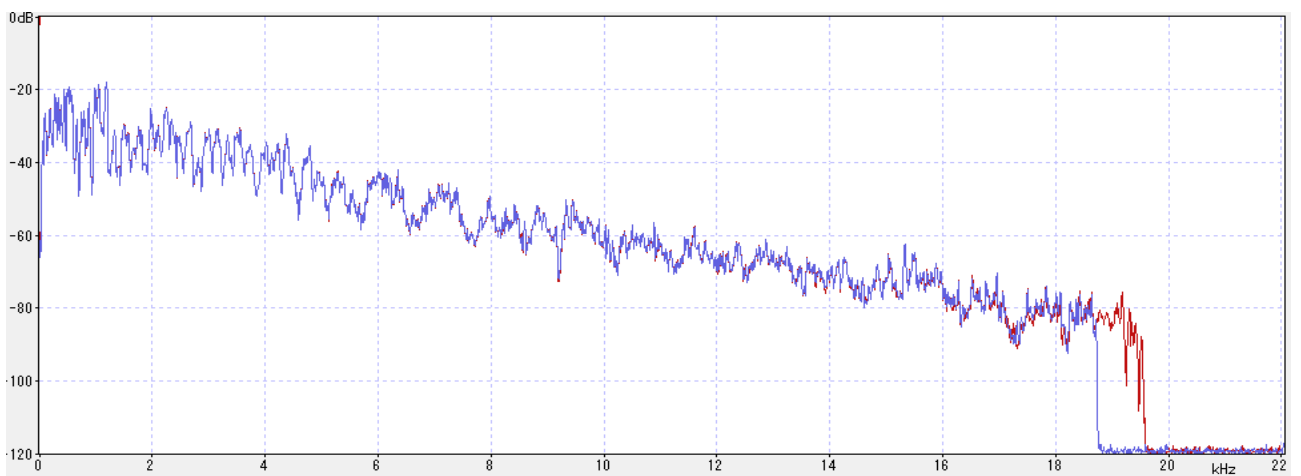


図 12 : ビットレート 192kbps(赤:OggVorbis,青:WMA)

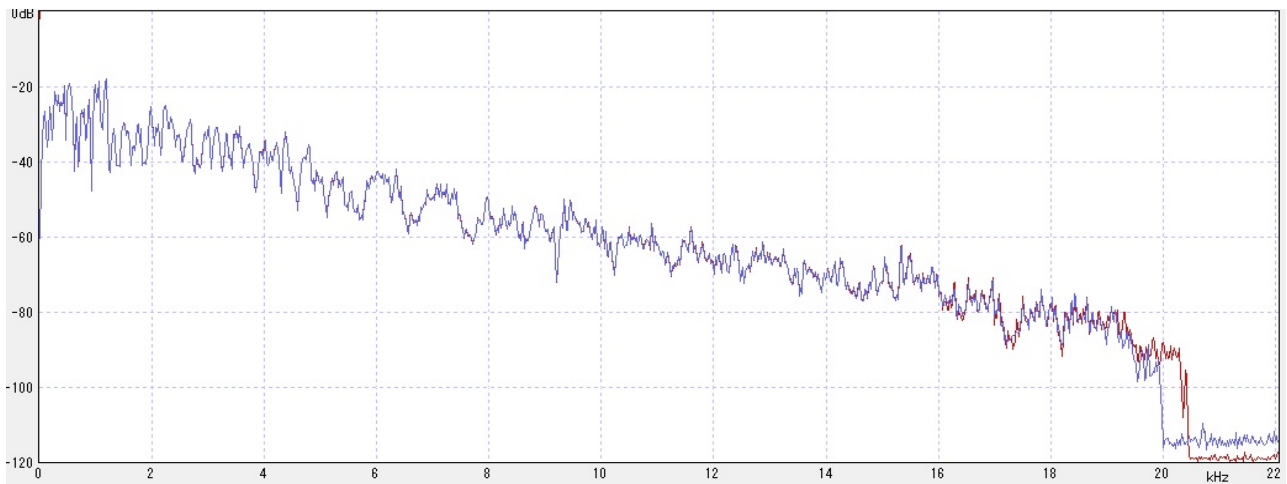


図 13 : ビットレート 320kbps(赤: AAC, 青: MP3)

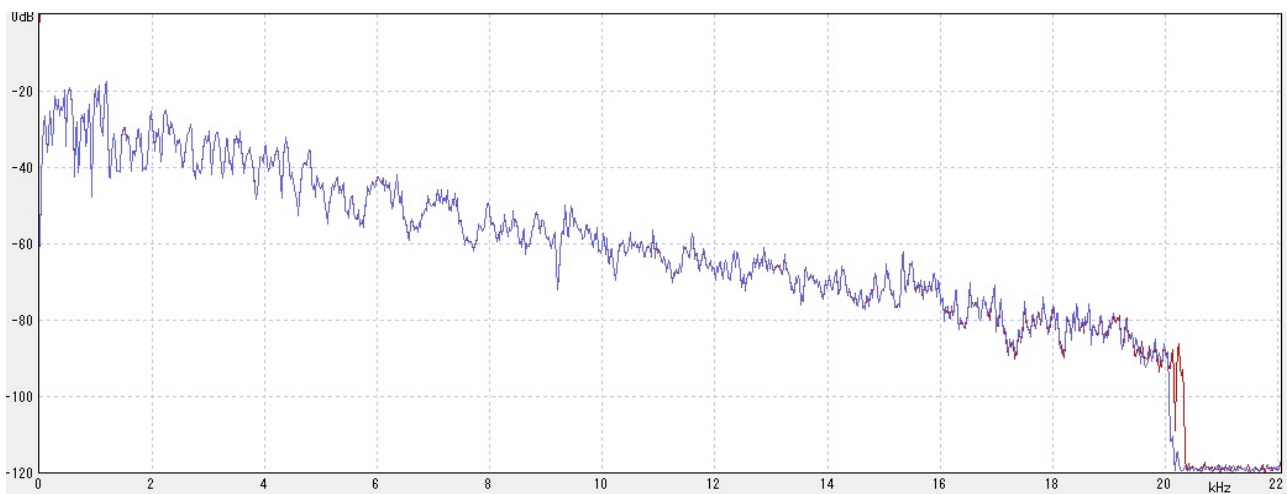


図 14 : ビットレート 320kbps(赤: OggVorbis, 青: WMA)

この結果より、ビットレートが 96kbps では MP3 と WMA は再生された周波数が 15kHz 前半までなのに対し、OggVorbis は 16kHz 後半と高い音域がでていることがわかる(図 9, 10). 192kbps では、MP3 と WMA は 18kHz 後半なのに対し、OggVorbis と AAC は 19kHz 以上の音域が出ている(図 11, 12). 320kbps では、成分の差は小さいものの、MP3 と WMA は約 20kHz, OggVorbis と AAC はどちらも 20kHz 中盤だったもののわずかに AAC の方が高い音域がでていた(図 13, 14). 以上のことから低ビットレートで圧縮する場合、OggVorbis, 高ビットレートで圧縮する場合、OggVorbis, AAC が原音に近い音域が出ているものと考えられる。また、それぞれビットレートを同じにして聴き比べたものの、フォーマット別の音質差は感じられなかった。

7 1/f ゆらぎ

次に 1/f ゆらぎについて説明をする。1/f ゆらぎとは、周波数成分の強さが周波数 f に反比例するゆらぎのことで、別名ピンクノイズとも言われている。“ゆらぎ”とは、寄せては返す波の音、人間の心拍、川のせせらぐ音のように周期や強さが一定ではなく、ある程度のばらつきがあることであり、具体的には、周波数成分が平均値を中心にして時間または空間的に変動する現象として定義されている。人間の心拍を例にあげると、1 分間の心拍数が平均で 60 回だったとして、それが 59 回になったり、61 回になったりすることである。音楽に関しては、1/f ゆらぎを持っている曲は人に快適感やヒーリング効果を与えるとされている。

ゆらぎを定量化する方法として、周波数・振幅をサンプリングし、フーリエ変換によってパワースペクトルを表示し、その分布からゆらぎの程度を表している直線の傾きを求める方法がある。つまり、周波数成分の強さが周波数とともにどのくらいの割合で減少していくかを数値として表すということになる。周波数を f 、周波数成分の強さを P 、減少の割合を λ とし、 $P=1/f^\lambda$ の対数、 $\log P = -\lambda \log f$ によりグラフ化して 1/f ゆらぎに近いかを表すことができる。

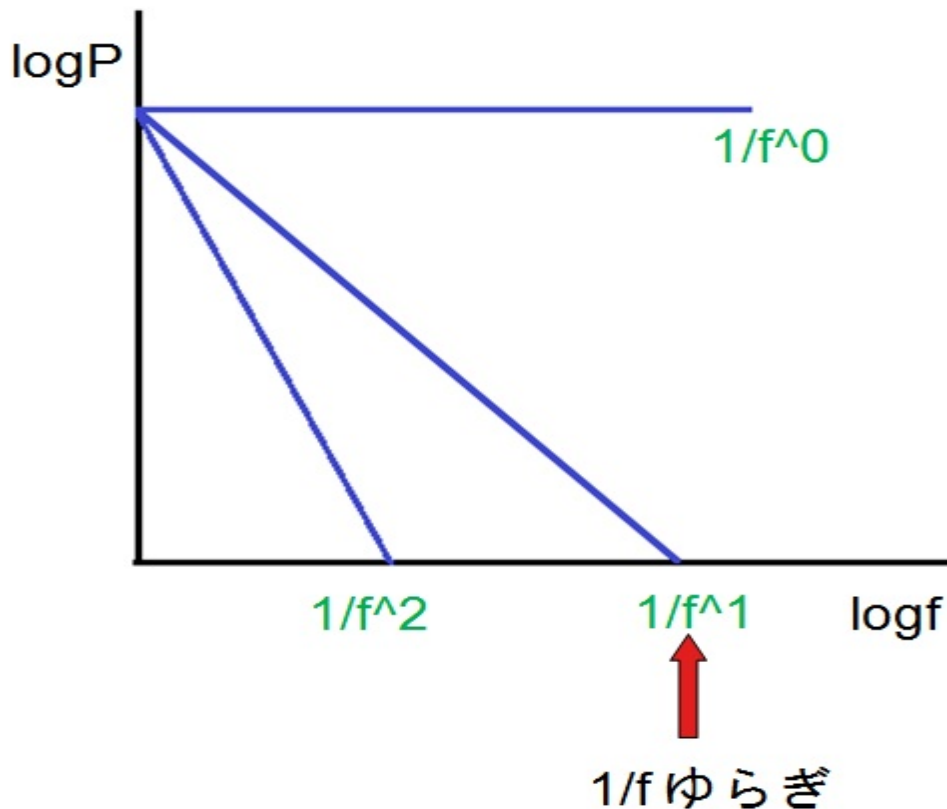


図 15 : 1/f ゆらぎ

7.1 ゆらぎの測定

先の実験でビットレートまたはフォーマットによって周波数成分が変化していることがわかったので、フォーマット別の比較と同じ条件でそれぞれ λ の値を求め、 $1/f$ ゆらぎとの関連を調べることにする。

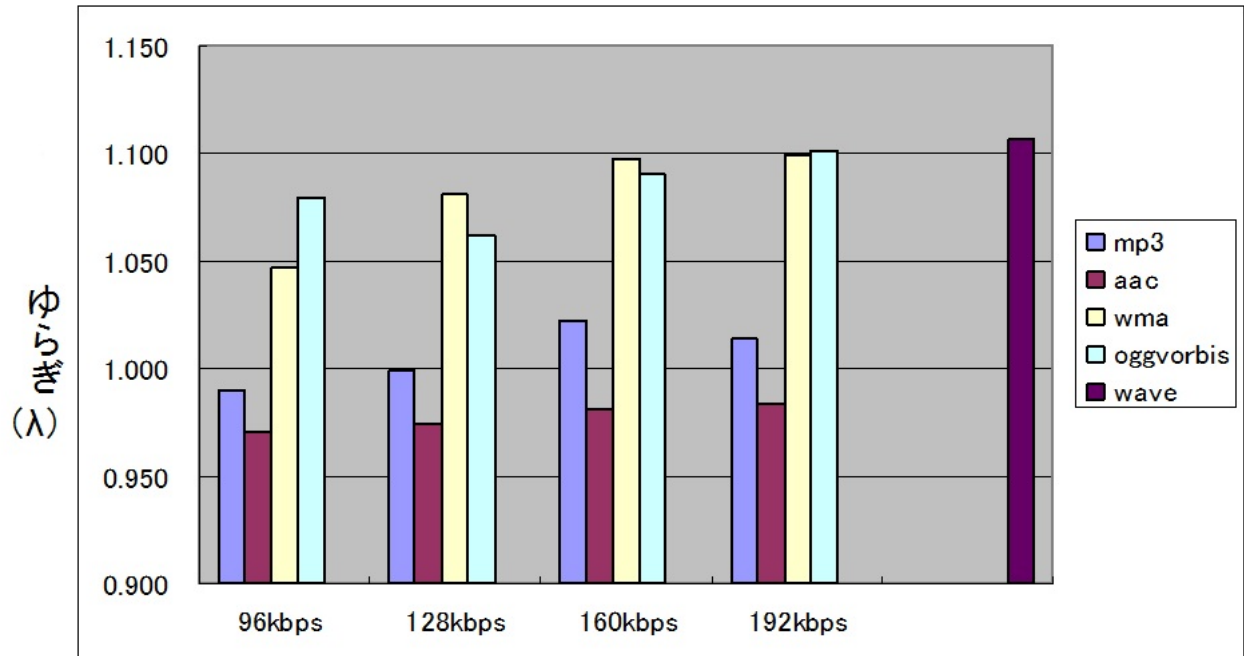


図 16 : ゆらぎの測定

ビットレートが低くなるほど λ の値が小さくなっている。数値が下がったということは、データ圧縮することでゆらぎが不明瞭になったということになる。ビットレート別で見たとき、96kbps が最も 1 に近い数値となり、192kbps が 1 番遠い数値となっている。フォーマット別に見ると、MP3 が最も 1 に近い数値となり、OggVorbis が一番遠い数値となっている。また、全体で最も 1 番 1 に近づいたのは MP3 を用い、128kbps でデータ圧縮した時のものとなった。

8 まとめ

ビットレート別に音質差を比較した結果によると、ビットレートが上がるにつれて高い音域が出ており、高ビットレートになるほど元の音に対する音域の変化の度合いは小さくなっている。また、主観による音質差の比較より、AACにおける音質とデータ量のバランスがとれているのは128kbpsで圧縮したときのものとなった。フォーマット別の比較では、低ビットレートのときは、OggVorbis、高ビットレートのときは、OggVorbisとAACが最も原音に近い音域がでていることがわかった。1/fゆらぎの差は、データ圧縮することでそれぞれゆらぎの度合いを示す数値は小さくなっており、1/fゆらぎは不明瞭になっていた。フォーマット別に比較をした場合、同じビットレートで最も1/fゆらぎに近づいたのはMP3となった。

今後の課題として、今回は一つのジャンル、一曲の音声データのみで比較を行ったが、曲またはジャンルによって、フォーマットによる優劣の差、ゆらぎの数値に変化があると思われることから、複数のジャンル、複数の音声データによる比較をすることが考えられる。

参考文献

- [1]よくわかるデジタルポータブルオーディオ
加賀章喜 霧島煌一 電波新聞者 (2005)
- [2]サウンド圧縮テクニカルガイド
藤本健 株式会社 BNN 新社 (2004)
- [3]ゆらぎアナライザー
<http://mahoroba.logical-arts.jp/>
- [4]WaveSpectra
<http://www.ne.jp/asahi/fa/efu/>
- [5]CDex
<http://www.surf.to/cdex/>
- [6]xrecode II
<http://xrecode.com>