

帯域制限した MPEG オーディオに対する 再符号化・再々符号化が符号化データと 再生音声に与える影響の考察

萩野谷 俊典^{†1} 田中 清^{†1} 高木 幸一^{†2}

本稿では、MPEG オーディオ信号に対する新たなデータハイディングを実現する目的から、帯域制限した MPEG オーディオに対する再符号化と再々符号化が、符号化データ (圧縮符号) と再生音声に与える影響について考察する。いくつかの音楽データを用いたシミュレーション実験により、帯域制限した圧縮符号とこれを再々符号化した圧縮符号の割り当てビット数の変化 (差異) は小さく、両者の再生音声もほとんど違いが認められないことを示す。これより、再符号化によって生じる冗長成分を利用する新しいデータハイディングの実現の可能性が示唆される。

An Examination on the Impact of Re-encoding and Re-reencoding for Band-passed MPEG Audio in Encoded Data and Decoded Audio

TOSHINORI HAGINOYA,^{†1} KIYOSHI TANAKA^{†1}
and KOICHI TAKAGI^{†2}

In this paper, to realize a new data hiding method for MPEG audio, we examine the impact of re-encoding and re-reencoding for band-passed MPEG audio in encoded data (compressed audio) and decoded audio. Through computer simulation using several music data, we found that the change (difference) in number of bits allocated to each sub-band between the band-passed compressed audio and the re-reencoded compressed audio is very small, and there is almost no perceptual difference between both decoded audios. From these results, the possibility of a new data hiding that uses redundancy caused by re-encoding is suggested.

1. はじめに

パソコン, iPod, 携帯電話などの電子機器やネットワークを介したストリーミング配信, 通信カラオケ, デジタル放送などで利用されるオーディオ信号は, 圧縮符号化した状態で蓄積・伝送される. 現在までに, MPEG-1 オーディオレイヤー I, II, III, MPEG-2, 4/AAC 等のいくつかの圧縮規格が開発され, 目的や用途に応じて様々な形で利用されている¹⁾⁻⁴⁾. 本稿では, MPEG オーディオ信号に対する新たなデータハイディング⁵⁾を実現する目的から, 帯域制限した MPEG オーディオに対する再符号化と再々符号化が, 符号化データ (圧縮符号) と再生音声に与える影響について考察する.

本稿では, MPEG-1 オーディオレイヤー II を用いて符号化した音声信号を利用し, これをまず指定したあるサブバンド番号で帯域制限して, 音声信号の高域成分をカットして劣化させる. 次に, この帯域制限した音声を, 帯域制限する前にサブバンド全体に割り当てられていた符号量で再符号化する. さらに, これを帯域制限した後の低域サブバンドだけに割り当てられていた符号量で再々符号化する. このとき, 帯域制限した圧縮符号とこれを再々符号化した圧縮符号について, 割り当てビット数の変化 (差異) を調査する. また, 両者の再生音声について, オリジナル音声との間の Segmental SNR^{6),7)} および周波数スペクトルを比較する. シミュレーション実験には, RWC データベースに含まれる特徴の異なる複数の音楽データを用いる.

2. MPEG-1 オーディオの概要

MPEG-1 オーディオは, サンプリング周波数 32, 44.1, 48[kHz] のモノラル, もしくはステレオ (2 チャンネル) の音声信号を符号化の対象としている. 符号化アルゴリズムは, その複雑度に応じてレイヤー I, II, III の 3 つのレイヤーから構成されている¹⁾⁻⁴⁾. 高いレイヤーほど高品質と高圧縮率が実現される反面, ハードウェアの規模も大きくなる. また, レイヤー I と II はフレームあたりのサンプル数が異なるだけで, 符号化方式が等しいため, まとめて取り扱われることが多い. 本稿では, ストリーミングでの利用頻度が高いレイヤー II を対象に検討する. この方式では, 16 ビット線形量子化された PCM 入力信号は, 時間

^{†1} 信州大学工学部
Faculty of Engineering Shinshu University
^{†2} KDDI 研究所
KDDI R&D Laboratories Inc.

領域から 32 個のサブバンドからなる周波数領域に写像される。限られたデータ量でできるだけ高品質な音声を再現するために、心理聴覚特性に基づく量子化誤差のマスクングレベルが計算される。各サブバンドの音声信号からマスクングレベルを差し引いた差分信号に対して、適応的なビット割り当てによる量子化および符号化が行われた後、アンシラリデータ(利用者が任意に定義できるデータ)を合成してフレーム(AAU:オーディオ・アクセス・ユニット)を構成する。本稿では、この動的ビット割り当てに注目して検討を進める。

3. 動的ビット割り当て処理

MPEG-1 オーディオレイヤー II の符号化器では、信号対マスク比 (SMR) の値に従い、各サブバンドに対して信号対雑音比 (SNR) と SMR の差であるマスク対雑音比 (MNR) の値が小さいサブバンドから順に、動的にビットが割り当てられていく。以下に、そのアルゴリズムを示す。

● 動的ビット割り当てアルゴリズム

Step1:式 (1) を用いて、1AAU に割り当て可能な総ビット数 (*size*) を算出する。

$$size = \frac{1152 \times bitrate}{samplingfrequency} \quad (1)$$

ただし、*bitrate* は符号化レート [kbps]、*samplingfrequency* はサンプリング周波数 [kHz] を示す。

Step2:1AAU の *size* から、符号を構成する Header(符号化レート、サンプリング周波数、チャンネル数を定義)、CRC(エラーチェック)、BitAllocation(サブバンド毎のサンプル値の符号長を定義)、Ancillary(アンシラリデータ)のビット数を減算する。

Step3:最小 MNR を示すサブバンドを探索する。

Step4:当該サブバンドの量子化ステップを 1 段階上の値に変更する。

Step5:新しい量子化ステップに対応する SNR を選択し、新たな MNR を算出する。

Step6:現在の割り当て可能ビット数から、現在の量子化ステップに対応するビット数を減算し、新たな割り当て可能ビット数を算出する。

Step7:最小 MNR に対応するサブバンドの量子化ステップが 0 の場合は、SCFSI(Scale Factor の伝送方法を指定)と Scale Factor(サブバンド毎の 12 サンプルの最大振幅を示す)のビット数を減算する。

Step8:割り当て可能ビットが正の最小値ならば、現在の AAU に対する処理を終了。そうでなければ Step3 へ。

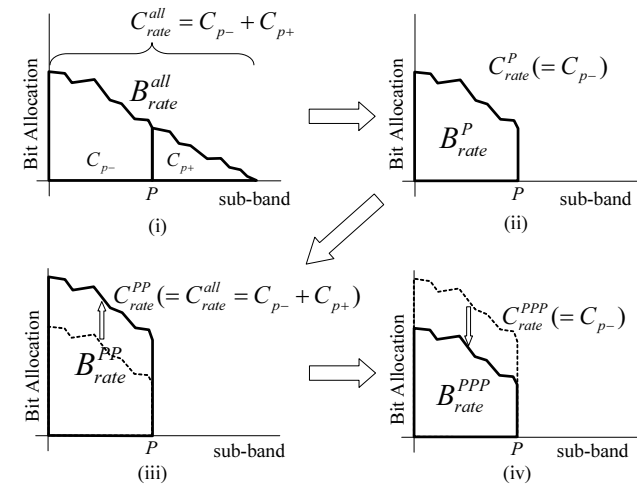


図 1 帯域制限から再符号化、再々符号化への概念
Fig. 1 Concept from band-limitation to Re-encoding and Re-reencoding

4. 再符号化・再々符号化の方法

本稿では、上記の動的ビット割り当て処理の一部に必要な修正(後述の 5.6 を参照)を加えて、再符号化、再々符号化を行う。その一連の処理の概念とブロック図による流れを、それぞれ図 1 および図 2 に示す。

Step1:オリジナル音声を符号化して、圧縮符号 B_{rate}^{all} を生成する(図 1(i))。ただし、*rate* は符号化レート、*all* はすべてのサブバンドを利用していることを示す。また、図中の C は符号量を表す。

Step2: B_{rate}^{all} のサイド情報とサンプル値を復号し、指定したあるサブバンド番号(遮断周波数 P)以降の割り当てビット数を 0 に書き換え、再びサイド情報とサンプル値を符号化して、帯域制限した圧縮符号 B_{rate}^P を生成する(図 1(ii))。このとき、遮断周波数 P よりも低域に割り当てられている符号量 C_{p-} を記録しておく。

Step3: B_{rate}^P を復号し、PCM 信号 G_{rate}^P を再生する。

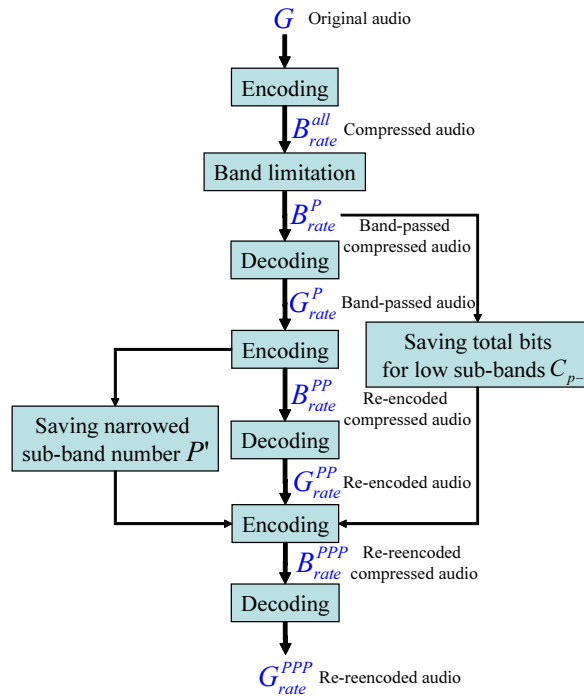


図2 帯域制限から再符号化、再々符号化に至る処理の流れ
 Fig. 2 Flow of series of processing from Band-limitation to Re-encoding and Re-reencoding

Step4: G_{rate}^P を帯域制限前の全体の符号量 C_{rate}^{all} で符号化^{*1}し、冗長な圧縮符号 B_{rate}^{PP} を生成する (図1(iii))。このとき、動的ビット割り当てはサブバンド番号 $0 \sim P-1$ の範囲で行うが、後述の方法 (6 を参照) を用いて $P' (< P)$ の範囲に収める。この P' の値も記録する。

Step5: 再符号化した圧縮符号 B_{rate}^{PP} を復号し、PCM 信号 G_{rate}^{PP} を再生する。

Step6: 記録しておいた P' と C_{p-} を取り出す。

*1 B_{rate}^{PP} のサンプル値の符号量は、 B_{rate}^{all} のそれよりも高域成分の SCFSI および Scale Factor 値がない分、若干ビット数が多くなる。(後述の図3 参照)

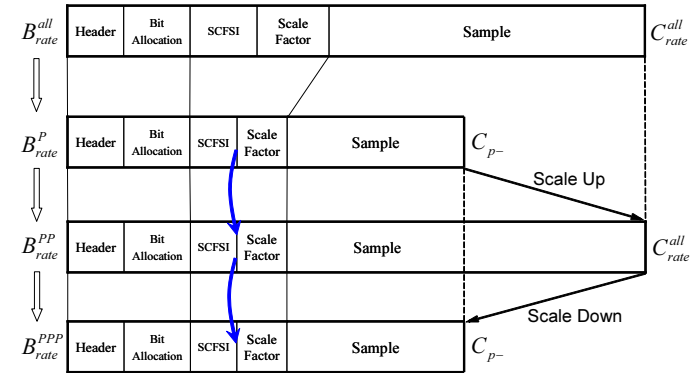


図3 圧縮符号と符号量の変化

Fig. 3 Transition of compressed audio and total amount of codes

Step7: G_{rate}^{PP} を低域成分の符号量 C_{p-} で符号化し、再々符号化した圧縮符号 B_{rate}^{PPP} を生成する (図1(iv))。

Step8: B_{rate}^{PPP} を復号し、PCM 信号 G_{rate}^{PPP} を再生する。

ここで、上記の一連の処理における符号化した圧縮符号と符号量の変化を図3に示す。帯域制限した圧縮音声 B_{rate}^P に含まれる SCFSI, Scale Factor の値は、その後の再符号化および再々符号化においてもそのまま利用するが、符号量が C_{p-} から C_{rate}^{all} に拡大され^{*1}、再び C_{p-} に縮小されることに注意する。

5. 最大ビット割り当ての制限

通常の符号化器を用いて高ビットレート (320, 384[kbps]) で符号化すると、第11サブバンド以降に最大ビット数 (符号量16ビット) が割り当てられる頻度が増加する。この場合、本来ならば符号長が5ビットで済むところを、限られた全体の割り当てビット数から11ビットも余分に消費してしまい、資源を有効に活用できない恐れがある。そこで、最大割り当てビット数の出現を制限し、この分の割り当てビットを他の成分に割り振るようにする⁹⁾。本稿では、この方法を用いて圧縮符号 B_{rate}^{all} 、再符号化した圧縮符号 B_{rate}^{PP} および再々符号化した圧縮符号 B_{rate}^{PPP} を生成する。帯域制限した再生音声 G_{rate}^P を再符号化して B_{rate}^{PP} を生成する際に、この方法で最大ビット割り当てに制限を加えた場合の例を図4に示す。符号長に上限値を設定したことにより、特定のサブバンドに過剰にビットが割り当てられること

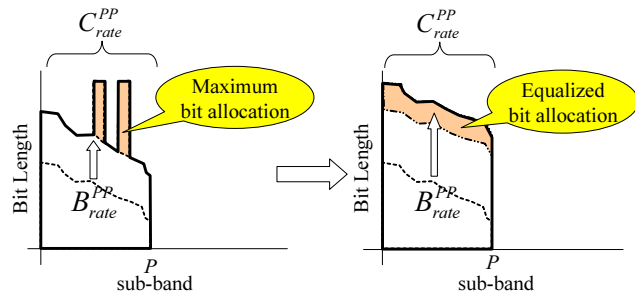


図 4 最大割り当てビット数の出現を制限した動的ビット割り当て⁹⁾の効果
Fig.4 Effects of the modified dynamic bit allocation⁹⁾ limiting the maximum bit allocation

がなくなり、指定したサブバンド内に均等にビットが割り当てられていることがわかる。

6. 割り当てビット拡大サブバンドの縮小

帯域制限後の再生音声 G_{rate}^{PP} を元の圧縮符号全体の符号量 C_{rate}^{all} で再符号化して B_{rate}^{PPP} を生成するとき、図 5 に示すように低域側の割り当てビットを増加させるサブバンド数 $P' (< P)$ を縮小する。遮断周波数 P 以降に存在する高域成分の符号量 C_{p+} を $0 \sim P' - 1$ サブバンドに完全に格納できるようにするため、第 0 サブバンドから順番に各サブバンドの最大割り当てビット数から実際の割り当てビット数を減算した未使用のビット数を加算してゆき、これが C_{p+} 以上になるサブバンド番号を P' として決定する。従って、図 5(a) のように遮断周波数 P が高域側に設定され、 P 以降に存在する高域成分の符号量 C_{p+} が少ない場合は、 P' は小さな値となり、狭い範囲のサブバンドに多数のビットが追加割り当てされる。逆に、図 5(b) に示すように P が低域側に設定され、高域成分の符号量 C_{p+} が多くなると、 P' の値も大きくなり、ビットが追加割り当てされるサブバンドの範囲も広がる。

この方法で割り当てビット数を増加するサブバンドの範囲を狭めた場合に、遮断周波数 P に対する P' の平均値 (“Pops”, 384[kbps] を利用) を調査した結果を表 1 に示す。上述の通り、遮断周波数 P の値が大きいかほど高域成分のビット数 C_{p+} が少なくなるため、 P' は次第に小さくなるのがわかる。遮断周波数 P をある値に固定した場合、 P 以降の高域成分の符号量 C_{p+} はフレーム (AAU) ごとに異なり、従って P' も変化する。このため、本稿で行う再符号化および再々符号化では、4 で示したように、それぞれの AAU に対する低域成

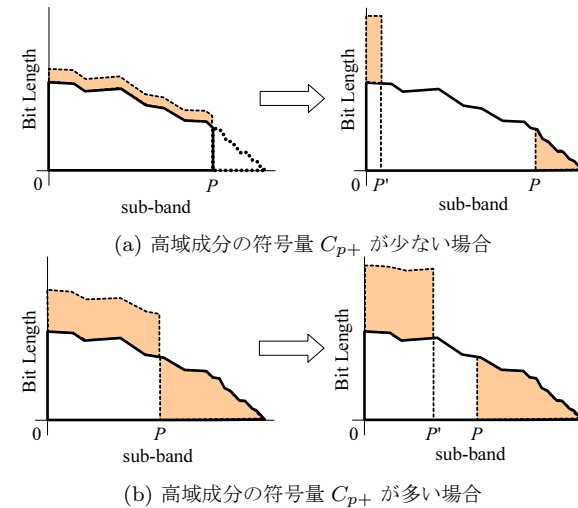


図 5 修正動的ビット割り当て処理による追加ビット割り当てサブバンドの縮小
Fig.5 Reduction of sub-bands to allocate extra bits by modified dynamic bit allocation

分の符号量 C_{p-} ($C_{p+} = C_{rate}^{all} - C_{p-}$ で容易に求まる) とそれに対する P' の値をテーブルに保存する。また、再符号化および再々符号化の動的ビット割り当てでは、表 2 に示すように、 $0 \sim P' - 1$ サブバンドまでは符号長に制限を設けず、 $P' \sim P - 1$ サブバンドではそれぞれ B_{rate}^P および B_{rate}^{PP} の符号長を用いるように修正する。

7. 実験結果と考察

実験には、RWC 研究用音楽データベース (音楽ジャンル)⁸⁾ から抜粋した “Pops”(No.1), “Jazz”(No.30), “Classic”(No.53) の 3 曲を利用し、必要な修正を加えた MPEG-1 オーディオレイヤー II 符号化器を用いた。帯域制限では、サブバンド番号 (遮断周波数) P を $1 \leq P \leq 26$ (112~160[kbps] の場合)、 $1 \leq P \leq 29$ (192~384[kbps] の場合) に変化させた。

まず、帯域制限した圧縮音声 B_{rate}^P とこれを再々符号化した圧縮音声 B_{rate}^{PPP} の各サブバンドにおける割り当てビット数の一致率を調査した結果を表 3 に示す。ただし、一致率の定義は、割り当てビット数が一致した回数を計数し、これを全フレーム数で除算した割合である。また、楽曲には “Pops” を利用し、遮断周波数は $P = 9$ に設定した。この結果から、ビットレートおよびサブバンド番号によって異なるが、誤差が生じていることがわかる。そ

表 1 遮断周波数 P を変化させた時の P' の平均値 (“Pops”, 384[kbps])

Table 1 Average of P' as we change cutoff frequency P

遮断周波数 P	狭めた P'	遮断周波数 P	狭めた P'
1	1.00	15	5.37
2	2.00	16	4.98
3	3.00	17	4.29
4	4.00	18	3.69
5	5.00	19	3.05
6	6.00	20	2.52
7	7.00	21	2.02
8	8.00	22	1.47
9	8.99	23	1.12
10	8.00	24	1.02
11	7.67	25	1.00
12	7.00	26	1.00
13	6.58	27	1.00
14	5.99	28	1.00

表 2 再符号化および再々符号化における割り当てビット数の上限値の設定

Table 2 Upper bound in dynamic bit allocation for re-encoding and re-reencoding

	サブバンド番号	符号長の上限値
再符号化	$0 \sim P' - 1$	制限無し (最大 16 ビット)
	$P' \sim P - 1$	B_{rate}^P の符号長
再々符号化	$0 \sim P' - 1$	制限無し (最大 16 ビット)
	$P' \sim P - 1$	B_{rate}^{PPP} の符号長

ここで、次にこの誤差の統計量を調査した結果を表 4 に示す。ここでは、単純にサブバンド毎に割り当てビット数の差を求め、全フレームでの平均値と分散を算出している。これより、誤差は全体的に 0 近傍の狭い範囲に収まっており、圧縮符号の割り当てビット数は、再符号化および再々符号化処理によって大きく変化していないことがわかる。

次に、帯域制限した圧縮符号 B_{rate}^P とこれを再々符号化した圧縮符号 B_{rate}^{PPP} をそれぞれ再生した音声 G_{rate}^P , G_{rate}^{PPP} について、オリジナル音声 G との間で Segmental SNR^(6),7) を計測した結果を図 6 に示す。ただし、それぞれのグラフには、参考値として各ビットレートで圧縮した音声 G_{rate}^{all} のオリジナル音声 G に対する Segmental SNR 値も示しておく。これより、 G_{rate}^P , G_{rate}^{PPP} とともに、遮断周波数 P を低域側に移動するに従い、圧縮音声 G_{rate}^{all} の達成する SNR 値から次第に減少し、帯域制限の効果が現れていることがわかる。また、 G_{rate}^P , G_{rate}^{PPP} いずれも、 P の減少に伴う SNR 値は同等な値を示している。このことから、 G_{rate}^P に対する再符号化および再々符号化による変化 (劣化) は小さく抑えられていると言

表 3 帯域制限音声 (B_{rate}^P) と再々符号化音声 (B_{rate}^{PPP}) の割り当てビット数の一致率 [%]

Table 3 Concordance rate for number of bits allocated in B_{rate}^P and B_{rate}^{PPP}

BitRate[kbps]	サブバンド番号								
	0	1	2	3	4	5	6	7	8
128	84.94	91.73	98.22	94.95	95.82	95.47	96.95	96.87	95.99
192	91.08	90.81	92.03	96.82	96.91	96.43	96.87	94.38	95.86
256	93.51	94.30	92.38	90.99	90.29	95.65	97.56	97.91	97.69
320	95.99	95.69	95.47	93.73	93.82	93.38	95.52	95.30	95.86
384	97.17	97.47	97.04	97.13	96.43	97.21	96.13	93.25	91.25

表 4 帯域制限音声 (B_{rate}^P) と再々符号化音声 (B_{rate}^{PPP}) の割り当てビット数の誤差の統計量

Table 4 Statistics for errors between number of bits allocated in B_{rate}^P and B_{rate}^{PPP}

BitRate[kbps]	サブバンド番号									
	0	1	2	3	4	5	6	7	8	
平均	128	0.13	0.06	-0.02	-0.03	-0.02	-0.02	-0.03	-0.03	-0.04
	192	0.05	0.03	0.03	-0.01	-0.02	-0.03	-0.02	-0.02	-0.02
	256	0.04	0.02	0.01	0.02	0.01	-0.04	-0.01	-0.01	-0.01
	320	0.03	0.03	0.02	0.02	0.01	0.00	0.01	-0.05	-0.04
	384	0.01	0.01	0.01	0.01	0.01	0.00	0.00	-0.02	0.01
分散	128	0.137	0.079	0.018	0.033	0.023	0.024	0.025	0.029	0.054
	192	0.092	0.091	0.080	0.016	0.023	0.026	0.015	0.016	0.017
	256	0.065	0.057	0.077	0.088	0.083	0.039	0.013	0.007	0.011
	320	0.042	0.042	0.045	0.062	0.062	0.066	0.045	0.045	0.040
	384	0.031	0.025	0.030	0.030	0.036	0.028	0.039	0.067	0.087

える。

このことをさらに詳細に調べるために、 G_{rate}^{all} , G_{rate}^P , G_{rate}^{PP} および G_{rate}^{PPP} の周波数スペクトルを調べた結果を図 7 に示す。ただし、384[kbps] で符号化した “Pops” を利用し、遮断周波数は $P = 9$ である。図 7(a) の G_{384}^{all} は約 20[kHz] 付近までスペクトルが存在するが、帯域制限によって図 7(b) に示すように約 6500[Hz] 以降の高域成分がカットされることがわかる。さらにこれを再符号化、再々符号化した音声のスペクトルを図 7(c), (d) に示す。これらは図 7(b) に示したスペクトルとほぼ同様の波形を示している。

これらの結果より、再符号化によってビット数を拡大した圧縮音声 B_{rate}^P の冗長部分をデータハイディングによる付加情報の埋め込みなどに利用しても、元の帯域制限した再生音声 G_{rate}^P にほぼ等しい G_{rate}^{PPP} を再現することができることが示された。

8. ま と め

本稿では、MPEG-1 オーディオレイヤー II で符号化した圧縮音声について、指定したサブバンド番号で帯域制限をした音声を帯域制限前の全体の符号量で再符号化し、さらにこれ

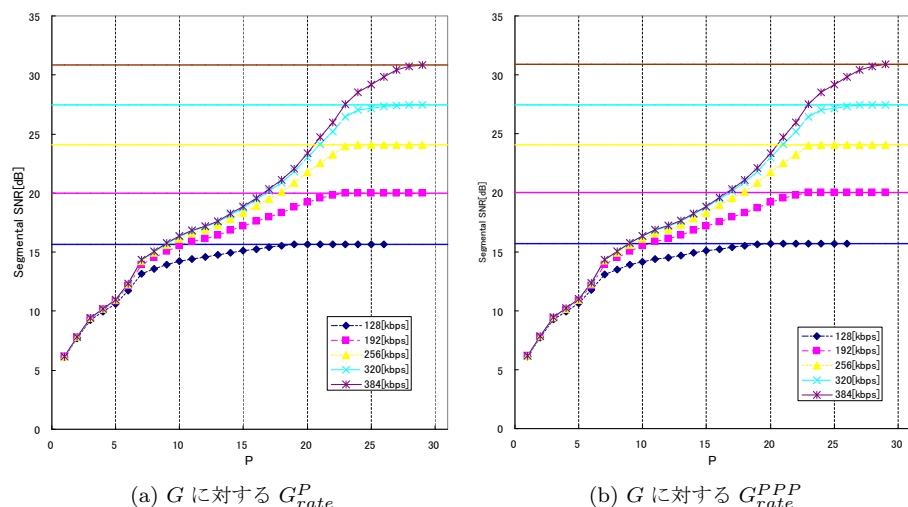


図 6 オリジナル音声に対する Segmental SNR の比較
Fig. 6 Comparison of Segmental SNR for original audio

を帯域制限後の符号量で再々符号化した場合の符号化データおよび再生音声変化について調査した。シミュレーション実験の結果、帯域制限した圧縮符号と再々符号化した圧縮符号のサブバンド毎の割り当てビット数(符号長)の変化は小さいこと。また、両者の再生音声は、オリジナル音声に対する Segmental SNR 値の計測および周波数スペクトルの比較によってほとんど差異がないことがわかった。

今後は、再符号化によって拡大した割り当てビット数の冗長成分を利用したデータハイディングの方法¹⁰⁾について検討する必要がある。

参考文献

- 1) 藤原 洋, 最新 MPEG 教科書, 株式会社アスキー, 2001.
- 2) 藤原 洋, 画像&音声圧縮技術のすべて, CQ 出版, 2000.
- 3) 安田 浩, 藤原 洋, デジタル放送インターネットのための情報圧縮技術, 共立出版, 1999.
- 4) 米満潤, MPEG 技術, トリケップス, 1997.
- 5) A. Matsuoka, K. Tanaka, A. Yoneyama and Y. Nakajima: "Data Embedding in MPEG-1/Audio Layer II Compressed Domain Using Side Information", *Proc. 2006 IEEE Int'l Conf. on Multimedia & Expo*, pp.1585-1588, 2006.

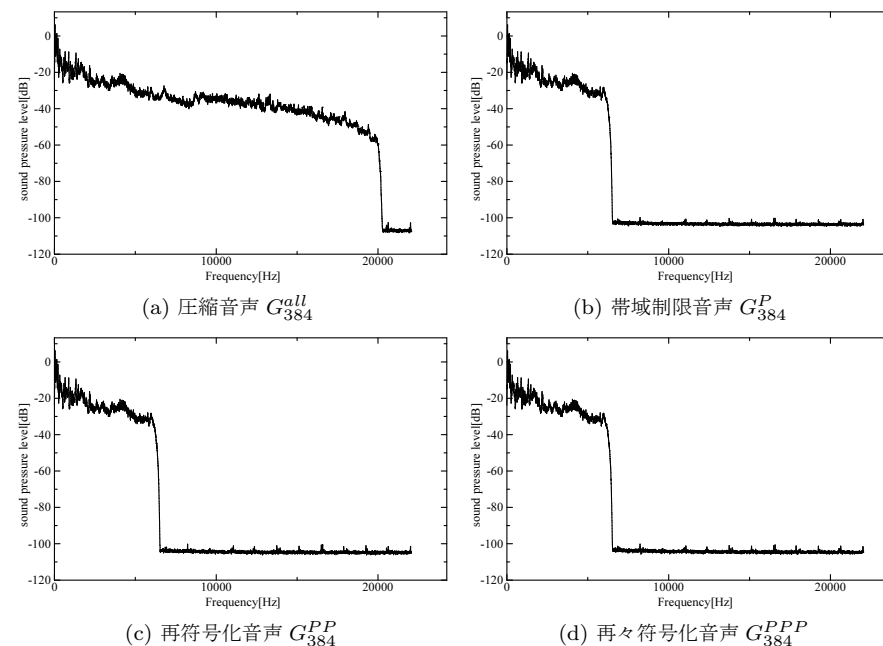


図 7 スペクトル波形の比較 ("Pops", $P = 9$, 44.1[kHz], 384[kbps])
Fig. 7 Comparison of frequency spectrum

- 6) 守谷 健弘, *Speech Coding*, 電子情報通信学会, 1998.
- 7) K. Takagi, S. Sakazawa, Y. Takishima: "Light Weight MP3 Watermarking Method for Mobile Terminals", *IEICE Trans. Fundamentals*, vol.E91-A, no.9, pp.2546-2554, 2008.
- 8) 後藤 真孝, 橋口 博樹, 西村 拓一, 岡 隆一: "RWC 研究用音楽データベース: 音楽ジャンルデータベースと楽器音データベース", 情報処理学会 音楽情報科学研究会, 2002-MUS-45-4, vol.2002, no.40, pp.19-26, May 2002.
- 9) 萩野谷 俊典, 田中 清, 高木 幸一: "MPEG-1 オーディオレイヤー II による高ビットレート符号化音声の品質向上", *ITE Technical Report*, vol.33, no.54, pp.5-8, ME2009-215, Dec. 2009.
- 10) 萩野谷 俊典, 田中 清, 高木 幸一: "データハイディングを用いる MPEG オーディオ向け準可逆ローパスフィルタ", 情報処理学会 オーディオビジュアル複合情報処理研究会, to appear, March 2010.