

楽音符号化品質に影響を及ぼす楽音信号の特徴量の検討*

松浦嶺, 山田武志, 牧野昭二, 北脇信彦 (筑波大)

1 はじめに

近年の楽音配信サービスの普及は目覚ましく, MP3 などにより高圧縮符号化された楽音を携帯端末で楽しむというスタイルが確立している. 一般に, 符号化楽音の品質は周波数帯域やビットレートなどの符号化条件により定まる. しかし, 実際には符号化条件が同じであっても楽曲(ソース信号)によって品質が変化することが知られている [1]. よって, 一定のサービス品質を保証するためには, ソース信号毎に適切な符号化条件を設定する必要がある. また, 符号化方式の品質評価を適切に実施するためには, 評価に用いるソース信号群をこのようならばつきを考慮した上で選定することが求められる.

高橋らは, 波形符号化方式であるサブバンド AD-PCM を対象として, 符号化楽音の主観品質に影響を及ぼす特徴量を提案した [1]. そして, ソース信号からこれらの特徴量を抽出することにより, ソース信号の違いによる主観品質の変動を予測できることを示した.

本稿では, まず高橋らが提案した特徴量が, 現在主に使われている MP3 などの変換符号化方式にも適用できるかどうかを検証する. 次に, MP3 などの聴覚特性を利用する変換符号化方式を対象として, マスキングに基づく特徴量を提案し, その有効性, 及び従来の特徴量との関係を検証する.

2 主観品質評価実験

2.1 実験条件

本稿では, 平栗らが実施した主観品質評価実験の結果を用いる [2]. 本実験は ITU-R BS.562-3[3] に準拠して行われており, 主観評価値は平均オピニオン評点 (MOS: Mean Opinion Score) により表される.

実験条件を表 1 に示す. ソース信号には, RWC 研究用音楽データベース [4] に含まれている 8 種類の楽曲を用いている. 以下では, これらを child(v), classic, jazz, rock(v), pop, opera(v), rock, pop(v) と呼ぶ. ここで, child は童謡であり, (v) は歌唱が含まれていることを表す. ソース信号の長さは 15 秒から 20 秒であり, サンプリング周波数は 44.1kHz, 量子化ビット数は 16 である. また, 符号化方式は Linear PCM と MP3 である. 被験者は成人男性 14 名, 成人女性

Table 1 Experimental conditions.

ソース信号	8 種類の楽曲
サンプリング周波数	44.1kHz
量子化ビット数	16
符号化方式	Linear PCM (4~13bit) MP3 (32~128kbps)
受聴環境	防音室におけるヘッドホン受聴
被験者数	男性 14 名, 女性 17 名
評価尺度	5 段階絶対品質評価尺度

17 名の計 31 名であり, 防音室内でヘッドホンによりソース信号, 及びそれらを符号化した信号を受聴し, 5 段階絶対品質評価尺度により評価している.

2.2 主観評価値の変動

同じ条件で符号化されたソース信号の主観評価値の間に有意差があるか否かを有意水準両側 5% の t 検定により調べた. PCM (10bit), PCM (7bit), MP3 (64kbps) の場合の検定結果を表 2~4 に示す. 表中の は, 対応するソース信号の主観評価値の間に有意差があることを示している. 高圧縮の符号化条件になるほど の数が増えており, ソース信号の違いが主観評価値に顕著な影響を及ぼしていることが分かる. 以下では, この 3 つの符号化条件を扱うことにする.

3 付加雑音に基づく特徴量とその有効性の検証

3.1 符号化による付加雑音に基づく特徴量

高橋らは, 波形符号化方式に特有の符号化ひずみである付加雑音に着目し, 中周波数帯域の雑音性 (cor), スペクトル概形の平坦性 (gain), スペクトルの調波性 (harm) の 3 つの特徴量を提案した [1].

中周波数帯域の雑音性 (cor)

cor は, 3~10kHz の周波数帯域における短時間自己相関関数のピーク値の全音フレームについての平均値である. 波形符号化方式による符号化ひずみは主に付加雑音として知覚される. よって, ソース信号の雑音性が低い, すなわち cor の値が大きいとき, こ

*A study on the characteristics of audio signals affecting the quality of coded audio signals. by Rei MATSUURA, Takeshi YAMADA, Shoji MAKINO, Nobuhiko KITAWAKI (University of Tsukuba)

Table 2 Results of t-test for PCM (10bit).

	child(v)	classic	jazz	rock(v)	pop	opera(v)	rock	pop(v)
child(v)	-							
classic	-	-						
jazz	-	-	-					
rock(v)	-	-	-	-				
pop	-	-	-	-	-			
opera(v)	-	-	-	-	-	-		
rock	-	-	-	-	-	-	-	
pop(v)	-	-	-	-	-	-	-	-

Table 3 Results of t-test for PCM (7bit).

	child(v)	classic	jazz	rock(v)	pop	opera(v)	rock	pop(v)
child(v)	-							
classic	-	-						
jazz	-	-	-					
rock(v)	-	-	-	-				
pop	-	-	-	-	-			
opera(v)	-	-	-	-	-	-		
rock	-	-	-	-	-	-	-	
pop(v)	-	-	-	-	-	-	-	-

Table 4 Results of t-test for MP3 (64kbps).

	child(v)	classic	jazz	rock(v)	pop	opera(v)	rock	pop(v)
child(v)	-							
classic	-	-						
jazz	-	-	-					
rock(v)	-	-	-	-				
pop	-	-	-	-	-			
opera(v)	-	-	-	-	-	-		
rock	-	-	-	-	-	-	-	
pop(v)	-	-	-	-	-	-	-	-

の付加雑音を検知することが容易になり、主観評価値が低くなると考えられる。

スペクトル概形の平坦性 (gain)

gain は、短時間フレーム毎に求めた 8 次線形予測利得の全有音フレームについての平均値である。波形符号化方式による符号化ひずみのスペクトル概形はほぼ平坦であるとみなせる。よって、ソース信号のスペクトル概形の傾きが急峻である、すなわち gain の値が大ききとき、パワーの小さい周波数で SNR が低下し、主観評価値が低くなると考えられる。

スペクトルの調波性 (harm)

harm は、短時間周波数スペクトルの調波成分と全スペクトル成分のパワー比の全有音フレームについての平均値である。gain と同様に、調波性が強い、すなわち harm の値が大ききとき、調波間の周波数で SNR が低下し、主観評価値が低くなると考えられる。

3.2 各特徴量と主観評価値の関係

cor, gain, harm と主観品質評価値の関係を図 1~3 にそれぞれ示す。ここで、色は符号化条件を表しており、個々のマーカはソース信号の一つに対応している。また、各特徴量と主観品質評価値の相関係数を表 5 に示す。PCM (10bit), PCM (7bit) の場合は、

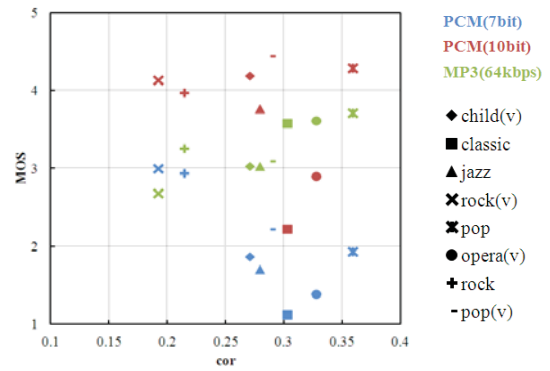


Fig. 1 Relationship between cor and MOS

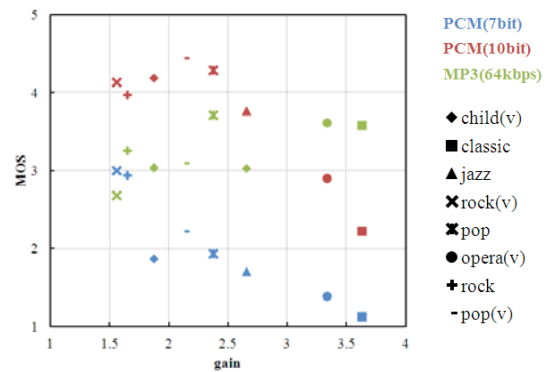


Fig. 2 Relationship between gain and MOS

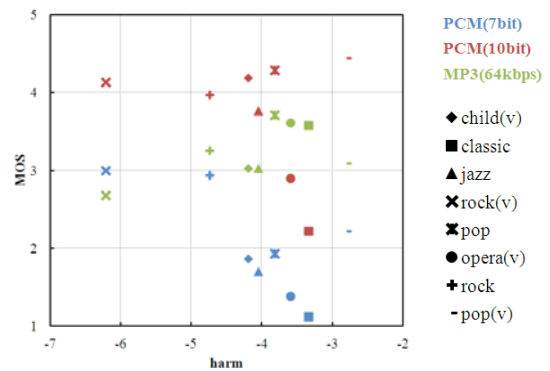


Fig. 3 Relationship between harm and MOS

cor, gain, harm の値が大きくなるにつれて、主観評価値が低下する傾向が見取れる。各特徴量と主観評価値の相関係数はいずれも負の値となっており、特に gain において強い相関が認められる。このように、波形符号化方式に対しては、各特徴量がソース信号の違いによる主観評価値の変動を良く捉えていることを確認できる。

一方、MP3 (64kbps) の場合は、cor, gain, harm と主観評価値の間には正の相関があり、各特徴量が想定している結果にはなっていない。しかし、相関係数の値は 0.6~0.8 と大きく、強い相関があると言える。

Table 5 Correlation coefficients between each feature and MOS.

	MP3	Linear PCM	
	64kbps	10bit	7bit
cor	0.81	-0.26	-0.77
gain	0.68	-0.86	-0.91
harm	0.61	-0.27	-0.70

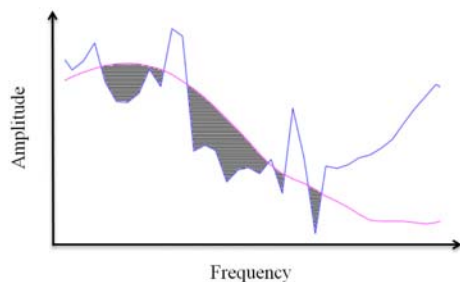


Fig. 4 Relationship between masking threshold (pink line) and spectrum (blue line).

これは、MP3 などの変換符号化方式において圧縮のために利用されている聴覚特性（マスキング）により説明できる可能性がある。次章では、マスキングに基づく特徴量を提案し、その有効性を検証すると共に、cor, gain, harm との関係进行分析する。

4 マスキングに基づく特徴量とその有効性の検証

4.1 マスキングに基づく特徴量

MP3 などの変換符号化方式においては、聴覚特性を圧縮のために利用している。例えば、聴覚特性の一つである同時マスキングは、ある周波数の音が他の周波数におけるより大きな音のために聞こえ難くなる現象であり、これを利用して効率的なビット割当を実現している。

本稿では、同時マスキングに基づく特徴量（SIM: Simultaneous Masking）を提案する。SIM の求め方は次の通りである。まず、短時間フレームごとにスペクトルとマスキングしきい値を求める [5]。次に、20Hz から 12,000Hz の間でマスキングしきい値がスペクトルを超える周波数について、その差を求めて合計する（図 4 の網掛け部の面積に相当する）。最後に、その合計値を全有音フレームについて平均する。SIM の値は、マスキングにより聞こえ難い成分が多く含まれているほど大きくなる。SIM の値が大きくなるほど効率的なビット割当が可能となる、すなわち主観評価値が高くなると考えられる。

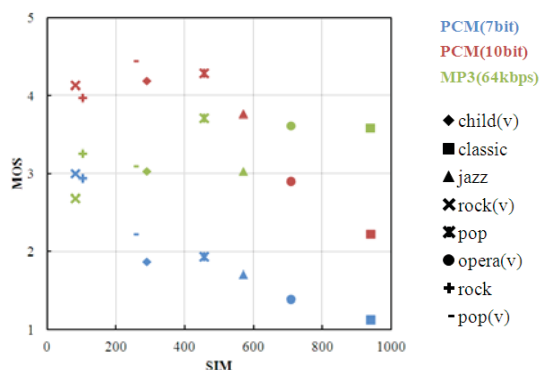


Fig. 5 Relationship between SIM and MOS.

Table 6 Correlation coefficients between SIM and MOS.

	MP3	PCM	
	64kbps	10bit	7bit
SIM	0.67	-0.85	-0.94

Table 7 Correlation coefficients between SIM and each feature.

	cor	gain	harm
SIM	0.68	0.98	0.58

4.2 SIM と主観評価値の関係

SIM と主観品質評価値の関係を図 5 に示す。また、SIM と主観品質評価値の相関係数を表 6 に示す。MP3 (64kbps) の場合は、SIM の値が大きくなるにつれて主観評価値が高くなるという傾向がある。実際、SIM と主観評価値の間には強い正の相関があることから、SIM がソース信号に違いによる主観評価値の変動を捉えていると言える。一方、PCM (10bit), PCM (7bit) の場合は、SIM の値が大きくなるにつれて主観評価値が低下する傾向が見取れる。SIM と主観評価値の間には強い負の相関があり、これは cor, gain, harm と同じ傾向であることから、SIM とこれら 3 つの特徴量に関係性があると考えられる。

SIM と cor, gain, harm の関係を図 6~8 にそれぞれ示す。また、特徴量間の相関係数を表 7 に示す。SIM は各特徴量と強い正の相関を示しており、特に SIM と gain の相関は極めて強いことが分かる。gain はソース信号のスペクトルの傾斜が急峻なときに大きくなるが、このときマスキングしきい値がスペクトルを超え易くなるために SIM も多くなると考えられる。cor と harm についても同様に関係性を説明することができる。よって、SIM とこれら 3 つの特徴量

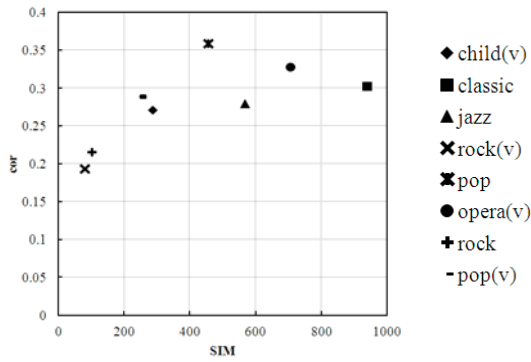


Fig. 6 Relationship between SIM and cor.

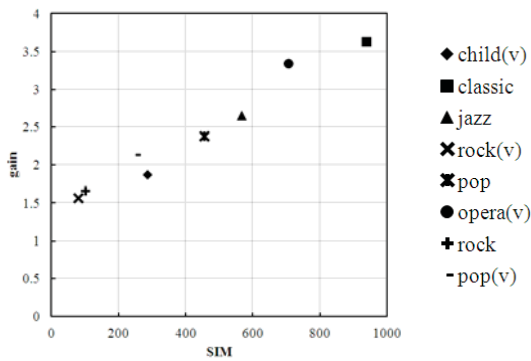


Fig. 7 Relationship between SIM and gain.

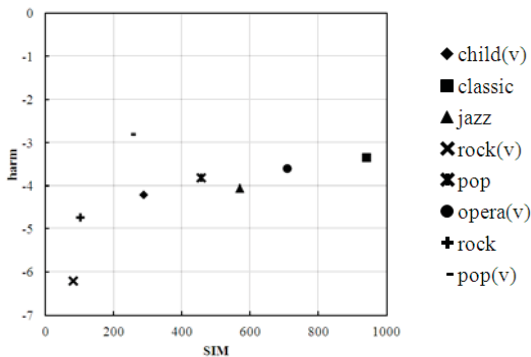


Fig. 8 Relationship between SIM and harm.

は算出方法が異なるものの、同種の現象を捉えているとみなすことができる。よって、SIM, cor, gain, harm の4つの特徴量は、PCMとMP3の両方に対して、ソース信号の違いによる主観評価値の変動を予測するのに有効であると言える。

5 おわりに

本稿では、符号化楽音の主観品質に影響を及ぼす特徴量について検討した。PCMを対象として提案された3つの特徴量と、MP3を対象として本稿で提案し

たマスキングに基づく特徴量は、算出方法が異なるものの同種の現象を捉えていること、及びPCMとMP3の両方に対してソース信号の違いによる主観評価値の変動を予測するのに有効であることを確認した。

参考文献

- [1] 高橋玲, 長淵裕美, “符号化した音楽信号の品質評価に用いるソース信号の選定尺度,” 電子情報通信学会論文誌, Vol. J76-A, No. 1, pp. 18–26, Jan. 1993.
- [2] 平栗陽介, 北脇信彦, 山田武志, “楽音符号化方式の客観品質評価法の検討,” 日本音響学会春季研究発表会, pp. 329–330, Mar. 2004.
- [3] ITU-R BS.562-3, “Subjective assessment of sound quality,” 1990.
- [4] 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: 研究目的で利用可能な著作権処理済み楽曲・楽器音データベース,” 情報処理学会論文誌, Vol. 45, No. 3, pp. 728–738, March 2004.
- [5] M. Bosi, R. E. Goldberg, “Introduction to digital audio coding and standards,” Kluwer Academic Publishers, 2003.