

短時間実効値による音楽CDのジャンル別変動特性

On a Property of Short-Time RMS Distribution of Music Genres

川瀬 翔[†]、 井 研 治^{††}
Shou KAWASE, Kenji INOMOTO

Abstract In the last decade, personal music players have become used by an increasing portion of the population, particularly young people. Music signal from such players has been a significant threat to hearing because it can reach very high sound volumes without quality loss. This paper describes music level characteristics with musical genres to discuss signal level variation. To estimate total quiet portion of music signal, short time root-mean-square values within 10ms are calculated every 5ms period from music CD. Frequency distribution curves of the short time rms is well expressed by a beta distribution. An index expressing music quietness is defined by the probability of strms going down below -6dB from its rms value. Some relationships between the -6dB probability and musical genres are indicated. It is concluded that the -6dB probability takes larger value in Classical, HipHop/Rap, Jazz and R&B genres than others.

1. はじめに

2001年にアップル社から携帯音楽プレーヤ iPod が発売され、若者を中心に通勤・通学や、ジョギング・ウォーキングにもこれを利用する光景が見られるようになった。日本レコード協会の調査によると、2009年のわが国の携帯音楽プレーヤの普及率は41.9%に上る¹⁾。

しかし、携帯音楽プレーヤの使用による事故、たとえば歩行者と救急車との接触事故や、自転車と列車による死亡事故も報告されており、公共交通機関内における音漏れの迷惑も懸念される。

また、大音量受聴による聴力減退も懸念され、EUの委員会は2008年に携帯音楽プレーヤ（携帯電話含む）を大音量に設定し、5年にわたって1日1時間以上聴き続けた場合、5~10%の人の聴力が永久に失われる恐れがあると警告している²⁾。さらに、米国医学会誌には、12~19歳の難聴者の割合は1994年には約20人に1人の割合であったが、2005~2006年には約5人に1人に増加したという報告がある³⁾。

これら大音量化の背景には携帯音楽プレーヤの普及や、音楽ジャンルによる特性、さらに音楽コンテンツに施されるデジタル編集技術があげられる。特に放送などで用いられた場合に購買層の注意を引きつける目的で、コ

ンプレッサ、つまり非線形振幅処理によるレベル上昇操作がCD作成時の編集段階で行われることがあり、これによって音楽コンテンツの静寂部分のレベルは持ち上げられる傾向にある⁴⁾。

本研究は、音楽コンテンツのレベル変動を測定し、その特徴を数値化することを目的としている。

まず、基本となる音楽コンテンツのレベル変動の測定法について検討し、10ms間の短時間実効値を用いる。音楽コンテンツは身近なCDを用い、短時間実効値の分布とコンテンツが分類されるジャンルとの関係に言及する。

さらに、短時間実効値の頻度分布をベータ分布で表現し、実効値から-6dB以下となる時間率を指標に選び、ジャンルとの関連性を追求する。また、ジャンルによる頻度分布の非対称度の違いを skewness (歪度) を用いて考察する。

2. 分析方法

CDコンテンツのおおまかな分析手順を図1に示す。まず、フリーソフト iTunes を用いて音楽CDから wave ファイルを作成する。wave ファイルには音楽コンテンツの瞬時値が 44.1kHz のサンプリング周波数で記録されている。瞬時振幅はステレオ 2ch でCDに記録されているが、分析では左右の平均を用いている。

[†] 愛知工業大学大学院工学研究科電気電子工学専攻(豊田市)

^{††} 愛知工業大学工学部電気学科(豊田市)

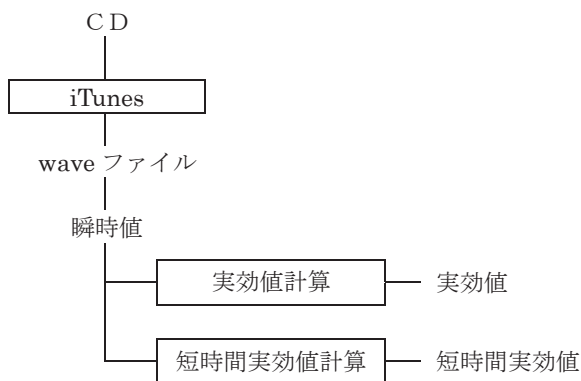


図 1 分析手順

分析の最初に、コンテンツ全体の平均レベルを表す実効値、すなわち瞬時値の 2 乗平均値の平方根を求める。次に、瞬時値のデータから一定の時間間隔で短時間実効値を順次計算し、これを用いてコンテンツのレベル変動を考察する。短時間実効値の計算方法については、次節でとりあげる。

2・1 実効値指示機器の動特性

CD コンテンツ内の録音レベル変動を測定するには、短時間における実効値がこの目的に合う。そのため、音楽を含む音声信号のレベル変化を測定する既存の機器について、その動特性を調査することからはじめる。

実効値指示計器の時間応答特性、つまり動特性は、一般にユニット応答が用いられる⁵⁾。特にこのシステムを 1 次遅れ系と見なした場合、指示値が最終値の 63.2% に達する時間である時定数は、動特性を評価する際の目安として身近なものである。

現在、一般に用いられている実効値測定機器を、この時定数を含めて調査すると次のようになる。

1) 騒音計 音響測定で一般に用いられる騒音計は、指示メータの動特性により、fast 特性、slow 特性と呼称され、それぞれの時定数はほぼ 125ms、1sec である⁶⁾。この機能は変動する騒音に対しては時定数を長くして指示の読み取りを容易にすること、つまり変動を抑えることが目的である。したがってこれらの時定数はそもそも変動を測定する目的には長すぎると思われる。

2) VU メータ 放送の現場で用いられる VU (Volume Unit) メータは、電話回線の音量監視用として米国のベル電話研究所、CBS、NBC の 3 社が共同で開発したもので、1939 年から米国工業規格に加えられた⁷⁾。以後、AM 放送の現場でも過変調を防止するため被変調波の監視に使用されるようになり、目盛りには変調度も併記されるようになる。VU メータの立ち上がり時定数

は 65ms であるが、放送の品質が高まるにつれ、短時間ピーク信号の見落としが指摘されるようになる。一方、指示を見やすくするため立ち下がり時定数は大きく選んである。

3) ピークプログラムメータ VU メータの欠点を補う目的で、動特性をさらに高めた PPM (Peak Program Meter) が開発され、主にヨーロッパの放送現場ではこれが従来の VU メータと併用されるようになる⁷⁾。PPM の時定数は 5ms、10ms、12ms など、統一されていないが、音声信号のピークを忠実に拾い上げるには時定数を 10ms 前後にすべきことがわかる。

以上を時定数の順番にまとめ、表 1 に示す。なお、遮断周波数 f_c は時定数 T_c から次式で計算した。

$$f_c = \frac{1}{2\pi T_c} \quad (1)$$

表 1 実効値表示機器の動特性の定義と時定数^{6,7)}

音響機器		時定数	遮断周波数[Hz]
Peak Program Meter	Nordic N10	5ms	31.8
	DIN 45 407	10ms	15.9
	IEC 268-11	10ms	15.9
	BBC 55428 Part 10	12ms	13.3
インパルス騒音計		35ms	4.55
VU メータ		65ms	2.45
騒音計	fast 特性	125ms	1.27
	slow 特性	1sec	0.159

2・2 短時間実効値

有限区間 T における実効値は式(2)で計算される。ここで T を積分時間という。

$$v_{rms} = \sqrt{\frac{1}{T} \int_0^T v(t)^2 dt} \quad (2)$$

この式を用いて求めたレベル変動から、楽器の演奏時間間隔を測定した研究⁸⁾では、 T が 5ms 以下のときは波形の基本周期成分が十分平滑されずに山谷の変動となってレベル波形に残ることが指摘されている。そして、積分時間 T を 10ms~20ms に選べば、立ち上がり鋭いピアノであっても十分追跡可能であると結論づけている。

$T=10ms$ の場合、後述される式(4)の議論から -3dB 遮断周波数 f_c は 44.3Hz になる。また、この f_c から求めた等価的な時定数は 3.59ms である。

本研究では積分時間 $T=10ms$ とした式(2)の値を短時間実効値(strms, Short Time RMS)と呼び、この式をデ

デジタル系で実現させて $strms$ を順次求め、音楽CDのレベル変動評価を試みる。

2・3 短時間実効値の特性

本研究で使用する積分時間 $T=10\text{ms}$ とした $strms$ はどのような性質を有しているか、最初に確認しておく。

まず、遮断周波数について整理しておく。 $strms$ 計算における積分の関係、つまり、

$$y(t) = \frac{1}{T} \int_0^T x(t) dt$$

において、 $x(t)$ と $y(t)$ の関係は、積分区間が矩形状パルスで表されるから、そのフーリエ変換から周波数応答が求められて次式で示される⁹⁾。

$$|H(f)| = \frac{\sin(2\pi fT)}{2\pi fT} \quad (3)$$

式(3)から遮断周波数 f_c を求めると、

$$f_c = 0.443 / T \quad [\text{Hz}] \quad (4)$$

となる。

これを確かめるため、実効値が変動する信号を変調度 m の AM 変調波で模擬し、式(3)、(4)を確認する。

図2に、その結果を式(3)の理論値とともに示す。理論では、 $strms$ 計算の -3dB 遮断周波数 f_c は 44.3Hz であったが、実験でもこれが確認される。そして、 f_c 以下の変動は -3dB 以内の精度で検出されるが、 f_c より高い周波数の変動は遮断されることがわかる。

音楽CDなど、 44.1kHz でサンプリングされたデジタルデータでは、 -3dB 遮断周波数 f_c と積分個数 p_1 の関係は次式で与えられる。

$$p_1 = \frac{0.443}{f_c} \times 44100 = \frac{19540}{f_c} \quad (5)$$

2・4 まとめ

本研究で用いる実効値 rms 、および短時間実効値 $strms$ の計算方法をまとめて次に示しておく。

$$rms = \sqrt{\frac{1}{N} \sum_{i=0}^{N-1} x_i^2} \quad (6)$$

$$strms(j) = \sqrt{\frac{1}{441} \sum_{i=j \times 221}^{j \times 221 + 441} x_i^2} \quad (7)$$

$$j = 0, 1, 2, \dots, \text{int}(N/441)$$

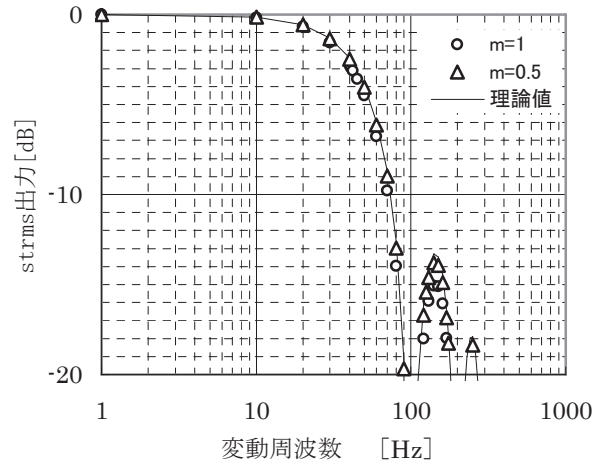


図2 $strms$ の変動特性

(m は信号に用いた AM 波形の変調度)

ここで N は音楽コンテンツの総サンプル個数、 441 は時間幅 10ms に相当するサンプル個数 p_1 である。また、

図2から変動周波数の上限を余裕を持って 100Hz とみなせば、 $strms$ の変動情報を損なわないようにサンプリングするナイキスト周波数は 200Hz で、その周期は 5ms となる。数字 221 はこのような観点から定めた $strms(j)$ の計算間隔 5ms (厳密には 5.0113ms) に相当するサンプル個数である。

測定された $strms(j)$ の性質は、次のようにまとめられる。

積分重み	矩形窓
積分時間	$T=10\text{ms}$
積分個数 (CDデータの場合)	$p_1=441$
-3dB 遮断周波数	$f_c=44.3\text{Hz}$

このようにして測定された rms と $strms$ の例を図3に

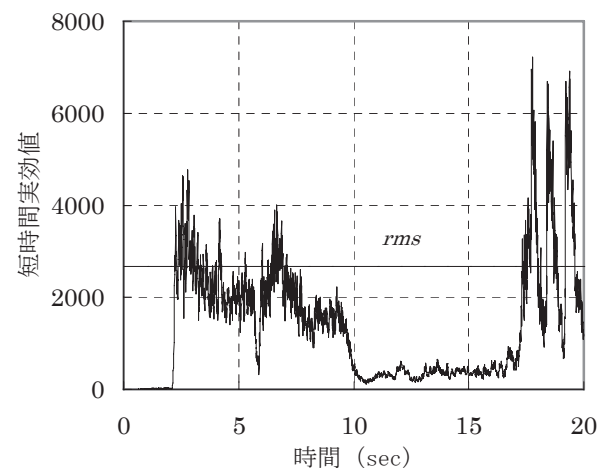


図3 rms と $strms$ の一例

示す。

wave ファイルは 16 ビットが用いられているため、縦軸の値の範囲は 0~32768 であり、この最大値、つまり、記録可能な最大値を K とする。図は、実際に記録されている値が K よりかなり小さい例である。

3. strms 変動のモデル化

変動の様子を説明するため、strms の分布形を考えてみる。CD は 16 ビットを用いたデジタル記録方式であるが、記録可能な最大値 K に比べて 1 ビットの変化幅は十分小さいため、連続分布とみなして考察する。

3・1 strms の分布

いくつかの音楽コンテンツについての分析結果を strms の出現した最大値で規格化後、表示すると図 4 のようになり、その分布形は対称、あるいは非対称が混在していることがわかる。

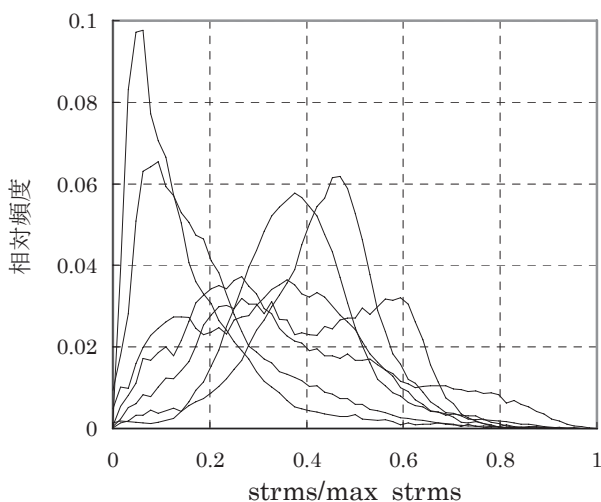


図 4 規格化した短時間実効値の頻度分布

次に測定された頻度分布曲線が、うまく表現される分布関数を見いだすことを考える。

分布形に多様性があること、および変域が有限区間であることから、ここではベータ分布を採用する。

文献によると、このベータ分布は、 α 、 β によって分布形はいろんな形をとること¹⁰⁾や、 α 、 β を動かすことによってほとんどの密度関数のグラフを書くことができるため、非常に貴重な分布である¹¹⁾、あるいは、きわめて融通がきくため、さまざまな観測データの頻度図に当てはめることができる¹²⁾、等の表現が用いられている。

3・2 ベータ分布

ベータ分布¹³⁾は、正の母数 α 、 β によって次式で定義される。ここで、 $f(x)$ は確率変数 x における確率密度で、本研究では strms を 0~1 に規格化したものに x をあてはめている。

$$f(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha) \cdot \Gamma(\beta)} \cdot x^{\alpha-1} \cdot (1-x)^{\beta-1} \quad (8)$$

ただし、 $n > 0$ のとき

$$\Gamma(n) = (n-1)! \quad n ; \text{整数}$$

$$\Gamma(n) = \int_0^\infty e^{-x} x^{n-1} dx \quad n ; \text{非整数}$$

図 4 で見たように短時間実効値は対称、あるいは小振幅側に偏った分布が見られるため、モデルの母数には、 $\alpha < \beta$ の条件をつける。このとき、ベータ分布の確率密度関数を描くと図 5 のようになる。

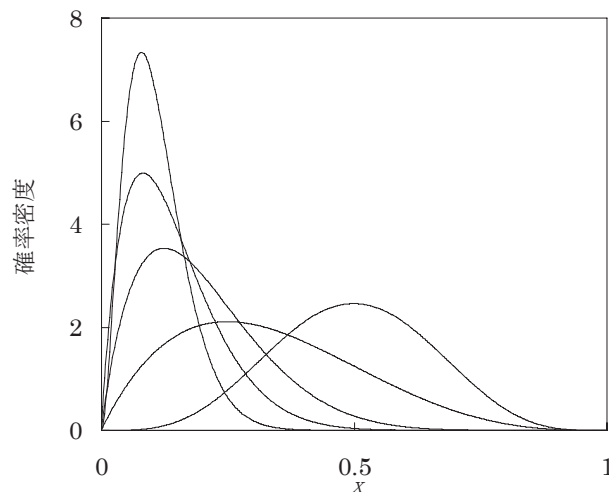


図 5 ベータ分布

(左より $\alpha, \beta = 3, 24 \quad 2, 12 \quad 2, 8 \quad 2, 4 \quad 5, 5$)

一方、母数 α 、 β が定まれば、ベータ分布の 2 乗平均の平方根が式(9)で定められるので、これを rms と関連づけることができる。同様にベータ分布の標準偏差を表す式(10)によって短時間実効値の(平均値周りの)標準偏差 sd を求めることができる。

$$rms = \sqrt{\frac{\alpha(\alpha+1)}{(\alpha+\beta) \cdot (\alpha+\beta+1)}} \quad (9)$$

$$sd = \sqrt{\frac{\alpha\beta}{(\alpha+\beta)^2 \cdot (\alpha+\beta+1)}} \quad (10)$$

これらを用いると、ベータ分布の確率密度から短時間実効値が $\text{rms}-6\text{dB}$ 以下になる確率 η を見積もることができ、測定結果と同じ座標軸での議論が可能になる。

4. 音楽コンテンツの分析

音楽コンテンツの分析を、データベースに登録されているジャンルに従って分類し、分析結果とジャンルの関係を考察する。

4・1 音楽コンテンツのジャンル

ジャンルは iTunes が採用している Gracenote¹⁴⁾ のデータベースに、世界中で出版された全CDのジャンルが整理されているので、これを用いた。

ジャンルは9つに分類され、以下の特徴がある。

- Classical ; 器楽の独奏から大編成オーケストラによる管弦楽曲、あるいはオペラなど、さまざまな音楽形態を含む。
- Jazz ; 4拍子の2,4拍にアクセントを持つオフビートにブルーノート音階を基本とし、即興性に富む音楽形態。
- Electronica ; 電子音楽に影響を受けた音楽で1拍または1小節程度の細かなリズムとフレーズが反復される。
- Jpop/Pop ; ポピュラーミュージック。
- Reggae ; ゆったりしたリズムで歌い、曲調が最後まで変化しない音楽。
- R&B ; リズム&ブルース。コンピュータで演奏し、歌唱がメイン。
- HipHop/Rap ; コンピュータ演奏に、喋るように歌うラップを乗せたもの。
- Rock ; ボーカル、ギター、ドラム、ベースで演奏される激しい音楽形態。
- Punk ; 3種類の和音であるスリーコードと、2音からなるパワーコードで演奏する簡素な Rock。

分析に用いた音楽コンテンツのジャンル別サンプル数

ジャンル	コンテンツ数
Classical	233
Hiphop/Rap	190
R&B	215
Jazz	154
Punk	53
Rock	303
Jpop/Pop	398
Electronica	109
Reggae	259

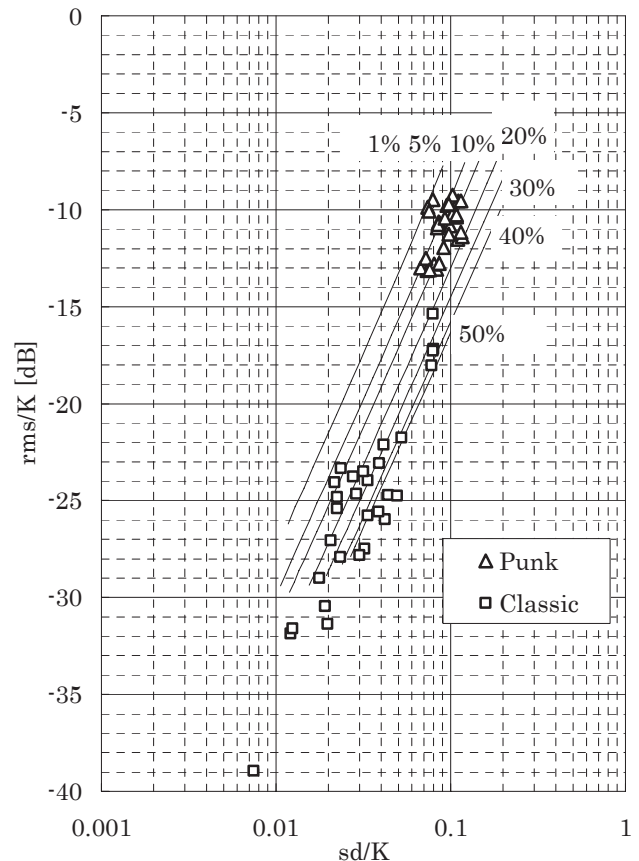


図6 2ジャンルの分析結果、および -6dB 時間率

を表2に示す。

次に、データベースに登録されている上記のジャンルへ、音楽コンテンツの短時間実効値分布を関連づける試みを以下で行う。

4・2 音楽コンテンツのレベル変動

分析結果は $K=32768$ を用いて次のように整理した。

縦軸 ; 平均レベル= rms/K

横軸 ; レベル変動= sd/K

ここで、 sd は strms の標準偏差を表す。

全9ジャンルを一度に表示すると多すぎるので、PunkとClassicalのジャンルのみを図6に表示した。図中の記号、 Δ や \square は、コンテンツごとの分析結果の一部である。

ジャンル Classical に比べて Punk は測定値の散らばりが少ない。そして Classical に比べて rms と sd がとも大きくになっている。

図中の直線はベータ分布から求めた確率 η で、これは短時間実効値 strms が実効値の半分以下、つまり $\text{rms}-6\text{dB}$ 以下になる確率(時間率)を示しており、百分率で示した η が図中に注記されている。

この η は、特定の時間率について求めた点を連ねた等

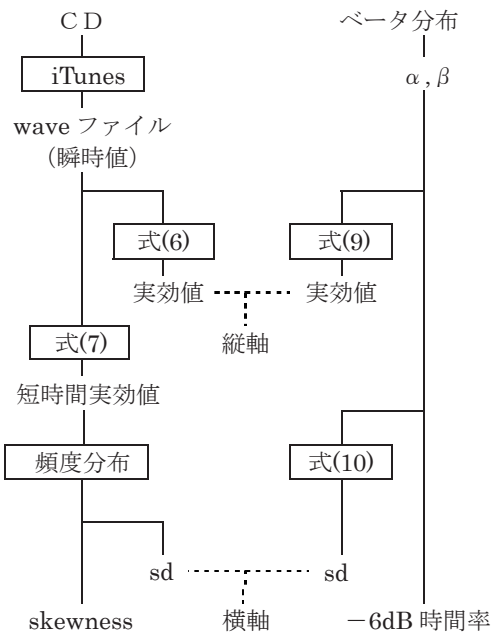


図 7 ベータ分布による分析手順

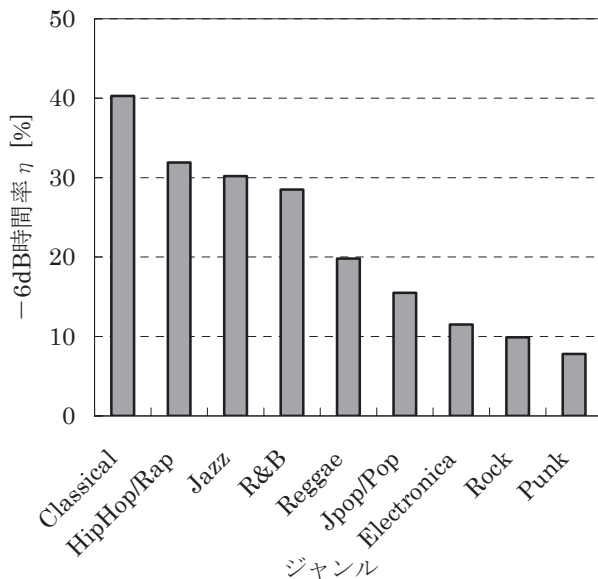


図 9 ジャンルごとの -6dB 時間率 η

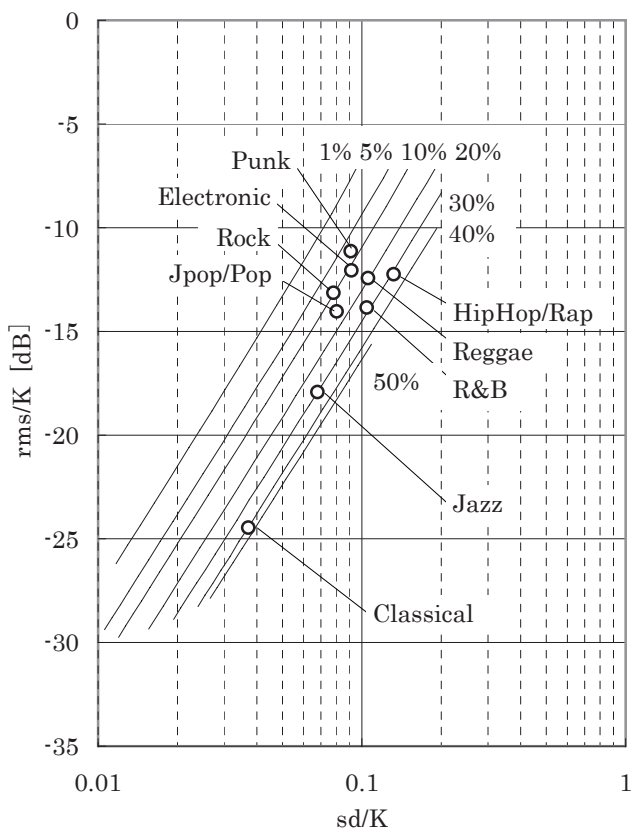


図 8 ジャンルの平均レベルとレベル変動

高線であり、この時間率を音楽コンテンツの静寂性を表す尺度とした。ここに述べた分析の手順を図 7 に示す。なお、図中の skewness については次節で述べる。

各ジャンルの分析結果を、その平均値で表示すると図 8 になり、Jazz と Classical が他の集団より離れた位置にあることがわかる。図 8 から、内挿して各ジャンルの -6dB 時間率 η を求め、大きさの順に並べると図 9 が得られる。η は 5% から 20% を越えるまで約 4 倍以上の変化幅を見せており、η が大きいことは、そのジャンルの短時間実効値が下がりやすく、比較的静寂な部分が多くみられるとしてよい。

したがって Classical や HipHop などコンテンツに比較的多くの静寂部分が存在するとみられる。

一方、図 8 でレベル変動が小さい Punk、Rock、Electronica、などは -6dB 時間率 η が小さくなっており、レベル変動と η との関連性を裏付けることができた。η が小さいこれらのジャンルは、音が小さくなりにくいので、rms をそろえて聞きくらべた場合、外部の音、例えばアラームや危険信号音を聴覚で検知することがおろそかになりやすいと推測される。

ここでは、短時間実効値が -6dB 以下になる時間特性に関しては取り扱わないので、変動速度に関する議論は行わない。

4・3 分布曲線の skewness

短時間実効値 $strms(j)$ の分布曲線が小振幅側に偏っていること、つまり分布曲線の非対称度を定量的に評価するため、以下で定義される skewness (歪度) を用いる¹⁵⁾。

まず、 $strms(j)$ の分布を表す確率密度を $f(x)$ とし、以下の平均 m を準備する。

$$m = \int_{-\infty}^{\infty} x \cdot f(x) dx$$

これを用いて、平均値の周りの2次モーメント $\mu_2 = \sigma^2$ と平均値の周りの3次モーメント μ_3 を、次式から求める。

$$\mu_2 = \int_{-\infty}^{\infty} (x-m)^2 f(x) dx \quad (= \sigma^2)$$

$$\mu_3 = \int_{-\infty}^{\infty} (x-m)^3 f(x) dx$$

これらを元にして、skewness は以下の計算で与えられる。

$$skewness = \frac{\mu_3}{\sigma^3}$$

ところで、ベータ分布の skewness は母数 α 、 β から次のように求められる¹³⁾。

$$skewness = \frac{2 \cdot (\beta - \alpha) \sqrt{\alpha + \beta + 1}}{(\alpha + \beta + 2) \cdot \sqrt{\alpha \beta}} \quad (11)$$

各ジャンル内の平均 rms と平均 sd から時間率を読みとった結果は表3のようにまとめられる。これには各コンテンツから求めた skewness の平均も記した。

また、図10に両者の平均的な関係をグラフで○印で表示した。図中の3本の曲線はベータ分布モデルに基づいて短時間実効値を計算した結果で、母数 β が 2.5α 、 5α 、および 30α としたものを示した。これを実際の分析データ○印と比べてみよう。

まず skewness が1以下の領域にあるジャンルは、その strms の分布はジャンルの中でも対称分布に近いといえる。そして、この5ジャンルの -6dB 時間率は20%以下であり、コンテンツの strms は比較的变化に乏しい。しかしこれらの5ジャンルは $\beta = 2.5 \sim 30\alpha$ に広がっており、 β 分布の母数との明確な傾向は認めがたい。

skewness が1以上となる4ジャンルは、 $\beta = 5\alpha$ の分

表3 ジャンルごとの実測値

ジャンル	-6dB 時間率 [%]	skewness
Classical	40.3	1.36
HipHop/Rap	31.9	1.16
Jazz	30.2	1.03
R&B	28.5	1.05
Reggae	19.8	0.50
Jpop/Pop	15.5	0.56
Electronica	11.5	0.53
Rock	9.9	0.71
Punk	7.5	0.79

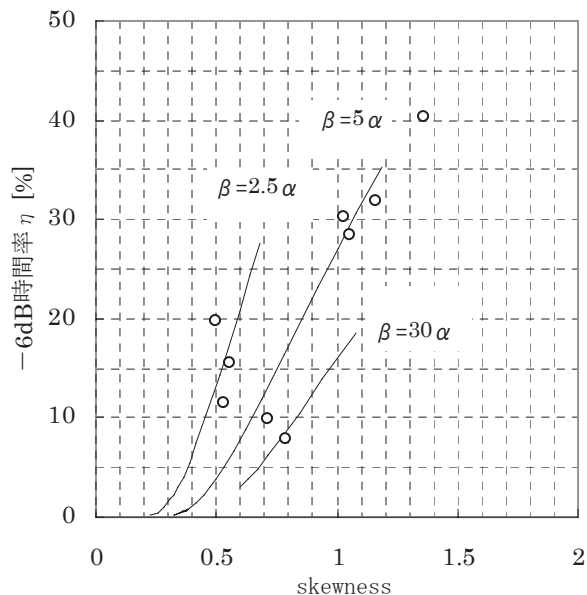


図10 skewness と -6dB 時間率 η

布に近い。このジャンルは、Classical、HipHop/Rap、Jazz、R&B であり、短時間実効値の分布は非対称になる。-6dB 時間率 η は25%以上で、これらは他のジャンルに比べて比較的静寂部分が多いことがわかる。

ここでは skewness=1 以上の値でジャンルの特徴付けを試みたが、そのような分布曲線の一例を図11に示す。横軸は最大値で規格化した strms に相当した量を表しており、それぞれのカーブから横軸に下ろした破線は、実効値の位置を示している。このように skewness=1 では strms の分布が低レベルに偏っており、その rms は最大値の7~8%付近にある。一方、skewness が0.5程度に下がるジャンルでは、strms の分布曲線は対称形に近づいている様子がわかり、その rms は37%付近にある。

本研究で分析した9ジャンルの strms の振幅分布形は、概ねこれら2つの分布形の間にあると思われる。

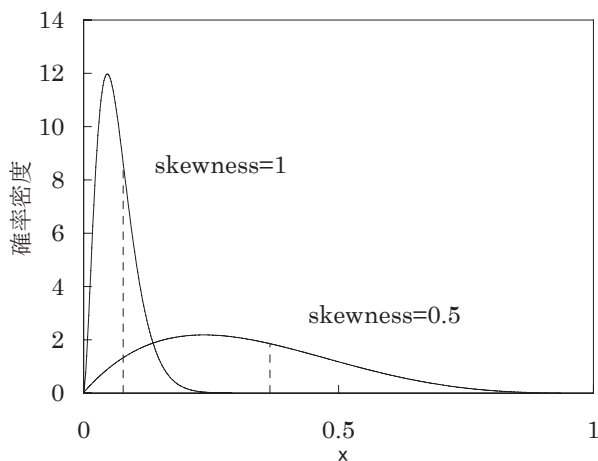


図11 skewness による分布形状の一例

5. まとめ

本研究では、短時間実効値を用いて音楽コンテンツの分析を行った。

最初に短時間実効値の分析方法を整理し、10ms 間の 2 乗積分値の平均を採用した。そしてこの方法の周波数特性を調べ、-3dB となる遮断周波数が 44.3Hz であることを確認した。

次に実際の音楽コンテンツを分析し、短時間実効値の標準偏差によって音楽コンテンツの揺らぎを表し、平均レベルである実効値と組み合わせて図に表し、各ジャンルごとの平均値で代表させて簡潔な表示を得た。

また、短時間実効値の振幅分布をベータ分布をあてはめ、実効値から-6dB 以下になる確率（時間率）によって特徴づけることを提案し、音楽コンテンツのジャンルによってこの-6dB 時間率が異なる傾向を明らかにした。

さらに、短時間実効値の分布曲線の skewness（歪度）を数値化し、-6dB 時間率との関連性を議論した。

一例として-6dB 時間率が約 25% の値でジャンルは 2 分され、小さい値のジャンルは静寂性が乏しいと推定した。

音楽コンテンツに対応しているジャンルは、データベースに登録されている内容を利用した。音楽コンテンツがどのジャンルに分類されるかは明確な基準で峻別されているのではなく、曖昧さがつきまとうものの短時間実効値 *strms* にジャンルによる違いが見られた。

本研究では-6dB 以下になる平均時間長に関しては言及していない。これについては *strms* が *rms* と交差する零交差時間の分析等によって、進んだ議論が可能になると思われ、今後の研究を待ちたい。

静寂部分が少ない音楽コンテンツのイヤホン受聴で

は、外部の警告音などが遮断されてしまうことが考えられるが、この問題の定量的考察は聴覚のマスクングと絡めて議論する必要がある。-6dB 以下になる平均時間長や、マスクングに関する追求などは、いずれも今後の課題である。

参考文献

- 1) 2009 年度音楽メディアユーザー実態調査報告書、日本レコード協会、2009.
- 2) Potential health risks of exposure to noise from personal music players and mobile phones including a music playing function. European Commission, 2008.
<http://europa.eu/rapid/pressReleasesAction.do?reference=IP/08/1492&format=HTML&aged=0&language=EN>
- 3) Josef Shargorodsky, et.al.; Change in Prevalence of Hearing Loss in US Adolescents, JAMA, Vol. 304, No. 7, 2010.
- 4) 葛巻善郎；マスタリングの全知識、p.11、リットーミュージック、2008
- 5) 近藤文治；基礎制御工学、p.70、森北出版、1977.
- 6) 子安勝編；騒音/振動(上)、p.73、コロナ社、1979.
- 7) Glen M. Ballou ed. Handbook for Sound Engineers, 3rd. ed. p.867, Focal Press, 2002.
- 8) 日本音響学会編；音楽と楽器の音響測定、p.226、コロナ社、2007.
- 9) F.R. コナー(鎌田一雄訳)；信号入門、p.35、森北出版、1985.
- 10) 脇本和昌；統計学、p.113、日本評論社、1984.
- 11) 白石高章；統計科学、p.97、日本評論社、2003.
- 12) 伊藤學；土木・建築のための確率統計の基礎、p.128、丸善、1977.
- 13) 箕谷千鳳彦；統計分布ハンドブック、p.625、朝倉書店、2003.
- 14) <http://www.gracenote.com/>
- 15) Harald Cramer; Mathematical Methods of Statistics, Almqvist & Wiksells, 1945.

(受理 平成 23 年 3 月 19 日)