

# 音の高さと音の長さの相対的な物理的関係性と 情報理論に基づいた音楽生成モデルの提案

大村 英史<sup>1,2,a)</sup> 柴山 拓郎<sup>3</sup> 高橋 達二<sup>3</sup> 澁谷 智志<sup>3</sup> 太原 育夫<sup>1</sup>

**概要:** 音楽は期待の実現や裏切りによって情動を引き起こすといわれている。本研究では、期待の実現・裏切りを情報理論に基づいて音楽を生成するモデルの提案を行い、実装システムの紹介をする。音楽はいくつかの音の物理的特性の関係性によって構造が成立している。本研究では、音の音高と音価と呼ばれる2つの特性に注目する。五度圏と呼ばれる音高の相対的な関係性と、音価と呼ばれる音の長さの相対的な関係性を用いて、情報理論に基づいた確率分布のコントロールによる音楽生成モデルを提案する。そして、このモデルを組み込んだ音楽生成システムを紹介し、生成された音楽の考察を行う。

## 1. はじめに

聴取者は音楽から、生き生きとした情景を感じ、情動的な感覚を得る。音楽と情動の関係は1950年ころから頻繁に行われるようになってきている [1]。音楽の構造は情動を引き起こす原因の一つとして考えられている。Meyer や Narmour によると、音楽は構造によって聴取者に期待感を感じさせ、その裏切りや実現によって情動を引き起こすといわれている [2], [3]。近年の音楽と情動の関係を表した Huron の ITPRA 理論や Juslin らの BRECVEM モデルでも期待は重要な位置づけである [4], [5], [6]。本研究では、音楽を情報として捉えることにより音楽構造を抽象的に扱いつつも情動との関係を定量的に表現することを可能にするモデルの設計を行う。本稿では、はじめに情報理論に基づく期待感が得られるであろう音（またはその構造要素）の出現確率を定量的に与える方法の説明をする。次に、音高（音の高さ）と音価（音の長さ）の音楽の構造に注目し、本モデルで利用するこれらの関係性の説明を行う。最後に、提案モデルを説明し、このモデルを実装したシステムの紹介をする。

## 2. 期待感と情報理論

### 2.1 期待感の生理的反応

音楽と情動の理論で人間の期待感が注目されているだけでなく、脳波を用いた生理実験でも人間が意識・無意識の両方で刺激に対して敏感であることが確認されている [7]。これはミスマッチ陰性電位 (MMN: mismatch negativity) とよばれており、ある特定の刺激イベント（例えば 400Hz の聴覚刺激）を反復し与え、任意のタイミングで他の刺激イベント（例えば 600Hz の聴覚刺激）を与えた場合、この2種類の ERP と呼ばれる刺激に対する脳波を加算平均し比較すると、その刺激のオンタイムから 300 ミリ秒後に負の変化が見られる、というものである。この現象は、人間が常に環境からの知覚の逸脱を監視しているということであり、芸術作品の鑑賞やアミューズメントでこの反応が複雑に生じていることが推測される。以上の知見は、期待の実現と裏切りの音楽聴取時の情動との深い関係性を示唆している。

### 2.2 不確実性と情報理論

期待からの逸脱は、あらかじめ期待された構造の規則性の減少や消失から引き起こされる。これは複雑性の増加として見なすことができる。複雑性については、Berlyne が最適複雑性モデルとして「快情動」と「複雑性」の関係性を示している [8]。横軸に複雑性 (complexity)、縦軸に快 (hedonic) 情動を設定すると、図 1 の実線で描かれているような逆 U 字で表すことができる。また、この関係は経験によりシフトしていく (図 1 の点線)。MMN も順化によっ

<sup>1</sup> 東京理科大学  
Tokyo University of Science

<sup>2</sup> 東京工芸大学  
Tokyo Polytechnic University

<sup>3</sup> 東京電機大学  
Tokyo Denki University

a) hidefumi.ohmura@gmail.com

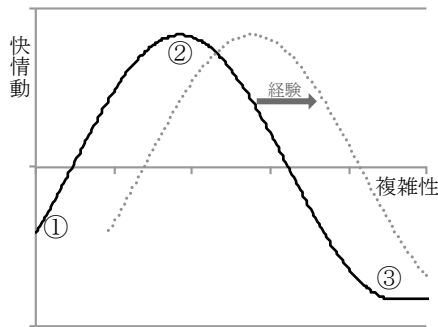


図 1 逆 U 字関数 ([8] を改変)  
Fig. 1 Inverse U function

て消失することが確認されており、この関係と合致する。

以上の知見より、期待の実現や逸脱のためには複雑性の定量化が必要であることがわかる。複雑を期待（予測）できない事象がより頻度が高く生じる見なすと、複雑性は不確実性として考えることができる。不確実性は情報理論のエントロピーによって定量的に表現可能である [9]。情報理論では、情報の伝達がどの程度意味があり価値のあるものであるかという指標を情報量という値として求めることができる。事象  $i$  が起きたときの情報量  $I$  は次式で表現する。

$$I = -\log p_i \quad (1)$$

ここでは、 $i$  が生じる確率を  $p_i$  とする。 $p_i$  が小さいほど  $I$  が大きくなり、滅多に起こらない事象は情報の価値が高い、つまり情報量が大いことを表している。そして、 $n$  個の事象がそれぞれ  $p_1, p_2 \dots p_n$  で生じるときの情報量の期待値は次式で求められる。

$$H = -\sum_{i=1}^n p_i \log p_i \quad (2)$$

この値が大きくなると、様々な事象が生じることを表している。逆に小さくなると、単一の（またはごく少ない）事象が生じることを表している。つまり、この値が大きくなると、不確実性が高まり、複雑な情報伝達が生じている言い換えることができる。この値をエントロピーまたは平均情報量と呼ぶ。

本研究では、情報理論に基づいた不確実性をコントロールするためにガウス関数を用いる。ガウス関数は、正規分布の確率密度関数であり、多くの自然現象や社会現象が従う分布の一つである。ガウス関数は以下の式で表される。

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad (3)$$

これは平均値  $\mu$ 、分散  $\sigma^2$  の正規分布の確率密度を表す関数である。

ガウス関数の分散  $\sigma^2$  の値を変えることにより、分布の尖度調整することができる。分散  $\sigma^2$  の値を小さくすると、関

数の曲線が扁平となり確率分布は一様分布に近づく。このとき、エントロピーは最大値に近づく。この状態は、生じる事象を予想することが困難となり、複雑性が上昇していることがわかる。一方、分散  $\sigma^2$  を大きくすると、関数の曲線は尖る。このとき、エントロピーは最小値に近づく。この状態は、特定の事象のみ生じないため予測することが容易になり、複雑性が低下していることがわかる。ガウス関数では  $\sigma^2$  の値とエントロピーの値が連動しており、容易に複雑性の設定が可能であるため提案モデルに採用する。

### 3. 本研究で用いる音楽構造

#### 3.1 音の高さの関係性：五度圏

音楽は空気の振動である音から成立している。一秒間に振動する回数を周波数としてあらわすと、この値が音の高さの絶対量となる。この音の高さを音高 (pitch) という。実際の音はいくつかの周波数の合成によって成立しており、最も低い周波数を基本周波数といい、聴覚上この音高の音として知覚される。

二つの音の音高の関係は周波数の比で表すことができる。この関係は音程と呼ばれている。この比が簡単であればあるほどよく響き合う。例えば、1:1 の音程は、同じ音高の音であり完全 1 度と呼ばれる。1:2 の音の関係性は、2 倍または半分の音を表しており、完全 8 度またはオクターブと呼ばれている。さらに 2:3 は 1.5 倍を表しており、完全 5 度と呼ばれている。この関係の逆数 3:2 の音は 2/3 倍であり、この音の 2 倍つまり 3:4 の関係は完全 4 度と呼ばれ、物理的にも近い関係にある。西洋の音楽理論では、これらの音を完全協和音程と呼び、最も響き合う音程であると定義している。1:2 と 2:3 比率を用いて、順に音を定義していく方法をピタゴラス音律という。ピタゴラス音律では、3/2 倍を 12 回繰り返した音高と、2/1 倍を 7 回繰り返した音高が微妙にずれる\*1。このずれを解消するためにオクターブ内の音を等間隔に 12 等分した 12 平均律がある。12 平均律は正確な比率ではないが、12 の音高が同等に扱えるメリットがある\*2。また、わずかなずれ\*3は人間の曖昧な処理で補正されるため、聴取者は純正な比率の音程と比べない限りはほとんど違和感を感じない。そこで、本研究では各音高を同等に扱いたいため 12 平均律を用いて議論を進める。そして、図 2 のように、時計回りを完全 5 度 (3/2 倍)、反時計回りを完全 4 度 (2/3 倍) として描いた五度圏の関係性を用いる。

#### 3.2 音の長さの関係性

音の空気振動の周波数は音高であったが、振動している

\*1 このずれをピタゴラスコンマという。

\*2 ピタゴラス音律の場合、常に起点となる音高を定義しなくてはならない。

\*3 ピタゴラスコンマを 12 の音程に分けるため。

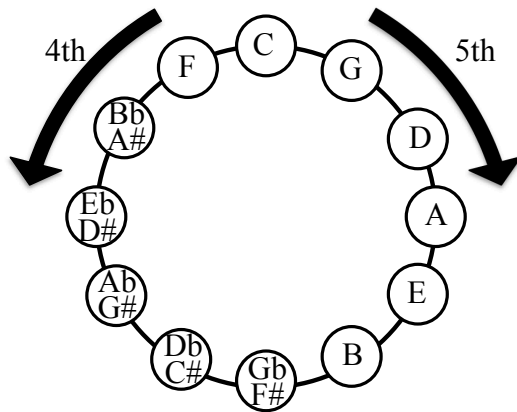


図 2 五度圏  
Fig. 2 circle of fifths

(発音している) 期間を音価 (value) という。音価も音程のように基準に対する相対的な比率で表す。基準となる長さは一般的には小節を用いる。例えば 4 分音符は小節を 4 分割した長さの発音を意味する。

本研究では、音は次の音が発音されるまでの時間を支配すると考え<sup>\*4</sup>、発音が開始される時間的位置に着目し、音の長さの関係性として扱う。この位置によりリズムが作り出される。ある特定の長さの間隔で発音される音は拍動 (pulse) を作り出す。拍動の間隔を拍節と呼ぶ。本研究では、拍節の中のどの位置に音を配置するかを決定することで、リズムのパターンを作り出す。これらの位置は拍節を分割することで得る。分割は素数での分割を繰り返す行いで、様々な位置を得ることが可能である。音楽では、2 分割、3 分割を用いることが多い。5 分割以上は複雑になるためほとんど用いられることはない。例えば 5 拍子の曲を 2 + 3 として捉えたり、7 連符の演奏を 4 + 3 として捉えたりする。これは、5 以上の分割は人間にとっての知覚は困難であるからだろう。また、分割が混合している場合も人間は正確に知覚ができないと言われている [10]。このことから、リズムパターン生成の基本は 2 または 3 分割を考えれば良いことが示唆されている。今回は、単純化のため、2 分割のみ扱う。

## 4. モデル化

音の高さの関係性と、音の長さの関係性それぞれにおいてモデル化を行う。

### 4.1 音の高さのモデル化

音の高さは音程によって関係性が作られる。前述の五度圏を用いてモデル化を行う。ガウス関数の横軸に、五度圏の関係性、つまり正方向に完全五度、負方向に完全四度の

\*4 例えば、スタカートで演奏した旋律と、レガートで演奏した旋律は、聴取者は異なるアーティキュレーションであると感じるが、同じ旋律であると認識することが可能であるため。

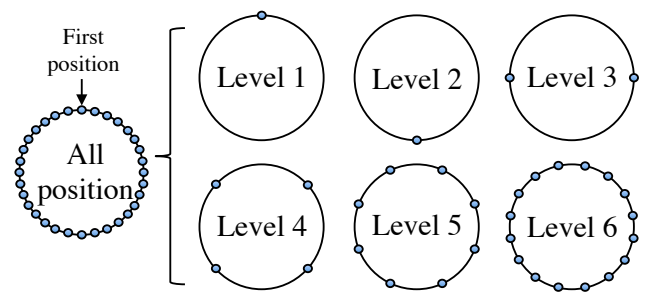


図 3 発音位置  
Fig. 3 Positions of Notes

順に配置する。五度圏は円環であるため、 $\mu$  の反対側の位置で関数の曲線を結合する。また、 $\mu$  の位置は最も出現頻度の高い音であり、旋法の主音として見なすことができる。反時計回りの方向にずらした場合は、ガウス関数が 4 度の方向にずれる、時計回りの方向にずらした場合は 5 度の方向にずれる。これは転調や移調に該当する。

### 4.2 音の長さのモデル化

音の長さの関係性で述べたように、拍節を 2 分割していくことで発音位置を定義する。図 3 に分割方法の概要を示す。円の一周は拍節を表し、最上部の位置をはじめの頭拍とする。Lv.1 は分割なしの位置を示す。Lv.2 は 1 回分割を行った位置を示す。Lv.3 は 2 回、Lv.4 は 3 回と示し、Lv.6 まで 5 回までの分割を示す。例えば、拍が 4/4 の場合、Lv. 1 は全音符を組み込む場合の位置であり、Lv.6 は 32 分音符を組み込む場合の位置である。Lv.7 以上も考えることが可能であるが、64 分音符以上は装飾音との区別がつかなくなり、リズム的要素として意味をほとんど持たないことが予想される。そこで、Lv.6 までを考慮することとする。これにより、拍節内に 32 個の位置が存在し、ここから音を選択する。分布は片側の正規分布を用いる。関数の  $x$  座標は、 $\mu$  の位置をレベル 1 とし、等間隔に順にレベルを配置する。

### 4.3 実装

実装は、HTML および JavaScript を用いて行った<sup>\*5</sup>。音の出力には Web Audio API を用いた。動作確認は Google Chrom で行っている。出力される音には純音 (正弦波) を用いている。音量の調節やテンポの設定ができる。操作画面のにある「Play/Stop」ボタンでシステムが起動/停止する。

音の高さのモデルの実装は、二つの正規分布の混合分布を用いている。これは、長調短調などの偏りのある旋法も表現できるようにしたためである [11]。二つのガウス関数の平均値  $\mu$ 、分散  $\sigma^2$  および関数の  $y$  軸方向のサイズを調

\*5 <https://sites.google.com/site/hidefumiohmura/home/program/sigmus109>

節することができる。

音の長さのモデルは、 $\mu$  の位置はつねに一定であるため  $\sigma^2$  のみ変更可能である。また、拍節内で複数の音を鳴らすこと（旋律の生成）を目的としているため、選択位置の可能性を作り出すサンプル数の設定と、そのサンプルから実際に鳴らす音を抽出数の設定をできるようになっている。

#### 4.4 考察

音の高さのモデルでは、旋法に従ったような旋律が得られる。一つの正規分布から得られる音は、分散を大きくするにつれ複雑性が増していく。旋法と比較すると、単純なオクターブから、5度と4度の響き、ペンタトニックスケールのような旋律、ダイアトニックスケールのような旋律、そして最大値まで上げるとクロマティックスケールのような旋律が得られる。これらは複雑性の増加として見なすことができ、分布から計算されるエントロピーの変化と合致する。

二つの分布を組み合わせるにより最大極値の位置を左右にずらすことができる。左にずらしていくと、だんだんと明るい雰囲気に変化し、右にずらしていくと暗い雰囲気に変化することがわかる。音楽的には、左にずらすと長調に近づき、右にずらすと短調に近づく。だいたい7つの音が発音する程度に分散を設定すると、極値が中央の時はドリア旋法に似た響きの旋律が得られる。右周りに一つ極値がずれるとエオリア旋法（短調）、二つずれるとフリギア旋法、三つずれるとロクリア旋法の響きに似た旋律を得ることができる。また、左回りに一つずれるとミクソリディア旋法、二つずれるとイオニア旋法（長調）、三つずれるとリディア旋法の響きに似た旋律を得ることができる。5つの音の場合も同様に、メジャーペンタトニックスケールや、マイナーペンタトニックに似た旋律が得られる。また、二つの分布の重ねずに離して配置した場合、沖縄旋法や都節音階に似た響きの旋律も得られる。

音の長さのモデルでは、分散が小さい場合低いレベルの音のみで、分散を大きくするにつれて、高いレベルの音も選択されるようになる。低いレベルばかりが選択される場合、単純なリズムとなり、聴取者は発音される場所の予測も容易である。これは、童謡やわらべうたなどを連想させるリズムである。一方高いレベルも選択されるようになると、複雑なリズムとなり、聴取者は発音される場所の予測が困難となる。さらに、高いレベルばかりになると、連続的に拍節を聞いていても聴取者は頭拍の特定が困難になることが予想される。このようなリズムは、拍節構造のない現代音楽のようなリズムを連想させる。

音の高さと音の長さの複雑性を調節して生成される旋律を聞き比べると、どちらも複雑性が低い方が聞きやすい。しかし、複雑性が低すぎる場合は、単純な旋律が生成され、連続的に聴取していると聞き苦しくなってくる。直感的に

は図1のBerlyneの最適複雑性モデルが正しいように感じられる。実際の音楽は、このような複雑性が途中で変化させることが情動喚起につながっているのかもしれない。

旋律の複雑性が提案モデルで定式化できたので、今後はこのモデルを使った心理実験および生理実験を検討していきたい。また、拍節間の構造関係も音楽と情動の関係には重要な要素だと考えられる。それらのモデル化も取り組んでいきたい。

## 5. おわりに

音楽は期待の実現や裏切りによって情動を引き起こすといわれている。音の音高と音価と呼ばれる2つの特性に注目し、五度圏と呼ばれる音高の相対的な関係性と、音価と呼ばれる音の長さの相対的な関係性を用いて、情報理論に基づいた確率分布をコントロールすることによる音楽生成モデルを提案した。そして、このモデルを組み込んだ音楽生成システムを紹介した。

## 参考文献

- [1] P. N. Juslin and J. A. Sloboda: *Handbook of Music and Emotion: Theory, Research, Applications*. Oxford University Press (2011).
- [2] L.B. Meyer: *Emotion and meaning in music*, University of Chicago Press (1956).
- [3] E. Narmour: *The analysis and cognition of basic melodic structures*, University of Chicago Press (1990).
- [4] D. Huron: *Sweet anticipation: Music and the psychology of expectation*, MIT Press (2006).
- [5] P. N. Juslin, S. Liljeström, D. Västfjäll, and L. O. Lundqvist: *How does music evoke emotions? Exploring the underlying mechanism*. in *Handbook of music and emotion, theory, research, application*, P. N. Juslin and J. A. Sloboda, Eds, Oxford University Press, Oxford. Chap. 22, pp. 605–644 (2011).
- [6] P. N. Juslin and D. Västfjäll. *Emotional responses to music: The need to consider underlying mechanisms*. *Behavioral and Brain Sciences*, Vol. 31, No. 5, pp. 559–621 (2008).
- [7] 宮田洋, 藤沢清, 山崎勝男, 柿木昇治: 新 生理心理学 1 巻 生理心理学の基礎, 北大路書房 (1998).
- [8] D. E. Berlyne: *Aesthetics and psychobiology*, Appleton Century Crofts (1971).
- [9] C. E. Shannon: *The Mathematical Theory of Communication*, The University of Illinois Press (1949).
- [10] P. Essens: *Hierarchical organization of temporal patterns*, *Perception & Psychophysics*, 40, pp. 69–73, (1986).
- [11] 大村英史, 柴山拓郎, 片上大輔, 湯浅将英, 太原育夫: 人工的雰囲気生成のための分布関数のデザイン, 第31回フェジシステムシンポジウム (FSS2015), TA4-1, (2015).