

## 音高と音価の情報量操作による音楽と言語の構造

## Musical and Linguistic Structures Consisting of Pitch and Rhythm

大村英史 \*1  
Hidefumi Ohmura柴山拓郎 \*2  
Takuro Shibayama\*1東京理科大学  
Tokyo University of Science\*2東京電機大学  
Tokyo Denki University

Music is consist of relationships between feature quantities of physical factors. We developed a music generation system based on information entropy and physical tone features. The system provides controlling frequency of tone features which are pitch and note value. In low entropy, output music has musical structures. We can regard the outputs as music. On the other hand, in high entropy, we can regard them as pronounced language. However the outputs do not have linguistic structures. This phenomenon shows that though music and language have auditory features, a quality of musical information is differ from a quality of linguistic information. In this study, we discuss them on semiotics, and consider a next step of the system.

## 1. はじめに

音楽は聴取者に生き生きとした情景を感じさせ、情動的な感覚をもたらす。情動と音楽の関係の研究は1950年頃から活発に行われている [1]。音楽の物理的な構造は、情動に関わる要因として考えられており、いくつかの理論が提案されている。現代でもよく引用される古いモデルである Meyer [2] の理論やそれを発展させた Narmour [3] の理論によると、音楽は構造によって聴取者に期待感を抱かせ、その期待感の裏切りや実現によって聴取者の情動を引き起こす。比較的新しい Huron の ITPRA 理論 [4] や、Juslin らの BRECVEM モデル [5, 6] においても、期待感は聴取者の情動において重要な位置づけをしている。

私たちは期待感に着目し、音楽の構造が聴取者に期待感を抱かせるためシステムの開発を行ってきた [7, 8]。このシステムは情報理論に基づいた「起こりやすさ」を制御するための情報理論に基づいたエントロピーを操作することで、音の高さ（音高）と音の長さ（音価）の物理的特性音楽構造を作り出す。本システムにおいて、エントロピーが低い状態ではシンプルな音楽が生成される。反対にエントロピーを徐々に高くしていくと、より複雑な音楽が生成され、最終的にランダムに音が発声される状態が作り出される。本システムによってエントロピーを極端に高めた状態で作り出された音の集合体は、音楽というよりもむしろ言語の発声に近い印象をもったものとなった。エントロピーが中庸であれば、音楽として感じられる音の集合体が生成され、その音の集合体は音楽的な構造を有していると考えられるが、エントロピーを極端に上げたときに生成される、言語の発声として感じられる音の集合体は、当然ながら言語的な構造を有していない。本研究では、このような減少が生じる理由を、言語や音楽を含めた記号論 [9, 10] を参考にしながら、音楽的な構造と言語的な構造の生成過程を共通した基盤上に捉えることを試みる。

本稿では、前半に提案したシステムと出力される音楽について説明をする。後半では音楽と言語について記号論的な観点から考察を行い、提案システムの今後の展望について述べる。

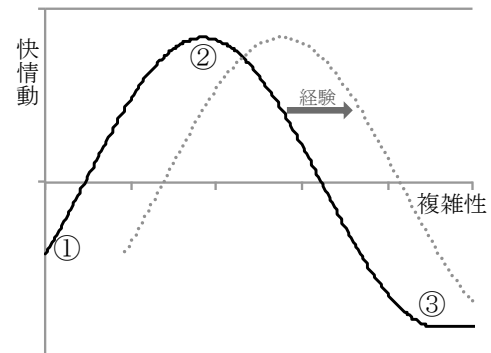


図 1: 最適複雑性モデル ([12] を改変)

## 2. エントロピーに基づいた音楽生成システム

## 2.1 期待感の裏切りと実現の定量化

期待感は音楽の情動理論において要素の一つとして考えられている。ミスマッチ陰性電位という現象として脳波の生理実験でも意識・無意識のいずれにおいても期待からの逸脱の反応が確認されており [11]、芸術作品の鑑賞やアミューズメントコンテンツの利用時などにおいてもこのような反応が生じていることが推測される。

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。これは規則によって構造化されており、規則がより複雑になることにより生じると考えることができる。複雑性については Berlyne が最適複雑性モデル [12] として「快情動」と「複雑性」の関係性を図 1 のように示している。さらにこの関係性は経験によって変化する。

この複雑性を定量的に扱うために、規則が複雑であることを、生じる事象の期待（予測）が困難な状態として見なすと、複雑性は不確実性としてとらえることができる。この不確実性は情報理論のエントロピー（平均情報量）で定量的に表現できる [13]。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、事象  $i$  が起きたときの情報量  $I$  は次式で求められる。

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

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、つまり情報量が大き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。この値が大きくなると不確実性が高まり、予測が困難な状況（期待感の逸脱）が生じている。この値をエントロピーと呼ぶ。

本システムでは、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。ガウス関数は、分散  $\sigma^2$  の値とエントロピーの値を同義に扱うことができ、容易に複雑性の設定が可能であるため本システムに用いた。

## 2.2 音の関係性

ガウス関数の縦軸は確率であるが、横軸は音の設定が必要である。本システムでは、音楽構造を作り出す物理的特性である音の高さ（音高）と音の長さ（音価）の関係性を踏まえた上で設定を行った。

### 2.2.1 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り音高（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、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。また、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

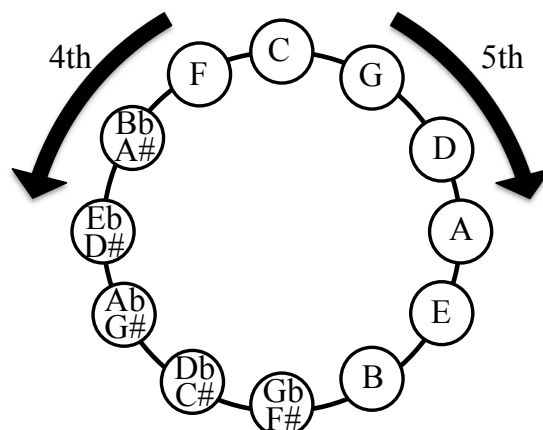


図 2: 音高の関係性：五度圏

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<sup>\*1</sup>。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。また、わずかなずれ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、本システムでは各音高を同等に扱うことを目的としているため12平均律を用いた。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。

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5度、負方向に完全4度の順に配置する。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、ガウス関数が4度の方向にずれ、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相当する。

### 2.2.2 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（発音している）期間を音価（note 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>\*2</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。そして、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とが可能となる。これらの位置は拍節を分割することで得られ。分割は素数での分割を繰り返すことで、様々な位置を得ることを可能としている。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

\*1 このずれをピタゴラスコンマという。

\*2 例えば、スタッ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。

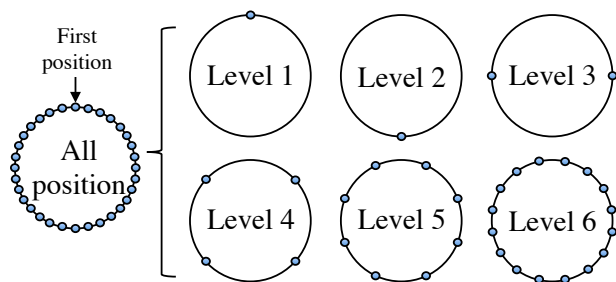


図 3: 発音位置

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いわれている [14]。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る。今回は、単純化のため、2 分割のみ扱う。

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る、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。

## 2.3 実装

実装は、HTML および JavaScript を用いて行った\*3。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。操作画面のある「Play/Stop」ボタンでシステムが起動/停止する。

## 2.4 出力される音列

エントロピーを下げていくと音楽的な音列が出力される。音高においては、頻度の高い主音や選択される音高が限定される魔法のような構造が生成される。音価においては、頭拍や裏拍がよく選択されるためサイクルがはっきりし、小節のような構造が生成される。このように、音楽的な構造を得ることができる。

一方で、エントロピーを上げていくとランダムな音列が出力される。各旋律の発音タイミングを同期させた場合、言語を話しているような響きが得られる。しかしながら、言語的な構造を得ることはできない。

次の章では、このような現象がなぜ起きるのか考察していく。

## 3. 記号的構造

本章では、言語と音楽において、情報量操作によって上記のように異なる結果が出た理由を検討していく。言語や音楽

を含めた記号体系に関する研究分野に記号論や記号学がある。音楽と言語はどちらも音を使った記号であり、どちらも聴覚を使って受信する。また、時間的変化を有しているため、過去との変化で現在が形成され論弁的に意味をなす。両者は類似点が多いため、同じ機能を祖としているのではないかとも考えもある [15]。以上のような類似点も踏まえつつ相違点を考えながら以下の三つの点について検討していく。

- なぜ音楽的な構造は得られ、言語的な構造は得られなかったのか。
- エントロピーを増加させたとき、なぜ言語的に聞こえるのか。
- なぜ音楽は記号的な振る舞いをするのか。

### 3.1 通信としての記号

言語も音楽も、発信者が意味されるものを記号化することで作成され、受信者が記号から意味を読み取って情報を獲得する。言語においては具体的な事象を記号化することが多いが、音楽においては情動のような抽象的な事象であることが多い。言語の場合、特に人工言語の場合には意味されるものが誤って伝達することはない。送受信者が人間である自然言語の場合は人工言語と比べると情報伝達の精度が落ちるが、ある程度の情報伝達の正確さは担保されていると考えられる。一方、音楽の場合は意味されるものが正しく伝達されることの方が少ない。そのため、解釈といった形で様々な意味理解があることがわかる。

このような違いの原因を探るためには、Peirce の 3 つの記号分類が役に立つ。一つ目は類似の関係で成立する「アイコン」、二つ目は因果関係や近接で成立する「インデックス」、三つ目は関連のない対象を恣意的に結びつきを作る「シンボル」である。アイコンとインデックスは指し示す理由があるため有契性といい、シンボルは指し示す理由がないため無契性という。言語は基本的にシンボルである。ただし、擬音語などの有契性がある記号も存在する。一方、音楽には無契性の記号は存在しない。すべてが事象の相対関係で成り立っており、その結果として意味をもつ。たとえば、頭拍はある時間的位置が頭拍のではなく、ある特定の時間間隔で発音される音が多い場合、その位置を頭拍であると聴取者（受信者）が意味づけする。また、カデンツは終了するための印ではなく、音程の関係性により聴取者が終了するように聞こえるからカデンツである。

以上のように、記号化するための規則が言語と音楽では異なっていることが原因として考えられる。音の相対的な物理的特徴量の関係性で構造化している提案システムでは音楽的な構造が作られていないことは一つの理由になるであろう。それに対して、言語的な構造が作られなかった理由もここにあると考えられる。

### 3.2 多層的な構造

言語の特徴として、より多くの表現を可能にしている理由として二重文節という特徴がある。「文」という単位と「語」という単位が階層的に成り立っていることである。言語全体を考えると、実際はもっと多層的な階層になっている。多層的特徴は音楽にも見られる。言語と音楽の階層の例を表 1 に示す。言語においても音楽においても最低階層は音である。言語の場合は記号として機能するのは「語」以上の階層である。たとえば、「りんご」という語は指し示す対象が存在するが、「り」と「ん」と「ご」は「りんご」を構成する要素であるだけでなぜそのような並びなのかは意味がない。これは言語が「シンボ

\*3 <https://sites.google.com/site/hidefumiohmura/home/program/JSAI2016>

表 1: 言語と音楽の階層の例

階層	言語	音楽
高階層	全体	全体
↑	テキスト	楽節
	文	楽句
↓	語	音形
低階層	音	音

ル」であるからに他ならない。さらに「り」にはその音を構成する周波数や発音時間などの要素も存在するが、そこも意味はない。同様に、音楽の場合も音には意味はない。たとえば、440Hz「ラ」の音から意味を見いだすことはできない。しかしながら、もしその後には 660Hz「ミ」の音があれば、「ラ」はより「低い音」という意味が現れるかもしれないし、音形の階層では「上昇する音形」という意味を作るであろう。このように概観すると、音楽は最下層である物理的特徴量も相対的に構造を作るための要素であるのに対して、言語は意味を作る上では重要ではない要素である。

ではなぜ、エントロピーを増加させたときに言語的に聞こえたのであろうか。これは語の要素は音高が関わるからであると考えられる。例えば、母音を形成するフォルマントは大きく二つの音高に依存しているし、アクセントも音高に依存している。一方で、「語」や「文」はものによって長さはばらばらであり、音価には全く依存しない。そのため、言語的構造は保有してないにもかかわらずいくつかの音高が混じり合うことで言語的に聞こえるのであろう。

### 3.3 美的機能

音楽は具体的に対象を指すことができない。それにもかかわらず、記号的な振る舞いをするということについて検討する。言語は階層的な構造をとる。この構造は、下層ほど規則が厳しく、上層ほど規則が緩い。例えば、「りんご」は音の並びを変えると意味をなさないが、「りんごが木から落ちる」という文は、「落ちるりんごが木から」という文でも意味を伝達することはできる。この文はあえて位置を入れ替える倒置法と呼ばれる手法であるが、ほかにも主語を省いたり、自明なことはあえて書かないという手法もある。文学的な文章に多い。さらに詩的な文章になると、文字数を制限したり、韻を踏んだり、意味をなさない語が登場するだけでなく、全体としても意味を理解するのが困難な作品もある。これらは明らかに「期待感からの逸脱」や聴覚的な「期待感の実現」を狙ったものであり、例で上げたような手法を用いると美的な感覚が生じる。以上のように考えると、音楽はシンボルのような恣意的な記号としては働かないため、有契性のみで構造化されている。有契性のみで構造化は恣意的な規則をから具体的な対象物読み取ることはできず、推論から作られる規則に基づいて抽象的な対象物を読み取ることしかできない。そこには相対的な関係性しか存在しないため、記号の美的機能の特徴で成立している。

## 4. おわりに

本研究では、音の物理的特徴量の音高と音価の相対的な関係性をもちいて情報量操作を行い音楽生成システムを作った。システムから音楽を出力すると、エントロピーを増大させると構造を持たない言語的な響きの音列が得られた。その原因を記号的に考察をした結果、音楽が有契的で言語が無契的であるこ

とが影響していると考えられた。また、音楽は有契的であるがゆえ、美的機能を多く含むことが考えられた。提案しているシステムは、階層的な音楽構造の最低層だけであるが、今後は上の層まで定量的に操作できるように改良していく予定である。

## 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] 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 回ファジィシステムシンポジウム (FSS2015), TA4-1, (2015).
- [8] 大村英史, 柴山拓郎, 高橋達二, 澁谷智志, 太原育夫: 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, 情報処理学会 第 109 回音楽情報科学研究会 (SIGMUS), 2015-MUS-109, 8, pp1–4, (2015).
- [9] 池上嘉彦: 記号論への招待, 岩波新書 (1984).
- [10] J. J. Nattiez: *Musicologie générale et sémiologie*, Christian Bourgeois Editeur (1987). (足立美比古訳: 音楽記号学, 春秋社 (1996)).
- [11] 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).
- [12] D. E. Berlyne: *Aesthetics and psychobiology*, Appleton Century Crofts (1971).
- [13] C. E. Shannon: *The Mathematical Theory of Communication*, The University of Illinois Press (1949). (植松友彦訳: 通信の数学的理論, ちくま学芸文庫 (2009)).
- [14] P. Essens: *Hierarchical organization of temporal patterns*, *Perception & Psychophysics*, 40, pp. 69–73, (1986).
- [15] S. Mizen: *The Singing Neanderthals*, Harvard University Press (2006) (熊谷淳子訳: 歌うネアンデルタール, 早川書房 (2006)).