

周波数比の素数指数表現に基づく調性理解モデルとその応用可能性の検討

A Tonality Understanding Model Based on Prime Factor Representation of Frequency Ratio and Its Application Potentiality

白松 俊, 大園 忠親, 新谷 虎松

Shun SHIRAMATSU, Tadachika OZONO, Toramatsu SHINTANI

名古屋工業大学 大学院工学研究科 情報工学専攻

Graduate School of Engineering, Nagoya Institute of Technology

{siramatu, ozono, tora}@nitech.ac.jp

Abstract

我々が提案する調性理解モデル PFG Tonnetz は、音名や階名といったシンボルからではなく、協和音程の知覚に関わる物理量（音の周波数比）から直接導かれる。本稿ではその工学応用として、音楽知識を必要としない身体動作により入力された旋律線 (pitch contour) から、調性感を損なわない音高を出力する手法を示す。さらに、即興合奏のような多人数音楽インタラクションへの参加支援という応用を検討する。

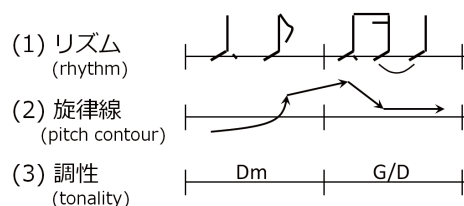


図 1: 旋律歌唱や旋律聴取における 3 つの処理側面



図 2: 想定する応用: 身体運動で入力された旋律線から、調性感を損なわない音高を出力

1 はじめに

調性感は、音楽を聴いたり演奏する上で重要な認知的要素の 1 つであり、主に協和音程／不協和音程の知覚や、更にその下位の物理現象（複数音間の周波数比や調波構造の重なり）によって説明される現象である。文献 [波多野 87] によると、旋律歌唱能力の発達過程や旋律聴取の認知的処理は、(1) リズム (rhythm) 構造の処理、(2) 旋律線 (pitch contour) 構造の処理、(3) 調性 (tonality) 構造の処理、という 3 つの処理側面から捉えられる (図 1)。本稿では特に (3) の調性構造に着目し、(音名や階名といったシンボルからではなく) 調の主音に対する周波数比から導かれる調性理解モデルについて述べる。

その前にまず、調性理解モデルが満たすべき要件を考える前提として、以下の 2 つの工学的応用を想定する。

1.1 多人数音楽パフォーマンスへの参加支援

近年、地域活性化のために住民参加型の音楽イベントが多く企画されている [堺都市 11]。地域活性化の効果を向上させるには、イベント参加者の裾野を広く、間口を広くする仕掛けが望まれる。例えば「音楽は好きだが、音楽知識／音楽経験は豊富でなく、そのイベントで演奏される楽曲にも詳しくない」という層が、情動の赴くまま自由かつ

能動的に演奏に参加できる仕掛けがあれば、地域活性化という趣旨に沿った新たな参加型音楽パフォーマンスや多人数音楽インタラクションの場をデザインできる可能性がある。そのためには、そのような人々の音楽知識や音楽経験の不足を補う技術が必要となる。

通常、そのような人々が演奏時に参加する手段は、手拍子や掛け声、あるいはリズムに乗って体を動かす程度である。これらは、上記 3 つの側面のうち (1) のリズム構造だけを使った参加形態であり、比較的自由度が高い。一方、持続的な調波音を用いた参加形態としては、例えばステージ上の演奏者が同じ旋律パターンを繰り返し提示した上で、聴衆にもその旋律の繰り返しを要求し、それに応えた聴衆がユニゾンで歌唱するといった形態が一般的であるが、これは自由度が低い。持続的な調波音を自由に発し、かつ演奏音との齟齬を生じさせないためには、(3) の調性構造の理解が不可欠である。しかし、これは音楽経験が豊富でない人々にとって敷居が高い。この解決のため、本稿では図 2 のように、(1) リズムと (2) 旋律線をユーザが自由に入力すると、(3) 調性の整合性が保証された旋律に変

換されるような機構を考える。

ここで(2)の旋律線 (pitch contour) とは、前ページ図1に示したように、音高の上昇、同音、下降の組み合わせによる運動のパターンである。なお本稿では、2度進行、3度進行のような旋法に関する要素は(3)の調性にも関係してくるため、(2)旋律線とは区別して考える。すなわち、運動のパターンのおおまかな概形として旋律線を捉える。この前提に立つと、リズムと旋律線の処理は、調性の処理よりも音楽知識や音楽経験の影響が少ない。リズムと旋律線は、音楽知識が乏しくても身体動作を介して生成しやすい構造であろう。特に旋律線と身体動作の親和性については文献[菅 08]でも指摘されている。そこで、自由な身体動作でリズムと旋律線を入力できるよう、加速度センサーやセンシング入力デバイスを用いて身体動作の上昇・下降を検出する。ここで入力された旋律線から音高への変換を担うのが調性理解モデルであるが、周囲の演奏音との調性感を損なわないために重要なのは調推定やコード名推定の精度向上ではなく、周囲の演奏音との協和/不協和を制御する性能であろう。つまり、音響信号から抽出された基本周波数からシンボル処理を介さずダイレクトに調性を処理するモデルが適していると考えられる。

1.2 即興合奏に参加可能な音楽ロボット

調性を損なわずに即興合奏に参加できるロボットを実現する上でも、上記3種の構造に対応する処理モジュールをどう構成すべきか考える必要がある。3つの構造のうちリズムと旋律線については、身体動作や情動の動きとの関係が指摘されており [Brown 00, Dogantan-Dack 13], そのような認知機構を考慮に入れた生成機構が必要となる。そこで生成されたリズムと旋律線を入力とし、調性感を損なわない音高を出力する調性理解モデルを考えると、やはり1.1節で示したような要件を満たす必要があると考えられる。

2 PFG Tonnetz: 素数指数表現に基づく調性理解モデル

ある調の主音の周波数 f_{tonic} と、協和音程の周波数 f の関係は、 $f = \frac{3}{2}f_{\text{tonic}}$ (完全5度) や $f = \frac{4}{3}f_{\text{tonic}}$ (完全4度) のような単純な比になることが知られている。この周波数比は、

$$f = \left(\prod_{p \text{ は } n \text{ 以下の素数}} p^{(z_p)} \right) \cdot f_{\text{tonic}} \quad (z_p \text{ は整数}) \quad (1)$$

のような素数の積で表すことができる。例えば素数の上限を $n = 5$ とおくと、完全4度の場合は

$$f = (2^2 \cdot 3^{-1} \cdot 5^0) \cdot f_{\text{tonic}} \quad (2)$$

となる。

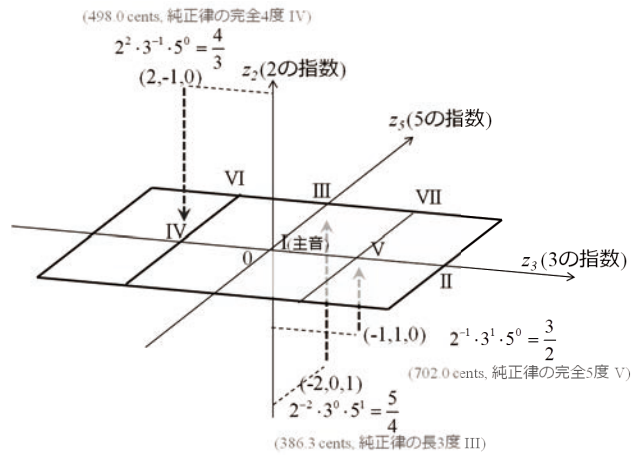


図3: オクターブ般化のための z_3z_5 平面への投影

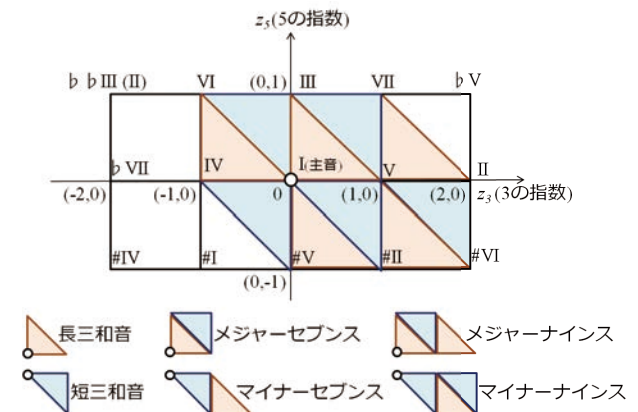


図4: 調性理解モデル PFG Tonnetz (5-limit)

このような主音 f_{tonic} と f の関係を、各素数 p の指数 z_p から成るベクトル (p_2, p_3, \dots, p_n) で表すことを考える。このベクトルは、整数論で用いられる素数指数表現 [Graham 93] を拡張し、指数 z_p が負の整数になるのを許容することで協和音程の単純な比を表現可能にしたものである。例えば上記の完全4度の場合、主音との関係を表す $n = 5$ の素数指数表現は $(2, -1, 0)$ となる。このベクトルを $z_2z_3z_5$ 空間にプロットすると、原点が主音 f_{tonic} を表し、原点近傍の整数格子点が主音との協和音程を表すことになる (図3)。

ここで、オクターブ隔たった音 ($f_1 = 2^2 f_2$ の関係にある f_1 と f_2) は音楽的に等価であることを考慮し、オクターブ関係にある音をまとめるオクターブ般化 (octave generalization) [Burns 87] を行う。具体的には、図3のように2の指数 z_2 を0にして z_3z_5 平面へ投影する。上記の完全4度の場合、 z_3z_5 平面上の $(-1, 0)$ へ投影される。

このように、主音から1オクターブ内におさまる原点近傍の整数格子点¹ (z_2, z_3, z_5) をオクターブ般化した (z_3, z_5) は、図4のように z_3z_5 平面に配置される。このとき、図中の赤い三角形、つまり $[(a, b), (a, b+1), (a+1, b)]$ の3

¹つまり $1 \leq 2^{z_2} \cdot 3^{z_3} \cdot 5^{z_5} < 2$ となる原点近傍の整数格子点

点を辿った三角形は明るい響きの長三和音の表現形式となり、青い三角形、つまり $[(a, b), (a + 1, b - 1), (a + 1, b)]$ の3点を辿った三角形は暗い響きの短三和音の表現形式となる。さらに、赤い三角形と青い三角形を交互に右方向 (z_3 軸の正の方向) へ積み重ねていくことで、セブンスコードやナインスコード、さらにその上のテンションコードの表現形式が得られる。これを定式化するために、根音 (a, b) に対して以下のような整数格子点列 $\text{code}(a, b, \delta, m)$ を考える。

$$\text{code}(a, b, \delta, m) = [(a, b) + \sum_{i=0}^k \delta(i)]_{k=0,1,\dots,m} \quad (3)$$

$$\delta_{\text{maj}}(i) = \begin{cases} (0, 1) & (i \text{ が奇数のとき}) \\ (1, -1) & (i \text{ が偶数のとき}) \end{cases} \quad (4)$$

$$\delta_{\text{min}}(i) = \begin{cases} (1, -1) & (i \text{ が奇数のとき}) \\ (0, 1) & (i \text{ が偶数のとき}) \end{cases} \quad (5)$$

このとき格子点列 $\text{code}(a, b, \delta_{\text{maj}}, m)$ は、 $m = 2$ のとき長三和音、 $m = 3$ のときメジャーセブンスコード、 $m = 4$ のときメジャーナインスコードの表現形式となる。これらメジャーコードの構成音は、根音 (a, b) の上側 $b \leq z_5 \leq b + 1$ に分布する。一方、格子点列 $\text{code}(a, b, \delta_{\text{min}}, m)$ は、 $m = 2$ のとき短三和音、 $m = 3$ のときマイナーセブンスコード、 $m = 4$ のときマイナーナインスコードの表現形式となる。これらマイナーコードの構成音は、根音 (a, b) の下側 $b - 1 \leq z_5 \leq b$ に分布する。根音 (a, b) を調の主音 $(0, 0)$ で置き換えて考えると、明るい長音階の構成音 (I, II, III, IV, V, VI, VII) が原点の上側 $0 \leq z_5 \leq 1$ に分布し、暗い短音階の構成音 (I, II, #II, IV, V, #V, #VI) が原点の下側 $-1 \leq z_5 \leq 0$ に分布するのが見て取れる。

ここまで、調性の表現形式を導く過程で用いた前提は、以下のような周波数比に基づく2つの認知的原理のみである。

1. 調性音程の周波数比は単純な整数比で表せる。よって、周波数比を素数指数表現 (z_2, z_3, z_5) で表せる (ただし $z_2, z_3, z_5 < 0$ も整数なら許容)。
2. 周波数比がちょうど 2^z (z は整数) である音程はオクターブ等価である。よって、素数指数表現 (z_2, z_3, z_5) の2の指数 z_2 を捨象し $z_3 z_5$ 平面上に投影することで、オクターブ般化できる。

つまり、音名、階名、コード名のようなシンボルを前提としては用いていない。まとめると、導出の過程で現れた以下の表現形式は、上記の認知的原理および導出されたモデルの観察から自然に導かれたものである。

- 整数格子点 (z_3, z_5) : 音階の構成音、階名
- 三角形 $[(a, b), (a, b + 1), (a + 1, b)]:$
整数格子点 (a, b) を根音とする長三和音

- 三角形 $[(a, b), (a + 1, b - 1), (a + 1, b)]:$
 (a, b) を根音とする短三和音
- 整数格子点列 $\text{code}(a, b, \delta_{\text{maj}}, m)$:
 (a, b) を根音とする長和音
- 整数格子点列 $\text{code}(a, b, \delta_{\text{min}}, m)$:
 (a, b) を根音とする短和音
- $a \leq z_5 \leq b + 1$: (a, b) を根音とする長和音の分布領域
- $a - 1 \leq z_5 \leq b$: (a, b) を根音とする短和音の分布領域
- $0 \leq z_5 \leq 1$: 原点を主音とする長音階の分布領域
- $-1 \leq z_5 \leq 0$: 原点を主音とする短音階の分布領域

この調性理解モデルは、著名な数学者 Leonhard Euler が 1739 年に考案し (図 5)、音楽学者 Hugo Riemann が 1880 年に図 6 のように発展させた Tonnetz [Behringer 10] というモデルに似ている。Tonnetz は 1980 年代以降、数学的に定式化された新リーマン理論 (Neo-Riemannian theory) へと発展し、様々な拡張が行われた [Hewlett 07, Tymoczko 12] が、基本的には図 5, 6 のように音名同士を繋いだモデルであり、調の主音は表現できていない。また、物理現象や認知機構に根ざした説明になっておらず、ヒューリスティックな前提が多い。

一方、提案する図 4 のモデルは、まず原点に調の主音という役割があり、そこからの相対的な位置関係によって調性が決まるという性質があるため、主音を変えても同じ座標系で表現でき、式 (3), (4), (5) のような定式化も容易となる。また、これまでは素数の上限 $n = 5$ としてきたが、 $n = 7$ にして $z_3 z_5 z_7$ 空間へ拡張すると 7-limit 純正律 [Partch 74] を表現でき、 $n = 11$ にすると 11-limit 純正

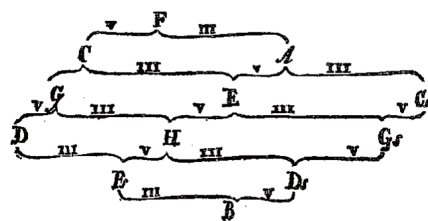


図 5: オイラーの Tonnetz

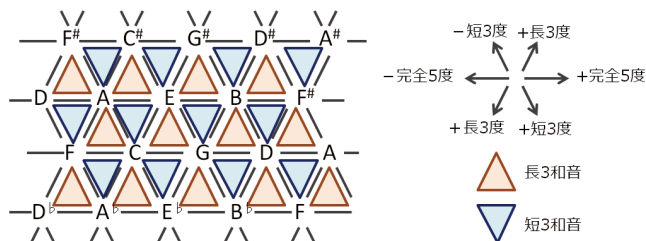


図 6: リーマンの Tonnetz

律を表現できる. さらに, 周波数比という物理量と協和音程に関するシンプルな原理から導かれたモデルであり, 従来の Tonnetz よりも前提が少ない.

これらのことから, 提案するモデルは従来の Tonnetz の一般化になっていると考え, 我々は過去の研究 [白松 12] で PFG Tonnetz (Prime-Factor-based Generalized Tonnetz) と名付けた. また, n -limit 純正律のアナロジーにより, $z_2 z_3 \cdots z_n$ 空間のモデルを n -limit PFG Tonnetz と呼ぶ. 特に $n = 5$ の 5-limit PFG Tonnetz (図 4) は, 通常の純正律 (5-limit 純正律) に対応しており, $z_3 z_5$ の 2次元平面で可視化でき, 従来の Tonnetz とのトポロジー的類似性が理解しやすい等の利点があり, 扱いやすい. よって, 以後 n に関して指定のない PFG Tonnetz は, 全て 5-limit PFG Tonnetz のことを指すものとする. また, 以後, $z_3 z_5$ 平面の整数格子点とその集合を $g = (z_3, z_5) \in G$ で表す.

3 PFG Tonnetz を用いた旋律線からの音高決定

1節で示したように, 身体操作により入力されたリズムと旋律線から, 周囲の演奏音の調性を損なわない音高を出力するために, 調性理解モデル PFG Tonnetz を用いる. そのためには, 周囲の演奏音の音響信号から調性の制約を決定する手法と, その制約と旋律線に合わせて音高を決定する手法が必要となる.

3.1 音響信号からの調性制約の決定

まず前処理として, 観測された周囲の多重奏音響信号に含まれる複数の基本周波数を推定する. この前処理を行うツールとして, Durrieu らによるオープンソースの Vamp プラグインである IMMF0salience² [Durrieu 11, Salamon 14] を用いる. IMMF0salience は, IMM (Instantaneous Mixture Model) と呼ばれる手法により, 多重奏音楽音響信号中の複数の基本周波数 F0 を推定し, 中間的な解析結果を可視化・出力する. 具体的には, 周波数ビン毎に基本周波数 F0 の顕著性スコア (F0 salience representation; F0 顕著性表現), 言い換えると F0 存在可能性の推定結果を出力する. これは北原らの Inrogram [Kitahara 07] による可視化と似た中間的表現であるが, IMMF0salience の出力は各楽器音に分離する前の F0 推定結果である. 通常, IMMF0salience の実行は Sonic Visualizer の GUI 上で行われるが, ここでは Sonic Annotator³ を用いて以下のように CUI 上で実行し, 前処理をバッチ処理化できる.

²<https://github.com/wslight/IMMF0salience>,
<http://www.durrieu.ch/research/jstsp2010.html>
³<http://www.vamp-plugins.org/sonic-annotator/>

```
$ sonic-annotator -d vamp:f0salience:f0salience
WAV_FILE -w default > OUTPUT_FILE
```

デフォルト設定では 100Hz から 800Hz までの帯域で F0 推定が行われ, その結果が図 7 のような形式で出力される. feature タグがフレーム毎の推定結果を表し, timestamp タグが当該フレームの時刻を, values タグが各周波数ビンの F0 顕著性スコアを表す.

```
:
:
<feature>
<name>Salience of F0s.</name>
<timestamp> 12.445895691R</timestamp>
<values>100 Hz:2629.86 100 Hz:192.319 101 Hz:40.9141
102 Hz:7.2439 102 Hz:2.62986 103 Hz:2.62986 104
Hz:2.62986 105 Hz:2.62986 ... 777 Hz:137.789 782
Hz:64.2944 788 Hz:164.495 0 Hz:25.5596</values>
</feature>
<feature>
<name>Salience of F0s.</name>
<timestamp> 12.469115646R</timestamp>
<values>100 Hz:26568.5 100 Hz:4487.88 101 Hz:116.993
102 Hz:116.993 102 Hz:116.993 103 Hz:116.993 104
Hz:116.993 ... 777 Hz:116.993 782 Hz:116.993 788
Hz:4033.32 0 Hz:116.993</values>
</feature>
:
```

図 7: IMMF0salience が出力する F0 推定結果

このように得られた, フレーム t における周波数ビン f の顕著性スコアを $s(t, f)$ とする. この $s(t, f)$ から, 以下のようなフレーム t における調性制約 tnl_t を推定する.

$$tnl_t = \langle f_{\text{tonic}}, h_t(f_{\text{tonic}}) \rangle \quad (6)$$

$$f_{\text{tonic}} = \arg \min_{f_{\text{cand}}} \sum_{g \in G, t=t_0, \dots, t} w(g) \cdot h_t(f_{\text{cand}})(g) \quad (7)$$

$$h_t(f_{\text{cand}})(g) = \sum_{f \in F(f_{\text{cand}}, g)} s(t, f) \quad (8)$$

$$F(f_{\text{cand}}, g) = \{(2^{z_2} \cdot 3^{z_3} \cdot 5^{z_5}) \cdot f_{\text{cand}}\}_{z_2 \in \mathbb{Z}} \quad (9)$$

$$w(g) = \min_{k_1, k_2, k_3 \in \mathbb{Z}} (|k_1| + |k_2| + |k_3|) \quad (10)$$

$$g = (z_3, z_5) = k_1(1, 0) + k_2(0, 1) + k_3(1, -1) \quad (11)$$

ただし, f_{tonic} は調の主音の周波数, $h_t(f_{\text{tonic}})$ はフレーム t の F0 顕著性スコアの G 上の分布, f_{cand} は主音の候補の周波数, $w(g)$ は格子点 g の重み, $F(g)$ は格子点 g にオクターブ般化される周波数の集合, t_0 は推定に用いるフレーム列の開始番号, \mathbb{Z} は整数の集合である. 格子点 g の重み $w(g)$ は, 完全五度に対応するベクトル $(1, 0)$, 長三度に対応するベクトル $(1, 0)$, 短三度に対応するベクトル $(1, -1)$ をネットワークのリンクと見なした時に, 原点から g を迎える最短ステップ数となる.

3.2 調性を損なわない音高決定

身体動作により入力された時刻 t の旋律線の高さを $x(t)$ とする. これを, 調性制約 tnl_t を考慮した周波数 $f(t)$ に変換する.

まず, $tnl_t = \langle f_{\text{tonic}}, h_t(f_{\text{tonic}}) \rangle$ のうち主音の周波数 f_{tonic} のみを考慮した $f'(t)$ への変換を行う.

$$f'(t) = f_{\text{tonic}} \exp(\alpha(x(t) - x_{\text{tonic}})) \quad (12)$$

ただし x_{tonic} は主音 f_{tonic} に対応付けた基準位置, α は位置変化と周波数変化の比率を調整する係数である. この $f'(t)$ を, $h_t(f_{\text{tonic}})$ を考慮して $f(t)$ に音高補正するには, 例えば以下のような手法が考えられる.

$$f(t) = \arg \min_{g \in \text{Scale}(t)} \left(|f(g) - f'(t)| + \beta \sum_{g' \in G} w(g - g') h_t(f_{\text{tonic}})(g') \right) \quad (13)$$

$$\text{Scale}(t) = \{g \mid \sum_{t=t_0, \dots, t} h_t(f_{\text{tonic}})(g) \geq \theta\} \quad (14)$$

$$f(g) = (2^{z_2} \cdot 3^{z_3} \cdot 5^{z_5}) \cdot f_{\text{tonic}}, g = (z_3, z_5), z_2 \in \mathbb{Z} \quad (15)$$

ここで, $\text{Scale}(t)$ は取りうる格子点の集合, θ は時刻 t までの F_0 顕現性スコア総和の閾値, β は調性をどれだけ重視するかという重みを表すパラメータである. ここでも式 (10) の重みを用いているが, 不協和度 [藤澤 06] に基づく重みに置き換えることも考えられる.

なお本手法は未検証であり, 今後, 実験による性能評価が必要である.

4 本研究の位置付けと関連研究

4.1 調性の表現手法

素数指数表現に基づく音の表現方法は, 従来研究でも提案されている [Monzo 99, Keislar 87]. しかし, これらの従来研究は Tonnetz との関連について触れておらず, 素数指数表現に基づいて Tonnetz を一般化したのは本研究の独自の着眼点である.

調性の表現形式に関しては, 五度圏 (Cycle of Fifth) など古くから多くのモデルや理論が存在し, 五度圏を用いた調推定手法 [Inoshita 09] も提案されている. また Tonnetz に関しては, 新リーマン理論への発展に伴いトラス面上の Tonnetz 表現形式 [Hewlett 07] 等へ拡張され, さらに isomorphic keyboard [Milne 07] 等の楽器インタフェースにも応用されている. しかし, 本稿で提案したような, 観測音響信号から推定される周波数比を直接用いて調性制約を得るという手法へは応用されていない.



図 8: PFG Tonnetz に基づくスマートフォンアプリ Tonal-ityTouch

4.2 スマートフォンアプリへの応用

図 8 は, 我々が過去の研究で開発した Tonal-ityTouch というスマートフォンアプリである [白松 12]⁴. これは, PFG Tonnetz を用い, スマートフォン上のタッチ動作やスワイプ動作を調性感ある音高へ補正して発音するという機能を持つ. マルチタッチによる協和音生成も可能である. これにより, 音楽知識や音楽経験に乏しいユーザでも調性感ある演奏が可能であるほか, PFG Tonnetz を用いて新たな音階を試行的に生成し, ユーザが気に入った新音階を用いて音高補正することができる. しかし, 周囲の演奏音を考慮に入れていないため, 1 節で述べたような即興合奏への参加支援のためには不十分である.

なお, スマートフォンを用いて参加型音楽パフォーマンスを支援するシステムの研究事例としては, massMobile [Weitzner 13] や SWARMED [Hindle 13] がある. これらは, 本研究で扱っているような無作為な身体動作によって調性感を伴う即興合奏参加が可能になる機構は有していない. この点が本研究独自の着眼点となっている.

4.3 地域振興音楽イベントへの応用

我々はこれまで, 地域住民向けのワークショップからの意見集約や, 公共圏での協働を支援する技術を研究してきた [Shiramatsu 13]. 地域コミュニティでの協働を円滑にするには, 課題解決のための議論だけが重要なわけではない. 緊張関係をほぐして場の空気を和ませるために, アイスブレイクと呼ばれる仕掛けが用意される場合がある. 音楽は, 身体動作の共有を通じてポジティブな情動を喚起する [寺澤 13] という社会的機能を持つので, 1.1 節で触れたような音楽イベントが地域コミュニティの円滑化に寄与し, 広い意味でのアイスブレイクの役割を果たす可能性は大いにある. 調性理解モデル PFG Tonnetz を用いる

⁴<https://play.google.com/store/apps/details?id=org.toralab.music.beta>

ことで自由な身体動作が許容されるが、この機能がコミュニケーション内でのアイスブレイク効果を向上させるかどうか、今後、改めて実証していく必要がある。

5 おわりに

本稿では、協和音程が単純な周波数比で表せるというシンプルな認知的原理から導かれる調性理解モデル PFG Tonnetz について述べた。これは、従来の Tonnetz の一般化になっており、また多重奏音楽音響信号からの基本周波数推定結果との親和性も高いモデルであると考えられる。さらにこの提案モデルを即興合奏への参加支援へ応用するために、身体動作によって入力された旋律線から、周囲の演奏音との調性感を損なわない音高を決定する手法を検討した。今後は提案手法の性能評価実験を行い、加速度センサーやセンシング入力デバイスを用いた旋律線入力機構と統合したシステムを開発する予定である。

参考文献

- [Behringer 10] Behringer, R. and Elliot, J.: *Linking Physical Space with the Riemann Tonnetz for Exploration of Western Tonality*, chapter 6, pp. 131–143, Nova Science Publishers (2010)
- [Brown 00] Brown, S.: The “musilanguage” model of music evolution, in *The origins of music*, pp. 271–300, MIT Press (2000)
- [Burns 87] Burns, E. M. and Ward, W. D.: 音程, 音階, 調律, 第 8 章, pp. 301–334, 西村書店 (1987)
- [Dogantan-Dack 13] Dogantan-Dack, M.: Tonality: the shape of affect, *Empirical Musicology Review*, Vol. 8, No. 3-4, pp. 208–218 (2013)
- [Durrieu 11] Durrieu, J., David, B., and Richard, G.: A musically motivated mid-level representation for pitch estimation and musical audio source separation, *IEEE Journal of Selected Topics in Signal Processing*, Vol. 5, No. 6, pp. 1180–1191 (2011)
- [Graham 93] Graham, R. L., Knuth, D. E., and Patashnik, O.: *コンピュータの数学* (1993)
- [Hewlett 07] Hewlett, W., Selfridge-Field, E., and Correia, E.: *Tonal Theory for the Digital Age*, Vol. 15 of *Computing in Musicology*, Center for Computer Assisted Research in the Humanities, Stanford University (2007)
- [Hindle 13] Hindle, A.: SWARMED: Captive Portals, Mobile Devices, and Audience Participation in Multi-User Music Performance, in *Proceedings of the 13th International Conference on New Interfaces for Musical Expression*, pp. 174–179 (2013)
- [Inoshita 09] Inoshita, T. and Katto, J.: Key Estimation Using Circle of Fifths, in *Advances in Multimedia Modeling*, pp. 287–297, Springer (2009)
- [Keislar 87] Keislar, D.: History and principles of microtonal keyboards, *Computer Music Journal*, pp. 18–28 (1987)
- [Kitahara 07] Kitahara, T., Goto, M., Komatani, K., Ogata, T., and Okuno, H. G.: Instrogram: probabilistic representation of instrument existence for polyphonic music, *IPJS Journal*, Vol. 48, No. 1, pp. 214–226 (2007)
- [Milne 07] Milne, A., Sethares, W., and Plamondon, J.: Isomorphic controllers and dynamic tuning: Invariant fingering over a tuning continuum, *Computer Music Journal*, Vol. 31, No. 4, pp. 15–32 (2007)
- [Monzo 99] Monzo, J.: *JustMusic: A New Harmony Representing Pitch as Prime Series*, J. Monzo, 4th edition (1999)
- [Partch 74] Partch, H.: *Genesis of a music: an account of a creative work, its roots and its fulfillments*, Da Capo Press (1974)
- [Salamon 14] Salamon, J., Gómez, E., Ellis, D. P., and Rechar, G.: Melody Extraction from Polyphonic Music Signals: Approaches, applications, and challenges, *IEEE Signal Processing Magazine*, Vol. 31, No. 2, pp. 118–134 (2014)
- [Shiramatsu 13] Shiramatsu, S., Ozono, T., and Shintani, T.: Approaches to Assessing Public Concerns: Building Linked Data for Public Goals and Criteria Extracted from Textual Content, in *Electronic Participation. Proceedings of the 5th IFIP WG 8.5 International Conference, ePart 2013*, Vol. 8075 of *Lecture Notes in Computer Science*, pp. 109–121, Springer (2013)
- [Tymoczko 12] Tymoczko, D.: The Generalized Tonnetz, *Journal of Music Theory*, Vol. 56, No. 1, pp. 1–52 (2012)
- [Weitzner 13] Weitzner, N., Freeman, J., Chen, Y.-L., and Garrett, S.: massMobile: towards a flexible framework for large-scale participatory collaborations in live performances, *Organised Sound*, Vol. 18, No. 01, pp. 30–42 (2013)
- [堺都市 11] 堺都市 政策研究所: 市民主体の地域活性化 (音楽イベントを通じたまちおこし) に関する調査研究業務報告書, <http://www.sakaiupi.or.jp/30.products/31.resarch/H22/H22.music.pdf> (2011)
- [寺澤 13] 寺澤 洋子, 星 柴 玲子, 柴山 拓郎, 大村 英史, 古川 聖, 牧野 昭二, 岡ノ谷 一夫: 身体機能の統合による音楽情動コミュニケーションモデル, *認知科学*, Vol. 20, No. 1, pp. 112–129 (2013)
- [菅 08] 菅 道子: 身体表現を取り入れた参加型音楽コンサートの可能性: カノンの理解を目指した「追いかけてこしよう」の事例から, *和歌山大学教育学部教育実践総合センター紀要*, Vol. 18, pp. 121–129 (2008)
- [藤澤 06] 藤澤隆史, ノーマン D. クック: 和音性の計算法と曲線の描き方: 不協和度・緊張度・モダリティ, *情報研究: 関西大学総合情報学部紀要*, Vol. 25, pp. 35–51 (2006)
- [波多野 87] 波多野 諠余夫: 音楽と認知, *認知科学選書*, 第 12 巻, 東京大学出版会 (1987)
- [白松 12] 白松 俊, 大園 忠親, 新谷 虎松: TonalityTouch: 素数指数表現に基づく一般化 Tonnetz を用いたスマートフォン楽器, *情報処理学会研究報告 音楽情報科学*, Vol. 2012-MUS-96, No. 20 (2012)