Vol.2019-MUS-122 No.11
Vol.2019-EC-51 No.11
2019/2/22

ベイズ推論により拡張されたスパイラルアレイモデル B-SAM による調性同定

名畑 皓正[1,a)]　竹川 佳成[1,b)]　平田 圭二[1,c)]

概要：調性音楽において、人間が認知する音どうしの関係の距離感を二次元上に表現した Tonnetz というダイアグラムがある。この Tonnetz を三次元にまとめ、螺旋状に表したものがスパイラルアレイモデル (E. Chew 2000)(以下 SAM) である。SAM では、ピッチ・コード・キーという次元の異なる概念を同一の三次元座標空間内で座標値として表すことができる。SAM では、Center of Effect Generator アルゴリズムを用いることで楽曲の調性同定を行うことができるが、距離の計算を用いるため結果は離散的である。本研究では、最小二乗法を用いた最適化とベイズ推論により、SAM のキーの点群を確率分布として表した B-SAM を提案する。B-SAM による確率分布を用いた調性同定と、従来の SAM による調性同定の比較評価を行った。

1. はじめに

現在私たちが普段耳にする楽曲は主に調性音楽であり，調（キー）を持っている．調性（トーナリティ）とは，広義においては，支配的な中心音を有する音体系を，狭義においては，長短調いずれかの主和音をもつ和声的な音体系を指す．現在のポピュラー音楽にまで用いられているのは後者であるとされている [3]．本稿ではその和声的な音体系を調性とする．そして，長音階あるいは短音階が特定の音を主音とするとき，その主音の音名をつけて表されるものを調と呼ぶ．ハ長調，ニ短調，変ホ長調など，長調・短調が各 12 個ずつある [8]．また，曲の特徴を捉え調性同定の研究には，音楽音響信号に対するものと楽譜に対するものがある．本研究は楽譜に対するものであり，記号処理の分野である．

楽曲の構造に対する人の認識をモデル化した研究に，Elaine Chew によるスパイラルアレイモデル [4](以下 SAM) がある．このモデルは，音楽の認知的な距離をそのまま配置したダイアグラムである Tonnetz[2] を円筒状に包み込んだものである．Tonnetz は Tonal pitch space[10] などと同じく，音同士の関係性を距離として表現する pitch space の考えが基になっている．SAM では図 2 のように，音高（ピッチ），和音（コード），調（キー）という次元の異なる三つの音楽的実体が一つの三次元座標空間内に示されるため，それらの関係性を幾何学演算で比較することができる．それぞれの音楽的実体は座標空間内に点として表現され，その点群は五度圏の順で螺旋状に配置される．また，SAM 内で幾何学的な座標演算である Center of Effect Generator アルゴリズム（CEG）を用いることにより，楽曲の調性同定を行うことが可能である．CEG は，図 2 のように，ある楽曲に含まれるピッチイベントの音高と音価の情報を SAM 内側の一点（Center of Effect，C.E.）に集約し，C.E. と最も距離の近い点の調がその楽曲の調性だと同定する方法である．

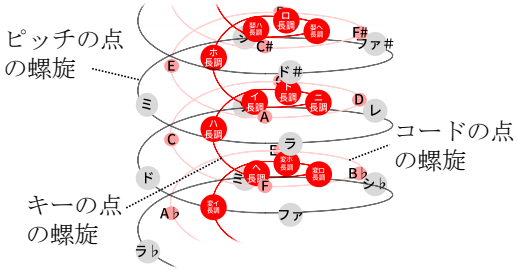

図 1　スパイラルアレイモデルの
一部（長調点のみ表示）

SAM は，発案者である E. Chew による調性同定や転調境界判定，自動ピッチスペリングなどの研究 [5] から，SAM のしくみを応用した即興演奏支援システムにおける調性同定手法 [15]，和音におけるテンションをテンションリボンとして定量化した研究 [7] など，様々な応用研究がなされ

[1] 公立はこだて未来大学大学院
Graduate School of Future University Hakodate
[a)] g2118029@fun.ac.jp
[b)] yoshi@fun.ac.jp
[c)] hirata@fun.ac.jp

ている．また，SAM は調性の動的な振る舞いを人間の直感に近い状態をそのまま視覚化できるモデルであるため，SAM 内を C.E. が動く様子やリアルタイムに調性同定を行う様子などを 3D モデルで視覚化するアプリケーションである MuSA_RT が存在する [1]．

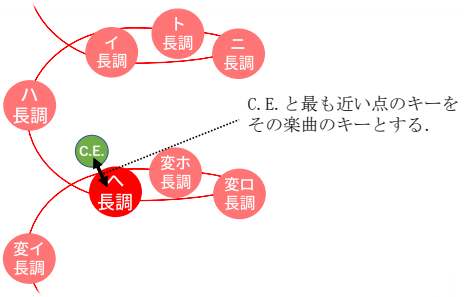

図 2 C.E. による調性同定

本研究では，この SAM を確率モデル化することを目指す．従来の SAM での調性同定では入力されたピッチクラスのみの情報から一意に同定するため頻度主義だと言えるが，それを確率化することは，ベイズ主義へと拡張する試みと言える．また，全ての調の確率が同じ形に分布しているのではなく，ハ長調の確率分布は狭く強く分布していたり，ト長調の確率分布は広く薄く分布しているといった可能性に考慮することができる．楽曲スタイルや作曲家など，様々な音楽的特徴の学習への応用が可能である．また，SAM をベースにしているため，人間の直感に近い状態で確率モデルを視覚化したアプリケーションへの応用も考えられる．

2. 関連研究

代表的な調性同定方法として，Krumhansl と Schmuckler による Probe Tone Profile Method（PTPM）[9] と Longuet-Higgins と Steedman の Shape Matching Algorithm（SMA）[12] がある．

PTPM は，対象楽曲に含まれる 12 個のピッチクラスのそれぞれの累積持続時間が入力ベクトルに記録され、そのベクトルと Probe Tone Profile との間の相関関係を計算し，相関の最も高い調を対象楽曲の調だと同定する方法である．Probe Tone Profile とは，12 個の各ピッチクラスが，ある調の文脈にどの程度うまく収まるかについて，聴取者の判断に基づいて作成されたものであり，図 3 のように表される．図中の average rating は，C major においてそのピッチクラスがどの程度うまく収まるかを示しており，0〜7 で聴取者が判断したものの平均である．

SMA は，図 4 のような Harmonic Network 上において，ある調に属するピッチの配置の形状を用いて調性同定を行う．ピッチイベントが読み込まれる度，そのピッチをカバーしていない形状が，それらが表す調と共に削除されていく．このプロセスが最後の一つの調になるまで繰り返され，残った最後の調が対象楽曲の調だと同定する方法である．

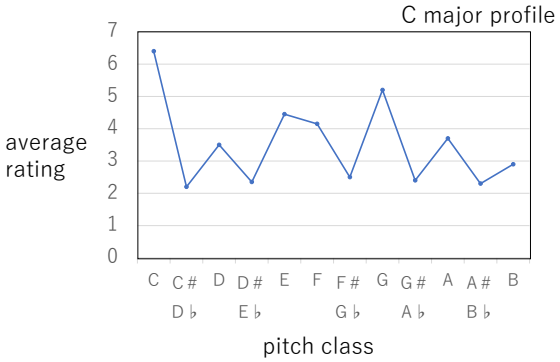

図 3 C major の Probe Tone Profile

E. Chew は CEG, PTPM, SMA の 3 つをバッハの Well-tempered Clavier を用いて比較テストした [5]．評価基準は，調を正しく同定するために必要なピッチイベントの数と設定された．それぞれ要した平均ピッチ数は，CEG が 3.75，PTPM が 5.25，SMA が 8.71 であった．CEG は従来の調性同定方法よりも少ない音数で正確に調を認識することが可能である．

調性同定の関連研究では，Specmurt 分析と Chroma Vector を用いた HMM(Hidden Markov model，隠れマルコフモデル) による音楽音響信号の調認識というものがある [14]．クロマベクトルとは，オクターブの違う同じ音階の成分を全て重ね合わせ，1 オクターブ内の半音階 12 音の成分に縮約したものである．長調と短調とでそれぞれ 12 個ずつ合計で 24 個ある調を HMM とする確率モデルを用いて，対象楽曲から得られたクロマベクトルの時系列に対し事後確率を最大にする HMM の探索を行い，最も尤度の大きかった HMM の調を対象楽曲の調として推定していた．性能評価実験では調認識率は 83.3%-93.8%という結果が得られている．

この方法では高い精度で調が推定されているが，調性の動的な振る舞いを人間の直感に近い状態で理解することには適していない．本稿執筆段階では未実装であるが，確率化した SAM と C.E. によって調性の動きが楽曲の流れと共に動いていく様子を可視化し，直感的な理解が容易なシステムの構築を目指している．

3. ベイズ推論を用いて拡張したスパイラルアレイモデル B-SAM

本研究では，ベイズ推論を用いて拡張した SAM である B-SAM を提案する．

3.1 B-SAM の概要

B-SAM は SAM を確率化したものである．頻度主義で

Vol.2019-MUS-122 No.11
Vol.2019-EC-51 No.11
2019/2/22

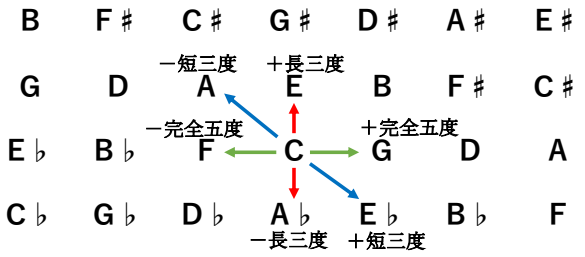

図 4　Harmonic Network

あった従来の SAM を確率化し，ベイズ主義へと拡張する試みである．あるデータ群から得られた過去の結果だけに基づいて予測するのが頻度主義であるが，ベイズ主義ではそこに信念が加わる．予測するモデルが持つ信念の確率分布を，起きた現象を学習することによって探り当てていくという主義である．

また，SAM を確率化することによって，全ての調の確率が同じ形に分布しているのではなく，ある調の確率分布は狭く強く分布していながら，ある他の調の確率分布は広く薄く分布しているといった可能性に考慮することができる．そのため，楽曲スタイルなどの特徴を取り入れながら学習することが可能である．本手法では 3 次元ガウス分布を使ってパラメータの学習を行う．ある楽曲のピッチイベントデータと正解の調の情報を用いて，その楽曲に最適化された B-SAM を導出する．

3.2　SAM 部分の実装

まずは，従来研究であり，本研究のベースとなる SAM について述べる．本研究では，SAM 部分の実装には以下に示す参考文献 [5] 中の数式を用いた．なお，本稿で述べる座標値とは，三次元空間における座標値であり，3 つの実数値の組である．

・各音高を表す座標値 $\mathbf{P}(\mathrm{k})$ の算出

この時，音高の螺旋の半径は 1，ある音高から次の音高への高さは $\sqrt{2/15}$ である．

$$\mathbf{P}(\mathrm{k}) \stackrel{\mathrm{def}}{=} \begin{bmatrix} x_{\mathrm{k}} \\ y_{\mathrm{k}} \\ z_{\mathrm{k}} \end{bmatrix} = \begin{bmatrix} \sin\frac{\mathrm{k}\pi}{2} \\ \cos\frac{\mathrm{k}\pi}{2} \\ \mathrm{k}\cdot\sqrt{2/15} \end{bmatrix}. \quad (1)$$

・各和音を表す座標値 $\mathbf{C}$ の算出

各和音を表す座標値は，各音高を表す座標値 $\mathbf{P}(\mathrm{k})$ を用いて導出される．$\mathbf{C}_{\mathrm{M}}$ は長三和音を表し，$\mathbf{C}_{\mathrm{m}}$ は短三和音を表す．$w1, w2, w3$ は三和音の点を導出する際のトニック音，ドミナント音，サブドミナント音の重みであり，同時に調の点を導出する際のトニックコード，ドミナントコード，サブドミナントコードの点それぞれにかかる重みを表す．本研究では従来の SAM として E. Chew の設定した値である $w1 = 0.536, w2 = 0.274, w3 = 0.19$ で設定した．

$$\mathbf{C}_{\mathrm{M}}(\mathrm{k}) = w1\cdot\mathbf{P}(\mathrm{k}) + w2\cdot\mathbf{P}(\mathrm{k}+1) + w3\cdot\mathbf{P}(\mathrm{k}+4) \quad (2)$$

$$\mathbf{C}_{\mathrm{m}}(\mathrm{k}) = w1\cdot\mathbf{P}(\mathrm{k}) + w2\cdot\mathbf{P}(\mathrm{k}+1) + w3\cdot\mathbf{P}(\mathrm{k}-3) \quad (3)$$

$$\mathbf{C} = \{\mathbf{C}_{\mathrm{M}}(\mathrm{k}) \text{ for all k}\} \cup \{\mathbf{C}_{\mathrm{m}}(\mathrm{k}) \text{ for all k}\} \quad (4)$$

・各調性を表す座標値 $\mathbf{T}$ の算出

各調性を表す座標値は，各和音を表す座標値 $\mathbf{C}_{\mathrm{M}}$，$\mathbf{C}_{\mathrm{m}}$ を用いて導出される．$\mathbf{T}_{\mathrm{M}}$ は長調を表し，$\mathbf{T}_{\mathrm{m}}$ は短調を表す．

$$\mathbf{T}_{\mathrm{M}}(\mathrm{k}) = w1\cdot\mathbf{C}_{\mathrm{M}}(\mathrm{k}) + w2\cdot\mathbf{C}_{\mathrm{M}}(\mathrm{k}+1) + w3\cdot\mathbf{C}_{\mathrm{M}}(\mathrm{k}-1) \quad (5)$$

$$\mathbf{T}_{\mathrm{m}}(\mathrm{k}) = \quad w1\cdot\mathbf{C}_{\mathrm{m}}(\mathrm{k}) + w2\cdot\mathbf{C}_{\mathrm{m}}(\mathrm{k}+1) + w3\cdot\mathbf{C}_{\mathrm{m}}(\mathrm{k}-1) \quad (6)$$

$$\mathbf{T} = \{\mathbf{T}_{\mathrm{M}}(\mathrm{k}) \text{ for all k}\} \cup \{\mathbf{T}_{\mathrm{m}}(\mathrm{k}) \text{ for all k}\} \quad (7)$$

・C.E. を表す座標値 CE の算出

P_i は各音高を示す座標値，d_i は各音高の音価，N は認識対象区間内のピッチイベント数を示す．

$$\mathrm{CE} = \sum_{i=1}^{N} \frac{d_i}{D}\cdot P_i \quad where D = \sum_{i=1}^{N} d_i \quad (8)$$

ここまでが，従来の SAM の設計である．ここから平均・精度パラメータの導出を行い確率化し，B-SAM となる．

3.3　平均パラメータの導出

平均パラメータの導出について述べる．この処理は，入力する楽曲データにおいて，調の点がどこに位置すればその楽曲データでの調性同定で最適な結果が得られるかを探るものである．

本手法における平均パラメータ μ は，対象楽曲に最適化された 24 個ある調の点の座標値である．導出は，図 5 のように行なっている．

SAM 内で調の点が定義されるときには (5)(6) の式を用いる．この式に含まれる $w1, w2, w3$ という重みを楽曲ごとに定義することで，その楽曲の平均パラメータとして用いることができる．楽曲ごとの重みの導出には，最小二乗法を用いる．楽曲を 8 小節ごとに区切り，正しい調の点と C.E. の距離を誤差とし，学習率は 0.1 で重みの値を更新し

Vol.2019-MUS-122 No.11
Vol.2019-EC-51 No.11
2019/2/22

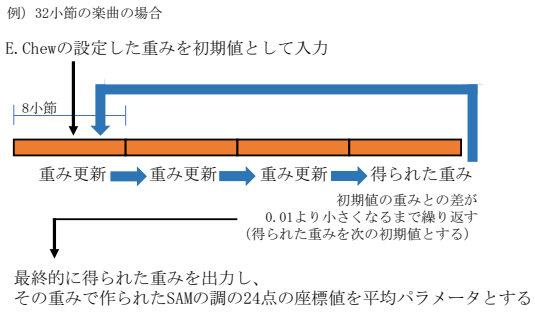

図 5 平均パラメータの導出

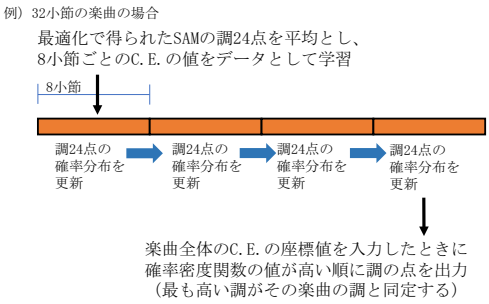

図 6 精度パラメータの学習・調性同定

た．重みの初期値には E. Chew の定義した重みの値である $w1, w2, w3 = 0.536, 0.274, 0.19$ を用いた．

具体的な計算は以下の式となる．r_1, r_2, r_3 は重み更新のためのパラメータ，CE は C.E. の座標値である．$\mathrm{C}_1, \mathrm{C}_2, \mathrm{C}_3$ は正しい調の主要三和音の座標値であり，長調の場合は (5) から $\mathrm{C}_1 = \mathbf{C}_{\mathrm{M}}(\mathrm{k}), \mathrm{C}_2 = \mathbf{C}_{\mathrm{M}}(\mathrm{k}+1), \mathrm{C}_3 = \mathbf{C}_{\mathrm{M}}(\mathrm{k}-1)$，短調の場合は (6) から $\mathrm{C}_1 = \mathbf{C}_{\mathrm{m}}(\mathrm{k}), \mathrm{C}_2 = \mathbf{C}_{\mathrm{m}}(\mathrm{k}+1), \mathrm{C}_3 = \mathbf{C}_{\mathrm{m}}(\mathrm{k}-1)$ が入力される．w_1, w_2, w_3 は更新前の重み，R_1, R_2, R_3 は更新後の重みの比率パラメータ，w'_1, w'_2, w'_3 は更新された重みである．

(9) で，C.E. の座標値と正解の調の点の座標値との距離を誤差として最小自乗法を行い，新たな重みを導出するためのパラメータを計算する．

$$\begin{cases} \frac{\partial}{\partial r_1}\left(-\frac{1}{2}\left(\mathrm{CE}-(r_1\mathrm{C}_1+r_2\mathrm{C}_2+r_3\mathrm{C}_3)\right)^2\right)=0 \\ \frac{\partial}{\partial r_2}\left(-\frac{1}{2}\left(\mathrm{CE}-(r_1\mathrm{C}_1+r_2\mathrm{C}_2+r_3\mathrm{C}_3)\right)^2\right)=0 \\ \frac{\partial}{\partial r_3}\left(-\frac{1}{2}\left(\mathrm{CE}-(r_1\mathrm{C}_1+r_2\mathrm{C}_2+r_3\mathrm{C}_3)\right)^2\right)=0 \end{cases} \quad (9)$$

(10)(11) で，新たな重みを導出する．

$$\begin{aligned} R_1 &= 0.9 * w_1 + 0.1 * r_1 \\ R_2 &= 0.9 * w_2 + 0.1 * r_1 \\ R_3 &= 0.9 * w_3 + 0.1 * r_1 \end{aligned} \quad (10)$$

$$\begin{aligned} w'_1 &= R_1 / (R_1 + R_2 + R_3) \\ w'_2 &= R_2 / (R_1 + R_2 + R_3) \\ w'_3 &= R_3 / (R_1 + R_2 + R_3) \end{aligned} \quad (11)$$

導出した新たな重みを用いて，(5)(6) の式で新たに 24 個の調の点の座標値を求める．得られた座標値が，平均パラメータの値となる．

3.4 精度パラメータの学習

次に，精度パラメータの学習について述べる．精度パラメータの学習は図 8 のように行なっている．精度行列の分布の学習では，まず観測モデルを次のようにおいた．

$$p(\mathbf{x}|\mathbf{\Lambda}) = \mathcal{N}(\mathbf{x}|\mu, \mathbf{\Lambda}^{-1}) \quad (12)$$

正定値行列である精度行列 $\mathbf{\Lambda}$ を生成するための確率分布としては，次のようなウィシャート事前分布がある．

$$p(\mathbf{\Lambda}) = \mathcal{W}(\mathbf{\Lambda}|v, \mathbf{W}) \quad (13)$$

ここで，$\mathbf{W}$ は 3×3 の正定値行列であり，自由度パラメータである v は $v > 2$ を満たす実数値として事前に与える．ベイズの定理を用いて計算すると，データ $\mathbf{X} = \{\mathbf{x}_1, ..., \mathbf{x}_N\}$ を観測して得られる確率分布は，次のようにパラメータをおいたウィシャート分布となる．

$$p(\mathbf{\Lambda}|\mathbf{X}) = \mathcal{W}(\mathbf{\Lambda}|\hat{v}, \hat{\mathbf{W}}) \quad (14)$$

ただし

$$\hat{\mathbf{W}}^{-1} = \sum_{n=1}^{N} (\mathbf{x}_n - \mu)(\mathbf{x}_n - \mu)^{\mathrm{T}} + \mathbf{W}^{-1} \quad (15)$$

$$\hat{v} = N + v \quad (16)$$

本研究は，$\mathbf{W}$ は 3×3 の単位行列，自由度 v は 10 に設定した．μ には学習を行う調の点の座標値，$\mathbf{X}$ には学習する 8 小節ごとの C.E. の座標値が入力される．この学習を 24 個の調の点全てに行い，確率化された SAM である B-SAM を導出する．

図 7，図 8 は，学習前と学習後の分布の変化の例である．幅 1 の三次元格子を設定し，それぞれの座標値における確率密度関数の大きさを図示している．この 2 つの図では，24 個ある調の確率分布のうち C major の確率分布のみを表示している．赤色に近く大きいものほど確率密度関数の値が大きく，青色に近く小さいものほど確率密度関数の値が小さい座標値であることを表している．

3.5 調性同定方法

ここまでで得られた B-SAM を用いて，調性同定を行う．B-SAM 内のある座標値を入力すると，その座標値における確率密度関数の値が大きい順に調を出力する．最も値が大きい，つまり可能性が高い調を同定された調とする．

Vol.2019-MUS-122 No.11
Vol.2019-EC-51 No.11
2019/2/22

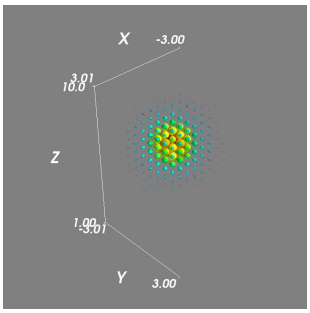

図 7　学習前の確率分布例

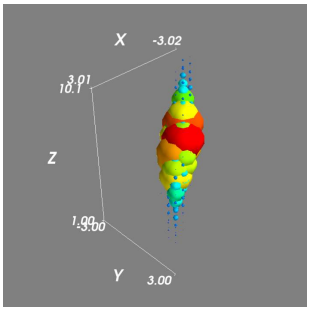

図 8　学習後の確率分布例

4.　調性同定の精度比較実験と考察

従来の SAM と提案する B-SAM の比較調査のため，調性同定精度を比較する実験を 20 曲で行なった．また，対象楽曲の中で入力に使用する音を変更することで結果がどう変化するかも調査するため，各曲 (ア) メロディ音のみ，(イ) ベース音のみ，(ウ) メロディ・ベース以外の音，(エ) 全ての音，の 4 パターンを入力データとしたときの調性同定精度の比較を行なった．本実験では，楽曲全体で一つの調を同定した．

4.1　実験に用いたデータ

対象楽曲には，RWC 研究用音楽データベースのポピュラー音楽データベース [6] の楽曲を用いた．データベース 1 番の楽曲から順に聞いていき，長調の曲を 10 曲，短調の曲を 10 曲採用した．なお，本実験では正解となる調を一つの楽曲につき一つに定めるため，本格的転調のない曲を選択した．採用した 20 曲の Standard MIDI File からそれぞれ実験に使用する音を変更した (ア) メロディ音のみ，(イ) ベース音のみ，(ウ) メロディ・ベース以外の音，(エ) 全ての音，の 4 パターン分の MusicXML を生成し，入力データとした．この際，ドラムやパーカッションなどの打楽器やサウンドエフェクト，効果音等の音階のない音のピッチイベントに関しては本研究では扱わないため，削除した．また，比較や考察のしやすさを考慮し，長調の曲はハ長調に，短調の曲はイ短調に移調し統一した．以下，ハ長調は C，イ短調は a のように，調をトニックの音名で表す．大文字は長調，小文字は短調とする．

4.2　実験結果

本実験では，24 個ある調のうち可能性が高い順に出力し，もっとも可能性の高いものとして出力された調を同定結果とする．

調性同定の結果をまとめると次のようになった．表 1 は，C が正解の調である楽曲 1〜10 のときの調性同定結果比較である．正しく C が同定できている部分は赤字で示している．表 2 は，a が正解の調である楽曲 11〜20 のときの調性同定結果比較である．正しく a が同定できている部分は赤字で示している．表 3 は，正解 C 調と正解 a 調をあわせた平均の結果である．

4.3　考察

まず，SAM では不正解であったが B-SAM で正解できた部分について例をあげて考察する．一つ目の例として，楽曲 1(エ) をあげる．この曲は，SAM では G だったが，B-SAM では正解の C が同定できた．この楽曲は，G メジャーコードや G におけるドミナントコードであり C のダイアトニックコードではない D メジャーコードが頻出するため，SAM では G と同定したと考えられる．二つ目の例として，楽曲 11(エ) をあげる．SAM では d だったが，B-SAM では a が同定できた．この曲は，トニックコードである A マイナーのコードが出てくる回数は多いが，ダイアトニックコードとは外れた和音も頻繁に出現する曲であり，従来の SAM のように単純な音の数だけでは a と同定することが難しかったと考えられる．上記二つの例において，B-SAM では，その外れた音の部分まで C である確率として表現し，正しく調性同定できたものと考えられる．

次に，B-SAM で正解できなかった部分について例をあげて考察する．一つ目の例として，楽曲 2(エ) をあげる．正解は C であるが，B-SAM は a と同定した．この曲は，A マイナーコード - D メジャーコード - G メジャーコード - C メジャーコードという和声進行が多く繰り返される楽曲である．この進行は，最後に C メジャーコードが出現するまで a なのか C なのかを感じ取りにくいものであり，B-SAM ではコードの順番などの細かい時系列情報は扱えないため，正しく同定できなかったと考えられる．

5.　おわりに

本研究では，ベイズ推論を用いて拡張した SAM である B-SAM を実装し，従来の SAM との調性同定精度の比較実験を行なった．今後の展望としては以下が挙げられる．

・重みを最適化し平均パラメータを求める際の学習率を変更して実験する．
・精度パラメータを求める際の自由度の数値を変更し，結果として同定される調の割合を強くして実験する．
・本研究では繰り返し学習を行なっていく範囲を 8 小節ずつと指定していたが，その範囲を調節したり，入力される音の時系列を考慮したものに改良する．
・実験に用いる曲数を増やし，考察をより深く行う．
・転調の扱いについて考慮した B-SAM の設計について考案する．
・認知的音楽理論である Generative Theory of Tonal Music（GTTM）[13] へ応用する．
・SAM を視覚化したアプリケーション MuSA_RT のように，B-SAM でも確率分布を視覚化したアプリケーションを開発する．

Vol.2019-MUS-122 No.11
Vol.2019-EC-51 No.11
2019/2/22

表 1 結果比較：正解が C の楽曲

手法	入力音	楽曲 1	楽曲 2	楽曲 3	楽曲 4	楽曲 5	楽曲 6	楽曲 7	楽曲 8	楽曲 9	楽曲 10	正解率
SAM	(ア)	C	a	a	C	C	C	a	C	C	C	70%
	(イ)	C	a	a	F	d	C	C	F	C	C	50%
	(ウ)	G	a	a	C	C	C	C	C	C	C	70%
	(エ)	G	a	a	C	C	C	C	C	C	C	70%
B-SAM	(ア)	c	C	C	C	C	C	C	C	C	C	90%
	(イ)	C	e	C	a	a♭	C	C	C	C	C	70%
	(ウ)	G	a	C	c	C	C	C	C	C	C	70%
	(エ)	C	a	C	a	C	C	C	C	C	C	80%

表 2 結果比較：正解が a の楽曲

手法	入力音	楽曲 11	楽曲 12	楽曲 13	楽曲 14	楽曲 15	楽曲 16	楽曲 17	楽曲 18	楽曲 19	楽曲 20	正解率
SAM	(ア)	C	a	C	C	a	C	a	d	a	a	50%
	(イ)	F	a	d	C	a	d	d	F	d	a	30%
	(ウ)	a	a	a	C	a	a	a	d	a	a	80%
	(エ)	C	a	a	C	a	a	a	d	a	a	70%
B-SAM	(ア)	F	a	F	a	a	a	a	C	A	a	60%
	(イ)	a	a	a	C	a	a	a	a	a	a	90%
	(ウ)	a	a	a	a	a	a	a	a	a	C	90%
	(エ)	a	a	a	a	a	a	a	a	a	a	100%

表 3 C・a をあわせた正解率

手法	入力音	正解率
SAM	(ア)	60%
	(イ)	40%
	(ウ)	75%
	(エ)	70%
B-SAM	(ア)	75%
	(イ)	80%
	(ウ)	80%
	(エ)	90%

謝辞 研究を通じて, 議論をしていただいた寺井あすか准教授 (公立はこだて未来大学) に感謝いたします. 本研究の一部は JSPS 科研費（16H01744，16K12560）の助成を受けたものです.

参考文献

[1] Alexandre R. J. Francois. MuSA_RT. https://itunes.apple.com/gb/app/musa-rt/id506866959?mt=12, 2012 (accessed January 30, 2019)

[2] Cohn, Richard. ”Neo-riemannian operations, parsimonious trichords, and their” tonnetz” representations.” Journal of Music Theory 41.1 (1997): 1-66.

[3] 海老澤敏，上参郷祐康，西岡信雄，山口修，新編・音楽中辞典，音楽之友社（2002）.

[4] Elaine Chew, Towards a mathematical model of tonality. Diss. Massachusetts Institute of Technology,（2000）.

[5] Elaine Chew: Mathematical and Computational Modeling of Tonality - Theory and Applications, Springer US(2014).

[6] 後藤真孝, et al. ”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 45.3 (2004): 728-738.

[7] Herremans, Dorien, Ching-Hua Chuan, and Elaine Chew. ”A functional taxonomy of music generation systems.” ACM Computing Surveys (CSUR) 50.5 (2017): 69.

[8] カワイ音楽研究部，すぐに役立つ 音楽用語ハンドブック—音楽・教育・保育に携わる人々に，カワイ出版（1999）.

[9] Krumhansl, Carol L. Cognitive foundations of musical pitch. Oxford University Press, 2001.

[10] Lerdahl, Fred. Tonal pitch space. Oxford University Press, 2004.

[11] Lerdahl, F. and Jackendoff, R.: A Generative Theory of Tonal Music, The MIT Press, Cambridge (1983).

[12] Longuet-Higgins, H. Christopher, and Mark J. Steedman. ”On interpreting bach.” Machine intelligence 6 (1971): 221-241.

[13] 澤田隼, 竹川佳成, and 平田圭二. ”音楽音響信号の階層的クラスタリングを用いた GTTM タイムスパン木の抽出法について.” 第 80 回全国大会講演論文集 2018.1 (2018): 173-174.

[14] 齋藤翔一郎，武山晴登，西本卓也，嵯峨山茂樹，Specmurt 分析と Chroma Vector を用いた HMM による音楽音響信号の調認識，情報処理学会研究報告音楽情報科学 (MUS), (2005): 85-90.

[15] 谷井章夫，片寄晴弘，演奏システム INSPIRATION の改良–SMF データのインポート機能と音高置換処理の検討，情報処理学会論文誌 46.3 (2005): 849-858.