

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号

特許第7122010号

(P7122010)

(45)発行日 令和4年8月19日(2022.8.19)

(24)登録日 令和4年8月10日(2022.8.10)

(51)国際特許分類

F I

G 1 0 H 1/00 (2006.01)

G 1 0 H

1/00

1 0 2 Z

請求項の数 24 (全39頁)

(21)出願番号	特願2020-12430(P2020-12430)	(73)特許権者	504132272
(22)出願日	令和2年1月29日(2020.1.29)		国立大学法人京都大学
(65)公開番号	特開2021-117434(P2021-117434 A)	(74)代理人	京都府京都市左京区吉田本町3 6 番地 1
(43)公開日	令和3年8月10日(2021.8.10)		100125117
審査請求日	令和2年11月20日(2020.11.20)	(74)代理人	弁理士 坂田 泰弘
特許法第30条第2項適用 研究集会での発表 研究集会名 : 44th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 論文タイトル: UNSUPERVISED MELODY STYLE CONVERSION 論文公開日 : 2019年4月17日 口頭発表日 : 2019年5月16日 [刊行物等] 出版物による公開 出版物名 : 研究報告音楽情報科学 発行者名 : 一般社団法人		(74)代理人	100086933
		(72)発明者	弁理士 久保 幸雄
			中村 栄太
		(72)発明者	京都府京都市左京区吉田本町3 6 番地 1
			国立大学法人京都大学内
		(72)発明者	吉井 和佳
			京都府京都市左京区吉田本町3 6 番地 1
			国立大学法人京都大学内
		審査官	菊池 智紀
	最終頁に続く		最終頁に続く

(54)【発明の名称】 音楽モデル生成システム、自動作曲システム、音楽モデル生成方法、およびコンピュータプログラム

(57)【特許請求の範囲】

【請求項1】

N個のスタイルそれぞれの、音楽的な特性を示す音楽モデルを生成する、音楽モデル生成システムであって、

前記N個のスタイルのうちのp番目のスタイルが出現するスタイル変数確率ならびに音高クラスおよび拍節位置を有する音符の列のp番目のスタイルにおける生成確率それぞれに関連するパラメータを有する時系列モデルであるN個の仮モデルを、当該各パラメータの初期値の組合せが相違するように用意する初期処理手段と、

前記N個のスタイルごとに複数用意された、連続する複数の音符それぞれの音高クラスおよび拍節位置が定まっている楽曲のそれぞれにおいて、前記N個の仮モデルそれぞれに示される特性が出現する出現確率を算出する算出処理を実行する出現確率算出手段と、前記N個の仮モデルそれぞれの、前記スタイル変数確率および前記生成確率それぞれに関連するパラメータを前記出現確率に基づいて調整する調整処理を実行する調整手段と、

所定の条件が満たされたか否かを判定する判定手段と、

を有し、

前記出現確率算出手段および前記調整手段が、それぞれ、前記所定の条件が満たされるまで前記算出処理および前記調整処理を交互に繰り返し行うことによって、前記N個のスタイルそれぞれの前記音楽モデルを生成する、

ことを特徴とする音楽モデル生成システム。

【請求項2】

10

20

前記初期処理手段は、 p 番目のスタイルである場合に最初の音符として音高クラスが q でありかつ拍節位置が b である音符が出現する初期確率および p 番目のスタイルである場合に音高クラスが q' でありかつ拍節位置が b' である音符から音高クラスが q でありかつ拍節位置が b である音符へ遷移する遷移確率それぞれのパラメータを前記生成確率に関連するパラメータとして有するマルコフモデルを、前記 N 個の仮モデルとして用意する、

請求項 1 に記載の音楽モデル生成システム。

【請求項 3】

前記初期処理手段は、 p 番目のスタイルである場合に音高クラスが q' でありかつ拍節位置が b' である音符から音高クラスが q でありかつ拍節位置が b である音符へ遷移する遷移確率のパラメータを前記生成確率に関連するパラメータとして有するマルコフモデルを、前記 N 個の仮モデルとして用意する、

10

請求項 1 に記載の音楽モデル生成システム。

【請求項 4】

N 個のスタイルそれぞれの、音楽的な特性を示す音楽モデルを生成する、音楽モデル生成システムであって、

前記 N 個のスタイルのうちの p 番目のスタイルが出現するスタイル変数確率および音高クラスを有する音符の列の p 番目のスタイルにおける生成確率それぞれに関連するパラメータを有する時系列モデルである N 個の仮モデルを、当該各パラメータの初期値の組合せが相違するように用意する初期処理手段と、

前記 N 個のスタイルごとに複数用意された、連続する複数の音符それぞれの音高クラスが定まっている楽曲のそれぞれにおいて、前記 N 個の仮モデルそれぞれに示される特性が出現する出現確率を算出する算出処理を実行する出現確率算出手段と、

20

前記 N 個の仮モデルそれぞれの、前記スタイル変数確率および前記生成確率それぞれに関連するパラメータを前記出現確率に基づいて調整する調整処理を実行する調整手段と、

所定の条件が満たされたか否かを判定する判定手段と、

を有し、

前記出現確率算出手段および前記調整手段が、それぞれ、前記所定の条件が満たされるまで前記算出処理および前記調整処理を交互に繰り返し行うことによって、前記 N 個のスタイルそれぞれの前記音楽モデルを生成する、

ことを特徴とする音楽モデル生成システム。

30

【請求項 5】

前記初期処理手段は、 p 番目のスタイルである場合に最初の音符として音高クラスが q である音符が出現する初期確率および p 番目のスタイルである場合に音高クラスが q' である音符から音高クラスが q である音符へ遷移する遷移確率それぞれのパラメータを前記生成確率に関連するパラメータとして有するマルコフモデルを、前記仮モデルとして用意する、

請求項 4 に記載の音楽モデル生成システム。

【請求項 6】

前記初期処理手段は、 p 番目のスタイルである場合に音高クラスが q' である音符から音高クラスが q である音符へ遷移する遷移確率それぞれのパラメータを前記生成確率に関連するパラメータとして有するマルコフモデルを、前記仮モデルとして用意する、

40

請求項 4 に記載の音楽モデル生成システム。

【請求項 7】

前記 N 個の仮モデルそれぞれのパラメータは、音高クラスそれぞれがトニックである上ニック確率が含まれる、

請求項 1 ないし請求項 6 のいずれかに記載の音楽モデル生成システム。

【請求項 8】

前記出現確率算出手段および前記調整手段は、前記楽曲を共通の調に揃えて前記出現確率を算出したまたは前記 N 個の仮モデルのパラメータを調整する、

請求項 1 ないし請求項 7 のいずれかに記載の音楽モデル生成システム。

50

【請求項 9】

N 個のスタイルそれぞれの、音楽的な特性を示す音楽モデルを生成する、音楽モデル生成システムであって、

前記 N 個のスタイルのうちの ρ 番目のスタイルが出現するスタイル変数確率および拍節位置を有する音符の列の ρ 番目のスタイルにおける生成確率それぞれに関連するパラメータを有する時系列モデルである N 個の仮モデルを、当該各パラメータの初期値の組合せが相違するように用意する初期処理手段と、

前記 N 個のスタイルごとに複数用意された、連続する複数の音符それぞれの拍節位置が定まっている楽曲のそれぞれにおいて、前記 N 個の仮モデルそれぞれに示される特性が出現する出現確率を算出する算出処理を実行する出現確率算出手段と、

前記 N 個の仮モデルそれぞれの、前記スタイル変数確率および前記生成確率それぞれに関連するパラメータを前記出現確率に基づいて調整する調整処理を実行する調整手段と、

所定の条件が満たされたか否かを判定する判定手段と、

を有し、

前記出現確率算出手段および前記調整手段が、それぞれ、前記所定の条件が満たされるまで前記算出処理および前記調整処理を交互に繰り返し行うことによって、前記 N 個のスタイルそれぞれの前記音楽モデルを生成する、

ことを特徴とする音楽モデル生成システム。

【請求項 10】

前記初期処理手段は、 ρ 番目のスタイルである場合に最初の音符として拍節位置が b である音符が出現する初期確率および ρ 番目のスタイルである場合に拍節位置が b' である音符から拍節位置が b である音符へ遷移する遷移確率それぞれのパラメータを前記生成確率に関連するパラメータとして有するマルコフモデルを、前記仮モデルとして用意する、

請求項 9 に記載の音楽モデル生成システム。

【請求項 11】

前記初期処理手段は、 ρ 番目のスタイルである場合に拍節位置が b' である音符から拍節位置が b である音符へ遷移する遷移確率それぞれのパラメータを前記生成確率に関連するパラメータとして有するマルコフモデルを、前記仮モデルとして用意する、

請求項 9 に記載の音楽モデル生成システム。

【請求項 12】

前記所定の条件は、前記算出ステップと前記調整ステップとを交互に繰り返し行った回数が所定の回数に達することである、

請求項 1 ないし請求項 11 のいずれかに記載の音楽モデル生成システム。

【請求項 13】

前記所定の条件は、前記 N 個の仮モデルのうちのいずれかのパラメータの変化量または変化率が所定の値以下になることである、

請求項 1 ないし請求項 11 のいずれかに記載の音楽モデル生成システム。

【請求項 14】

請求項 1 ないし請求項 3 のいずれかに記載する音楽モデル生成システムと、

前記 N 個のスタイルのうちの指定されたスタイルの新たな楽曲を生成する作曲手段と、

を有し、
前記音楽モデル生成システムの前記判定手段は、前記所定の条件が満たされるまで前記算出処理および前記調整処理が行われた前記 N 個の仮モデルを前記 N 個のスタイルそれぞれの前記音楽モデルとして記憶手段に記憶させ、

前記作曲手段は、前記指定されたスタイルの前記音楽モデルに基づいて前記新たな楽曲を生成する、

ことを特徴とする自動作曲システム。

【請求項 15】

2 つのメロディの類似度を算出するための編集モデルを前記 N 個のスタイルのうちの第一のスタイルの前記音楽モデルおよび第二のスタイルの前記音楽モデルに基づいて生成す

10

20

30

40

50

る編集モデル生成手段と、

前記第二のスタイルの前記音楽モデルに基づいて複数の候補メロディを生成するメロディ生成手段と、

前記類似度として、前記複数の候補メロディそれぞれの、前記第一のスタイルを有しかつ編曲する対象である対象メロディとの類似度を前記編集モデルに基づいて算出する類似度算出手段と、

前記複数の候補メロディのうちの前記類似度が最も高いものまたは所定の値以上のものを前記対象メロディの編集メロディとして選出する選出手段と、

を有する、

請求項 14 に記載の自動作曲システム。

10

【請求項 16】

前記N個のスタイルのうちの第一のスタイルの第一のメロディと第二のスタイルの第二のメロディとの類似度を算出する類似度関数および前記第二のスタイルへの前記第二のメロディの適合度を算出する適合度関数によって編曲の適切度を算出するための編集モデルを前記第一のスタイルの前記音楽モデルおよび前記第二のスタイルの前記音楽モデルに基づいて生成する編集モデル生成手段と、

前記第二のスタイルの前記音楽モデルに基づいて複数の候補メロディを生成するメロディ生成手段と、

前記複数の候補メロディそれぞれの、前記第一のスタイルを有しかつ編曲する対象である対象メロディとの前記適切度を前記編集モデルに基づいて算出する適切度算出手段と、

20

前記複数の候補メロディのうちの前記適切度が最も高いものまたは所定の値以上のものを前記対象メロディの編集メロディとして選出する選出手段と、

を有する、

請求項 14 に記載の自動作曲システム。

【請求項 17】

前記N個のスタイルのうちの第一のスタイルの第一のメロディと第二のスタイルの第二のメロディとの類似度を算出する類似度関数および前記第二のスタイルへの前記第二のメロディの適合度を算出する適合度関数によって編曲の適切度を算出するための編集モデルを前記第一のスタイルの前記音楽モデルおよび前記第二のスタイルの前記音楽モデルに基づいて生成する編集モデル生成手段と、

30

前記第一のスタイルを有する対象メロディを前記編集モデルに基づいて動的計画法によって前記第二のスタイルに編曲する編曲手段と、

を有する、

請求項 14 に記載の自動作曲システム。

【請求項 18】

請求項 4 ないし請求項 13 のいずれかに記載する音楽モデル生成システムと、

前記N個のスタイルのうちの指定されたスタイルの新たな楽曲を生成する作曲手段と、
を有し、

前記音楽モデル生成システムの前記判定手段は、前記所定の条件が満たされるまで前記算出処理および前記調整処理が行われた前記N個の仮モデルを前記N個のスタイルそれぞれの前記音楽モデルとして記憶手段に記憶させ、

40

前記作曲手段は、前記指定されたスタイルの前記音楽モデルに基づいて前記新たな楽曲を生成する、

ことを特徴とする自動作曲システム。

【請求項 19】

N個のスタイルそれぞれの、音楽的な特性を示す音楽モデルを生成する、音楽モデル生成方法であって、

前記N個のスタイルのうちのp番目のスタイルが出現するスタイル変数確率ならびに音高クラスおよび拍節位置を有する音符の列のp番目のスタイルにおける生成確率それぞれに関連するパラメータを有する時系列モデルであるN個の仮モデルを、当該各パラメータの

50

初期値の組合せが相違するように用意し、

1または複数のコンピュータのうちのいずれかが、前記N個のスタイルごとに複数用意された、連続する複数の音符それぞれの音高クラスおよび拍節位置が定まっている楽曲のそれぞれにおいて、前記N個の仮モデルそれぞれに示される特性が出現する出現確率を算出する算出処理を実行し、

前記1または複数のコンピュータのうちのいずれかが、前記N個の仮モデルそれぞれの、前記スタイル変数確率および前記生成確率それぞれに関連するパラメータを前記出現確率に基づいて調整する調整処理を実行し、

前記1または複数のコンピュータのうちのいずれかが、所定の条件が満たされたか否かを判定し、

前記1または複数のコンピュータのうちのいずれかが、前記所定の条件が満たされるまで前記算出処理および前記調整処理を交互に繰り返し行うことによって、前記N個のスタイルそれぞれの前記音楽モデルを生成する、

ことを特徴とする音楽モデル生成方法。

【請求項20】

N個のスタイルそれぞれの、音楽的な特性を示す音楽モデルを生成する、音楽モデル生成方法であって、

前記N個のスタイルのうちのp番目のスタイルが出現するスタイル変数確率および音高クラスを有する音符の列のp番目のスタイルにおける生成確率それぞれに関連するパラメータを有する時系列モデルであるN個の仮モデルを、当該各パラメータの初期値の組合せが相違するように用意し、

1または複数のコンピュータのうちのいずれかが、前記N個のスタイルごとに複数用意された、連続する複数の音符それぞれの音高クラスが定まっている楽曲のそれぞれにおいて、前記N個の仮モデルそれぞれに示される特性が出現する出現確率を算出する算出処理を実行し、

前記1または複数のコンピュータのうちのいずれかが、前記N個の仮モデルそれぞれの、前記スタイル変数確率および前記生成確率それぞれに関連するパラメータを前記出現確率に基づいて調整する調整処理を実行し、

前記1または複数のコンピュータのうちのいずれかが、所定の条件が満たされたか否かを判定し、

前記1または複数のコンピュータのうちのいずれかが、前記所定の条件が満たされるまで前記算出処理および前記調整処理を交互に繰り返し行うことによって、前記N個のスタイルそれぞれの前記音楽モデルを生成する、

ことを特徴とする音楽モデル生成方法。

【請求項21】

N個のスタイルそれぞれの、音楽的な特性を示す音楽モデルを生成する、音楽モデル生成方法であって、

前記N個のスタイルのうちのp番目のスタイルが出現するスタイル変数確率および拍節位置を有する音符の列のp番目のスタイルにおける生成確率それぞれに関連するパラメータを有する時系列モデルであるN個の仮モデルを、当該各パラメータの初期値の組合せが相違するように用意し、

1または複数のコンピュータのうちのいずれかが、前記N個のスタイルごとに複数用意された、連続する複数の音符それぞれの拍節位置が定まっている楽曲のそれぞれにおいて、前記N個の仮モデルそれぞれに示される特性が出現する出現確率を算出する算出処理を実行し、

前記1または複数のコンピュータのうちのいずれかが、前記N個の仮モデルそれぞれの、前記スタイル変数確率および前記生成確率それぞれに関連するパラメータを前記出現確率に基づいて調整する調整処理を実行し、

前記1または複数のコンピュータのうちのいずれかが、所定の条件が満たされたか否かを判定し、

10

20

30

40

50

前記 1 または複数のコンピュータのうちのいずれかが、前記所定の条件が満たされるまで前記算出処理および前記調整処理を交互に繰り返し行うことによって、前記 N 個のスタイルそれぞれの前記音楽モデルを生成する、

ことを特徴とする音楽モデル生成方法。

【請求項 2 2】

N 個のスタイルそれぞれの、音楽的な特性を示す音楽モデルを生成するためのコンピュータプログラムであって、

コンピュータに、

前記 N 個のスタイルのうちの p 番目のスタイルが出現するスタイル変数確率ならびに音高クラスおよび拍節位置を有する音符の列の p 番目のスタイルにおける生成確率それぞれに関連するパラメータを有する時系列モデルである N 個の仮モデルを、当該各パラメータの初期値の組合せが相違するように用意する初期化処理を実行させ、

前記 N 個のスタイルごとに複数用意された、連続する複数の音符それぞれの音高クラスおよび拍節位置が定まっている楽曲のそれぞれにおいて、前記 N 個の仮モデルそれぞれに示される特性が出現する出現確率を算出する算出処理を実行させ、

前記 N 個の仮モデルそれぞれの、前記スタイル変数確率および前記生成確率それぞれに関連するパラメータを前記出現確率に基づいて調整する調整処理を実行させ、

所定の条件が満たされたか否かを判定する判定処理を実行させ、

前記所定の条件が満たされるまで前記算出処理および前記調整処理を交互に繰り返し実行させる、

ことを特徴とするコンピュータプログラム。

【請求項 2 3】

N 個のスタイルそれぞれの、音楽的な特性を示す音楽モデルを生成するためのコンピュータプログラムであって、

コンピュータに、

前記 N 個のスタイルのうちの p 番目のスタイルが出現するスタイル変数確率および音高クラスを有する音符の列の p 番目のスタイルにおける生成確率それぞれに関連するパラメータを有する時系列モデルである N 個の仮モデルを、当該各パラメータの初期値の組合せが相違するように用意する初期化処理を実行させ、

前記 N 個のスタイルごとに複数用意された、連続する複数の音符それぞれの音高クラスが定まっている楽曲のそれぞれにおいて、前記 N 個の仮モデルそれぞれに示される特性が出現する出現確率を算出する算出処理を実行させ、

前記 N 個の仮モデルそれぞれの、前記スタイル変数確率および前記生成確率それぞれに関連するパラメータを前記出現確率に基づいて調整する調整処理を実行させ、

所定の条件が満たされたか否かを判定する判定処理を実行させ、

前記所定の条件が満たされるまで前記算出処理および前記調整処理を交互に繰り返し実行させる、

ことを特徴とするコンピュータプログラム。

【請求項 2 4】

N 個のスタイルそれぞれの、音楽的な特性を示す音楽モデルを生成するためのコンピュータプログラムであって、

コンピュータに、

前記 N 個のスタイルのうちの p 番目のスタイルが出現するスタイル変数確率および拍節位置を有する音符の列の p 番目のスタイルにおける生成確率それぞれに関連するパラメータを有する時系列モデルである N 個の仮モデルを、当該各パラメータの初期値の組合せが相違するように用意する初期化処理を実行させ、

前記 N 個のスタイルごとに複数用意された、連続する複数の音符それぞれの拍節位置が定まっている楽曲のそれぞれにおいて、前記 N 個の仮モデルそれぞれに示される特性が出現する出現確率を算出する算出処理を実行させ、

前記 N 個の仮モデルそれぞれの、前記スタイル変数確率および前記生成確率それぞれに關

10

20

30

40

50

連するパラメータを前記出現確率に基づいて調整する調整処理を実行させ、

所定の条件が満たされたか否かを判定する判定処理を実行させ、

前記所定の条件が満たされるまで前記算出処理および前記調整処理を交互に繰り返し実行させる、

ことを特徴とするコンピュータプログラム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、A I (Artificial Intelligence) を適用した作曲または編曲の技術に関する。

10

【背景技術】

【0002】

従来、コンピュータによる自動作曲の技術が提案されている。例えば、Orpheus は、日本語の歌詞が入力されスタイルが指定されると、入力された歌詞に応じかつ指定されたスタイルを有する楽曲を自動的に生成する（非特許文献1）。

【先行技術文献】

【非特許文献】

【0003】

【文献】東京大学 大学院 システム情報学 第一研究室 ウェブサイト、<http://www.orpheus-music.org/>

20

【発明の概要】

【発明が解決しようとする課題】

【0004】

しかし、従来の技術によると、スタイルのデータを用意するためには、人間がスタイルの特徴を分析しなければならない。したがって、指定可能なスタイルを増やすには、コストが嵩んでしまう。

【0005】

本発明は、このような課題に鑑み、従来よりも容易にA I による作曲を実現することを目的とする。

【課題を解決するための手段】

【0006】

本発明の一形態に係る音楽モデル生成システムは、N個のスタイルそれぞれの、音楽的な特性を示す音楽モデルを生成する、音楽モデル生成システムであって、前記N個のスタイルのうちのp番目のスタイルが出現するスタイル変数確率ならびに音高クラスおよび拍節位置を有する音符の列のp番目のスタイルにおける生成確率それぞれに関連するパラメータを有する時系列モデルであるN個の仮モデルを、当該各パラメータの初期値の組合せが相違するように用意する初期処理手段と、前記N個のスタイルごとに複数用意された、連続する複数の音符それぞれの音高クラスおよび拍節位置が定まっている楽曲のそれぞれにおいて、前記N個の仮モデルそれぞれに示される特性が出現する出現確率を算出する算出処理を実行する出現確率算出手段と、前記N個の仮モデルそれぞれの、前記スタイル変数確率および前記生成確率それぞれに関連するパラメータを前記出現確率に基づいて調整する調整処理を実行する調整手段と、所定の条件が満たされたか否かを判定する判定手段と、を有し、前記出現確率算出手段および前記調整手段が、それぞれ、前記所定の条件が満たされるまで前記算出処理および前記調整処理を交互に繰り返し行うことによって、前記N個のスタイルそれぞれの前記音楽モデルを生成する。

40

【0007】

好ましくは、前記所定の条件は、前記算出ステップと前記調整ステップとを交互に繰り返し行った回数が所定の回数に達することである。

【0008】

または、前記N個のスタイルのうちの第一のスタイルの第一のメロディと第二のスタイ

50

ルの第二のメロディとの類似度を算出する類似度関数および前記第二のスタイルへの前記第二のメロディの適合度を算出する適合度関数によって編曲の適切度を算出するための編集モデルを前記第一のスタイルの前記音楽モデルおよび前記第二のスタイルの前記音楽モデルに基づいて生成する編集モデル生成手段と、前記第一のスタイルを有する対象メロディを前記編集モデルに基づいて動的計画法によって前記第二のスタイルに編曲する編曲手段と、を有する。

【発明の効果】

【0009】

本発明によると、従来よりも容易にAIによる作曲を実現することができる。特に、様々な楽曲を用意しNの値を大きくすることによって、多様なスタイルでの作曲のための音楽モデルを提供することができる。

10

【図面の簡単な説明】

【0010】

【図1】自動音楽生成装置を含むネットワークシステムの例を示す図である。

【図2】自動音楽生成装置のハードウェア構成の例を示す図である。

【図3】自動音楽生成装置の機能的構成の例を示す図である。

【図4】自動音楽生成装置の全体的な処理の流れの例を説明するフローチャートである。

【図5】作曲学習処理の流れの例を説明するフローチャートである。

【図6】作曲処理の流れの例を説明するフローチャートである。

【図7】編曲学習処理の流れの例を説明するフローチャートである。

20

【図8】編曲処理の流れの例を説明するフローチャートである。

【発明を実施するための形態】

【0011】

〔1 自動音楽生成装置1の全体〕

図1は、自動音楽生成装置1を含むネットワークシステムの例を示す図である。図2は、自動音楽生成装置1のハードウェア構成の例を示す図である。図3は、自動音楽生成装置1の機能的構成の例を示す図である。図4は、自動音楽生成装置1の全体的な処理の流れの例を説明するフローチャートである。

【0012】

図1に示す自動音楽生成装置1は、ユーザの好みのスタイルの楽曲を自動的に作曲し、または、ある原曲をユーザの好みのスタイルに編曲するシステムである。「スタイル」は、クラシック、ポピュラー、カントリー、ロック、演歌、およびジャズなどの音楽スタイルである。または、中世西洋音楽、ルネサンス音楽、バロック音楽、古典派音楽、およびロマン派音楽のような、より具体的なスタイルであってもよい。

30

【0013】

従来の自動作曲システムおよび自動編曲システムは、スタイルごとの音楽的な特徴を人間が解析してルール化したデータに基づいて作曲または編曲する。つまり、ルールベースの人工知能によって作曲または編曲する。

【0014】

これに対し、自動音楽生成装置1は、機械学習によってモデル（学習済モデル）を生成し、生成したモデルに基づいて作曲または編曲する。機械学習のアルゴリズムとして、統計的でありかつ教師なしの機械学習のアルゴリズムが用いられる。以下、この仕組みを説明する。

40

【0015】

なお、楽曲には、主旋律および対旋律など複数のメロディーが含まれていることがあるが、本実施形態では、説明の簡単のため、単音の主旋律のみからなりかつ4分の4拍子の楽曲を作曲または編曲する場合を例に説明する。

【0016】

自動音楽生成装置1は、図2に示すように、メインプロセッサ10、RAM（Random Access Memory）11、ROM（Read Only Memory）12、補助記憶装置13、ネッ

50

トワークアダプタ 14、キーボード 15、ポインティングデバイス 16、データ入出力ボード 17、ディスプレイ 18、および音声再生ユニット 19 などによって構成される。

【0017】

ROM 12 または補助記憶装置 13 には、オペレーティングシステムのほか種々のプログラムがインストールされている。特に、本実施形態では、プログラムの 1 つとして音楽生成プログラム 2 (図 3 参照) がインストールされている。補助記憶装置 13 として、ハードディスクまたは SSD (Solid State Drive) などが用いられる。

【0018】

RAM 11 は、自動音楽生成装置 1 のメインメモリである。RAM 11 には、オペレーティングシステムのほか音楽生成プログラム 2 などのプログラムがロードされる。

10

【0019】

メインプロセッサ 10 は、RAM 11 にロードされたプログラムを実行する。メインプロセッサ 10 として、GPU (Graphics Processing Unit) または CPU (Central Processing Unit) などが用いられる。

【0020】

ネットワークアダプタ 14 は、TCP/IP (Transmission Control Protocol/Internet Protocol) などのプロトコルで通信回線 42 を介して端末装置 41 など他の装置と通信するための装置である。ネットワークアダプタ 14 として、NIC (Network Interface Card) または Wi-Fi 用の子機が用いられる。

【0021】

20

キーボード 15 およびポインティングデバイス 16 は、コマンドまたはデータなどをオペレータが入力するための入力装置である。

【0022】

データ入出力ボード 17 は、周辺機器とデータのやり取りを行うための装置である。データ入出力ボード 17 として、例えば、USB (Universal Serial Bus) または Bluetooth (登録商標) などに準拠した入出力ボードが用いられる。

【0023】

ディスプレイ 18 は、コマンドもしくはデータを入力するための画面またはメインプロセッサ 10 による演算の結果などを表示する。音声再生ユニット 19 は、音声ボードおよびスピーカなどによって構成され、楽曲を再生する。

30

【0024】

音楽生成プログラム 2 には、図 3 に示すメインモジュール 20、作曲モジュール 21 および編曲モジュール 25 が含まれる。作曲モジュール 21 は、ユーザの好みのスタイルの楽曲を自動的に作曲するためのモジュールである。編曲モジュール 25 は、ある原曲をユーザの好みのスタイルに編曲するモジュールである。

【0025】

〔2 メインモジュール〕

メインモジュール 20 は、全体的な処理を規定するモジュールである。メインモジュール 20 によると、自動音楽生成装置 1 は、図 4 に示す手順で全体的な処理を実行する。

【0026】

40

メインモジュール 20 の起動後、自動音楽生成装置 1 は、メニュー画面を表示する (S101)。

【0027】

作曲のための学習のコマンドがメニュー画面において入力されると (S102 で Yes)、自動音楽生成装置 1 は、作曲学習処理を実行する (S103)。または、作曲のコマンドが入力されると (S104 で Yes)、作曲処理を実行する (S105)。作曲学習処理および作曲処理は、後述するように、作曲モジュール 21 に基づいて行われる。

【0028】

または、編曲のための学習のコマンドが入力されると (S106 で Yes)、自動音楽生成装置 1 は、編曲学習処理を実行する (S107)。または、編曲のコマンドが入力さ

50

れると (S 1 0 8 で Y e s)、編曲処理を実行する (S 1 0 9)。編曲学習処理および編曲処理は、後述するように、編曲モジュール 2 5 に基づいて行われる。

【 0 0 2 9 】

〔 3 作曲モジュール 〕

作曲モジュール 2 1 は、図 3 に示すように学習データ記憶部 2 1 1、初期化処理部 2 1 2、トニック推定部 2 1 3、事後確率算出部 2 1 4、パラメータ更新部 2 1 5、収束判定部 2 1 6、スタイルモデル記憶部 2 1 7、音符決定部 2 1 8、系列変換部 2 1 9、および作曲データ生成部 2 2 0 などによって構成される。上述の通り、作曲モジュール 2 1 は、作曲学習処理および作曲処理を実行するためのモジュールである。作曲学習処理が機械学習のフェーズに相当し、作曲処理が推論のフェーズに相当する。

10

【 0 0 3 0 】

以下、作曲モジュール 2 1 の各部の処理を機械学習のフェーズおよび推論のフェーズに大別して説明する。

【 0 0 3 1 】

〔 3 . 1 機械学習のフェーズ 〕

〔 3 . 1 . 1 概要 〕

図 5 は、作曲学習処理の流れの例を説明するフローチャートである。機械学習のフェーズにおいては、作曲用の学習済モデルが、図 5 に示す手順で生成される。

【 0 0 3 2 】

図 3 の初期化処理部 2 1 2 ないし収束判定部 2 1 6 は、学習器を構成し、推論つまり自動作曲のための学習済モデルを教師なし機械学習によってスタイルごとに生成する。

20

【 0 0 3 3 】

オペレータは、モデルを必要とするスタイル（以下、「対象スタイル」と記載する。）ごとに、機械学習用の学習データとして多数の楽曲の楽曲データ 6 0 を用意し自動音楽生成装置 1 へ入力する。これらの楽曲データ 6 0 は、学習データ記憶部 2 1 1 に記憶される（図 5 の S 7 0 1）。楽曲データ 6 0 の構成および用意については、後述する。さらに、オペレータは、対象スタイルの数を入力する。以下、対象スタイルの数を「対象スタイル数 N_M 」と記載する。例えば、対象スタイルがロックおよび演歌の 2 つである場合は、対象スタイル数 N_M として「2」を入力する。すると、対象スタイル数 N_M が初期化処理部 2 1 2 によって受け付けられる（S 7 0 2）。

30

【 0 0 3 4 】

なお、本実施形態では、教師なし機械学習によって学習済モデルを生成するので、それぞれの楽曲データ 6 0 がどのスタイルに属するのかが入力する必要がない。

【 0 0 3 5 】

初期化処理部 2 1 2 は、対象スタイル数 N_M のマルコフモデルを用意し、各マルコフモデルのパラメータにランダムな値を設定する（S 7 0 3）。この時点において、オペレータは、どのマルコフモデルがどの対象スタイルに対応するのか分からない。なお、これらのマルコフモデルのうちの 2 つ以上が偶然、一致することがある。このような場合は、各マルコフモデルのパラメータが相違するように、ステップ S 7 0 3 の処理をやり直す。

【 0 0 3 6 】

40

事後確率算出部 2 1 4 およびパラメータ更新部 2 1 5 は、E M (Expectation-Maximization) アルゴリズムによって各マルコフモデルのパラメータを調整する。

【 0 0 3 7 】

事後確率算出部 2 1 4 は、E ステップの処理を行い（S 7 0 4）、パラメータ更新部 2 1 5 は、M ステップの処理を行う（S 7 0 5）。E ステップの処理および M ステップの処理を交互に繰り返すことによって、各マルコフモデルにいずれかの対象スタイルの特徴が表われるように徐々に調整される。

【 0 0 3 8 】

収束判定部 2 1 6 は、収束の判定を行う（S 7 0 6）。例えば、マルコフモデルのパラメータのうちの少なくとも 1 つまたは全部が、E ステップの処理および M ステップの処理

50

を繰り返してもほとんど変化しなくなった場合に、理想的な値に収束したと判定する。または、所定の回数、Eステップの処理およびMステップの処理を繰り返した場合に、収束したと判定する。そして、収束したと判定した場合は（S707でYes）、その時点の各マルコフモデルを学習済モデルとしてスタイルモデル記憶部217に記憶させる（S708）。

【0039】

収束していないと判定された場合は（S707でNo）、ステップS704に戻って、Eステップの処理およびMステップの処理が再度、実行される。

【0040】

〔3.1.2 楽曲データ60の構成および用意〕

10

楽曲データ60は、次の式（1）のように、楽曲のメロディーすなわち楽曲を構成する音符ごとの音高および発音時刻を示す。

【0041】

【数1】

$$(p_{mn}, s_{mn})_{m=1, n=1}^{M, N_m} \cdots (1)$$

【0042】

「M」は楽曲に含まれる小節の数であり、「N_m」はm番目の小節に含まれる音符の数である。p_{mn}およびs_{mn}は、それぞれ、m番目の小節のn番目の音符の音高および発音時刻である。

20

【0043】

音高は、MIDI（Musical Instrument Digital Interface）規格におけるノートナンバと同様の値によって表わされる。つまり、音高は、半音上がるごとに「1」上がり、半音下がるごとに「1」下がる。音高の範囲は、0～127の各整数値および休符を表わす「-1」からなる。つまり、p_{mn} ∈ {-1, 0, 1, 2, ..., 127}であり、中央のド（C）の値は「60」である。

【0044】

本実施形態では、楽曲の拍子が4分の4拍子であるものとする。また、1つの音の長さの最短が16分音符の1/3であり、最長が1小節（つまり、全音符）である。したがって、例えば4分音符の長さは、「12」である。また、楽曲の先頭の時刻が「0」である。したがって、発音時刻s_{mn} ∈ {0, 1, ..., 47}である。

30

【0045】

または、楽曲データ60は、次の式（2）のように、楽曲を小節ごとに区切らずに楽曲を構成する音符ごとの音高および発音時刻を示してもよい。

【0046】

【数2】

$$(p_i, s_i)_{i=1}^I \cdots (2)$$

40

【0047】

「I」は、楽曲に含まれる音符の数である。「p_i」および「s_i」は、それぞれ、i番目の音符の音高および発音時刻である。楽曲がj小節からなる場合は、s_i ∈ {0, 1, ..., (48j - 1)}である。

【0048】

楽曲のMIDIまたはMusicXML（Extensible Markup Language）のフォーマットのファイルがあれば、このファイルを変換することによって楽曲データ60が得られる。そこで、オペレータは、MIDIまたはMusicXMLのファイルを楽曲データ

50

60に変換し、ネットワークアダプタ14またはデータ入出力ボード17を介して自動音楽生成装置1へ入力し学習データ記憶部211に記憶させればよい。または、変換を自動音楽生成装置1が実行してもよい。

【0049】

なお、逆に、楽曲データ60をMIDIまたはMusicXMLのファイルに変換することもできる。この方向の変換は、後述するように、作曲および編曲の際に用いられる。

【0050】

ところで、以下に説明する、音楽生成のための統計学習手法において、式(1)または式(2)で表わす値をそのまま使用すると、モデルのパラメータが多すぎるので効果的な学習ができないことがある。

10

【0051】

そこで、楽曲データ60において、自由度をより低くするために、次の式(3)のように、オクターブ内のドを基準とする相対値(音高クラス、ピッチクラス)によって音高を表わし、かつ、小節内の先頭を基準とする相対位置(拍節位置)で発音時刻を表わしてもよい。

【0052】

【数3】

$$(q_{mn}, b_{mn})_{m=1, n=1}^{M, N_m} \cdots (3)$$

20

【0053】

「 q_{mn} 」および「 b_{mn} 」は、それぞれ、 m 番目の小節の n 番目の音符の音高クラスおよび拍節位置である。以下、音高クラスの個数を「 Q 」と記載し、音符の長さのパターン数を「 B 」と記載する。本実施形態では、 $Q = 12$ 、 $B = 48$ 、である。

【0054】

または、楽曲データ60は、(4)式のように各音符の音高を音高クラスによって表わし発音時刻を拍節位置によって表わしてもよい。

【0055】

【数4】

$$(q_i, b_i)_{i=1}^I \cdots (4)$$

30

【0056】

「 q_i 」および「 b_i 」は、それぞれ、 i 番目の音符の音高クラスおよび拍節位置である。

【0057】

音符が休符である場合を除くと、 $q_i \equiv p_i \pmod{12}$ $\{0, 1, \dots, 11\} = \{C, C\#, \dots, B\}$ の関係が成り立つ。音符が休符である場合は、 $q_i = p_i$ 、である。さらに、 $b_i \equiv s_i \pmod{48}$ $\{0, 1, \dots, 47\}$ の関係が成り立つ。これらの関係に基づいて、式(1)から式(3)へ変換し、または式(2)から式(4)へ変換することができる。

40

【0058】

したがって、楽曲のMIDIまたはMusicXMLのファイルがあれば、式(3)または式(4)形式でメロディーを示すデータが楽曲データ60として得られる。後述する各処理においては、式(3)または式(4)の形式の値が適宜、用いられる。

【0059】

以下、 H ($H \geq 2$)曲の楽曲の楽曲データ60が用意された場合を例に説明する。各楽曲には、1、2、3、...、 H のいずれかのシーケンス番号が付いている。説明の簡単のため、 H 曲の楽曲すべてが4分の4拍子であるものとする。また、説明の簡単のため、休符を扱わない場合を例に説明する。なお、 $h \in \{1, \dots, H\}$ である。式(4)で表わされる

50

系列を「q b 系列」と記載する。また、式(2)で表わされる系列を「p s 系列」と記載する。

【0060】

特にh番目の楽曲のメロディーを表わしたい場合は、式(3)および式(4)をそれぞれ式(5)または式(6)式のように記載することができる。

【0061】

【数5】

$$(q_{mn}^{(h)}, b_{mn}^{(h)})_{m=1, n=1}^{M_h, N_{m_h}} \dots (5)$$

10

$$(q_i^{(h)}, b_i^{(h)})_{i=1}^{I_h} \dots (6)$$

【0062】

〔3.1.3 マルコフモデルの構成〕

各マルコフモデルには、スタイル変数確率、初期確率、および遷移確率の3つのパラメータによって構成され、それぞれ、次の式(7)、式(8)、および式(9)によって表わされる。

20

$$\theta(\rho) \dots (7)$$

$$\pi_{ini}(q, b | \rho) \dots (8)$$

$$\pi(q, b | q', b', \rho) \dots (9)$$

式(7)は、 ρ 番目の対象スタイルが出現する確率を表わしている。したがって、 N_M 個の要素からなる確率ベクトルである。なお、 $\rho \in \{1, \dots, N_M\}$ である。

【0063】

式(8)は、スタイルが ρ 番目の対象スタイルである場合の最初の音符として (q, b) が出現する確率を表わしている。したがって、 $Q \times B$ 個の確率ベクトルが N_M 個、集まったものである。

【0064】

30

式(9)は、スタイルが ρ 番目の対象スタイルである場合において (q', b') の直後に (q, b) が出現する確率すなわち (q', b') から (q, b) へ遷移する確率を表わしている。したがって、 $Q \times B \times Q \times B$ 個の要素からなる確率行列が N_M 個、集まったものである。

【0065】

ところで、これらの楽曲データ60のそれぞれに示される楽曲(メロディー)の音階(調)がすべて同じであれば、自動音楽生成装置1は、音階を揃える処理を行うことなく後述の方法で機械学習を行うことができる。または、それぞれの音階が異なっても、それぞれの音階が分かっている場合は、自動音楽生成装置1は、すべての楽曲を共通のトニックの音階に移調し、後述の方法で機械学習を行う。

【0066】

40

しかし、音階が揃っておらず、かつ、音階が分からない場合は、楽曲データ60をそのまま使用して機械学習を行うと、精度のよい学習済モデルが得られないことがある。

【0067】

そこで、自動音楽生成装置1は、楽曲の調が分からない場合であっても機械学習を行うことができるように、トニックの自由度を持った移調対称マルコフモデルを使用する。なお、式(7)～(9)が表わすような、トニックの自由度のないマルコフモデルは、「単純マルコフモデル」と呼ばれる。

【0068】

移調対称マルコフモデルは、単純マルコフモデルと同じ初期確率 $\pi_{ini}(q, b)$ および遷移確率 $\pi(q, b | q', b')$ を有し、さらに、トニック確率 $P(k)$ を有する。「 k 」はトニッ

50

ク変数であって、 $k \in \{0, 1, \dots, 11\} = \{C, C\#, \dots, B\}$ である。ただし、楽曲を移調しても生成確率が変わらないという移調対称性を仮定する。すると、次の式(10)、式(11)、および式(12)が成り立つ。

$$P(k) = 1/12 \dots (10)$$

$$\pi^{\text{ini}}(q, b | k) = \pi^{\text{ini}}(q - k, b) \dots (11)$$

$$\pi(q, b | q', b', k) = \pi(q - k, b | q' - k, b') \dots (12)$$

このように、式(11)および式(12)それぞれの右辺に示すパラメータは、単純なマルコフモデルのパラメータである。以下、それぞれのパラメータを「ベース初期確率」および「ベース遷移確率」と記載する。また、ベース初期確率およびベース遷移確率を有する単純マルコフモデルを「ベースマルコフモデル」と記載する。

【0069】

なお、トニック変数 k および q, b 系列(式(4)で表わされる系列)の同時生成確率とベースマルコフモデルとの関係は、次の式(13)の通りである。

【0070】

【数6】

$$P(k, (q_i, b_i)_{i=1}^I) = \frac{1}{12} \pi^{\text{ini}}(q_1, b_1 | k) \prod_{i=2}^I \pi(q_i, b_i | q_{i-1}, b_{i-1}, k) \dots (13)$$

【0071】

さらに、トニック変数 k が与えられた場合の q, b 系列の条件付き生成確率は、次の式(14)の通りである。

【0072】

【数7】

$$P((q_i, b_i)_{i=1}^I | k) = \pi^{\text{ini}}(q_1, b_1 | k) \prod_{i=2}^I \pi(q_i, b_i | q_{i-1}, b_{i-1}, k) \dots (14)$$

【0073】

そして、式(13)および式(14)より、次の式(15)が成り立つ。

【0074】

【数8】

$$P(k, (q_i, b_i)_{i=1}^I) = \frac{1}{12} P((q_i, b_i)_{i=1}^I | k) \dots (15)$$

【0075】

さらに、式(10)、式(11)、および式(12)にスタイル変数 p を、それぞれ $P(p, k)$ 、 $\pi^{\text{ini}}(q, b | p, k)$ 、および $\pi(q, b | q', b', p, k)$ のように導入することによって、 N_M 個の移調対称マルコフモデルを混合させる。そして、移調対称性を仮定すると、これらの式を式(16)、式(17)、および式(18)のように書き換えることができる。

$$P(p, k) = (1/12) P(p) = (1/12) \theta(p) \dots (16)$$

$$\pi^{\text{ini}}(q, b | p, k) = \pi^{\text{ini}}(q - k, b | p) \dots (17)$$

$$\pi(q, b | q', b', p, k) = \pi(q - k, b | q' - k, b', p) \dots (18)$$

各スタイル(スタイル変数 p が 1、...、 N_M であるそれぞれのスタイル)に対するこの移調対称マルコフモデル(移調混合対称マルコフモデル)も、ベース初期確率およびベース遷移確率を有する単純マルコフモデルなので、ベースマルコフモデルである。なお、トニック変数 k および q, b 系列の同時生成確率とこのベースマルコフモデルとの関係は、次

の式 (1 9) の通りである。

【 0 0 7 6 】

【 数 9 】

$$P(\rho, k, (q_i, b_i)_{i=1}^I) = \frac{1}{12} \theta(\rho) \pi^{\text{ini}}(q_1, b_1 | \rho, k) \prod_{i=2}^I \pi(q_i, b_i | q_{i-1}, b_{i-1}, \rho, k) \quad \cdots (19)$$

【 0 0 7 7 】

10

後半の項は、トニック変数が与えられた場合の、q b 系列の条件付生成確率である。この条件付生成確率は、次の式 (2 0) ~ (2 1) のように表わされる。

【 0 0 7 8 】

【 数 1 0 】

$$P((q_i, b_i)_{i=1}^I | \rho, k) = \pi^{\text{ini}}(q_1, b_1 | \rho, k) \prod_{i=2}^I \pi(q_i, b_i | q_{i-1}, b_{i-1}, \rho, k) \quad \cdots (20)$$

$$= \pi^{\text{ini}}(q_1 - k, b_1 | \rho) \prod_{i=2}^I \pi(q_i - k, b_i | q_{i-1} - k, b_{i-1}, \rho) \quad \cdots (21)$$

20

【 0 0 7 9 】

これを用いると、式 (1 9) の左辺に示す同時生成確率は、次の式 (2 2) のように表わされる。

【 0 0 8 0 】

【 数 1 1 】

$$P(\rho, k, (q_i, b_i)_{i=1}^I) = \frac{1}{12} \theta(\rho) P((q_i, b_i)_{i=1}^I | \rho, k) \quad \cdots (22)$$

30

【 0 0 8 1 】

そして、スタイル変数確率、ベース初期確率、およびベース遷移確率の3つのパラメータとして、式 (1 6)、式 (1 7)、および式 (1 8) それぞれの右辺の値を有するベースマルコフモデルが、学習用のベースマルコフモデルとして用いられる。

【 0 0 8 2 】

ところで、学習済の移調対称マルコフモデルが存在する場合は、図3のトニック推定部213は、楽曲データ60に示される楽曲のq b 系列の調(トニック)を次のように推定することができる。

40

【 0 0 8 3 】

ベイズの公式により、次の式 (2 3) が成り立つ。

【 0 0 8 4 】

【 数 1 2 】

$$P(k | (q_i, b_i)_{i=1}^I) = P(k, (q_i, b_i)_{i=1}^I) / P((q_i, b_i)_{i=1}^I) \quad \cdots (23)$$

【 0 0 8 5 】

式 (2 3) の右辺の分母は k に依らないので、次の式 (2 4) が成り立つ。

【 0 0 8 6 】

【 数 1 3 】

$$P(k|(q_i, b_i)_{i=1}^I) \propto P(k, (q_i, b_i)_{i=1}^I) \propto P((q_i, b_i)_{i=1}^I | k) \quad \cdots (24)$$

【 0 0 8 7 】

10

ここで、トニック変数条件付き確率は、式 (1 1)、式 (1 2)、式 (1 4)、および式 (1 5) より、次の式 (2 5) 式のように表わすことができる。

【 0 0 8 8 】

【 数 1 4 】

$$P((q_i, b_i)_{i=1}^I | k) = \pi^{ini}(q_1 - k, b_1) \prod_{i=2}^I \pi(q_i - k, b_i | q_{i-1} - k, b_{i-1}) \quad \cdots (25)$$

20

【 0 0 8 9 】

左辺がトニック変数条件付き確率である。式 (2 5) によると、 q, b 系列をトニック変数 k だけ移調して、ベース初期確率およびベース遷移確率を用いて計算した生成確率と見なせる。よって、与えられた系列のトニックがトニック変数 k に対応する音高クラスである確率は、この系列をトニック変数 k だけ移調した時のベースマルコフモデルに対する生成確率に比例する。そこで、この確率を最大にする k を求めることによって、この系列の最も尤もらしいトニックを推定することができる。

【 0 0 9 0 】

トニックを推定することができる場合は、式 (1 6) ~ (1 8) それぞれの右辺の値をパラメータとして有するベースマルコフモデルの代わりに、式 (7) ~ (9) それぞれの値をパラメータとして有するベースマルコフモデルを、学習用のベースマルコフモデルとして用いることができる。

30

【 0 0 9 1 】

以下、式 (7) ~ (9) それぞれの値をパラメータとして有するベースマルコフモデルを学習用のモデル 6 1 (図 3 参照) として用いる場合を例に説明する。

【 0 0 9 2 】

〔 3 . 1 . 4 学習用のモデルの準備 〕

初期化処理部 2 1 2 は、対象スタイルごとのモデル 6 1 のスタイル変数確率 $\theta(p)$ 、ベース初期確率 $\pi^{ini}(q, b | p)$ 、およびベース遷移確率 $\pi(q, b | q', b', p)$ の各パラメータの初期値をランダムに設定する。

40

【 0 0 9 3 】

スタイル変数確率、ベース初期確率、およびベース遷移確率は、上述の通り、E ステップの処理およびM ステップの処理が交互に繰り返されることによって調整される。以下、 t 回目のE ステップの処理およびM ステップの処理の直後のスタイル変数確率、ベース初期確率、およびベース遷移確率それぞれの値を次の式 (2 6)、式 (2 7)、および式 (2 8) のように表わす。

【 0 0 9 4 】

【 数 1 5 】

50

$$\theta_t(\rho) \cdots (26)$$

$$\pi_t^{\text{ini}}(q-k, b | \rho) \cdots (27)$$

$$\pi_t(q-k, b | q'-k, b', \rho) \cdots (28)$$

【0095】

なお、 $t = 0$ のときの式(26)、式(27)、および式(28)の各値がスタイル変数確率、ベース初期確率、およびベース遷移確率の初期値である。また、式(26)、式(27)、および式(28)をそれぞれ θ_t 、 π_t^{ini} 、および π_t と記載する。このように、本実施形態では、ベクトルの記号として下線を用いる。

10

【0096】

〔3.1.5 EMアルゴリズムによる処理〕

事後確率算出部214は、EMアルゴリズムのうちのEステップの処理を次のように行う。具体的には、次の式(29)によって表わされる確率を算出する。

【0097】

【数16】

$$P(\rho_h, k_h | (q_i^{(h)}, b_i^{(h)})_{i=1}^{l_h}, \underline{\theta}_t, \underline{\pi}_t^{\text{ini}}, \underline{\pi}_t) \cdots (29)$$

20

【0098】

「 ρ_h 」および「 k_h 」は、それぞれ、 h 番目の楽曲のスタイル変数およびトニック変数である。

【0099】

この確率は、モデルパラメータの値に依存するので、 t ステップ目の各パラメータが変数として記されている。これは、次の式(30)に従って計算することができる。

30

【0100】

【数17】

$$P(\rho, k | (q_i, b_i)_{i=1}^l) \propto P(\rho, k, (q_i, b_i)_{i=1}^l) \propto \theta(\rho) P((q_i, b_i)_{i=1}^l | \rho, k) \cdots (30)$$

【0101】

なお、式(30)は、次のように導かれたものである。まず、学習済の移調対称マルコフモデルがある場合に、与えられた楽曲の q, b 系列(式(4)参照)に合致するスタイルおよびトニックは、次の式(31)で表わされる条件付き確率を計算することによって推定することができる。

40

【0102】

【数18】

$$P(\rho, k, (q_i, b_i)_{i=1}^l) \cdots (31)$$

【0103】

この条件付き確率を最大にする ρ および k が、この楽曲に最も合致するスタイルおよび

50

トニックである。さらに、ベイズの公式により、次の式(32)のように表わすことができる。

【0104】

【数19】

$$P(\rho, k | (q_i, b_i)_{i=1}^I) = P(\rho, k, (q_i, b_i)_{i=1}^I) / P((q_i, b_i)_{i=1}^I) \quad \cdots (32)$$

【0105】

10

式(32)の右辺の分母は、スタイル変数 ρ およびトニック変数 k に依らない。したがって、上記の式(30)が成り立つ。なお、スタイルおよびトニックを推定する方法は、後述する原曲属性推定部256においても用いられる。

【0106】

パラメータ更新部215は、EMアルゴリズムのうちのMステップの処理を行う。具体的には、次の式(33)～(35)式に従ってベースマルコフモデルの各パラメータを更新する。

【0107】

【数20】

20

$$\theta_{t+1}(\rho) \propto \sum_{h=1}^H \sum_{k_h=0}^{11} P(\rho_h = \rho, k_h | (q_i^{(h)}, b_i^{(h)})_{i=1}^{I_h}, \underline{\theta}_t, \underline{\pi}_t^{ini}, \underline{\pi}_t) \quad \cdots (33)$$

$$\pi_{t+1}^{ini}(q, b | \rho)$$

$$\propto \sum_{h=1}^H \sum_{k_h=0}^{11} \delta_{q_i^{(h)}(q+k)} \delta_{b_i^{(h)}b} P(\rho_h = \rho, k_h | (q_i^{(h)}, b_i^{(h)})_{i=1}^{I_h}, \underline{\theta}_t, \underline{\pi}_t^{ini}, \underline{\pi}_t) \quad \cdots (34)$$

30

$$\pi_{t+1}(q, b | q', b', \rho)$$

$$\propto \sum_{h=1}^H \sum_{k_h=0}^{11} \sum_{i=2}^{I_h} \delta_{q_i^{(h)}(q+k)} \delta_{b_i^{(h)}b} \delta_{q_{i-1}^{(h)}(q'+k)} \delta_{b_{i-1}^{(h)}b} P(\rho_h = \rho, k_h | (q_i^{(h)}, b_i^{(h)})_{i=1}^{I_h}, \underline{\theta}_t, \underline{\pi}_t^{ini}, \underline{\pi}_t) \quad \cdots (35)$$

【0108】

δ_{ij} は、はクロネッカーデルタである。つまり、 $i = j$ であれば $\delta_{ij} = 1$ であり、 $i \neq j$ であれば $\delta_{ij} = 0$ である。式(33)～(35)それぞれの比例係数は、それぞれの確率値の総和が「1」になるように選択される。

40

【0109】

収束判定部216は、事後確率算出部214およびパラメータ更新部215がそれぞれ1回、Eステップの処理およびMステップの処理を行うごとに、パラメータの変化が収束したか否かを判定する。例えば、 $\underline{\theta}_t$ に対する $\underline{\theta}_t$ と $\underline{\theta}_{t-1}$ との差の絶対値の比が所定の値(例えば、 10^{-6})以下になったら、収束したと判定する。または、 $\underline{\pi}_t^{ini}$ に対する $\underline{\pi}_t^{ini}$ と $\underline{\pi}_{t-1}^{ini}$ との差の絶対値の比が所定の値以下になったら、収束したと判定してもよい。または、 $\underline{\pi}_t$ に対する $\underline{\pi}_t$ と $\underline{\pi}_{t-1}$ との差の絶対値の比が所定の値以下になったら、収束したと判定してもよい。

50

【 0 1 1 0 】

または、収束判定部 2 1 6 は、上記の比に関わらず、E M アルゴリズムを所定の回数（例えば、5 0 回）繰り返したら、収束したと判定してもよい。

【 0 1 1 1 】

そして、スタイルモデル記憶部 2 1 7 は、収束したと収束判定部 2 1 6 によって判定されたら、E M アルゴリズムの処理を終了し、E M アルゴリズムの処理が施された対象スタイルそれぞれのモデル 6 1 を学習済モデル 6 2 として記憶する（図 3 参照）。以下、T 回目収束したと判定された場合を例に説明する。

【 0 1 1 2 】

なお、以上の処理だけでは、オペレータは、どの学習済モデル 6 2 がどの対象スタイルのものであるのか判別することができない。そこで、オペレータは、次に説明する作曲処理をこれらの学習済モデル 6 2 を用いて 1 回ずつ自動音楽生成装置 1 に実行させ、自動作曲された楽曲を聴くことによって、それぞれの学習済モデル 6 2 の対象スタイルを判別する。そして、それぞれの学習済モデル 6 2 に、判別した対象スタイルをラベリングする。例えば、対象スタイルがロックおよび演歌の 2 つである場合は、オペレータは、生成された 2 つの学習済モデル 6 2 それぞれによって自動作曲された楽曲を聴き、ロックおよび演歌それぞれの学習済モデル 6 2 を判別し、ラベリングする。

10

【 0 1 1 3 】

〔 3 . 2 推論のフェーズ〕

図 6 は、作曲処理の流れの例を説明するフローチャートである。推論のフェーズにおいては、新しい楽曲が、図 6 に示す手順で自動的に生成される。

20

【 0 1 1 4 】

自動音楽生成装置 1 は、作曲の条件として音符数 I およびスタイルを受け付けると（図 6 の S 7 1 1）、受け付けたスタイルの学習済モデル 6 2 に基づいて 1 番目の音符を決定する（S 7 1 2、S 7 1 3）。2 ~ I 番目それぞれの音符を、その学習済モデル 6 2 および直前の音符に基づいて決定する（S 7 1 4 ~ S 7 1 6）。そして、1 ~ I 番目それぞれの決定した音符を所定のフォーマットのデータに変換する（S 7 1 7）。以下、図 6 の各処理をさらに具体的に説明する。

【 0 1 1 5 】

図 3 の音符決定部 2 1 8 ないし作曲データ生成部 2 2 0 は、推論器であって、指定されたスタイルの楽曲をそのスタイルの学習済モデル 6 2 に基づいて生成する。

30

【 0 1 1 6 】

学習済モデル 6 2 には、 π_{ini_T} および π_T が含まれている。 π_{ini_T} および π_T には、スタイル（対象スタイル）ごとの初期確率 $\pi_{ini}(q, b)$ および遷移確率 $\pi(q, b | q', b')$ が含まれる。

【 0 1 1 7 】

オペレータは、自動音楽生成装置 1 に生成させたい楽曲のスタイルおよび音符の数 I を自動音楽生成装置 1 に対して指定する。

【 0 1 1 8 】

すると、音符決定部 2 1 8 は、これらの指定を受け付け、楽曲の I 個の音符を次のように決定する。

40

【 0 1 1 9 】

音符決定部 2 1 8 は、I 個の区間 (0 , 1) の一様乱数を発生させる。以下、i 番目の一様乱数を「乱数 r_i 」と記載する。なお、 $i = 1, \dots, I$ である。そして、離散確率分布 $P(x)(x = 1, \dots, X)$ および乱数 r に基づいて x の値を生成（サンプル）する方法で、各音符をサンプルする。つまり、次の式 (3 6) を満たす x を選ぶことによってサンプルする。

【 0 1 2 0 】

【 数 2 1 】

$$\sum_{x'=1}^{x-1} P(x') < r \leq \sum_{x'=1}^x P(x') \quad \cdots (36)$$

【 0 1 2 1 】

具体的には、音符決定部 2 1 8 は、1 番目の音符つまり (q_1, b_1) を、乱数 r_1 および指定されたスタイルの初期確率 $\pi^{ini}(q, b)$ に基づいてサンプルする。次に、2 番目の音符つまり (q_2, b_2) を、乱数 r_2 および指定されたスタイルの遷移確率 $\pi(q, b | q_1, b_1)$ に基づいてサンプルする。以下、同様に、 i 番目の音符を、乱数 r_i および指定されたスタイルの遷移確率 $\pi(q, b | q_{i-1}, b_{i-1})$ に基づいてサンプルする。これにより、 q, b 系列のデータが楽曲のメロディーのデータとして得られる。

10

【 0 1 2 2 】

系列変換部 2 1 9 は、楽曲のメロディーのデータを式 (4) のフォーマットつまり q, b 系列から式 (2) のフォーマットつまり p, s 系列へ次のように変換する。

【 0 1 2 3 】

系列変換部 2 1 9 は、小節線の位置を楽曲のメロディーのデータのうちの各音符の拍節位置つまり次の式 (37) の値より特定する。

【 0 1 2 4 】

20

【 数 2 2 】

$$(b_i)_{i=1}^I \quad \cdots (37)$$

【 0 1 2 5 】

すなわち、系列変換部 2 1 9 は、 b_i, b_{i-1} 、を満たすときの $(i - 1)$ 番目の音符と i 番目の音符との間に小節線があると特定する。そして、小節線の位置に基づいて、式 (37) の値を式 (38) の値に変換する。

【 0 1 2 6 】

30

【 数 2 3 】

$$(s_i)_{i=1}^I \quad \cdots (38)$$

【 0 1 2 7 】

さらに、系列変換部 2 1 9 は、次のように、楽曲のメロディーのデータのうちの音高すなわち式 (39) の値を式 (40) の値に変換する。

【 0 1 2 8 】

【 数 2 4 】

$$(q_i)_{i=1}^I \quad \cdots (39)$$

40

$$(p_i)_{i=1}^I \quad \cdots (40)$$

【 0 1 2 9 】

系列変換部 2 1 9 は、最初の音高 p_1 のオクターブを適当な方法で定める。そして、後続の音符の音高 p_i を、直前の音符の音高 p_{i-1} および音高クラス q_i, q_{i-1} から決定する。具体的には、

50

$q_i - q_{i-1} < -6$ ならば、 $p_i = p_{i-1} + q_i - q_{i-1} + 12$
 $-6 \leq q_i - q_{i-1} < 6$ ならば、 $p_i = p_{i-1} + q_i - q_{i-1}$
 $q_i - q_{i-1} \geq 6$ ならば、 $p_i = p_{i-1} + q_i - q_{i-1} - 12$
 とする。

【0130】

この方法によると、ある音符が想定之音域から外れる場合があるが、この場合は、想定之音域の中に収まるようにオクターブ修正を行えばよい。これにより、想定之音域内に収まる音符列が生成される。

【0131】

作曲データ生成部220は、系列変換部219によって得られた楽曲のps系列のデータをMIDIまたはMusicXMLのフォーマットのファイルに変換することによって、作曲データ63を生成する。作曲データ63は、自動音楽生成装置1の所定の記憶領域に保存される。または、端末装置41などへ送信される。または、ps系列のデータに従って楽曲を音声再生ユニット19（図2参照）によって再生してもよい。

10

【0132】

なお、オクターブ情報が学習済モデルに含まれるように機械学習を行ってもよい。つまり、オクターブ情報に対してもマルコフモデルなどの確率モデルを生成してもよい。すると、推論のフェーズにおいて、オクターブの確率モデルに基づいて式(39)の系列のデータから式(40)の系列のデータを生成することができる。

【0133】

以上説明したように、作曲モジュール21によると、教師なし機械学習によって学習済モデルを生成するので、楽曲データ60ごとにタグ付けすることなく従来よりも容易に自動作曲のAI(Artificial Intelligence)を構築することができる。

20

【0134】

本実施形態では、初期化処理部212は、 $\theta(p)$ の初期値としてランダムな値を設定したが、用意した楽曲の総数に対する対象スタイルごとの楽曲の比率を $\theta(p)$ の初期値として設定してもよい。

【0135】

本実施形態では、初期化処理部212は、各モデル61の各パラメータにランダムな値を与えたが、モデル61ごとに異なる値を与えることができれば、ランダムな値でなくてもよい。例えば、それぞれに異なる所定の値を与えてもよい。

30

【0136】

〔4 編曲モジュール〕

編曲モジュール25は、図3に示すように、初期化処理部251、事後確率算出部252、パラメータ更新部253、収束判定部254、編集モデル記憶部255、原曲属性推定部256、候補楽曲生成部257、最適候補選出部258、および編曲データ生成部259などによって構成され、ある楽曲を、目標のスタイルの楽曲に変換する。つまり、編曲を行う。上述の通り、編曲モジュール25は、編曲学習処理および編曲処理を実行するためのモジュールである。編曲学習処理が機械学習のフェーズに相当し、編曲処理が推論のフェーズに相当する。

40

【0137】

以下、変換対象の楽曲を「原曲」と記載し、原曲のスタイルを「原曲スタイル」と記載する。目標のスタイルを「目標スタイル」と記載し、原曲が目標スタイルに変換されたものを「編曲」と記載する。また、原曲スタイルおよび目標スタイルそれぞれのシーケンス番号を「 p_S 」および「 p_T 」とする。原曲スタイルと目標スタイルとが同じであっても構わない。また、スタイルモデル記憶部217に、 N_M 個のスタイルの学習済モデル62が記憶されているものとする。

【0138】

〔4.1 概要〕

図7は、編曲学習処理の流れの例を説明するフローチャートである。図8は、編曲処理

50

の流れの例を説明するフローチャートである。

【 0 1 3 9 】

図 3 の初期化処理部 2 5 1 ないし収束判定部 2 5 4 は、学習器であって、編集モデルを、図 7 に示す手順で生成する。なお、編集モデルは、推論のフェーズにおいて原曲を編曲するための学習済モデルとして用いられる。

【 0 1 4 0 】

初期化処理部 2 5 1 は、ランダムに初期化した編集モデルを予め用意する (S 7 2 1) 。事後確率算出部 2 5 2 およびパラメータ更新部 2 5 3 は、編集モデルに対して E ステップの処理および M ステップの処理を実行する (S 7 2 2 、 S 7 2 3) 。収束判定部 2 5 4 は、収束の判定を行う (S 7 2 4) 。そして、収束したと判定した場合は (S 7 2 5 で Y e s) 、その時点の編集モデルを学習済モデルとして編集モデル記憶部 2 5 5 に記憶させる (S 7 2 6) 。

10

【 0 1 4 1 】

収束していないと判定された場合は (S 7 2 5 で N o) 、ステップ S 7 2 2 に戻って、E ステップの処理および M ステップの処理が再度、実行される。

【 0 1 4 2 】

このように、編集モデルは、E M アルゴリズムを繰り返し実行されることによって、学習済モデルになる。なお、 $p_S = \{ 1, \dots, N_M \}$ かつ $p_T = \{ 1, \dots, N_M \}$ なので、 $N_M \times N_M$ 個の編集モデルを用意し、それぞれについてステップ S 7 2 2 ~ S 7 2 6 の処理が行われる。

20

【 0 1 4 3 】

原曲属性推定部 2 5 6 ないし編曲データ生成部 2 5 9 は、生成器であって、原曲の編曲を、図 8 に示す手順で生成する。

【 0 1 4 4 】

原曲属性推定部 2 5 6 は、編曲の条件として原曲および目標スタイルなどを受け付けると (図 8 の S 7 3 1) 、原曲スタイルを推定する (S 7 3 2) 。候補楽曲生成部 2 5 7 は、原曲と音符の数が同じである目標スタイルの候補楽曲を複数、作曲モジュール 2 1 によって生成する (S 7 3 2) 。最適候補選出部 2 5 8 は、候補楽曲ごとに編集モデルに基づいて適切度を算出し (S 7 3 3) 、最も適切度の高い候補楽曲を原曲の編曲として選出する (S 7 3 4) 。そして、編曲データ生成部 2 5 9 は、選出した候補楽曲を所定のフォーマットのデータに変換する (S 7 3 5) 。

30

【 0 1 4 5 】

なお、適切度は、 $Q(X \sim | X)$ と表わされ、次の式 (4 1) の通り、候補楽曲の目標スタイルへの合致度および候補楽曲の原曲との類似度の 2 つの尺度によって算出される。

【 0 1 4 6 】

【数 2 5 】

$$Q(X \sim | X) := P_L(X \sim)^{\alpha_L} P_E(X | X \sim)^{\alpha_E} \quad \dots (41)$$

40

【 0 1 4 7 】

「 X 」は原曲の系列であって次の式 (4 2) のように表わされ、「 $X \sim$ 」は候補楽曲の系列であって次の式 (4 3) のように表わされる。

【 0 1 4 8 】

【数 2 6 】

$$X=(q_i, b_i)_{i=1}^I \quad \cdots (42)$$

$$X^{\sim}=(q_i^{\sim}, b_i^{\sim})_{i=1}^I \quad \cdots (43)$$

【0149】

「 $P_L(X^{\sim})$ 」は、候補楽曲の目標スタイルへの合致度を表わす確率である。「 $P_E(X | X^{\sim})$ 」は、原曲と候補楽曲との類似度を表わす確率である。「 α_L 」および「 α_E 」は、合致度と類似度とのバランスを調整するためのパラメータであって、目的に応じてオペレータによって指定される。

10

【0150】

適切度、合致度、および類似度は、原曲スタイルまたは目標スタイルに依るので、式(41)を次の式(44)のように記述することができる。

$$Q(X^{\sim} | X; \rho_S, \rho_T) := P_L(X^{\sim}; \rho_T) \alpha^L P_E(X | X^{\sim}; \rho_S, \rho_T) \alpha^E \quad \cdots (44)$$

〔4.2 編集モデル〕

上の式(41)の右辺を構成する因子のうち、 $P_E(X | X^{\sim}; \rho_S, \rho_T)$ に編集モデルが用いられる。 $P_E(X | X^{\sim}; \rho_S, \rho_T)$ は、次の式(45)のように記述される。

【0151】

20

【数27】

$$\begin{aligned} P_E(X | X^{\sim}; \rho_S, \rho_T) \\ = \frac{1}{Z} P_F(X | X^{\sim}; \rho_S, \rho_T) \prod_{i=1}^I \left[\exp\left(-\frac{(q_i^{\sim} - q_i)^2}{2\sigma_p^2}\right) \exp\left(-\frac{(b_i^{\sim} - b_i)^2}{2\sigma_r^2}\right) \right] \quad \cdots (45) \end{aligned}$$

【0152】

右辺の2番目の因子は原曲および候補楽曲それぞれのi番目の音高クラスの二乗距離であり、 σ_p は、距離尺度の基準を与えるパラメータである。右辺の3番目の因子は原曲の系列Xおよび候補楽曲の系列 $X^{\sim}$ それぞれのi番目の拍節位置の二乗距離であり、 σ_r は、距離尺度の基準を与えるパラメータである。本実施形態では、 $\sigma_p = 0.7$ 、 $\sigma_r = 3$ がデフォルトで設定されている。これら2つの因子は、すべての ρ_S および ρ_T に対して一定である。つまり、 ρ_S および ρ_T に依らない。

30

【0153】

右辺の最初の因子 $P_F(X | X^{\sim}; \rho_S, \rho_T)$ は、音高クラス系列の統語構造に基づいた類似度を表わし、音符の統語機能が潜在変数で表現されるように隠れマルコフモデル(HMM: Hidden Markov Model)を用いて記述される。「Z」は、確率分布の規格化係数であるが、候補楽曲の生成には用いられない。

40

【0154】

ところで、HMMは、自然言語の単語列からその統語機能を表わす品詞を教師なしで学習するために用いられている(S. Goldwater and T. Griffiths, "A Fully Bayesian Approach to Unsupervised Part-of-Speech Tagging," in Proc. ACL, pp. 744-751, 2007. 参照)。音楽においても、コード記号の系列からその統語機能を表わす和声機能を学習するために用いられている(H. Tsushima et al., "Generative statistical models with self-emergent grammar of chord sequences," J. New Music Res., vol. 47, no. 3, pp. 226-248, 2018. 参照)。本実施形態では、音高の統語機能を学習するためにHMMが用いられる。

【0155】

50

また、HMMは、一般に、マルコフモデルに従って生成される潜在変数の系列と、この系列から観測される記号列とを生成するモデルである。本実施形態においては、観測される記号列は、音高クラスの系列 $\underline{q}$ および系列 $\underline{q}^{\sim}$ であってそれぞれ次の式(46)および式(47)のように表わされる。

【0156】

【数28】

$$\underline{q} = (q_i)_{i=1}^I \quad \dots (46)$$

10

$$\underline{q}^{\sim} = (q_i^{\sim})_{i=1}^I \quad \dots (47)$$

【0157】

潜在変数 z は、 $\{1, \dots, N_F\}$ の中の値を取る離散確率変数であり、音符ごとに1つずつ生成される。つまり、潜在変数 z_i は、 i 番目の音符の離散確率変数である。「 N_F 」は、仮定する統語機能の数であり、目的に応じて決められる変数である。本実施形態では、 $N_F = 7$ とする。

【0158】

20

そして、本実施形態において編集モデルとして用いられるHMMは、次の式(48)～(51)のそれぞれの左辺に示されるパラメータを有する。

$$\psi_{ini}(z_1) = P(z_1) \quad \dots (48)$$

$$\psi(z_i | \underline{q}_{-1}) = P(z_i | \underline{q}_{-1}) \quad \dots (49)$$

$$\phi(q | z, \underline{p}) = P(q_i = q | \underline{q} = z) \quad \dots (50)$$

$$\phi(q^{\sim} | z, \underline{p}) = P(q_i^{\sim} = q^{\sim} | \underline{q} = z) \quad \dots (51)$$

式(48)の $\psi_{ini}(z_1)$ は、潜在変数 z の初期確率を表わす、 N_F 個の確率の確率ベクトルである。式(49)は、潜在変数 z の遷移確率を表わす、 $N_F \times N_F$ 個の要素を持つ確率行列である。

【0159】

30

式(50)の $\phi(q | z, \underline{p})$ は、原曲スタイルにおける、音高クラスごとの(つまり、 Q 個の)出力確率の確率ベクトルを N_F 個分、表わしている。式(51)は、目標スタイルにおける、音高クラスごとの出力確率の確率ベクトルを N_F 個分、表わしている。式(50)と式(51)とを比較して分かるように、音高クラス q_i および音高クラス $q_i^{\sim}$ は、各潜在変数 z から、互いに異なる離散確率分布に従って生成される。

【0160】

また、系列 $\underline{q}$ 、系列 $\underline{q}^{\sim}$ 、および $\underline{z}$ について、次の式(52)～(54)の関係が成り立つ。

【0161】

【数29】

40

$$P(\underline{z}) = \psi^{\text{ini}}(z_1) \prod_{i=2}^I \psi(z_i | \psi_{i-1}) \quad \cdots (5.2)$$

$$P(\underline{q} | \underline{z}) = \prod_{i=1}^I \phi(q_i | z_i, \rho_S) \quad \cdots (5.3)$$

10

$$P(\underline{q}^{\sim} | \underline{z}) = \prod_{i=1}^I \phi(q_i^{\sim} | z_i, \rho_T) \quad \cdots (5.4)$$

【 0 1 6 2 】

パラメータが定まっている場合に、このHMMを用いて因子 $P_F(X | X^{\sim}; \rho_S, \rho_T)$ を計算する手順は以下の通りである。この因子をHMMに基づいて、 $P_F(X | X^{\sim}; \rho_S, \rho_T) = P(\underline{q} | \underline{q}^{\sim})$ と定めることにより次の表現が得られる。

20

【 0 1 6 3 】

【 数 3 0 】

$$P_F(X | X^{\sim}; \rho_S, \rho_T) \quad \cdots (5.5)$$

$$= P(\underline{q} | \underline{q}^{\sim}) = \frac{1}{Z'} P(\underline{q}, \underline{q}^{\sim}) \quad \cdots (5.6)$$

30

$$= \frac{1}{Z'} \sum_{z_1=1}^{N_F} \cdots \sum_{z_I=1}^{N_F} P(\underline{q} | \underline{z}) P(\underline{q}^{\sim} | \underline{z}) P(\underline{z}) \quad \cdots (5.7)$$

$$= \frac{1}{Z'} \sum_{z_1=1}^{N_F} \cdots \sum_{z_I=1}^{N_F} \phi(q_1 | z_1, \rho_S) \phi(q_1^{\sim} | z_1, \rho_T) \psi^{\text{ini}}(z_1)$$

40

$$\prod \phi(q_i | z_i, \rho_S) \phi(q_i^{\sim} | z_i, \rho_T) \psi^{\text{ini}}(z_i | z_{i-1}) \quad \cdots (5.8)$$

【 0 1 6 4 】

Z' は規格化定数であるが、編曲の生成には用いられない。式 (5.8) により左辺の因子が原始的に求まるが、すべての和記号を同時に計算することは組合せの数が多すぎて困難である。そこで、動的計画法（フォワードアルゴリズム）を用いて逐次計算することにより効果的に計算することができる。

50

【 0 1 6 5 】

フォワードアルゴリズムの手順は、次の通りである。まず、各 $i \in \{1, \dots, I\}$ に対してフォワード変数を次のように定義する。

$$F_i(z_i) := P(q_{1:i}, q_{\sim 1:i}, z_i) \dots (59)$$

$q_{1:i} = (q_1, \dots, q_i)$ である。以下、同様の記法を用いる。 $F_i(z_i)$ は、 $i = 1$ から順に以下の通り逐次的に計算する。

【 0 1 6 6 】

【数 3 1】

$$F_1(z_1) = P(q_1, q_{\sim 1}, z_1) = P(q_1|z_1) P(q_{\sim 1}|z_1) P(z_1) \quad 10$$

$$= \phi(q_1|z_1, \rho_S) \phi(q_{\sim 1}|z_1, \rho_T) \psi^{\text{ini}}(z_1) \dots (60)$$

$$F_{i+1}(z_{i+1}) = P(q_{1:(i+1)}, q_{\sim 1:(i+1)}, z_{i+1})$$

$$= \sum_{z_i=1}^{N_F} P(q_{1:(i+1)}, q_{\sim 1:(i+1)}, z_{i+1}, z_i) \dots (61) \quad 20$$

$$= \sum_{z_i=1}^{N_F} P(q_{i+1}|z_{i+1}) P(q_{\sim i+1}|z_{i+1}) P(z_{i+1}|z_i) P(q_{1:i}, q_{\sim 1:i}, z_i) \dots (62)$$

$$= \sum_{z_i=1}^{N_F} \phi(q_{i+1}|z_{i+1}, \rho_S) \phi(q_{\sim i+1}|z_{i+1}, \rho_T) \psi(z_{i+1}|z_i) F_i(z_i) \quad 30$$

$$\dots (63)$$

【 0 1 6 7 】

そして、次の式 (64) を用いて確率 $P(\underline{q}, \underline{q}_{\sim})$ が得られる。

【 0 1 6 8 】

【数 3 2】

40

$$P(\underline{q}, \underline{q}_{\sim}) = \sum_{z_I=1}^{N_F} P(q_{1:I}, q_{\sim 1:I}, z_I) = \sum_{z_I=1}^{N_F} F_I(z_I) \dots (64)$$

【 0 1 6 9 】

〔 4 . 3 機械学習のフェーズ〕

上述の通り、 $\rho_S \in \{1, \dots, N_M\}$ かつ $\rho_T \in \{1, \dots, N_M\}$ なので、 $N_M \times N_M$ 通りの

50

編集モデルが生成される。以下、機械学習の内容を、ロックから演歌に編曲するための編集モデルを生成する場合を例に説明する。

【 0 1 7 0 】

初期化処理部 2 5 1 は、機械学習のためのデータを次のように準備する。学習に使用する系列 $\underline{x}$ および系列 $\underline{x} \sim$ を生成する。

【 0 1 7 1 】

系列 $\underline{x}$ は、ロックの音高クラスの系列であって、 $(x_1, \dots, x_N)$ である。「 N 」は、機械学習に必要な音符の数であって、十分に大きな整数である。初期化処理部 2 5 1 は、作曲モジュール 2 1 の推論のフェーズと同様に、ロックの学習済モデル 6 2 に基づいてロックの楽曲を作曲する方法で系列 $\underline{x}$ を生成する。

10

【 0 1 7 2 】

系列 $\underline{x} \sim$ は、演歌の音高クラスの系列であって、 $(x_1, \dots, x_{N'})$ である。初期化処理部 2 5 1 は、演歌の学習済モデル 6 2 に基づいて演歌の楽曲を作曲する方法で系列 $\sim$ を生成する。

【 0 1 7 3 】

さらに、初期化処理部 2 5 1 は、式 (4 8) ~ (5 1) それぞれの左辺に示した 4 つのパラメータ $\psi_{ini}(z_1)$ 、 $\psi(z_i | z_{i-1})$ 、 $\phi(q | z, \beta)$ 、および $\phi(q \sim | z, p)$ を有する編集モデル 6 4 を生成し、これらのパラメータを初期化する。ランダムな値を与えることによって初期化してもよいが、統語機能が音階音と類似した構造を表すことを期待して、 $\phi_0(q | z, \beta)$ は確率行列の対角に近い成分が大きな値を持つように初期化するのが望ましい。具体的には次の式 (6 5) のように初期化する。ただし、比例係数は適切に正規化されるように定める。

20

$$\phi_0(q | z, \beta) = \exp(-(q - 1/2 z / N_F)^2 / 2) \dots \dots (65)$$

また、 $\psi(z_i | z_{i-1})$ および $\phi(q \sim | z, p)$ は、一様分布になるように初期化する。

【 0 1 7 4 】

事後確率算出部 2 5 2 およびパラメータ更新部 2 5 3 は、HMM (Hidden Markov Model) の教師なし学習方法として一般的である Baum と Welch の方法 (C. M. Bishop, Pattern Recognition and Machine Learning, Springer, 2006. 参照) によって、編集モデルを生成する。Baum と Welch の方法は、E M アルゴリズムの一種であり、E ステップでは潜在変数の事後確率を算出し、M ステップではパラメータの更新を行う。E ステップの処理および M ステップの処理が交互に繰り返される。

30

【 0 1 7 5 】

事後確率算出部 2 5 2 は、E ステップの処理として、次の式 (6 6) および式 (6 7) でそれぞれ定義する潜在変数の t ステップ目の事後確率 $y_t(z_i)$ および $z_t(z_{i-1}, z_i)$ を計算する。

$$y_t(z_i) = P(z_i | \underline{x}, \underline{x} \sim, Q_t) \dots \dots (66)$$

$$z_t(z_{i-1}, z_i) = P(z_{i-1}, z_i | \underline{x}, \underline{x} \sim, Q_t) \dots \dots (67)$$

「 Θ 」は、4 つのパラメータのうち $\psi(z_i | z_{i-1})$ 、 $\phi(q \sim | z, \beta)$ 、および $\phi(q | z, p)$ の集合を表わす。「 Θ_t 」は、 t ステップ目における、これら 3 つのパラメータの集合である。以下、これらの式の計算方法を説明する。

40

【 0 1 7 6 】

まず、次の式 (6 8) および式 (6 9) でそれぞれ定義するフォワード変数 $A_t(z_i)$ およびバックワード変数 $B_t(z_i)$ を計算する。

$$A_t(z_i) := P(x_{1:i}, x_{\sim 1:i}, z_i | Q) \dots \dots (68)$$

$$B_t(z_i) := P(x_{(i+1):N}, x_{\sim (i+1):N} | z_i, \Theta_t) \dots \dots (69)$$

これらは、次の式 (7 0) ~ (7 3) に手順に従って再帰的に計算することができる。

【 0 1 7 7 】

【数 3 3】

$$A_t(z_1) = \phi_t(x_1|z_1, \rho_S) \phi_t(\tilde{x}_1|z_1, \rho_T) \psi_t^{\text{ini}}(z_1) \quad \cdots (70)$$

$$A_t(z_{i+1}) = \sum_{z_i} \phi_t(x_{i+1}|z_{i+1}, \rho_S) \phi_t(\tilde{x}_{i+1}|z_{i+1}, \rho_T) \psi_t(z_{i+1}|z_i) A_t(z_i) \quad \cdots (71)$$

$$B_t(z_N) = 1 \quad \cdots (72)$$

10

$$B_t(z_{i-1}) = \sum_{z_i} \phi_t(x_i|z_i, \rho_S) \phi_t(\tilde{x}_i|z_i, \rho_T) \psi_t(z_i|z_{i-1}) B_t(z_i) \quad \cdots (73)$$

【0178】

さらに、計算された $A_t(z_i)$ およびバックワード変数 $B_t(z_i)$ を用いて、事後確率を次の式(74)および式(75)に従って計算する。

20

$$\gamma_t(z_i) = A_t(z_i) B_t(z_i) \quad \cdots (74)$$

$$\xi_t(z_{i-1}, z_i) = \frac{A_t(z_{i-1}) \psi_t(z_i|z_{i-1}) \phi_t(x_i|z_i, \rho_S) \phi_t(\tilde{x}_i|z_i, \rho_T) B_t(z_i)}{A_t(z_i) B_t(z_i)} \quad \cdots (75)$$

ただし、比例係数は、各フォワード変数が潜在変数 z_i に関して、足して1になるように適切に与える。

【0179】

なお、初期確率すなわち $\psi^{\text{ini}}(z_1)$ は、学習に用いる系列が1ペアのみであるので、学習を行わない。初期確率の算出方法については、後述する。

【0180】

パラメータ更新部253は、Mステップの処理を実行する。具体的には、次の式(76)～(78)に従って編集モデルの各パラメータを更新する。この際に、各式の比例係数を、それぞれの確率値の合計が「1」になるように選択する。

30

【0181】

【数34】

$$\psi_{t+1}(z|z') \propto \sum_{i=2}^N \xi_t(z_{i-1} = z', z_i = z) \quad \cdots (76)$$

40

$$\phi_{t+1}(q|z, \rho_S) \propto \sum_{i=1}^N \delta_{qx_i} \gamma_t(z_i = z) \quad \cdots (77)$$

$$\phi_{t+1}(q^{\sim}|z, \rho_T) \propto \sum_{i=1}^N \delta_{q^{\sim}x_i^{\sim}} \gamma_t(z_i = z) \quad \cdots (78)$$

50

【 0 1 8 2 】

収束判定部 2 5 4 は、E ステップの処理および M ステップの処理を繰り返しても編集モデルの各パラメータがほとんど変化しなくなった場合に、各パラメータが理想的な値に収束したと判定する。そして、収束したと判定された時点の編集モデル 6 4 を学習済モデル 6 5 として編集モデル記憶部 2 5 5 に記憶させる。または、E M アルゴリズムを所定の回数だけ繰り返したら、収束したと判定してもよい。

【 0 1 8 3 】

以上、ロックから演歌に編曲するための学習済モデル 6 5 (編集モデル) を生成する方法の例を説明したが、他の方向に編曲するための学習済モデル 6 5 も同様の方法によって生成される。

10

【 0 1 8 4 】

〔 4 . 4 推論のフェーズ 〕

原曲属性推定部 2 5 6 ないし編曲データ生成部 2 5 9 は、指定された原曲を、指定されたスタイルに編曲する。

【 0 1 8 5 】

オペレータは、複数の楽曲の中から編曲したい楽曲を原曲として指定するとともに、どのスタイルに編曲したいかを指定する。

【 0 1 8 6 】

すると、原曲属性推定部 2 5 6 は、その原曲のスタイルおよびトニックを推定する。推定の方法は、前に式 (3 0)、式 (3 2) を用いて説明した通りである。

20

【 0 1 8 7 】

つまり、原曲の系列 X (式 (4 2) 参照) に合致するスタイルおよびトニックは、式 (3 2) の左辺の条件付き確率 P を最大にする p および k を求めることによって推定することができる。

【 0 1 8 8 】

候補楽曲生成部 2 5 7 は、指定されたスタイルの楽曲を多数、候補楽曲として生成する。生成する方法は、音符決定部 2 1 8 が楽曲を生成する方法と同様である。つまり、式 (3 6) を満たす x を選出することによって各音符をサンプルする。

【 0 1 8 9 】

ただし、ここでは、「 I 」は、原曲の音符数である。また、指定されたスタイルの学習済モデル 6 2 が用いられる。また、1 2 個のトニック変数すべてについて、候補楽曲の生成が実行される。

30

【 0 1 9 0 】

最適候補選出部 2 5 8 は、候補楽曲生成部 2 5 7 によって生成された多数の候補楽曲の中から原曲に最も類似するものを編曲として選出する。具体的には、式 (4 4) の左辺すなわち $Q(X \sim | X; \rho, \rho_T)$ を最大にする候補楽曲を編曲として選出する。ただし、「 ρ_S 」は、原曲スタイルのスタイル変数であり、「 ρ_T 」は、オペレータが指定したスタイル (目標スタイル) のスタイル変数である。したがって、この原曲スタイルおよびこの目標スタイルに対応する学習済モデル 6 5 が選出のために用いられる。

【 0 1 9 1 】

40

ところで、原曲の音符数が多いと、候補楽曲の生成可能なパターン数が膨大になる。そこで、いわゆる部分サンプリング法によって編曲を生成してもよい。すなわち、候補楽曲生成部 2 5 7 は、原曲を複数の部分 (例えば、小節) に分け、部分ごとに多数の系列を生成する。そして、最適候補選出部 2 5 8 は、部分ごとに、生成した系列の中から最も類似するものを選出し、選出した系列を連結することによって編曲を生成する。

【 0 1 9 2 】

または、動的計画法によって編曲を生成してもよい。動的計画法によると、 $Q(X \sim | X; \rho_S, \rho_T)$ を最大にする $X \sim$ が、Viterbi アルゴリズムによって探索される。

【 0 1 9 3 】

音高クラスおよび発音時刻がそれぞれ 1 2 および 4 8 なので、 i 番目の音符は 5 7 6 (

50

= 1 2 × 4 8) 通りある。そして、音符の数が I であるので、5 7 6 I 通りの音符の並びがある。

【 0 1 9 4 】

これらの楽曲の中から $Q(X \sim | X; \vartheta, \rho_T)$ を最大にするもの (ビタビ経路) を次のように探索する。 $Q(X \sim | X; \vartheta, \rho_T)$ は正の値なので、 $Q(X \sim | X; \vartheta, \rho_T)$ を最適化したものは、 $Q(X \sim | X; \vartheta, \rho_T)$ の対数をとったものの最適化と等価である。 $Q(X \sim | X; \vartheta, \rho_T)$ の対数は、式 (21)、(44)、(45)、(57) を用いることによって次のように与えられる。

【 0 1 9 5 】

【数 3 5 】

10

$$\begin{aligned} \ln Q(X \sim | X; \rho_S, \rho_T) \\ = \alpha_L \ln P(X \sim; \rho_T) + \alpha_E \ln P_F(X \sim; \rho_S, \rho_T) - \alpha_E \sum_i \left[\frac{(q_i - \tilde{q}_i)^2}{2\sigma_p^2} + \frac{(b_i - \tilde{b}_i)^2}{2\sigma_r^2} \right] \end{aligned}$$

… (79)

$$\begin{aligned} = \alpha_L \{ \ln \pi^{\text{ini}}(\tilde{q}_1, \tilde{b}_1 | \rho_T) + \sum_{i=2}^I \ln \pi(\tilde{q}_i, \tilde{b}_i | \tilde{q}_{i-1}, \tilde{b}_{i-1}, \rho_T) \} \\ + \alpha_E \ln \left(\sum_{z_1} \cdots \sum_{z_I} P(\underline{q} | \underline{z}) P(\underline{q} \sim | \underline{z}) P(\underline{z}) \right) - \alpha_E \sum_i \left[\frac{(q_i - \tilde{q}_i)^2}{2\sigma_p^2} + \frac{(b_i - \tilde{b}_i)^2}{2\sigma_r^2} \right] \\ + \text{const} \end{aligned}$$

… (80)

【 0 1 9 6 】

30

ただし、const は、最適化には関係しない定数項を表わしている。以下、同様である。この関数は、 $\ln P_F(X \sim | X; \vartheta, \rho_T)$ を含んでいるので、直接は、最適化することができない。そこで、入力楽曲 $\underline{q}$ に対して事後確率 $P(\underline{z} | \underline{q}) = P(\underline{q} | \underline{z}) P(\underline{z}) / P(\underline{q})$ を最大とする潜在変数の系列 $\underline{z}^\wedge$ をまず求めて、和をこの系列で置き換える。これにより次の式 (81) が得られる。

【 0 1 9 7 】

【数 3 6 】

$$\begin{aligned} \ln Q(X \sim | X; \rho_S, \rho_T) \\ = \alpha_L \{ \ln \pi^{\text{ini}}(\tilde{q}_1, \tilde{b}_1 | \rho_T) + \sum_{i=2}^I \ln \pi(\tilde{q}_i, \tilde{b}_i | \tilde{q}_{i-1}, \tilde{b}_{i-1}, \rho_T) \} \\ + \alpha_E \sum_{i=1}^I \ln \phi(\tilde{q}_i | z_i^\wedge, \rho_T) - \alpha_E \sum_{i=1}^I \left[\frac{(q_i - \tilde{q}_i)^2}{2\sigma_p^2} + \frac{(b_i - \tilde{b}_i)^2}{2\sigma_r^2} \right] + \text{const} \end{aligned}$$

… (81)

【 0 1 9 8 】

系列 $\underline{z}^\wedge$ は、Viterbi アルゴリズムを用いて以下の手順で求められる。まず、事後確率の

50

最大化は次の式 (8 2) で表わされ、その対数をとったものの最大化と等価である。

【 0 1 9 9 】

【数 3 7 】

$$\ln P(\underline{z} | \underline{q}) = \ln [(P(\underline{q} | \underline{z}) P(\underline{z}) / P(\underline{q}))]$$

$$\ln \pi^{\text{ini}}(z_1) + \sum_{i=2}^i \ln \pi(z_i | z_{i-1}) + \sum_{i=1}^i \ln \phi(q_i | z_i, \rho_s) + \text{const} \quad \cdots (82)$$

10

【 0 2 0 0 】

Viterbi変数を次の式 (8 3) および (8 4) のように定義する。

【 0 2 0 1 】

【数 3 8 】

$$u_i(z_i) = \max_{z_1, \dots, z_{i-1}} [\ln \pi^{\text{ini}}(z_1) + \sum_{i'=2}^i \ln \pi(z_{i'} | z_{i'-1}) + \sum_{i'=1}^i \ln \phi(q_{i'} | z_{i'}, \rho_s)] \quad \cdots (83)$$

20

$$v_i(z_i) = \operatorname{argmax}_{z_1, \dots, z_{i-1}} [\ln \pi^{\text{ini}}(z_1) + \sum_{i'=2}^i \ln \pi(z_{i'} | z_{i'-1}) + \sum_{i'=1}^i \ln \phi(q_{i'} | z_{i'}, \rho_s)] \quad \cdots (84)$$

【 0 2 0 2 】

【数 3 9 】

$$u_1(z_1) = \ln \pi^{\text{ini}}(z_1) + \ln \phi(q_1 | z_1, \rho_s) \quad \cdots (85)$$

30

$$u_{i+1}(z_{i+1}) = \max_{z_i} [u_i(z_i) + \ln \pi(z_{i+1} | z_i) + \ln \phi(q_{i+1} | z_{i+1}, \rho_s)] \quad \cdots (86)$$

$$v_{i+1}(z_{i+1}) = \operatorname{argmax}_{z_i} [u_i(z_i) + \ln \pi(z_{i+1} | z_i) + \ln \phi(q_{i+1} | z_{i+1}, \rho_s)] \quad \cdots (87)$$

【 0 2 0 3 】

式 (8 3) および (8 4) は、次の式 (8 5) ~ (8 7) によって再帰的に計算することができる。

次の式 (8 8) の値は、式 (8 9) ~ (9 0) のように後ろから再帰的に求めることができる。

【 0 2 0 4 】

【数 4 0 】

40

50

$$\mathbf{z}^{\wedge} = (\mathbf{z}_i^{\wedge})_{i=1}^I \quad \cdots (88)$$

$$\mathbf{z}_i^{\wedge} = \underset{\mathbf{z}_i}{\operatorname{argmax}} v_i(\mathbf{z}_i) \quad \cdots (89)$$

$$\mathbf{z}_{i-1}^{\wedge} = \mathbf{v}_i(\mathbf{z}_i^{\wedge}) \quad \cdots (90)$$

10

【 0 2 0 5 】

式 (8 1) の const を除いた部分は、 q_{i-1} および b_{i-1} の関数として次の (9 1) のように書き換えられる。

【 0 2 0 6 】

【 数 4 1 】

$$f_1(q_1^{\sim}, b_1^{\sim}) + \sum_{i=2}^I f_i(q_i^{\sim}, b_i^{\sim} | q_{i-1}^{\sim}, b_{i-1}^{\sim}) \quad \cdots (91)$$

20

【 0 2 0 7 】

ただし、

【 0 2 0 8 】

【 数 4 2 】

$$\begin{aligned} f_1(q_1^{\sim}, b_1^{\sim}) &= \alpha_L \ln \pi^{\text{ini}}(q_1^{\sim}, b_1^{\sim} | \rho_T) \\ &+ \alpha_E \left[\ln \phi(q_1^{\sim} | \mathbf{z}_1^{\wedge}, \rho_T) - \frac{(q_1^{\sim} - q_1)^2}{2\sigma_p^2} - \frac{(b_1^{\sim} - b_1)^2}{2\sigma_r^2} \right] \quad \cdots (92) \end{aligned}$$

30

$$\begin{aligned} f_i(q_i^{\sim}, b_i^{\sim} | q_{i-1}^{\sim}, b_{i-1}^{\sim}) &= \alpha_L \ln \pi(q_i^{\sim}, b_i^{\sim} | q_{i-1}^{\sim}, b_{i-1}^{\sim}, \rho_T) \\ &+ \alpha_E \left[\ln \phi(q_i^{\sim} | \mathbf{z}_i^{\wedge}, \rho_T) - \frac{(q_i^{\sim} - q_i)^2}{2\sigma_p^2} - \frac{(b_i^{\sim} - b_i)^2}{2\sigma_r^2} \right] \quad \cdots (93) \end{aligned}$$

40

【 0 2 0 9 】

である。これらを最大化する q, b 系列 (式 (4) 参照) は、Viterbi アルゴリズムによって求められる。Viterbi 変数を次の式 (9 4) で定義する。

【 0 2 1 0 】

【 数 4 3 】

50

$$u'_i(q_i, b_i) := \max_{q_1, b_1, \dots, q_{i-1}, b_{i-1}} [f_1(q_1, b_1) + \sum_{i'=2}^i f_{i'}(q_{i'}, b_{i'} | q_{i'-1}, b_{i'-1})] \quad \dots (94)$$

$$v'_i(q_i, b_i) := \operatorname{argmax}_{q_1, b_1, \dots, q_{i-1}, b_{i-1}} [f_1(q_1, b_1) + \sum_{i'=2}^i f_{i'}(q_{i'}, b_{i'} | q_{i'-1}, b_{i'-1})] \quad \dots (95)$$

10

【 0 2 1 1 】

これらは、次のように再帰的に計算する。

【 0 2 1 2 】

【 数 4 4 】

$$u'_1(q_1, b_1) = f_1(q_1, b_1) \quad \dots (96)$$

$$u'_{i+1}(q_{i+1}, b_{i+1}) = \max_{q_i, b_i} [u'_i(q_i, b_i) + f_{i+1}(q_{i+1}, b_{i+1} | q_i, b_i)] \quad \dots (97)$$

20

$$v'_{i+1}(q_{i+1}, b_{i+1}) = \operatorname{argmax}_{q_i, b_i} [u'_i(q_i, b_i) + f_{i+1}(q_{i+1}, b_{i+1} | q_i, b_i)] \quad \dots (98)$$

【 0 2 1 3 】

これらの式に基づいて、式 (8 9) および式 (9 0) による方法と同様の方法によって、最適な編曲が X ~ として得られる。

30

【 0 2 1 4 】

図 3 の編曲データ生成部 2 5 9 は、最適候補選出部 2 5 8 によって原曲の編曲として選出された候補楽曲の系列 X ~ を、M I D I または M u s i c X M L のフォーマットのファイルに変換することによって、編曲データ 6 6 を生成する。編曲データ 6 6 は、自動音楽生成装置 1 の所定の記憶領域に保存される。または、端末装置 4 1 などへ送信される。または、p s 系列のデータに従って候補楽曲を音声再生ユニット 1 9 (図 2 参照) によって再生してもよい。

【 0 2 1 5 】

以上説明したように、編曲モジュール 2 5 によると、作曲モジュール 2 1 によって生成された学習済モデル 6 2 (作曲用のモデル) を用いて学習済モデル 6 5 (編曲用のモデル) を生成するので、従来よりも容易に自動編曲の A I を構築することができる。

40

【 0 2 1 6 】

本実施形態では、自動音楽生成装置 1 は、適切度を算出するモデル (学習済モデル 6 5) として、類似度および適合度の両方を因子として含むモデルを生成したが、類似度および適合度のうちの類似度のみを因子として含むモデルを生成してもよい。

【 0 2 1 7 】

本実施形態では、自動音楽生成装置 1 は、適切度の最も高い候補楽曲を原曲の編曲として選出したが、適切度が所定の値以上である候補楽曲を複数、原曲の編曲として選出してもよい。

50

【 0 2 1 8 】

〔 5 変形例 〕

ところで、一般に、楽曲の途中でスタイルまたはトニックなどが変わる場合がある。例えば、スタイルがクラシックからロックに変わったり、八長調からホ長調へ転調したり、長調から短調へ移調したりすることがある。そこで、自動音楽生成装置 1 は、途中でスタイルまたはトニックなどが変わるように構成したモデルを生成しておき、このようなモデルを用いて作曲してもよい。

【 0 2 1 9 】

このようなモデルは、楽曲の各小節 m ごとにスタイル変数 p_m およびトニック変数 k_m を導入することによって定式化することができる。曲中のいくつかの場所において転調したりスタイルが変わったりしても、その他の多くの場所では調やスタイルが継続していることを表現するため、スタイル変数 p_m およびトニック変数 k_m は、それぞれ、1 つ前のスタイル変数 p_{m-1} およびトニック変数 k_{m-1} に依存する確率により生成されるものとする。つまり、スタイル変数とトニック変数に対して次の式 (9 9) および式 (1 0 0) に示すマルコフモデルを考えればよい。

$$\theta_{ini}(p, k) = P(p_1 = p, k_1 = k) \dots (99)$$

$$\theta(p', k', p, k) = P(p_m = p, k_m = k \mid p_{m-1} = p', k_{m-1} = k') \dots (100)$$

各音符の音高クラスおよび拍節位置は、生成された各小節のスタイル変数 p_m およびトニック変数 k_m に条件付けられたマルコフモデルにより、次の式 (1 0 1) および式 (1 0 2) に基づいて生成される。

$$\pi_{ini}(q_{11}, b_{11} \mid p_1, k_1) = P(q_{11}, b_{11} \mid p_1, k_1) \dots (101)$$

$$\pi_{ini}(q_{mn}, b_{mn} \mid q'_{mn}, b'_{mn}, p_m, k_m) = P(q_{11} = q, b_{11} = b \mid p_m, k_m) \dots (102)$$

ただし、 q'_{mn} および b'_{mn} は、それぞれ q_{mn} および b_{mn} の 1 つ前の音符の音高クラスおよび拍節位置を表わす。

【 0 2 2 0 】

このようなモデルを用いれば、学習データとして与えられた楽曲に対して、小節ごとの合致するスタイルとトニックとを推定することができる。また、このモデルも、上述のモデルと同様に EM アルゴリズムを用いて学習することができる。推定および学習のために、次の式 (1 0 3) の、スタイル変数およびトニック変数の事後確率を計算する必要がある。これは、隠れマルコフモデルに適用されるフォワード・バックワードアルゴリズムを用いて計算することができる。

【 0 2 2 1 】

【 数 4 5 】

$$P(p'_m, k'_m \mid (p_{mn}, s_{mn})_{m=1, n=1}^{M, N_m}) \dots (103)$$

【 0 2 2 2 】

また、原曲が、途中でスタイルまたはトニックなどが変わる場合は、自動音楽生成装置 1 は、原曲をサンプリング法によって編曲すればよい。サンプリング法によると、部分 (例えば、小節) ごとに最も類似するメロディーを採用するので、原曲の変化に応じて原曲を編曲することができる。

【 0 2 2 3 】

本実施形態では、学習済モデルを生成する際に、マルコフモデルをベースとした移調対称性を持つ混合時系列生成モデルを用いたが、移調対称性の導入、混合モデルによるスタイル、およびトニックの自由度を持たせたモデルであれば、マルコフモデル以外の時系列生成モデルを用いてもよい。つまり、記号列 $Y = (y_1, \dots, y_l)$ の生成確率 $P(Y)$ を与える生成モデルであり、生成確率の計算、系列データからのモデルパラメータの学習、およびサンプリングによる系列の生成ができるモデルであれば、上記の定式化により移調対称性を

10

20

30

40

50

持った混合モデルを構成し、学習および生成（推論）が可能である。このような生成モデルの例として、音楽情報学および計算言語学などで広く用いられるモデル、例えば、高次のマルコフモデルおよび可変次マルコフモデル、隠れマルコフモデル、確率的文脈自由文法モデル、長短期記憶(long short-term memory; LSTM)ネットワークを含む再起型ニューラルネットワーク(recurrent neural network; RNN)、変分自己符号化器(variational autoencoder; VAE)および敵対的生成ネットワーク(Generative Adversarial Network; GAN)に基づく生成モデルなどが挙げられる。

【0224】

本実施形態では、音符の時間的な要素を拍節位置によって表わしたが長さで表わしてもよい。

10

【0225】

本実施形態では、音高と発音時刻の系列として表されるメロディーあるいは単旋律の楽曲を学習に用いるデータおよび生成する対象として用いたが、記号列として表されるデータであればこれ以外の形式の音楽データを用いてもよい。このような音楽データの例として、コード進行、ドラム楽譜、ギターやピアノなど多声部を含む楽曲、および複数の楽器または声部からなる合奏曲の形式のデータなどが挙げられる。

【0226】

本実施形態では、自動音楽生成装置1を1台のコンピュータによって実現する場合を例に説明したが、複数のコンピュータによって実現してもよい。例えば、メインモジュール20、作曲モジュール21、編曲モジュール25のそれぞれを別々のコンピュータによって実行してもよい。または、メインモジュール20、作曲モジュール21の学習フェーズ、作曲モジュール21の推論フェーズ、編曲モジュール25の学習フェーズ、および編曲モジュール25の推論フェーズのそれぞれを別々のコンピュータによって実行してもよい。

20

【0227】

その他、自動音楽生成装置1の全体または各部の構成、処理の内容、処理の順序、データの構成などは、本発明の趣旨に沿って適宜変更することができる。

【符号の説明】

【0228】

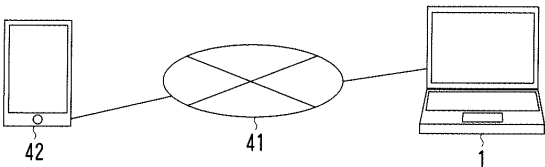
- 1 自動音楽生成装置（音楽モデル生成システム、編曲システム）
- 212 初期化処理部（初期処理手段）
- 214 事後確率算出部（出現確率算出手段）
- 215 パラメータ更新部（調整手段）
- 216 収束判定部（判定手段）
- 217 スタイルモデル記憶部（記憶手段）
- 218 音符決定部（作曲手段）
- 251 初期化処理部（編集モデル生成手段）
- 252 事後確率算出部（編集モデル生成手段）
- 253 パラメータ更新部（編集モデル生成手段）
- 254 収束判定部（編集モデル生成手段）
- 257 候補楽曲生成部（メロディ生成手段）
- 258 最適候補選出部（類似度算出手段、選出手段、編曲手段）
- 61 モデル（仮モデル）
- 62 学習済モデル（音楽モデル）
- 65 学習済モデル（編集モデル）

30

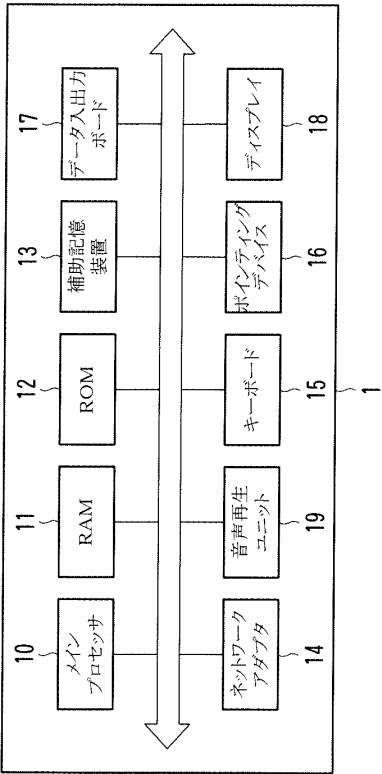
40

【図面】

【図 1】



【図 2】



10

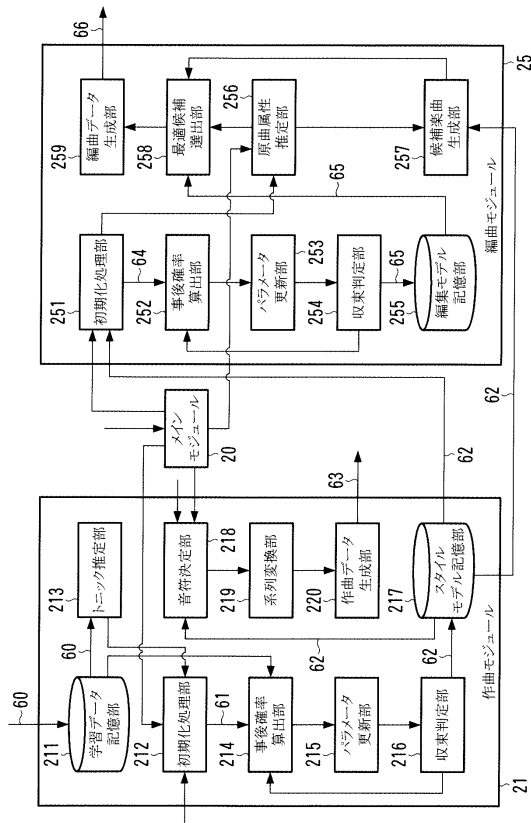
20

30

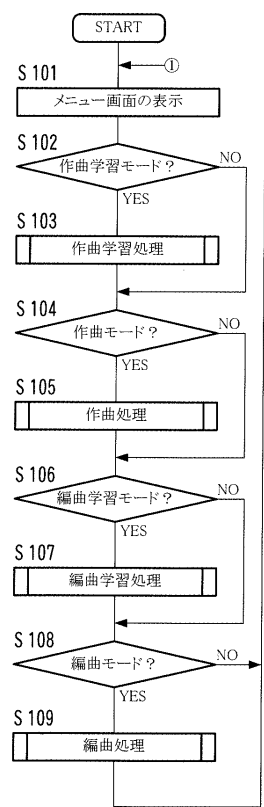
40

50

【図 3】



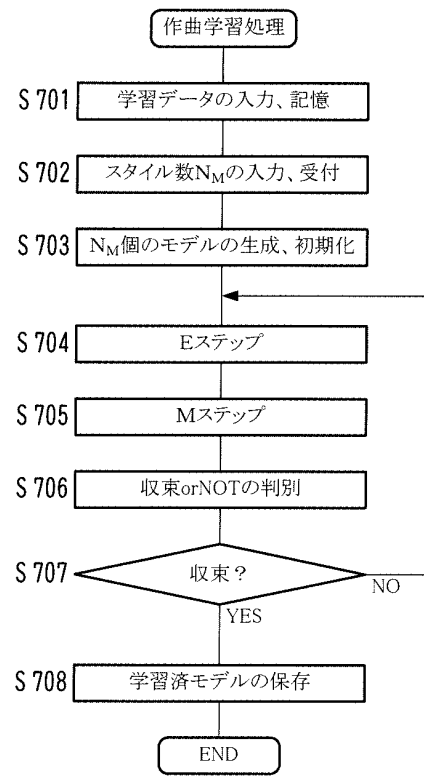
【図 4】



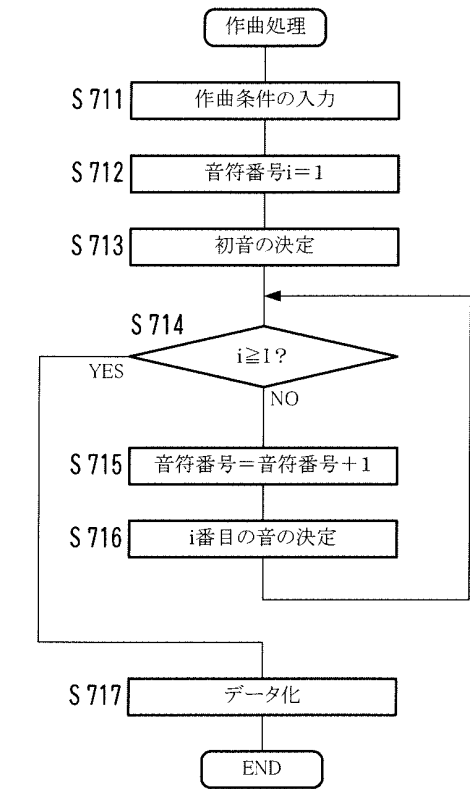
10

20

【図 5】



【図 6】

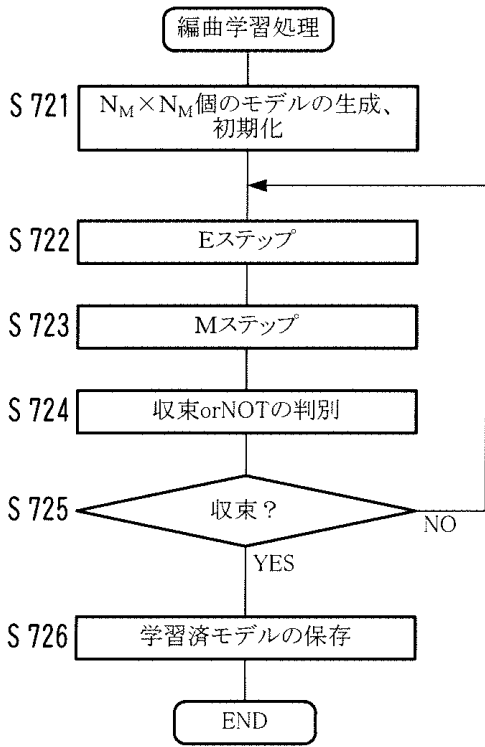


30

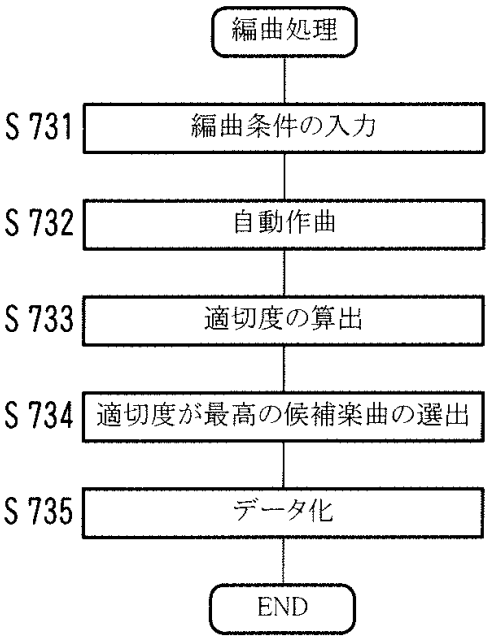
40

50

【 図 7 】



【 図 8 】



10

20

30

40

50

フロントページの続き

情報処理学会 論文名 : 教師なしスタイル変換によるメロディーの自動生成 発行日 : 2019年8月20日
〔刊行物等〕 研究集会での発表 研究集会名 : 第124回音楽情報科学研究会 タイトル : 教師なしスタイル
変換によるメロディーの自動生成 スタイル変換による多様なスタイルのメロディー自動生成(デモンストレー
ション) 発表日 : 2019年8月27日 〔刊行物等〕 ウェブサイトでの公開 サイト名 : 多様なスタイルに
よる自動音楽生成 URL : [https://melodyarrangement.github.io/d
emo-ja.html](https://melodyarrangement.github.io/demo-ja.html) [https://melodyarrangement.github.io/de
mo-en.html](https://melodyarrangement.github.io/de
mo-en.html) 公開日 : 2019年6月12日

(56)参考文献 米国特許出願公開第2009/0071315(US,A1)

国際公開第2021/106694(WO,A1)

(58)調査した分野 (Int.Cl., DB名)

G10G 1/00-7/02

G10H 1/00-7/12