4T-05

デジタル信号処理を用いた歌声採譜の検討

常山航輝†　松井くにお††
金沢工業大学 大学院 工学研究科 情報工学専攻†　金沢工業大学 大学院 工学研究科††

1 はじめに

歌声採譜は音響信号を用いて歌声の楽譜を推定することである. 歌声採譜が実現すれば手間をかけずに合成音声ソフトに歌わせることや, ハミング検索など様々な場面で応用できる. 現状では, 人間が歌った歌を合成音声ソフトに歌唱させるためには, 一音ずつ手動で入力しなければならない. さらに音感がなければ精度の高い楽譜を作成するのは難しい. よってカバー音源を作成するには手間と能力が必要になるという問題がある. そこで本研究では, デジタル信号処理を用いた歌声の採譜を行う. 歌声を人手で譜面化した音源と, デジタル信号処理を用いて作成した音源を比較し, 精度を検証した.

2 関連研究

歌声に限らず自動採譜は様々な研究がなされている. 柴田ら[1]は, 楽曲のセクションの境界推定と各セクション(イントロ, A メロ, B メロ, サビ)に対するラベル付けを行う音楽構造解析を, LSTM と隠れセミマルコフモデルのハイブリットモデルを用いてポピュラー音楽に対する構造解析手法を提案していた. 平松ら[2]は, ピアノ採譜のために音価と声部を同時に推定する手法について述べている. 錦見ら[3]は, ポピュラー音楽を対象に楽曲と歌声のピアノロールを入力として, 歌声の音高と音価を推定する方法について述べている.

3 実験方法

3.1 学習・評価用データ作成

先行研究では人間が理解しやすい楽譜形式での出力がなされているが, 本研究では従来の出力に多いピアノロール推定を行い MIDI 形式で出力する. 本実験で用いるツールを表 1 に示す.

表 1　実験に使用するツール

機能	ツール名
歌声抽出	spleeter[4]
音声モデル構築	Train_Autoencoder[5]
音声モデルの音声と楽器音に変換	DDSP[6]
MIDI へ変換	VariAudio[7]

歌声採譜に用いる楽曲は魔王魂[8]から収集し, 歌曲全てに対して歌声抽出を行った. 抽出した歌声は音声モデルの構築と, DDSP を用いて音声モデルの音声と楽器音の変換に使用する. 人手で作成する MIDI は有志が作成したものを編曲し, それぞれワンコーラスだけを 5 曲分用意した. 音声モデルの構築には歌声を抽出した 27 曲を学習用データとして用いた. 前処理として音声全体を 4 秒に変換し, CREPE で基本周波数を推論し, ラウドネスを計算する. 学習ステップは 13200, spectral_loss が 8.79, total_loss は 8.79 だった. DDSP を用いた音声変換は, 音声モデルと 2 種類のフルート, サックス, トランペット, バイオリンの 6 種類あり, 人手で MIDI を作成した 5 曲を対象としたので合計 30 曲作成した. この 30 曲に対して VariAudio を使用して MIDI へ変換する. 人手で作成した MIDI を図 1 に, 音声モデルの音声を MIDI にしたものを図 2 に示す.

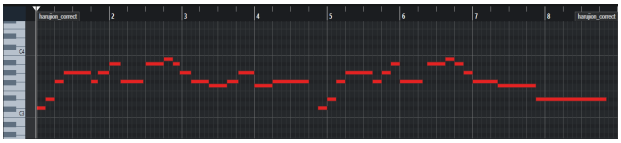

図 1　人手で作成した MIDI

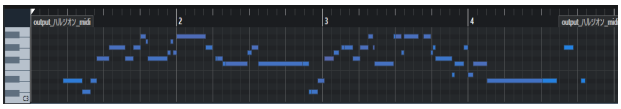

図 2　音声モデルの MIDI

Copyright ©2022 Information Processing Society of Japan.
All Rights Reserved.

3.2 採譜の精度比較

本研究では精度を出すのに平均二乗誤差(以後 MSE と表記)を用いた.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y^{(i)} - y'^{(i)})\text{^}2$$

$y^{(i)}$は i 番目の実際の値で，$y'^{(i)}$は i 番目の予測値を表す. MSE は実際の値とモデルによる予測値との誤差の二乗の平均値である．値が小さいほど精度が良く，値が大きいほど精度が悪い. MSE で計算するために人手による MIDI と DDSP で変換した MIDI の数値化を行った．比較のために抽出した歌声を何の処理もせず MIDI に変換し，数値化した．音程は C2 から C5 までを対象とし，それ以外の音程はノイズとして扱わなかった. 休符や音符が配置されていない場合には-1 を入れる．1 オクターブは 12 音なので 4 オクターブで 48 音，C2 を 0 とし，C5 を 48 とする．数値化した MIDI を図 3 として示す.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH
1	correct	2	4	7	9	9	9	7	9	11	7	7	7	11	11	12	11	9	7	7	6	6	7	9	9	6	6	7	7	7	7	-1	2	4
2	model	-1	-1	-1	4	4	4	2	7	7	9	7	9	11	7	10	8	11	11	11	9	7	6	6	6	8	9	7	7	7	7	-1	2	4
3	flute	-1	-1	-1	16	16	16	15	19	-1	20	20	20	23	22	19	20	-1	-1	23	23	-1	11	18	19	20	-1	-1	19	19	19	18	17	19
4	original	14	18	4	4	4	4	4	2	4	7	7	9	9	7	9	11	7	7	7	11	11	11	12	12	9	7	7	7	6	6	8	9	6

図 3　数値化した MIDI

correct が人手で作成した MIDI を数値化したもので，original が抽出した歌声そのままを MIDI にし，数値化したものである．MSE は correct と original，correct と flute のように correct と各楽器，original が組み合わさるように計算した．correct か楽器，original のどちらかに-1 が入っている場合にはどちらも数値を削除し, データ数が同じになるようにした.

4 評価結果と考察

MSE の評価結果を表 2 に示す.

表 2　評価結果

	original	model	flute	flute2	sax	trumpet	violin
A	3.183	5.050	540.563	145.180	4.569	141.264	135.136
B	15.619	6.836	126	128.549	9.096	8.571	129.346
C	11.604	6.776	452.66	146.821	4.389	135.296	133.768
D	2.444	9.778	143.919	158.4	13.611	11.147	186.595
E	3.105	5.877	494.268	126.9	7.246	126.569	137.869

A から E は使用した楽曲であり，A から BPM が高い順になっている．B と C を除き，model に比べて original は精度が高かった．これは，model と楽器音は変換時に音価情報が失われるのに対し，original は変換して音価情報が失われていないため正確に数値化できたからだと推測する. 楽器音は精度が悪いが, A と C の sax だけは精度が高かった. 歌声と楽器音の音程が同じだったからだと考えられる. A から E まで model の精度は安定していた.

5 おわりに

本研究では, デジタル信号処理を用いた音源を使って歌声採譜する方法と，採譜結果の精度について述べた．今回の実験から, 音声モデルを使用することで一定の精度を出せると分かった．今後は先行研究の方法に倣うとともに，音程の精度向上と音価情報の欠落の改善を進める.

参考文献

[1]柴田剛，錦見亮，中村栄太，吉井和佳:LSTM-HSMM ハイブリットモデルに基づく音楽構造解析, 研究報告音楽情報科学，2020-MUS-128，Vol10，pp.1-8(2020)

[2]平松祐紀，中村栄太，吉井和佳:ピアノ採譜のための音価推定と声部分離のマルチタスク学習, 研究報告音楽情報科学，2021-MUS-132，Vol18，pp.1-7(2021)

[3]錦見亮，中村栄太，吉井和佳:ビート同期注意機構に基づく歌声のリズム採譜，研究報告音楽情報科学，2019-MUS-124，Vol.8，pp.1-6(2019)

[4]Deezer，spleeter:https://github.com/deezer/spleeter(2021 年 12 月 17 日)

[5]JesseEngel，Train_Autoencoder: https://github.com/magenta/ddsp/blob/main/ddsp/colab/demos/train_autoencoder.ipynb(2021 年 12 月 17 日)

[6]JesseEngel, DDSP:https://github.com/jesseengel/ddsp(2021 年 12 月 17 日)

[7]Cubase，VariAudio(CubasePro のみ): https://steinberg.help/cubase_pro_artist/v9/ja/cubase_nuendo/topics/sample_editor_variaudio/sample_editor_variaudio_c.html(2021 年 12 月 20 日)

[8]森田交一, 魔王魂 https://maou.audio(2021 年 12 月 17 日)

Copyright ©2022 Information Processing Society of Japan.
All Rights Reserved.