

## 歌声を用いたDTM向け演奏表現パラメータの入力\*

虻川内 努 (法政大学 情報科学部), 伊藤 克亘 (法政大学 情報科学部)



図 1. 譜面: グリーク/ホルベアの時代より

### 1 まえがき

歌声や楽器の奏法にはピラートやピッチベンド、音の強弱などがあり、実際の演奏ではこれらをコントロールする事で音楽的な表現を実現している。しかしこの楽器操作は演奏する楽器を十分に習熟していなければ行なう事ができない。コンピュータに演奏させる打ち込みならば楽器が演奏ができなくとも演奏音を出力することができるが、演奏表現を再現するには弾く強さ・音の立ち上がりなどの要素を考慮し、音量やピラートの推移を設定しなければならず、煩雑であったり、知識や経験がなければスムーズに設定する事は困難である。本稿では声に含まれる音のニュアンスを奏法情報として利用することを検討する。具体的には表現したい楽器らしく歌う事で、歌声から演奏表現の推移を抽出し利用者に提案する。これにより打ち込みを用いる音楽制作または編曲において、演奏表現パラメータの調整に”歌による演奏意図”を与えることができ、調整に対する思考時間の短縮を提供する。

### 2 扱うパラメータ

扱う要素は発音の強さを決定するベロシティと音量の時間的変化を表現するエクスプレッション (共に 128 段階) である。この二つをともなって表現することができる楽器は主に吹奏楽器、擦弦楽器である。この表現情報は楽譜においても表記がなされることがあり、ピアノ・フォルテなどの強弱記号、またクレッシェンド・デクレッシェンドにより示される (図 1)。実際の演奏では演奏者の楽譜の解釈・表現の意図によって音量を変動させるため単調な変化とはならず、表現の個性が現れる。図 1 の譜面をプロのバイオリン奏者により演奏した場合の音量の時間変化を図 2 に示す。この音量変化を作曲ソフトを用いて音源に与える場合、音符の発音時のベロシティ設定・音符内やスラー記号で結ばれる間のエクスプレッション設定の 2 点によって制御できると仮定する。この 2 点を踏まえ、実例として実際の楽器演奏の音量変化 (図 2) を模倣し MIDI 演奏音を手作業で作成した。音量の時間変化を図 3、与えたベロシティ・エクスプレッションを図 4 に示す。これらの図より、音量変化と与えたパラメータの概形が似ていると判断した。そこで本手法では音量の時間変化を歌声を用いて入力、ベロシティ・エクスプレッションの値に変換し出力することを試みた。

### 3 システムの概要

本システムの構成を図 5 に示す。入力は音量変動の演奏意図を意識して録音したサンプリング周波数 44.1kHz

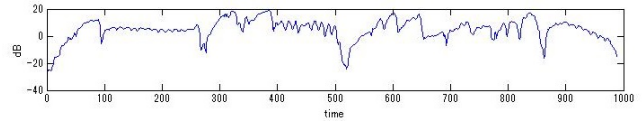


図 2. プロ演奏データの音量

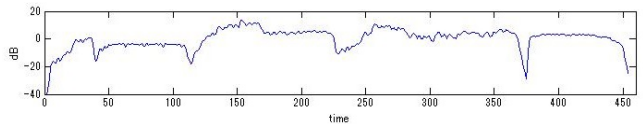


図 3. 手作業により作成した MIDI 演奏の音量

の歌声ファイル、出力は譜面データに対応するベロシティ、エクスプレッションの値である。入力ファイルにはサンプリング周波数 8000 へのダウンサンプリングを行い、編集の対象音をバイオリンと想定し、実際に演奏音を作成した。

### 4 音符区間セグメンテーション

楽譜情報と入力データの基本周波数 ( $f_0$ ) を DP マッチングにより対応付ける事により、入力データ上での音符区間を決定した。 $f_0$  は自己相関を用いて取得する。窓幅 30 ミリ秒、シフト幅 15 ミリ秒にて周期のピークを調べ算出した。次に、入力データ上の音符区間を楽譜情報の音符の比率に伸縮する事で、譜面と入力データ間の音符長の正規化を行った。伸縮の度合いは JIS で定められる寸法変化率を用い、その倍率で入力データ上の各音符区間の長さを楽譜情報と同じ比率になるよう伸縮した。

### 5 演奏パラメータ変換

入力データの音量を演奏パラメータに変換する。音量値は入力データに対し窓幅 80 ミリ秒、シフト幅 40 ミリ秒にてパワーを算出し  $10\log_{10}(P)$  の対数を取ることでデシベルに変換し利用する。大局的な音量変化に寄与しない音量の揺れは 5 点のメディアンフィルタにより平滑化した。

演奏パラメータ変換の関数を決定するために、音のデシベルとエクスプレッション値の相関関係を調査した。バイオリン音源 soundfont”rolandviolinsolo.sf2”を対象に、BPM120 のもと、エクスプレッションを 1~127 の値でそれぞれ一定に与えた全音符の MIDI 命令により鳴らされた 127 個の音声データを用意しパワー

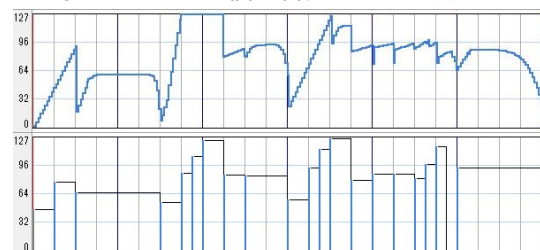


図 4. 手作業により作成した MIDI 演奏のエクスプレッション (上)/ベロシティ (下)

\* Articulation parameter values proposal for DTM with using the voice.: Tsutomu Kerakawachi (Hosei Univ.) et al.

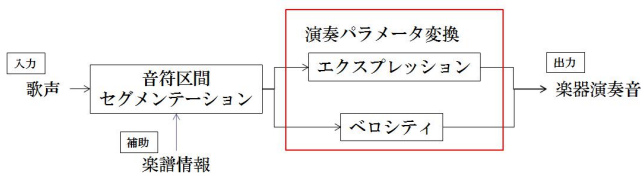


図 5. システムの構成

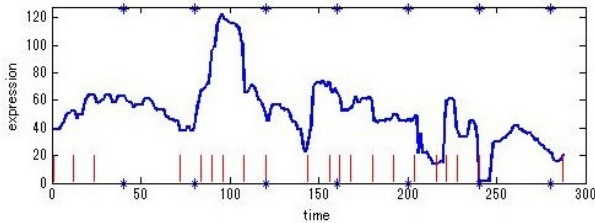


図 6. 歌声から変換されたエクスペッション値, 赤線は音符の発音時刻

の比較を行った. 観測地点は開始地点付近は音の立ち上がり部分になってしまうことを考慮し, 音の持続部分である中央付近の時間で比較を行う. パワーの推移を近似する関数の式を求め, 変換関数を決定する. 近似先の式の形は  $y$  をパワーの値,  $x$  をエクスペッションの値として

$$y = Ae^{Bx} + Ce^{Dx} \quad (1)$$

となると想定し, 係数  $A, B, C, D$  を求めた. 初期値  $[A, B, C, D] = [0, 1, 0, 0]$  として最小二乗近似を行った結果係数の値は

$$A = 3.4523, B = -0.0137, C = -3.3986, D = -0.0120 \quad (2)$$

と決定され, これを最大値で割ることで正規化し, 関数で扱える  $y$  軸の範囲をおよそ  $0 \sim 1$  にした. この関数により値の範囲を  $0 \sim 1$  に正規化したパワー情報を  $y$  に代入し, エクスペッション値  $1 \sim 127$  に変換する事が可能となる. ベロシティに関しては, エクスペッションの値に伴って変化するとし, 音符の開始地点における音量をエクスペッションに変換した値と同値に設定した

## 6 出力

最後に, 一つの音符に対して何段階の時間変化を与えるかを決定し, 全体の長さを伸縮し出力の分解能を決定する. 今回は 4 分音符一つの長さに対して 12 段階の時間変化を与えるとして全体を伸縮し出力した. 楽譜情報と同時に表示したものを図 6 に示す.

## 7 評価実験

実験においては, まず 2 章にて取り上げたプロ奏者の演奏音を目標とすることで音量変化の意図を固定する. 出力音はバイオリン音源 (5.1 章参照) を用いて生成する. 生演奏の音と比較できない事を考慮し, 目標演奏音を音源を用いて手編集によって再現した出力音を正解データ, 本システムを用いて作成した演奏音を評価対象とした上で, 正解データにどれだけ近いかわ

することを検討した. 作成した比較対象音は以下のものを用意する. 1. システムによる値をそのまま反映したものの, 2. MIDI 編集ソフトに用意されている曲線, 直線描画ツールのみを使い, それぞれ作業時間 1 分 40 秒で概形をなぞる編集をしたもの, 3. バイオリンの立ち上がり方を考慮した編集をしたもの, 4. 最大の値が 127 になるように上下移動編集をしたものである. その他, 特に演奏意図を持たせていない演奏音としてエクスペッション値をすべて 127 にしたもの, 適当な値を手作業で入れたもの, 乱数を与えたものの 3 データを比較のために用意した.

演奏の近さ算出 目標演奏とシステム利用の演奏がどの程度近いかを評価するために演奏ファイルのデシベルの値を正規化誤差 (ND) により計算した.

$$ND = \frac{\sum_{n=2}^N |CP_n - TP_n|^2}{\sum_{n=2}^N |FP_n - TP_n|^2} \quad (3)$$

式 3 の  $CP_n$  はシステムで作った演奏音の  $n$  番目の値,  $TP_n$  は目標演奏音の  $n$  番目の値,  $FP_n$  はエクスペッション操作情報を与えていない演奏音の  $n$  番目の値であり,  $ND$  の値が 0 に近いほど目標演奏に近く, 1 に近づくほど何も演奏情報を与えていない事に近い事を意味する. その結果を表 1 に示す.

表 1. 各データの ND 値

データ	ND 値	データ	ND 値
1	1.1442	5	0.8231
2	0.6829	6	0.9529
3	0.6834	7	1.0330
4	0.4969		

## 8 結論

演奏音のデシベルとエクスペッション値の関係を明らかにし, 求められた変換関数に基づき歌声の音量変化を奏法情報に変換した. その結果を単純に反映するだけでは目標音に近づく事できなかったが, 概形をなぞるだけでおよそ 32% 目標に近づいた事から, 歌声の意図情報はこの手法により求められた音量変化に含まれている事が確認できた. また全体の音量基準を適切に設定する事ができればおよそ 50% 目標に近づける事ができた.

## 参考文献

- [1] Janer, J., Maestre E. "Phonetic-based mappings in voice-driven sound synthesis" ICSPMA 2007
- [2] 赤本仁史, 武田正之 "DP マッチングを用いた演奏の現在位置解析手法の提案" 第 9 回情報科学技術フォーラム 第 2 分冊 289-292, 2010
- [3] 田原佳代子他 "歌唱音声制御に伴うスペクトル変動の主成分分析と合成への応用について" 信学技報 SP, 音声 105(198), 19-24, 2005-07-14
- [4] 寺村 佳子, 前田 新一 "演奏者の個性を表す特徴に関する考察" 情処学音楽情報科学研報] 2011-MUS-89(11), 1-6, 2011-02-04
- [5] 鈴木, 泰山他 "Kagurame Phase - I 事例ベースの演奏表情生成システム" 情処学音楽情報科学研報 1997-MUS-024(14), 61-68, 1998-02-13