歌声を用いたＤＴＭ向け演奏表現パラメータの入力

Articulation parameter values proposal for DTM with using the voice

蝮川内　努
Tsutomu Kerakawachi
法政大学情報科学部コンピュータ科学科
E-mail: 10k0015@stu.hosei.ac.jp

Abstract

In playing musical instruments, it can control strength pitch-bend and vibrato. These controls give articulation to sound. In composition with computer software, these elements must be set in consideration of the actual instruments. But it is difficult because these value is not quantitative. In this study, the goal is to get the articulation from the human voices that singing aware of the instrument. System makes and propose articulation transition graph to the user. That means provide guidance, user can adjustment parameters easier in music production. In this method, tries to get information of velocity and expression from voice that imitation violin sound. The way is 3steps. First, identify the note-position information in singing voice using DP matches. Second, estimates a transform function that represents the relationship between the value and volume, then the volume of singing voice converted to a value of 127 steps. Finally, set the velocity from max value of inter-note, and expression value set from actual singing-voice and model of instruments attack. As a result, distance that between target sound and system-made sound was calculated by the Normalized difference shows about 0.58 to target sound(0 is Exact match and 1 is distance from mean value.). And this distance is not change because diffelent instlument sounds.

1　序論

歌声や楽器の奏法にはビブ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弾く強さ・音の立ち上がりなどの要素が，演奏者の意図によって変化するため，定量的に扱う事ができないことから定数としてパラメータ設定を行う事は困難である．

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，調整に有用な設定値を提供する．

Supervisor: Prof. Katunobu Itou

図 1. 譜面:グリーグ/ホルベアの時代より

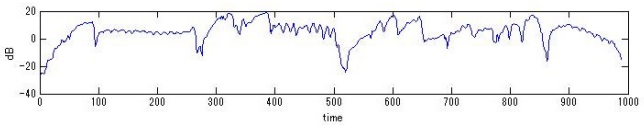

図 2. プロ演奏データの音量

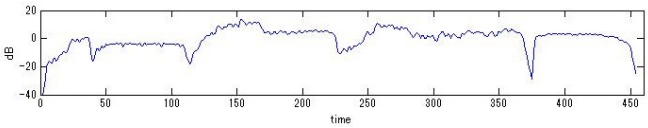

図 3. 手作業により作成した MIDI 演奏の音量

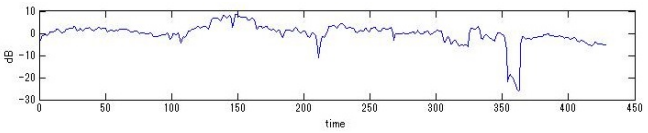

図 4. 入力した歌声の音量

1.1　扱うパラメータ

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 2 つをともなって設定する上で効果的な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示す (図 2).

1.2　パラメータ値決定の方針

この音量変化を作曲ソフトを用いて音源に与える場合，大局的な音量変化をベロシティ設定・楽器の表現力の再現をエクスプレッション設定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，与えたベロシティ・エクスプレッションを示す．(図 3・図 5) このパラメータで出力された演奏音は 4 章で後述する評価実験において実際の演奏音と近いものができていると言え，パラメータ設定による再現度がこの程度では行えるという前提のもと．本手法では歌声を用いて二つのパラメータを設定し，良い演奏音を生成する事を目標とする．

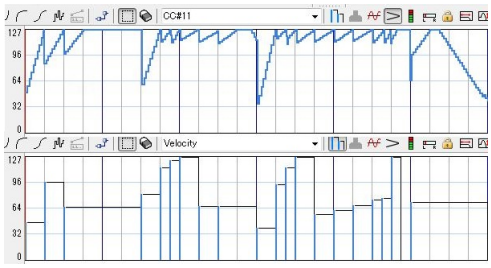

図 5. 手作業により作成した MIDI 演奏のエクスプレッション (上)/ベロシティ (下)

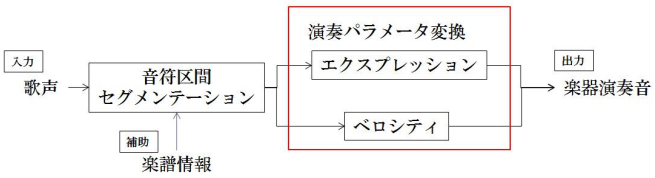

図 6. システムの構成

2 システムの概要

本システムの構成を示す (図 6)．入力は音量変動の演奏意図を意識して録音した歌声ファイル (音量変動の様子は図 4)，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．編集の対象音はバイオリンを想定し，実際に演奏音を作成した．

入力ファイルはサンプリング周波数 44.1kHz の録音データであり，8000 点へのダウンサンプリングを行った．

2.1 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 (f0) 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．f0 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．

周波数の推移情報は MIDI から取得する．MIDI 情報には演奏するノートナンバーと発音する時間が記載されている．これを元に周波数と時間の推移情報列を生成する．周波数はノートナンバーより計算可能であり．ノートナンバーを N，周波数を f とすると

$$f = 440 \times 2^{\frac{(N)-69.0}{12.0}} \tag{1}$$

の式により周波数を計算する事ができ，データの長さは発音時間の比率を保って任意に設定し，情報列として保持する．

楽譜の音高情報と入力音声データの周波数情報において類似度の高い経路を探索する．経路決定は以下の式で表される．

2 つの 1 次元パターン $X = x_1, x_2, ...x_i, ...x_I, Y = y_1, y_2, ...y_i, ...y_{II}$ における X の第 i 要素 x_i と Y の第 j 要素 y_j との対応付け $j = u_j(i = 1, ..., I)$ の最適化について

$$\text{局所距離} \quad d_i(u_i) = \| x_i - y_{u_i} \| \tag{2}$$

$$\text{最小コスト選択 } argminF = \sum_{i=1}^{I} d_i(u_i) \tag{3}$$

対応付けの経路探索に設定したコスト付けの詳細は以下の 2 点である．

■音高の差異によるコスト　それぞれのデータ点を単純比較した物が局所距離となり，この値が離れているか否かによってコストの重み付けを行う．

入力データのある点での周波数の cent 距離を観察すると，音程での発音を意図した部分での距離は 40 を超えず，その他の全く違う部分では距離が数百単位で離れている事が確認できた．

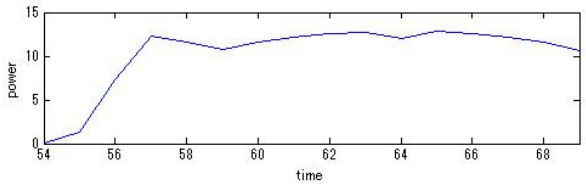

図 7. バイオリン音源 soundfont”rolandviolinsolo.sf2” 無編集の音量変動

そこで誤差を 40cent 以内まで許可し，その範囲内のコストを 1 として，それ以上の誤差があるときにコストを 10 倍にする設定を行った．

■経路移動でのコスト　経路探索の問題であるため，データ間の移動コストの設定も必要である．同じ値のデータが続く場合，MIDI・入力データ両方向への時間軸移動が行われており，コスト 1 を与える．同じ値が続かない場合は，状態が維持されているデータ方向へのみ移動となり，コスト 2 を設定した．これらのコストは，移動先の点における局所距離コストの倍率により重み付けが行われる．

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比率を統一した．JIS で定められる寸法変化率を用い歌声データの音符長 VT を譜面の音符長 MT に伸縮する．伸縮の度合いは JIS で定められる寸法変化率を用いて

$$\text{寸法変化率} = (MT - VT)/VT \tag{4}$$

で表され，この倍率で歌声データの各音符の長さをＭＩＤＩの音符の長さと同じ比率になるよう修正した．

2.2 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 ミリ秒にて算出し $10log_{10}(P)$ の対数を取ることでデシベルとして利用する．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．

2.2.1 エクスプレッション

楽器の表現力を与える値であり，音源自体にもそれぞれ備わっている (例:図 7)．しかしこの状態では実際の楽器音で与えられるような表現を再現できない．そこでエクスプレッションを設定する事で表現を再現する．

エクスプレッションの値を設定することで適切な変化を与えられるか，またどのような値を与えればよいかを，楽器の実演奏音・MIDI によって再現したもの・歌声の 3 つの音量変化の比較により調査した．また歌声を値に変換するにあたって楽器を考慮したモデルを適用すべき部分と歌声の表現を反映すべき部分もこの比較により調査する．比較グラフ (図 8，図 9) は上から生楽器演奏音，MIDI 編集音，歌声入力音である．音符の長さを 3 種に分類・定義し考察を行った．

■開始音符　音楽の開始音符，休符の後，また 4 分音符よりも大きい音符の後ろかつ小節の開始音である音符をこの分類に分けられると定義する．

連続して発音する音符群の始まりの音符 (図 1：1,4,10 番目) では，強い立ち上がりが観察され，声ではこれを再現できていない．これは楽器と声の発音方法に物理的な違いがあるからである．MIDI 編集ではある程度の再現が可能で，S 字曲線を用いてゼロから最大になるような変化与え再現した.(比較図：図 8) この編集をモデルとする．

■要素の音符　直前の音符との距離が短い，またスタッカートなどの演奏記号内の音符群を構成する音符 (例：図 1：13〜16

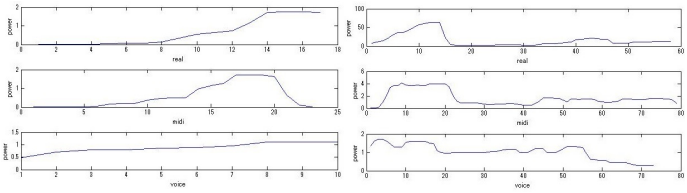

図 8. 右：開始音符，左：要素の音符 (楽譜 13〜16 番目)

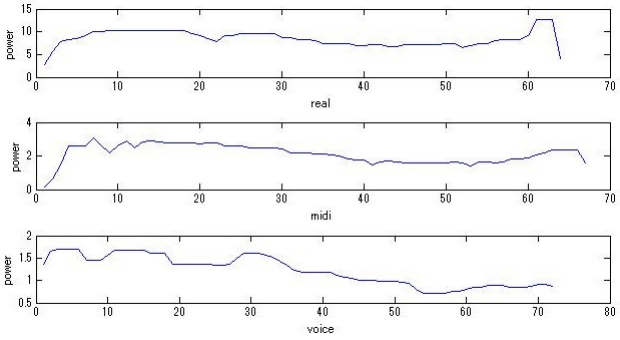

図 9. 長音符 (3 番目)

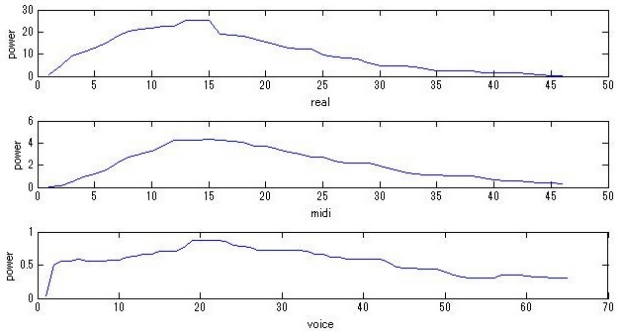

図 10. 長音符 (22 番目)

番目) のなかで付点 4 分音符未満の音符がこの分類に分けられると定義する．

音符の長さが短い事に起因する発音時の楽器特性が顕著であり，声での再現が不可能である．この音の再現に当たっては，エクスプレッション値を無音部分からの発音ではないため 96 から始め，最大値までの一次関数的な値を与える事で再現した．(比較図：図 8) この編集をモデルとする．

■長音符　全音符や付点音符等の長い音符 (例：図 1：3,9,22 番目) がこの分類に分けられると定義する．

音の立ち上がりの再現を除く音量の変化が似ている事が比較図 (図 9, 図 10) から観察された．これより声での再現が ADSR 法における S(sustain) 部分に限って可能であると考え，楽器音の立ち上がりのみ前述の短い音符に対してのものと同様の処理を行い，その値からつながるように入力した歌声の音量変化を適用する．

2.2.2 ベロシティ

音量・パラメータ変換関数により入力データの音量変動から 128 段階表現のベロシティへと変換する．

■音量・パラメータ変換関数　入力音量を 128 段階で表現するために，音量と演奏パラメータ設定値の相関関係を調査した．バイオリン音源 soundfont”rolandviolinsolo.sf2” とストリングス音源”KORG M1 le(Strings)” を対象に BPM120 のもと，一つの音符に対しベロシティを 0〜127 の値で与えたものに，エクスプレッションを 0〜127 で与えた音量変動・ベロシティ・エクスプレッションの関係を示す (図 11).

観測された音量の推移を近似する関数の式を求めることで変

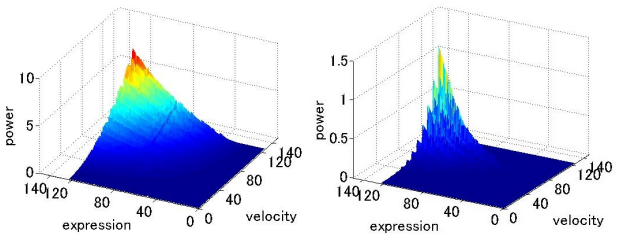

図 11.　音量変化の様子 (右”rolandviolinsolo”，左”KORG M1le”

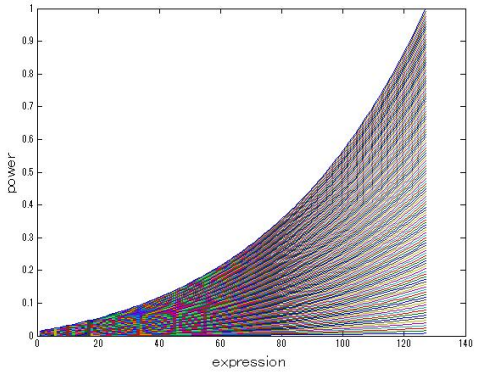

図 12. ベロシティ 128 段階分の 128 本の音量近似曲線

換関数を決定する．近似先の式の形は y を入力音量の値，x を 128 段階の値として

$$y = Ae^{Bx} + Ce^{Dx} \tag{5}$$

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．その結果，. 例として soundfont”rolandviolinsolo.sf2” に対する係数の値は

$$A = 3.4523, B = -0.0137, C = -3.3986, D = -0.0120 \tag{6}$$

と決定された近似結果を図 12 の線に示す．これを最大値で割ることで正規化しｙ軸の範囲をおよそ 0〜1 にして関数として扱う．この関数により値の範囲を 0〜1 に正規化したパワー情報を y に代入し 128 段階の値に変換する事が可能となる．

音符区間ごとの最大値を観察し，この場所ではエクスプレッションがデフォルト値 127 を取るとして関数に代入し，ベロシティを決定した．

3 処理結果

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 (図 13).

4 評価実験

意図した表現を与えられたかを検証する．演奏音の生成に関しては，1 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．以下の 5 つの演奏音を評価対象とし，目標としたプロ奏者の演奏音にどれだけ近いかを評価した．

■作成した演奏データ

1. rolandviolinsolo.sf2 Violin Solo 1
2. rolandviolinsolo.sf2 Violin Solo 2

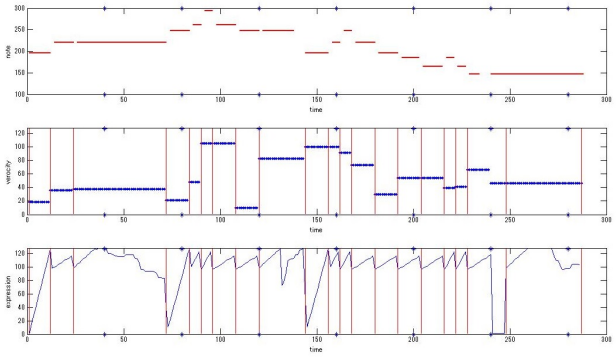

図 13. 処理結果：上から楽譜情報．歌声から変換されたベロシ

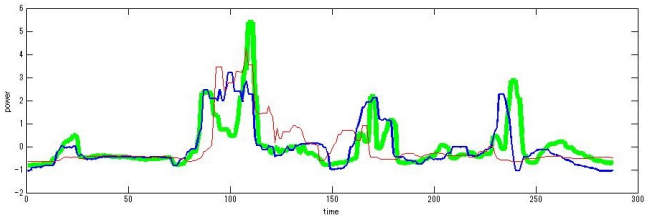

図 14. 太線：目標，中線：手編集，細線：データ 2

3. M1 le Strings
4. パワー変化を観察し手作業で編集した演奏音
5. 無編集音

1・2・3 については音源の変更にかかわらず同じ意図を与えられているかを評価する．

4.1 正規化誤差

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，音量の値を正規化誤差 (Normalized difference, 以下 ND と略) により計算した．平均が 0，分散が 1 になるよう，線形変換し正規化した上で比較を行った．

$$ND = \frac{\sum_{n=2}^{N} |CP_n - TP_n|^2}{\sum_{n=2}^{N} |\overline{TP_n} - TP_n|^2}, \overline{TP_n} = \frac{1}{N-1}\sum_{n-2}^{N} TP_n \quad (7)$$

式 7 の CP_n はシステムにより出力した演奏音の n 番目の値，TP_n は目標演奏音の n 番目の値であり，N は評価に用いたデータの点数とする．$ND = 0$ は目標演奏と比較対象の演奏の音量変化が完全に一致する事を意味し，$ND = 1$ は実演奏の平均値を出力した場合の誤差と同じ誤差である事を意味する．その結果を表 1 に示す．

表 1. 各データの ND 値

データ	ND 値
1	0.6530
2	0.5853
3	0.5366
4	0.1561
5	1.1883

音量変化の概形の例は図 14 に示す．

4.2 考察

無編集データと比べ，すべての音源で ND 誤差の減少が見られることから，本システムにより表現が付与されていることが確認できる．また，音源 2・3 の値が近い事から、音源に適応して値設定のための変換関数が有効にはたらいている．音源 1 に関して音源 2・3 と比べ誤差が離れてしまった原因としては，音源自体の音量変化の度合いが比較的緩やかな音源であるため，それによってこの場合ではベロシティ値の変化量が正確に変換されなかった事が考えられる．

歌声の変動に影響されると考えた長音符の部分に関しては音量変動を観察すると (図 14) 変化量に乏しい事が考えられる．またクレッシェンド部分の立ち上がりの再現度が足りないこともグラフから見て取れる．これより演奏記号間の音量変動のモデル化，変化量の強調などを考慮する必要があると考える．

また単純に誤差の値を小さくすることを考えると，エクスプレッションの設定がまだ適切でない事があげられる．音源をさらに観察し，音源による変化も考慮した適切なモデルの設計が必要だと考えられる．

また実験対象楽曲を増やし, 楽曲の違いによる一般性の有無を調査する必要がある．

5 結論

歌声の音量変化を演奏表現のパラメータに変換するシステムを構築した．モデルと入力歌声を併用したエクスプレッションの設定と，演奏音の音量と作曲ソフトで用いられる 127 段階の音量変化との関係を明らかにし求められた変換関数に基づき，使用する音源を考慮したベロシティ変換を行う事で，調査したすべての音源において無編集状態の音源よりも誤差を減らせた．また音量変動の様子が違う音源でも同じような表現付加ができたことから，音源を考慮した変換関数の有効性が確認できた．

参考文献

[1] 中野倫靖, 後藤真孝, 平賀 譲 ”楽譜情報を用いない歌唱力自動評価手法”情報処理学会論文誌. 48(1),227-236,2007-01-15

[2] 中野倫靖, 後藤真孝 ”VocaListener：ユーザ歌唱の音高および音量を真似る歌声合成システム”情報処理学会論文誌. 52(12),3853-3867,2011-12

[3] Janer, J., Maestre E. ”Phonetic-based mappings in voice-driven sound synthesis” International Conference on Signal Processing and Multimedia Applications 2007

[4] 亀岡弘和, 篠田浩一, 嵯峨山茂樹” スペクトル領域の DP マッチングによる自然楽器演奏の多重音解析”

[5] 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

[6] 田原 佳代子, 高橋 徹, 森勢 将雅, 坂野 秀樹, 河原 英紀 ”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”電子情報通信学会技術研究報告. SP, 音声 105(198), 19-24, 2005-07-14

[7] 安部 武宏, 糸山克寿, 吉井 和佳, 駒谷 和範, 尾形 哲也, 奥乃博 ”音色の音高依存性を考慮した楽器音の音高操作手法” 情報処理学会論文誌 50(3), 1054-1066, 2009-03-15

[8] 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”情報処理学会研究報告. [音楽情報科学] 2011-MUS-89(11), 1-6, 2011-02-04

[9] 鈴木, 泰山, 宮本, 朋範, 西田, 深志, 徳永, 健伸, 田中, 穂積 ”Kagurame Phase - I - 事例ベースの演奏表情生成システム - ”情報処理学会研究報告. [音楽情報科学] 1997-MUS-024(14), 61-68, 1998-02-13