ハミングを入力とする類似音楽検索システムにおける自動採譜手法の検討

辻　紗千† 獅々堀正幹†† 北　研二††

† 徳島大学大学院 先端技術科学教育部 システム創生工学専攻
†† 徳島大学大学院ソシオテクノサイエンス研究部 〒 770-8506 徳島県徳島市南常三島町 2-1
E-mail: †{tsujis,bori,kita}@is.tokushima-u.ac.jp

あらまし　我々はハミングを入力キーとする類似音楽検索システムの提案を行ってきた．提案手法では検索キーから特徴量を得るため，ユーザが入力したハミングデータから自動的に採譜を行う．本手法は音価の系列と周波数比の間隔で並んだくし型のテンプレートを用いて音価と音程の推定を行う．更にテンプレートの形状を従来のものから拡張し精度の向上を図った．その上で自動採譜の精度向上による類似音楽検索システムの精度向上を確認した．
キーワード　類似音楽検索，ハミング入力，MIDI，くし型テンプレート

Automatic Music Transcription on Similar Music Retrieval System for the Query-by-humming

Sachi TSUJI†, Masami SHISHIBORI††, and Kenji KITA††

† Systems Innovation Engineering, Graduate School of Advanced Technology and Science, Tokushima University.
†† Institute of technology and science,Tokushima University.
E-mail: †{tsujis,bori,kita}@is.tokushima-u.ac.jp

1. はじめに

近年，通信技術の発展は著しく，スマートフォン等の携帯端末も広がりネットはもはや生活の一部となりつつある．また，それに合わせて音楽や映像等の配信サービスも更に充実してきている．しかし，膨大なデータの中から必要なデータのみを見つけ出すのは困難であり，より効率的な検索手法が必要であると考えられる．また，各メディアデータにあった検索方法の必要性も高まっている．

そこで我々は楽曲検索手法の一つとして，ハミングを入力とし，MIDI 形式 (SMF フォーマット形式) の音楽データを検索するシステムを開発してきた．距離尺度に Earth Mover's Distance(EMD) を用いた類似音楽検索手法 [1] によるシステムの構築を行い，後に入力ハミングを特徴量へと変換するための機構を追加した．しかし入力されるハミングは人によるもののため揺らぎが多く，常に正しい音高，音程であるとは限らない．また構築したシステムは，入力される音楽データに含まれる誤りにより精度が大きく変化する．つまり入力ハミングの採譜精度が検索システム全体の精度を制限するというのである．

そのため，ハミングの揺らぎを考慮した上で，採譜を行う必要性がある．そこで揺らぎの要素の一つである基準キー，テンポの違いを考慮した採譜システムとして，くし型テンプレートを用いた採譜手法 [2] に注目した．この手法は基準となるテンポ，音程を求めて推定に用いることで，相対的な採譜を行うことが可能である．

更に音階や音価の幅が一定していないハミングへの対策として，我々は基本となるくし型テンプレートをベースに，音階や音価の幅のぶれを考慮した新しいくし型のテンプレートを提案する．

2. くし型テンプレートを用いた採譜手法

ここでは，くし型テンプレートを用いた採譜システムの手法について説明する．採譜は以下の流れで行う．

（1） 音程情報を抽出する

一定時間フレーム辺りの音程情報を得る．

尚，ここでいう音程は基本周波数のことである．これを用いることで，階名の推定や，音符の時間区間の検出等を行う．

今回のシステムでは自己相関ピッチ検出と，検出した周波数

候補から特定する際に DP マッチングを組み合わせた手法を利用している．

（2） 取り出した音程情報から，各音符に対応する時間区間に区切る

抽出した音程情報を元に，時間区間の検出を行う．音高が大きく変化した時のフレームを音の立上りとみなし，各音符の時間区間を検出する．

具体的には，有音時の音程情報の過去 f フレーム分の基本周波数の平均値と，現在位置から過去 r フレーム分の基本周波数の平均値を取り，50 セント (半音の半分) 以上差があった場合，現在のフレームは別の音の立上りであると判断する．音程情報を $F0$，フレーム番号を t，過去 f フレーム分の平均を $\bar{f}$，過去 r フレーム分の平均を $\bar{r}$，とすると，音の立上りを判断する式は式 1 から 3 のようになる．

尚，無音区間があった場合，音の切れ目であると判断し，次の有音区間の始まりを立上りとして検出する．

無音区間は本来，休符として扱われるべきだが，残響や雑音等の影響を大きく受けるため，正確な認識を行うのは困難である．また，検索においても，特徴量は音符の立上りを重視し，休符の情報は用いていない．そこで前の音符の音価に無音部分も含めることにする．

$$\bar{f} = \frac{\sum_{i=t-f}^{t} F0(i)}{f} (t > f > 0) \tag{1}$$

$$\bar{r} = \frac{\sum_{j=t-r}^{t} F0(j)}{r} (f > r > 0) \tag{2}$$

$$50 \leqq |1200 * \log_2(\frac{\bar{r}}{\bar{f}})| \tag{3}$$

（3） IOI(オンセット間間隔) から音価を推定する．

まず，最も短い音符の音価に値する IOI，τ_0 を求める．IOI の頻度分布を取り，分布に見られる複数のピークの中で，2 のべき乗系列での IOI が最も短いものの IOI を基準の IOI とする．

具体的には式 4 から 6 を用いて，最も整合度が高い IOI を求め，その時の τ を基準 IOI とする．尚，$T_t(t,\tau)$ は τ の 2 のべき乗系列であれば 1，付点音符であれば 0.5，を返す関数，$F_t(t)$ は IOI の頻度分布，$C_t(\tau)$ は整合度である．

$$T_t(t,\tau) = \begin{cases} 1 & t = 2^n\tau (n = 0,1,2\cdots) \\ 0.5 & t = 1.5 * 2^n\tau (n = 0,1,2\cdots) \\ 0 & (\text{上記以外}) \end{cases} \tag{4}$$

$$C_t(\tau) = \int T_t(t,\tau) \cdot F_t(t) dt \tag{5}$$

$$C_t(\tau_0) = \max C_t(\tau) \tag{6}$$

次に各音に対応する IOI を，求めた τ_0 のおよそ何倍かを調べ，音価の推定を行う．これは単純に τ_0 の 2 のべき乗系列のうち最も近いものを選んでいくだけである．

尚，$T_t(t,\tau)$ は音価推定のテンプレートであり，図 1 のような形になる．

（4） 階名を推定する．

まず，基本周波数の頻度分布を取り，分布に見られる複数のピークが全音，あるいは半音間隔である時の C(ドの音) を基準周波数，f_c とし，求める．

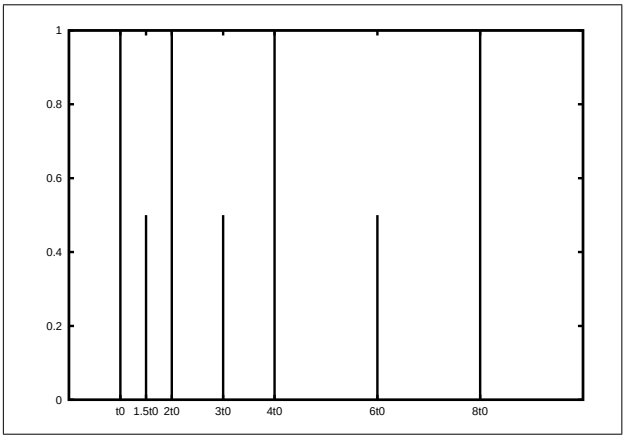

図 1　音価推定テンプレート

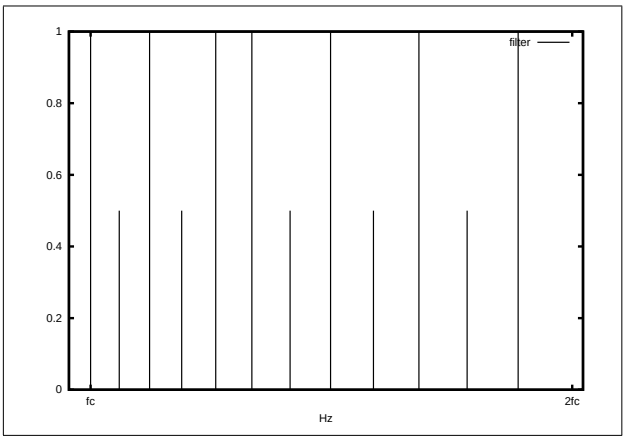

図 2　音程推定テンプレート

具体的には式 7 から 9 を用いて，最も整合度が高い f_c を求め，その時の f_c を基準周波数 f_{c0} とする．尚，$T_f(f,f_c)$ は f_c を C の音とした時，白鍵の音の周波数であれば 1，黒鍵であれば 0.5，を返す関数，$F_f(f)$ は IOI の頻度分布，$C_f(f_c)$ は整合度である．

$$T_f(f,f_c) = \begin{cases} 1 & f = f_c * r^k (k = 0,2,4,5,7,9,11) \\ 0.5 & f = f_c * r^k (k = 1,3,6,8,10) \\ 0 & (\text{上記以外}) \end{cases} \tag{7}$$

但し ($r = 2^{\frac{1}{12}}$) である．

$$C_f(f_c) = \int T_f(f,f_c) \cdot F_f(f) df \tag{8}$$

$$C_f(f_{c0}) = \max C_f(f_c) \tag{9}$$

次に f_{c0} から半音で k 個上の階名の周波数 f_k を求め，1 つの音と判断された時間区間内の基本周波数の平均 (但し後から足し合わされた無音区間は含まない) と比較し周波数が近いものをその階名として推定する．

尚，f_k は式 $f_k = f_{c0} * r^k (k = 0,1,2,\cdots 11)$ で求める．

また $T_f(f,f_c)$ は音階推定のテンプレートであり，図 1 のような形になる．

（5） 推定した音符情報を MIDI ファイルとして出力する

出力前に音符情報に以下の処理を加えておく．

- 移動ドの補正
- ノイズ音の消去
- テンポと最小音価の推定

移動ドの補正は f_{c0} が正規音階でどの階名に当たるのか求め，C との差分，固定値を加えるだけで良い．これは MIDI 形式の階名表現が C4 を 64 とし，そこから半音毎に数値を加えた数で表現されるからである．

ノイズ音は隣接する音から大きく外れた音であることが多い．そこで，ノイズ音と判断する閾値を定め，閾値より隣接する音との差が大きい音をノイズと判断し，前の音に吸収，あるいは消去する．尚，閾値は必要に応じて，自由に変えることが出来る．

テンポは取得した基本周波数情報 1 フレーム辺りの時間と τ_0，最小の音価 (初期状態では八分音符) から求める．この時，テンポが入力すると想定されるテンポの倍以上速い場合，最小の音価が合っていないと考える．そこで最小の音価を現在の半分の長さの音価に変え，再びテンポを計算し直す．遅い場合は逆の処理を行う．

この処理を，想定されるテンポの範囲内に収まるか，最小の音価が表現できる最小の長さになるまで繰り返す．

尚、この処理は入力されるハミングのテンポに制限があることを前提としたものであり，変速，極端に遅い，あるいは速いテンポに対して，正確な最小の音価は得られない．

最後に，得られた音符情報やテンポを MIDI 形式のファイルとして出力する．

3. 追加テンプレート

推定に用いるテンプレートの形状と実際に取得したヒストグラムの形状 (図 3) は異なる．またハミングは常に一定の音価や音程に対し，同じ時間区間や基本周波数を取るとは限らない．

特にビブラートによる周波数の揺らぎや発音区間の区切りのぶれ等，考慮すべき部分は多い．

そこで従来のくしの周辺に周波数の揺らぎや発音区間区切りのぶれの影響を受けた値があることを前提としたくしを追加することにより，基準推定の精度を高められるのではないかと考え，新しい形のテンプレートを提案した．

以下では，新しく提案した音階や音価の幅のぶれを考慮したくし型のテンプレートについて述べる．

図 4 に示すテンプレートは従来のテンプレートを基準とし，基準からの離散距離の逆数を取ったものを，それぞれのくしについて付加したテンプレートである．

基準からの距離に反比例する形で，わずかにずれた音程幅を考慮した評価を行うことができる．但し，わずかな差で大きく加算される重みが変わるため，大きい音程幅のずれに対する補償はあまり期待できない．

尚，このテンプレートはずれの発生がずれの距離に対し大きく減衰していく傾向である可能性を試すために考案したものである．

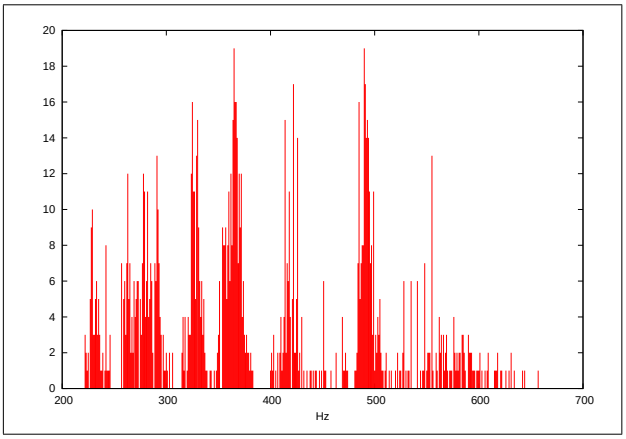

図 3　取得したヒストグラムの例

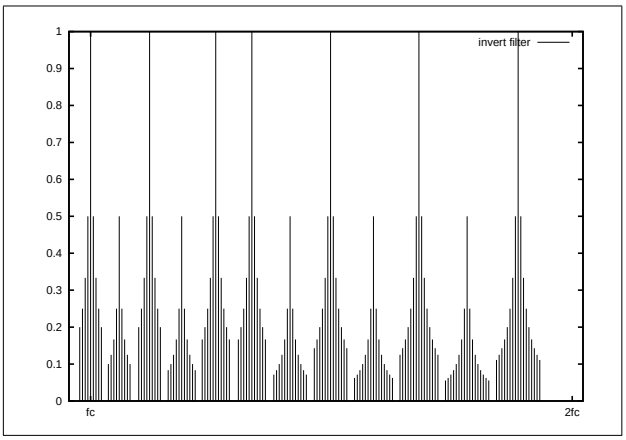

図 4　反比例型テンプレート

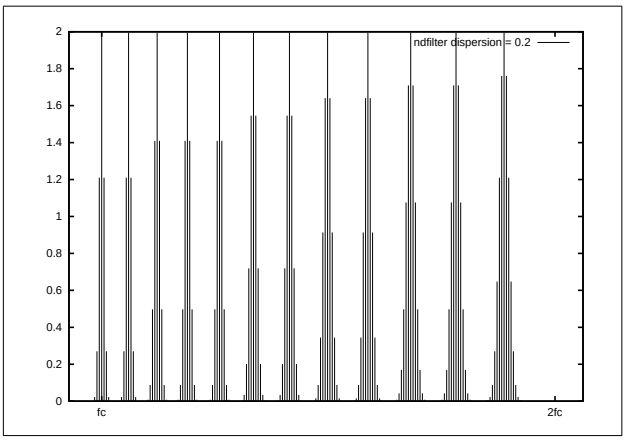

図 5　正規分布型テンプレート (分散値 0.2)

図 5 に示すテンプレートは従来のテンプレートを中心とする分散値 σ の正規分布を利用したテンプレートである．尚，図 5 は分散値 0.2 を取る時の物である．

このテンプレートはずれの発生がずれの距離に対し，正規分布に従う傾向を持つ可能性を試すために考案した．

また，分散値を調整することで，任意の音程幅のずれを考慮できる特性を持ったテンプレートに変えることが可能である．

反面，精度は分散値と音程幅のずれの大きさが噛み合うかどうかに左右される可能性がある．

4. 実　　験

採譜手法による検索精度の違いを検証するため，以下のような実験を行った．

4.1 実験方法

検索手法は Earth Mover's Distance(EMD) を用いた類似音楽検索手法 [1] を用いる．検索対象の楽曲データベースは全 4,050 曲の MIDI データで構成され，童謡，ポップス，演歌等，幅広いジャンルを扱っている．

特徴量データベースは各楽曲から主旋律を抽出し，スライディング・ウィンドウ方式によって，メロディ片を生成することで別途作成する．また，分割は八分音符を 1 拍，ウインドウ長 16 拍 (四分の四拍子における 2 小節分)，スライド長 4 拍として行った．

入力には事前に録音した 177 個のハミングデータを用いる．録音したハミングは不特定多数の男女のもので，保存形式は wav 形式，サンプリング周波数は 44.1kHz である．

ハミングには特に制限を設けておらず，長さ，音量，テンポ，正確さ等は不揃いである．従来の採譜手法との比較のため，テンプレートを用いない従来の採譜，くし型のテンプレートを用いた採譜，拡張したテンプレートを用いた採譜 (反比例型，正規分布型，分散率 0.1，0.2，0.3，0.4，0.5 の 5 つ)，の 3 つの方法で採譜を行う．

採譜の後，特徴量データベースと同じ手法で特徴量を抽出，それを入力キーとして検索を行う．

尚，特徴量抽出の際，スライド長のみ 1 拍として行う．これはハミングの開始位置の差による誤りを少なくするためである．また採譜に必要な幾つかのパラメータは 3 つの手法とも全て同じ値で固定してある．

更に音価推定と音程推定のそれぞれに有効なテンプレートを検証するため，複数のテンプレートの組み合わせでも採譜を行う．

4.2 結　　果

結果の図は横軸は順位を，縦軸は横軸の順位までに正解曲が検索できたハミングの割合を表している．

まず，提案したテンプレートのうち正規分布型の分散率の違いによる結果を示す．

図 6 のグラフは上からそれぞれ，分散率 0.1，0.2，0.3，0.4，0.5 を音価推定に適用した結果である．尚，音程推定はテンプレートを用いない従来の手法を適用した場合のものである．

分散率が 0.3 の場合を除けば，基本的に分散率が低いほど精度が高くなる傾向が見られる．

この傾向は音程推定のテンプレートの種類を問わず同じ結果であった．

図 7 のグラフは上からそれぞれ，分散率 0.1，0.2，0.3，0.4，0.5 を音程推定に適用した結果である．尚，音価推定はテンプレートを用いない従来の手法を適用した場合のものである．

20 位付近で分散率が高いほど精度が高くなる傾向が見られたものの，他の部分では入り交ざりはっきりとした差が現れていない．

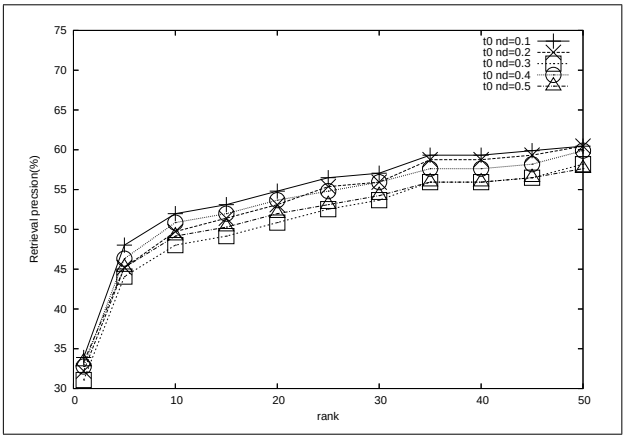

図 6　正規分布型テンプレートを用いた音価推定における分散率毎の検索精度

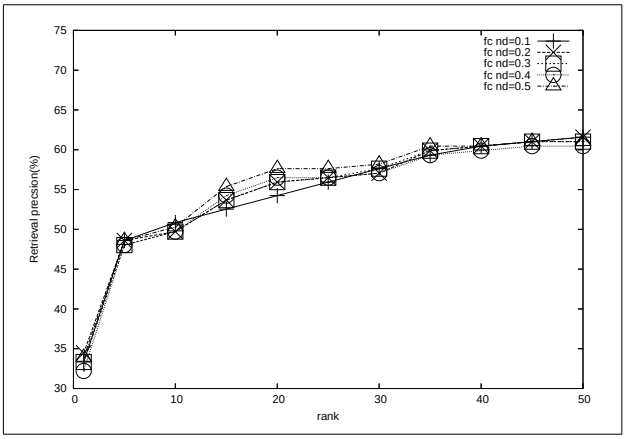

図 7　正規分布型テンプレートを用いた音程推定における分散率毎の検索精度 (音価推定はテンプレートなし)

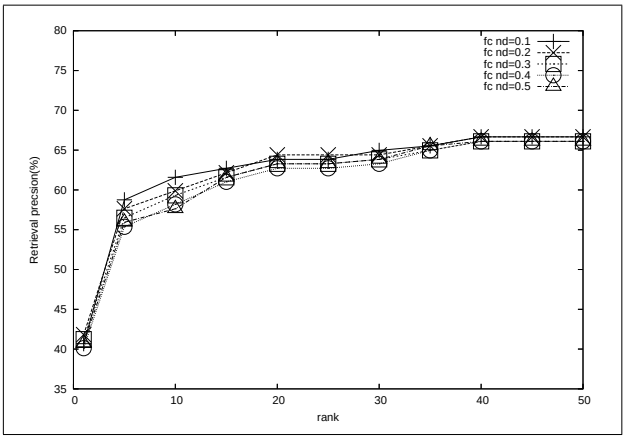

図 8　正規分布型テンプレートを用いた音程推定における分散率毎の検索精度 (音価推定は従来テンプレート)

また，音価推定に従来のテンプレートを用いた時の場合，違った傾向が見られた．音価推定に従来のテンプレート，音程推定に正規分布型のテンプレートを用いた時の結果を図**??**に示す．

こちらは 10 位付近で分散率が低いほど精度が高くなるという逆の傾向が見られた．

次に音価推定，音程推定，それぞれのテンプレートの違いによる検索精度の違いを図 9,10 に示す．

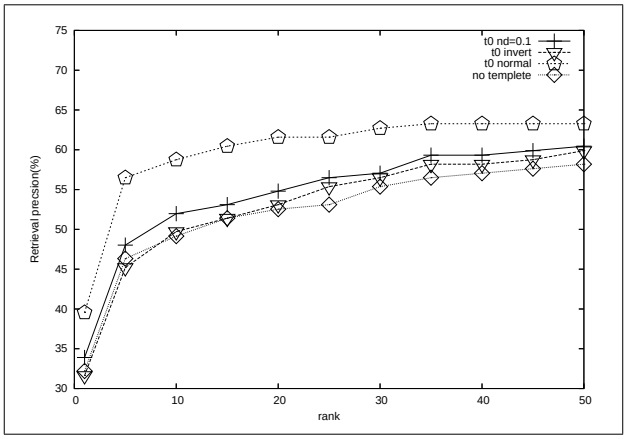

図 9　音価推定のテンプレートの違いによる検索精度

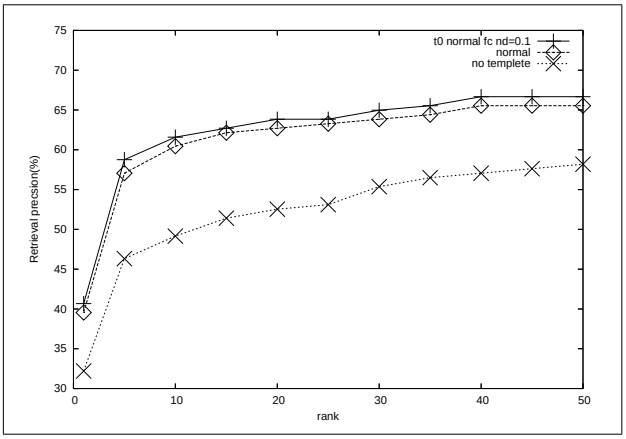

図 11　音程推定のテンプレートの違いによる検索精度

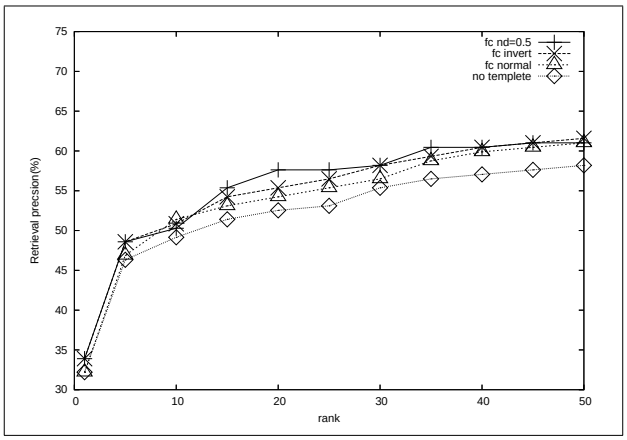

図 10　音程推定のテンプレートの違いによる検索精度

図 9 のグラフは上からそれぞれ，音価推定に反比例型テンプレート，正規分布型テンプレート，従来のテンプレート，テンプレート不使用，を適用した結果である．尚，どのテンプレートにおいても音程推定にはテンプレートを用いておらず，正規分布型テンプレートの分散値は最も良い結果であった 0.1 の時を示した．

従来のテンプレートを用いた場合大幅な精度向上が見られた．反比例型テンプレートが 20 位未満においてテンプレートを用いない場合以下の精度となったが，概ねテンプレートを用いることによって音価推定の精度向上が見られた．

図 10 のグラフは上からそれぞれ，音価推定に反比例型テンプレート，正規分布型テンプレート，従来のテンプレート，テンプレート不使用，を適用した結果である．音価推定の精度比較時と同様に，音価推定にテンプレートを用いていない時のものを示した．尚，正規分布型テンプレートの分散値は音価推定にテンプレートを用いない時最も良い結果であった 0.5 を用いた．

テンプレートを用いた場合，精度が向上した．用いたテンプレートの中でも正規分布型のテンプレートにおいて最も精度の向上が見られた．

最後に今までの結果からそれぞれの推定において最も効果的だったテンプレートの組み合わせを用いたもの，従来のテンプレートを両方の推定に用いたもの，テンプレートを全く用いないものの結果を図 11 に示す．

尚，最も効果的なテンプレートは，音価推定においては従来のくし型テンプレート，音程推定においては正規分布型テンプレート (但し分散値は 0.1) を選択した．

テンプレートを用いることにより大幅に精度が向上し，更に音程推定において新しく提案したテンプレートを用いることで僅かに精度が向上した．

4.3　考　　察

音価推定において，従来のテンプレートを用いることで大幅な精度向上が見られた一方，提案した拡張テンプレートで精度が伸び悩んだ理由について考える．

テンプレートを用いて最適な基準を求めること自体は精度向上に繋がっている．これは従来のテンプレートによる結果から明らかである．提案したテンプレートはこの最適な基準を求める際に誤りがあったのではないかと考察する．

実際に求めた基準の値を各手法で比べてみたところ，テンプレートを用いない場合は，最小の IOI の値そのままのため小さい値が並んだ．一方拡張テンプレートを用いた場合，従来のテンプレートの推定値よりも大きい値が並んだ．これはぶれを許容しようとするあまり，飲み込んではいけない差分も誤差として判断してしまったということである．そのため本来は 16 分音符や 8 分音符といった音価として採譜されるべきものが，ノイズと判断されより大きな音価として採譜されたのである．

実際に正規分布型テンプレートにおいて分散率を小さくし，ぶれの許容量を少なくすることで精度が向上したことからも，ぶれの許容が過ぎていたことが読み取れる．

このことからぶれの許容には限度があり，特に頻度分布の持つ情報が少なく偏りやすい音価の推定にはあまり適していなかったことが分かる．

音程推定においても概ね望む結果が得られた．しかし，正規分布型テンプレートにおける分散率の違いによる傾向を一度考察する必要がある．

本来，音程推定の結果が類似度に与える影響は音価推定の結果が与える影響に比べると低い．これは元々検索システムがリズムに対して重みをおいていることに起因する．

そのため推定の精度が上がると，類似度が全体的に僅かに上昇した分検索順位が順次繰り上がり，音程推定の結果に添う形

のグラフが形成されることが予測される．

しかし実際の結果を見るとかなり変動が激しい．

細かく見ていくと 15 位前までは分散率の低いものが高い精度を取っている．しかし，15 位を境に分散値 0.5 が急に伸びていることが分かる．

ここから考えられるのがハミングによって相性の良い分散率の違いである．

元から上位に検索されていたハミングはある程度正確であるものが多い．

そのため揺らぎの考慮がなくても採譜がきちんと行われていた可能性が高い．逆に音価推定であったように許容が過ぎてしまうことによる下落の発生が考えられる．

一方順位が後半になるにつれ，ハミングは揺らぎの多いものが増えていく．

そのため揺らぎを考慮している分散率の高いテンプレートが伸びたのだと予想できる．

最終的に順位が下がる程，精度が収束していることから音程推定の精度が全体に与える影響はやはり当初の予測通り，音価推定が精度に与える影響に比べて小さいことが分かる．

このことから 15 位前後の変動は分散率によって類似度に影響を与えたハミングが異なっていたことが原因であると予想される．

これを応用し，ハミングの基本周波数のヒストグラムの形状に最適な分散率を自動的に選択することが可能になれば，より上位に検索出来る，つまり更に精度を向上させることが可能である．

全体を通して見ると推定にテンプレートを全く用いない場合に比べ，最も効果的であったテンプレートの組み合わせを用いた場合，10 %程の精度向上が見られた．このことからハミング毎に異なる基準を考慮し推定する手法が検索精度の向上に繋がることが分かった．

一方検索精度そのものに関しては，採譜の精度がそのままボトルネックとして表に出る形となった．

検索システムそのものの検索精度は [1] によると，入力時にテンポや長さの制限を設けたハミングに対して 80 %から 90 %の値を得られている．今回はユーザの自然な入力を考慮しハミングに制限を設けなかったため，検索キーとなるハミング片の数の違いによる検索精度の変化も考えられるが，検索に支障があるほどの短いハミングはごく少数であり，他の違いは採譜部分で補う形になっている．つまり新しく追加した採譜部の誤りにより検索精度が下落していることが伺える．

このことから採譜部の更なる改善が必要である．

5. おわりに

本稿では類似音楽検索における入力ハミングの採譜手法としてくし型のテンプレートを用い，基準のずれを考慮した採譜を行うことを述べた．

音価と音程，各推定に異なるテンプレートを適用し，基準を固定した採譜手法との比較実験を行った結果，音価の推定には従来のテンプレートが，音程の推定には拡張したテンプレートが有効であった．

今後更なる改善を行うならば，音価推定においては，発音区間の区切り方に関する手法の検討が必要である．

今回はテンプレートによって音価の基準を推定することで精度向上を図ったが，それ以前の問題として存在する正確に発音区間が区切れないことによって発生する誤りに対しての補償は全く行われていない．

この問題を解決することによって更なる精度の向上が望めるはずである．

音程推定においては，考察でも述べたように基本周波数のヒストグラムに合わせたテンプレートを自動的に選択する手法を確立させる必要がある．

また今回は固定していた有音区間を区切るパラメータの最適値の推定や，3 連符の考慮，テンポ推定方法等の改善も採譜の正確性を高めるために後々必要であると考えられる．

文　献

[1] 大西泰代，獅々堀正幹，柘植覚，北研二：Earth Mover's Distance を用いたハミングによる類似音楽検索の改良手法，2007.

[2] 清水純，丸山剛志，三浦雅展，柳田益造：ハミングによる単旋律の自動採譜，情報処理学会研究報告，MUS-57，2004.

[3] 佐倉卓馬，尾崎昭剛，原尾政輝：拍節認識を用いた自動採譜システム，火の国情報シンポジウム 2005 論文集，A-1-1，2005.

[4] 櫻庭洋平，奥乃博：自動採譜におけるパート形成処理のための特徴量の検討，情報処理学会研究報告，MUS-51，2003.

[5] 亀岡弘和，齊藤翔一郎，西本卓也，嵯峨山茂樹：Specmurt における準最適共通調波構造パターンの反復推定による多声音楽信号の可視化と MIDI 変換，情報処理学会研究報告，MUS-56，2004.