音程変化に基づく歌唱音声の音符区間検出

岡松 竜徳 † 鈴木 基之 † 任 福継 †
†徳島大学

1 はじめに

近年，実際に自分で歌を歌って音楽を検索できるようなシステムの研究が行われている [1, 2]．従来のメロディー検索システムの多くはハミングを入力とし，歌唱を「タタタ」といった鼻唄に限定している．そのため，無声区間に音声パワーがほとんどないことから，音声パワーの差分を使って音符区間を切り出し，音程特徴量を計算している．しかし，この歌唱法は「タタタ」といった鼻唄に限定されるため，ユーザーに負担を強いることとなり，望ましくない．そこで本稿では，歌詞で歌って楽曲を検索するために，任意の歌唱中の音符区間検出を行う．

歌唱の音程変化に注目すると，ひとつの音符に対応する区間のピッチはほぼ同じとなることが予想される．そのため，ピッチが大きく変化した時刻が音符の区切りにあたると考えられることから，音符区間の区切り推定を行う．ここで，ピッチを直接計算するのは誤差が大きいことから，ピッチを抽出しない方法として，相互相関関数を用いた音程推定法を利用する [3]．

2 音程変化に基づく音符区間検出法

2.1 相互相関関数を用いた音程推定法

任意の歌唱中の音符区間を検出するために，音符区間の区切りの推定を行う．歌唱中の音程変化に注目すると，ひとつの音符に対応する区間のピッチはほぼ同じ値となることが予想される．そのため，ピッチ変化が大きく変化した時刻が音符の区切りにあたると考えられる．ピッチを直接計算するのは誤差が大きいことから，ピッチを抽出しない方法として，相互相関関数のピークを用いた音程推定法を利用する．ひとつの音符に対応する区間のフレームを切り出し，フレームの対数周波数領域のパワースペクトルを求め，隣り合うフレームの相互相関関数を計算し，ピークをとることで音程を計算する．パワースペクトルは対数周波数の変化量だけ対数周波数スペクトルの並行移動した形でほぼ表現されるため，ピッチを直接計算しなくても音程特徴量を抽出することができる．

Detection note are based on the musical Interval changes
† Tatsunori Okamatsu
† Motoyuki Suzuki
† Fuji Ren
Tokushima University (†)

2.2 提案方法

入力を任意の歌唱で音符区間を検出する場合，音符区間が連続音声となってしまうので，従来の音声パワー差分のみで音符区間を検出するのは困難である．

そこで，我々は歌唱中の音程変化に着目した．ひとつの音符区間のピッチは，ほぼ同じピッチとなるので，ピッチが大きく変化した時刻が音符の区切りにあたると予想される．しかし，ピッチの推定には誤差が伴うことから，相互相関関数を用いた音程推定法を利用する．音符の計算をすると，同じ音符なら相互相関関数のピークの値は0となり，変化すると値が変わる．変化した値でも，同じ音符区間同士の組合せなら計算した値は同じとなる．1曲をフレームに分割し，全てのフレームの組合せについて音程を推定する．推定結果をx軸，y軸をそれぞれフレームとし，z軸方向に音程をプロットすると，音符の区切り付近で段差のある市松模様になる．このことは前回の実験で証明している [4]．音声パワーのみで区切れなかった区間に対して，y軸方向に微分フィルタを通し，二乗した値をx軸方向に加算して，合計で変化の大きい値が音符の区切りである可能性がある．

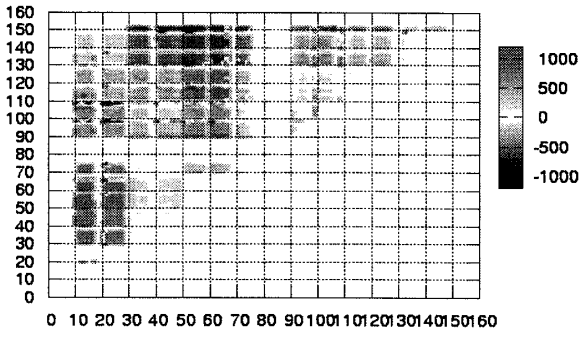

図1: 市松模様の図

このグラフから音符区間を検出する具体的アルゴリズムは以下のとおりである．

Step1 1曲間を一定間隔でフレームにわけたあと，音声パワーの小さなフレームを閾値により無声区間と判定し，計算から除外する．連続したパワーの小さいフレームの最後の時刻を音符の区切りとする

Step2 音声パワーの高いフレームに対し，すべての組合せについて音程を計算

Step3 市松模様のグラフを書き (図 1)，音程の変化を強調するために y 軸方向に微分する

Step4 微分された値を二乗して，x 軸方向にすべて加算し，そのピークが閾値以上であれば，音符の区切りとする

3 評価実験

提案方法の有効性を示すために従来方法であるパワー差分を用いた音符区間検出法と提案方法の比較実験を行う．歌唱者データベースを用いて，以下の条件で実験を行う．

3.1 実験条件

Music Database	about 404 singing/humming clips from about 26 persons
Sampling Frequency	16kHz
frame size	64ms (Hanning window)
frame shift	8ms

相互相関関数のピークの値が-1200～1200 の間の値にならない場合は，計算から削除する．その理由は，相互相関関数のピークの値がうまく計算されないことがあるからである．また，音符の区切りが前後 2 フレームにあたる部分を正解とし，認識実験を行う．

3.2 実験結果

前回の実験結果から，音声パワーが一定値以下のフレームには，必ず音符の区切りが存在した [4]．図 2，3 のグラフは x 軸をフレーム，y 軸方向に微分フィルタを通して，x 軸方向に二乗して加算した値を y 軸として，グラフ化したものである．緑線のグラフは音符の始まりの部分であり，正解の区切りの位置である．音声パワーが低いために計算されてないフレームについては，0 を代入している．

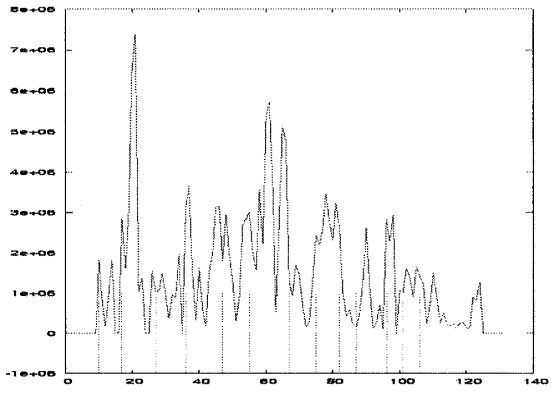

図 2: どこかで春が

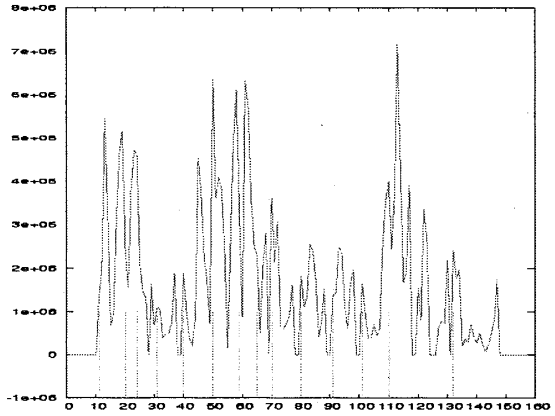

図 3: 春が来た

このグラフから y 軸の値 0 が連続し，その次の値には必ず音符の区切りがある．このことから，音声パワーが一定値以下のフレームの次には必ず音符の区切りがあることがわかる．しかし，図 3 のように，音声パワーが一定値以下の次に区切りがない場合がある．これは 120～123 フレーム目には息継ぎがあり，息継ぎを音符区間と間違えて認識してしまっている．音符の区切りの正解部分に着目すると，音符の始めと終わりのフレームには必ずピークがあることがわかる．

4 おわりに

本稿では音程変化に基づく歌唱音声の音符区間検出法について提案した．また，図 2，3 のグラフから閾値の設定が困難なことがわかる．今回の実験を基に，最適な閾値の設定，Δ幅を変更して評価実験をしていく予定である．

参考文献

[1] 獅々堀正幹，大西泰代，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手法，” 情報処理学会論文誌，vol.48，no.1，pp.300-311，2007.

[2] 後藤真孝，平田圭二，“音楽情報処理の最近の研究，” 日本音響学会誌，60 巻 11 号，pp.675-681，2004.

[3] 市川拓人，鈴木基之，伊藤彰則，牧野正三，“音程特徴量の確立分布を考慮したハミング入力楽曲検索システム，” 情報処理学会研究報告，vol.2007，no.81，pp.33-38，Aug 2007.

[4] 岡松竜徳，鈴木基之，任福継，“フレーム間の相互相関関数を用いた歌唱中の音符の区切り推定，” 電気学会電子・情報・システム部門大会講演論文集，平成 21 年 9 月，p.734，2009.