

音符区切り情報を用いた高精度歌唱音声認識

鈴木 基之^{1,a)} 杉田 裕亮^{1,†1}

概要：歌唱音声を入力として歌詞を認識しようとした時、通常の朗読音声に対する音声認識精度と比較して大きく劣化する。その原因のひとつは、歌詞中の各モーラがそれぞれ音符と対応づけられているため、一部の母音が長音化し、挿入誤りが増加してしまう事である。そこで本論文では、別途推定した音符の区切り情報を用い、認識仮説のモーラ区切り時刻に制約を加えることで歌詞認識の高精度化を目指す。27名が歌唱した童謡歌唱データを用いて認識実験を行ったところ、単語認識精度が85.7%→92.0%と大きく改善した。

キーワード：歌唱音声認識、音符区切り、特殊フレームの挿入、楽曲検索システム。

1. はじめに

近年音楽圧縮技術が向上し、大量の音楽データをmp3プレーヤーなどの小型デバイスに保存、再生することが可能となった。これらのデバイスは小型であるがゆえに、一般的にインターフェイスが貧弱である。その一方で保存可能な曲数は何千曲と膨大な数になっているため、保存してある曲の中から特定の1曲を指定して再生する、といった操作が非常にやり難い仕様となっている。

こうした問題を解決するため、歌声やハミングを用いて楽曲を検索するシステム(Query-by-Singing/Humming)が開発されてきた。特に歌声を検索キーとして用いる検索システムでは、音の高さや長さといったメロディ情報に加えて、歌声から別途歌詞情報を抽出することで、両方の情報を用いて検索精度を大幅に高めることができる[1]。

しかし、一般に歌唱音声から歌詞情報を高精度に抽出することは困難である。音声波形から歌詞を認識する、という意味においては、通常の音声認識と全く同じ枠組みであるが、通常の(話し声用に開発された)音声認識システムをそのまま用いても、歌声からの歌詞認識精度は非常に低いことが知られている[2]。

その原因のひとつは、歌唱することによる発声の違いである。一般に歌唱音声は通常の話し言葉と比べて音高の変化が多く、また歌唱フォルマントと呼ばれる独特のフォルマントが出現することがある。そのため、話し声から学習した音響モデルを用いた音声認識システムではミスマッチ

がおこり、認識率の低下を招いてしまう。

この問題に対しては、音響モデルを歌唱音声用に適応させることで改善することが可能である。MLLR等の話者適応用アルゴリズムを用い、話し声用の音響モデルを歌唱音声に適応させることで、認識率が大幅に向上することが報告されている[3]。

もうひとつの問題点は、歌唱によって通常ではあり得ないほど長く発声する音素が出現する、ということである。歌唱においては、各モーラは音符と対応付いているため、長い音符に対応付けられたモーラは長音として発話される。通常はHMMが音響モデルとして用いられているため、多少の時間変動は自己ループによる遷移で吸収することができるが、あまりにも長い発話の場合、その区間に別の音素が挿入されてしまう場合が多い。尾関ら[2]は、4分音符以上の長さに対応付けられたモーラを末尾に含む単語についての認識精度が33.95%であり、全体の認識精度である58.76%と比べて大幅に低下していると報告している。そのため、歌詞認識を高精度に行うためには、長く発音されたモーラを含む単語をいかに高精度に認識するか、という問題を解決する必要がある。

この問題に対し、主に言語モデルの観点から、いくつか対処法が提案されている。細谷ら[3]は、言語モデルとしてn-gramではなく有限状態文法を用いた。楽曲検索システムは自身が持つデータベース中で最も類似した楽曲を検索することから、入力データベース中に存在する曲に限られ、歌詞もデータベース中の曲だけを考えればよい。そのため、事前にデータベースに登録されているすべての曲の歌詞を並列に並べた有限状態オートマトンを構築し、そ

¹ 大阪工業大学 情報科学部

^{†1} 現在、エヌ・ティ・ティシステム開発

^{a)} moto@m.ieice.org

れを言語モデルとして用いている。こうすることで、不要な単語の挿入を抑えることができ、 n -gram を用いた歌詞認識と比較して、おおよそ 15 ポイント程度の精度向上を実現している。

また川井ら [4] は、音響モデルの歌唱音声適応に加え、発音辞書に長音化した単語を追加で登録（例えば「蝶」という単語に対し、/ch o u/だけでなく、/ch o u u/ や /ch o o u u/といった発音を登録する）することで、20 ポイント程度の精度向上を果たしている。

しかし、こうした方法では根本的な解決にはなっておらず、また精度もまだまだ十分とは言えない。そこで本論文では、音符の区切り時刻の情報を用い、明示的に各音素長を制限することで、挿入誤りを低減させた歌唱音声認識法を提案する。

2. 認識仮説におけるモーラ長を制限した音声認識アルゴリズム

2.1 基本的な考え方

歌唱音声における各モーラの継続時間長は、メロディに制約される。各モーラはメロディを表す音符に対応しているため、長い音符に対応したモーラは長く、短い音符に対応したモーラは短く発話されることとなる。特に日本語の場合、ひとつの音符にひとつのモーラが対応付くことが多いため、音符の長さがわかればモーラの長さが推定でき、音声認識時に余計なモーラが挿入されている仮説を棄却することが可能となる。

そこで、入力歌唱中の各音符の区切り時刻を推定し、その時刻でのみモーラ間の遷移を許すようにした音声認識法を提案する。音符の区切り時刻を推定する方法は“On-set detection”と呼ばれ、様々な方法が提案されている [5]。その結果を用い、音符の区切り時刻ではないフレームにおいては、次のモーラへの遷移を許さない（ひとつのモーラ区間であるように制限する）ことで、長い音符に対応した区間もひとつだけのモーラと対応付くことになり、挿入誤りを大幅に低減できると思われる。

2.2 実装の詳細

前節で提案した基本的な考え方は、既存の認識アルゴリズムを変更することなく、実装のレベルで実現可能である。

まず、通常の音声データから得られる特徴量ベクトルとはかけ離れた値を持つ特殊な特徴量ベクトル（以下、「マーカーベクトル」と呼ぶ）を定義する。このマーカーベクトルを、入力された歌唱音声から計算された特徴量ベクトル系列中の、音符の区切り時刻に対応するフレームにひとつづつ挿入していく。こうすることで、特徴量ベクトル系列でどの位置に音符の区切りがあるか、判別可能としておく。

また、音声認識を行う音響モデルに、マーカーベクトルに対応した特殊な音素 HMM（以下、「マーカー HMM」と

呼ぶ）をひとつ追加する。このマーカー HMM は自己ループを持たない状態 1 つだけからなる。その状態が持つ出力確率分布の平均値はマーカーベクトルが持つ特徴量と一致させ、分散は小さな値を定義する。こうすることで、マーカー HMM はマーカーベクトルに対しては非常に高い尤度を出力するが、その他の（通常の音声から得られる）特徴量ベクトルに対しては非常に低い尤度を出力することとなる。

最後に発音辞書を変更する。各単語について、モーラの切れ目となる位置にマーカー HMM を挿入する。例えば「音声」という単語であれば、/o N sei/ という定義を変更し、/o # N # se # i #/ とする（“#” はマーカー HMM を表す音素記号）。

このようにすることで、認識仮説中のモーラの時間長を、音符長と一致させることが可能となる。発音辞書でマーカー HMM をモーラ間に挿入することで、モーラ間を遷移する際には、必ずマーカー HMM が使用されることとなる。しかし、マーカー HMM は特殊な出力確率分布を持つため、マーカーベクトルの位置と対応付けないと尤度が極端に低くなり、仮説が棄却されてしまう。結果として、モーラ間の時刻にはマーカーベクトル（= 音符の区切り時刻）が対応している仮説のみが残るため、長音化による不要な単語挿入を抑制することが可能となる。

3. 1 対多に対応している歌詞への対処

3.1 1 対多対応への対処法

前節で提案したアルゴリズムでは、歌唱音声中のひとつの音符は、必ずひとつのモーラと対応することとなる。しかし一般には、ひとつの音符に複数のモーラが対応していたり、逆に複数の音符にひとつのモーラが対応していたりする曲も存在する（図 1）。こうした曲の歌唱が入力されると、前節のアルゴリズムでは必ず誤認識が起きてしまう。

そこで、マーカーベクトルが持つ特徴量の値や、マーカー HMM が持つ出力確率分布の値を適切に設定することで、こうした 1 対多対応をした単語に対しても誤認識させない方法を提案する。

ひとつのモーラに複数の音符が対応している曲（例えば、図 1 の右側の曲）の場合、ひとつのモーラ区間（マーカー HMM を含まず、すべて通常の音素 HMM が対応すべき区間）中にマーカーベクトルが出現することとなる。そのため、こうした仮説の尤度が極端に低くなることを避

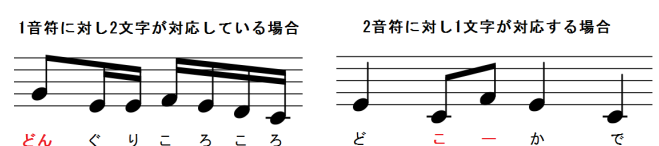


図 1 音符と歌詞が 1 対多に対応している楽曲の例

けるためには、通常の HMM がマーカーベクトルに対してもそれなりの尤度を出力するようにする必要がある。そこで、マーカーベクトルが持つ特徴量の値を通常の音声とかけ離れた値とするのではなく、それなりに類似した値に設定する。

具体的には、事前に大量の音声データから i 次元目の特徴量の平均 μ_i と標準偏差 σ_i を計算しておき、マーカーベクトルが持つ i 次元目の特徴量の値を $\mu_i + n\sigma_i$ と設定する。ここで、 n は実験的に決められるパラメータである。

一方、ひとつの音符に複数のモーラが対応している曲（例えば、図 1 の左側の曲）の場合、音符区切りがない区間（マーカーベクトルが存在しない区間）にマーカー HMM を対応付ける必要がある。そのためには、マーカー HMM が通常の音声から得られた特徴量ベクトルに対してもそれなりの尤度を出力する必要がある。そこで、マーカー HMM が持つ出力確率分布を変更し、通常の特徴量ベクトルに対してもそれなりの尤度を与えるようにする。

具体的には、出力確率分布の平均値（＝マーカーベクトルが持つ特徴量の値）を通常の特徴量ベクトルの値にそれなりに近づけた上で、分散の値を大きくし、平均値から離れたベクトルに対してもそれなりの尤度を出力するように変更する。

3.2 マーカー HMM 等に設定する値と認識性能の関係

すべての曲において、音符とモーラが 1 対 1 の関係であれば、マーカーベクトルの特徴量の値は通常の特徴量ベクトルとはかけ離れた値とし、またマーカー HMM が持つ出力確率分布の分散は狭くしておいた方がよい。しかし、3.1 節で指摘したように、実際には 1 対多の対応があるため、ある程度値を通常の特徴量ベクトルに近づけ、分散も拡くする必要がある。

そうした場合、逆に 1 対 1 の対応をしている部分においてはモーラ間の遷移時刻の制約が弱くなってしまいうため、誤認識を発生させる可能性がでてくる。ここでは、マーカーベクトルが持つ特徴量とマーカー HMM の出力確率分布の分散の設定を変更すると、認識精度がどのように変化するかを定性的に検討する。

特徴量と分散の設定によって認識精度がどのように変化するか、音符とモーラの対応ごとにまとめた結果を表 1 に示す。ここで上矢印は認識率が向上する方向であることを、下矢印は低下する方向であることを示している。

まず、1 音符に対して 1 モーラの対応の場合、マーカーベクトルが区切りの役割を果す時に精度が向上すると思われる。つまり、マーカーベクトルが持つ特徴量の値が通常の特徴量ベクトルからかけ離れている場合と、マーカー HMM が持つ出力確率分布の分散が極端に狭く、通常の特徴量ベクトル（のほとんど）に対して低い尤度を出力する場合に認識精度が向上する。

表 1 特徴量と分散を変えることによる認識精度の変動

| 平均ベクトル | 通常音声に近い | | かけ離れた値 | |
|-------------|---------|----|--------|----|
| | 狭い | 広い | 狭い | 広い |
| 1 音符対 1 モーラ | ↑ | ↓ | ↑ | ↑ |
| 1 音符対多モーラ | ↓ | ↑ | ↓ | ↓ |
| 多音符対 1 モーラ | ↑ | ↑ | ↓ | ↓ |

一方、1 音符に対して多モーラが対応する曲の場合、マーカーベクトルのない区間にマーカー HMM が対応づけられることが必要なため、マーカー HMM の出力確率分布が通常の特徴量ベクトルをそれなりに高い尤度で出力する必要がある。つまり、平均値が比較的近く、また分散も拡くする必要がある。

最後に多音符に対して 1 モーラが対応する曲の場合、マーカーベクトルを含む区間と通常の音素 HMM を対応づける必要があるため、マーカーベクトルに対して通常音素 HMM がそれなりに高い尤度を出力する必要がある。そこで、マーカーベクトルが持つ特徴量の値を通常の特徴量ベクトルの値に近づける必要がある。

このようにまとめてみると、すべての曲に対して認識性能が向上すると思われる設定は存在しない。そこで、マーカーベクトルが持つ特徴量をどの程度通常の特徴量ベクトルに近づけるのか、またマーカー HMM が持つ出力確率分布の分散をどの程度拡げるのかについて、実験的に検討する。

4. 歌唱音声の認識実験

4.1 実験条件

4.1.1 歌唱音声データベース

実験には、徳島大学で収録された歌唱音声データベース [6] を使用した。これは、童謡 48 曲について 27 名（男性 19 名、女性 8 名）がアカペラ歌唱を行ったもので、おおよそ 1 名あたり 7 曲、全 198 データを用いた。これらのデータについて、音符と歌詞中の各モーラがどのように対応しているかを調べた結果を表 2 に示す。童謡というジャンルの特性上、音符とモーラが 1 対 1 で対応している部分が非常に多くあった。一方で、1 対 2 や 2 対 1 に対応している部分が所々見られた。1 対 3 以上の対応関係は存在しなかった。

また、音符の区切り時刻情報については、本来であれば入力歌唱から自動推定を行うべきであるが、今回は提案方法の有効性を確認するため、別途事前に正解となる区切り時刻情報を求め、それを用いた。正解となる区切り時刻情

表 2 音符とモーラの対応別のデータ数内訳

| | 2 音符対 1 モーラの対応 | | |
|----------------|----------------|-----|----|
| | 含まない | 含む | |
| 1 音符対 2 モーラの対応 | 含まない | 103 | 21 |
| | 含む | 61 | 13 |

表 3 マーカーベクトルの有無による認識結果の例

| 音符とモーラの対応 | マーカーベクトル | 認識結果 |
|-------------|----------|----------------------------|
| 1 音符対 1 モーラ | なし | ド は ドーナツ の ドーナツ では レモン の レ |
| | あり | ド は ドーナツ の ド レ は レモン の レ |
| 1 音符対 2 モーラ | なし | どんぐり コロコロ どんぶりこ |
| | あり | ド 栗 コロコロ どこ か と |
| 2 音符対 1 モーラ | なし | どこ か で 春 が |
| | あり | どこ お か で 春 る が |

報は、まずデータベース中の全データを用いて monophone を学習し、それを用いて歌唱データの強制アライメントを行った。得られた結果（歌唱データ中の各音素区切り時刻）と楽譜の情報（歌詞と音符の対応関係）から、各音符の区切り時刻を人手によって求めた。

4.1.2 使用した音声認識システム

音声認識システムは julius[7] を用いた。音響モデルは、評価用歌唱者 1 名のデータを除いた 26 名分の全歌唱データから monophone を学習し、評価用歌唱者を変えながら 27 回の認識実験を行い、結果を平均した。言語モデルは 48 曲の歌詞データのみから trigram を学習して用いた。使用した単語数は 314 単語であり、未知語なしの条件である。

また、マーカー HMM の出力確率分布は単一正規分布とし、その平均ベクトルは 3.1 節で説明したように、全歌唱データの特徴量ベクトルの平均とその標準偏差を元に、いくつかの場合を設定した。共分散行列は対角共分散とし、各次元の分散の値はすべての次元で 10^{-4} としたもの（以下、「狭い分散」と表記）と、すべての歌唱音声から学習したすべての音素 HMM のすべての共分散行列の値の平均値を各次元ごとに計算したもの（以下、「平均的分散」と表記）の 2 種類を設定した。

4.2 マーカーベクトル挿入による効果

まず、音符の区切り時刻位置にマーカーベクトルを挿入することによる効果を検証した。マーカーベクトルの特徴量（= マーカー HMM の出力確率分布の平均ベクトルの

値）は、通常の特徴量としては表れないような値（ 10^5 ）に設定し、またマーカー HMM の出力確率分布の共分散行列は「狭い分散」とした。

認識精度を図 2 に、また認識結果の例を表 3 に示す。これを見ると、1 音符に対して 1 モーラの対応しか含まない曲（“1 vs 1”）に対しては目論見どおり認識精度が大幅に向上していることがわかる。認識結果の例を見ても、長音化している「ド」が、「ドーナツ」と認識されるのではなく、正しく「ド」と認識されている事がわかる。

一方で、1 音符に対して 2 モーラの対応を含む曲（“1 vs 2”）や 2 音符に対して 1 モーラの対応を含む曲（“2 vs 1”）、また、それら両方の対応を含む曲（“Both”）については、大幅な認識精度の低下が見られる。これも 3.2 節で予想したとおりの結果であり、1 音符に対応している「どん」が「ド」と認識されたり、2 音符に対応している「こ」が「こお」と 2 モーラに認識されてしまったりしている。

最終的に 198 データ全体（“All data”）としては、対応付けたことによる効果よりも多対 1 の対応部分における弊害の方が大きな影響があり、認識精度はマーカーベクトルを挿入することで低下してしまっている。

4.3 マーカー HMM の出力確率分布のパラメータ調整による精度変化

多対 1 対応の部分における誤認識を減らすため、マーカー HMM の出力確率分布の平均ベクトルと共分散行列の値を変化させ、その時の認識精度を調べた。 i 次元目の平均ベクトルの値は、通常の特徴量としては表れないような値（ 10^5 ）に加え、 $\mu_i + n\sigma_i$ において n を 0 ~ 3 まで変化させた値で実験を行なった。一方共分散行列の値は、「狭い分散」と「平均的分散」の 2 通りを設定した。

4.3.1 1 音符と 2 モーラが対応している曲の認識精度

まず、1 音符に対して 2 モーラが対応している部分を含む曲（61 データ）についての認識精度を図 3 に示す。この場合、マーカー HMM と通常の特徴量ベクトルが対応付く必要があるため、表 1 で示したとおり、分散を拡くとり、平均ベクトルを通常の特徴量 HMM に近づけることで対処することができる。結果を見ると、「平均的分散」を設定した上で、平均ベクトルの値が $\mu + 2\sigma$ 程度までであれば、高い認識性能を示すことがわかった。

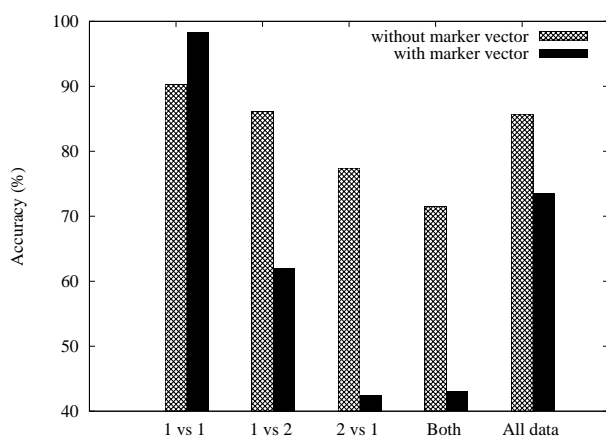


図 2 マーカーベクトルの有無による認識精度比較

4.3.2 2 音符と 1 モーラが対応している曲の認識精度

次に、2 音符に対して 1 モーラが対応している部分を含む曲 (21 データ) についての認識精度を図 4 に示す。こちらは、マーカーベクトルと通常の音素 HMM を対応付ける必要があるため、表 1 で示したとおり、マーカーベクトルの値を通常の特徴量ベクトルに近づける必要がある。結果を見ると、平均ベクトルの値が $\mu + 2\sigma$ 程度までであれば、高い認識性能を示すことがわかった。また共分散行列については、平均ベクトルの値が平均値に近ければ、「狭い分散」でもそれなりの認識率を示すが、平均値から離れていくに従って認識率が低下していくことがわかった。これは、分散が狭いことでマーカーベクトルに対してマーカー HMM が与える尤度が非常に高いものとなるため、平均ベクトルの値が平均値から離れていくに従って通常の音素 HMM がマーカーベクトルに与える尤度との差が大きくなり、マーカーベクトルとマーカー HMM が対応付く (1 音素に 1 モーラを対応付ける) 仮説の方が尤度が高くなっていくためと思われる。そのため、分散の値は「平均的分散」とし、過度にマーカー HMM からの尤度が高くないようにしておく必要があることがわかった。

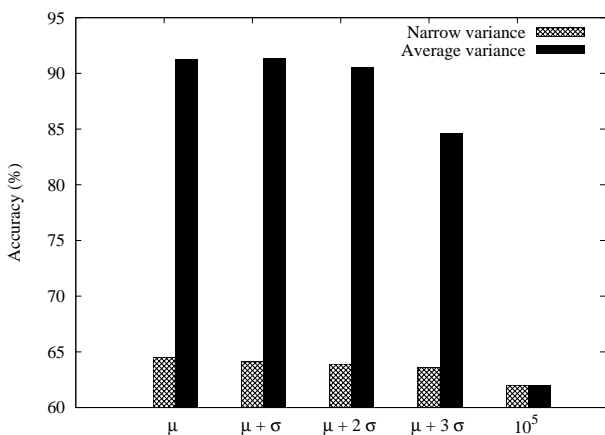


図 3 1 音符対 2 モーラの対応を含む曲の精度変化

4.3.3 1 音符対 1 モーラの対応のみである曲への影響

平均ベクトルや共分散行列の値を変化させるのは、1 対多対応している部分への対処のためであり、1 対 1 対応している部分については、逆に悪影響が及ぶ可能性がある。そこで、1 対 1 対応しか含まない曲について、平均ベクトル等の値を変化させた時の影響を調べた。

1 対 1 対応しか含まない曲 (103 データ) についての認識精度を図 5 に示す。これを見ると、「狭い分散」を設定することで、平均ベクトルの値にかかわらず高い認識精度を示すことがわかる。「狭い分散」にすることでマーカーベクトルに対して非常に高い尤度をマーカー HMM が出力するため、マーカーベクトルが通常の音素 HMM と対応付くこともなく、安定して高い認識精度を出していると思われる。

一方で、「平均的分散」に設定すると、マーカー HMM が出力する尤度が通常の音素 HMM が出力する尤度と大きな差がなるため、平均ベクトルの値を通常の特徴量ベクトルと異なる値にしていけないと、マーカーベクトルが区切りとしての役割を果たせなくなる。実験結果から、 $\mu + 2\sigma$ 程度離せば区切りとしての役割を果たすことができ、高い認識精度を示すことがわかった。

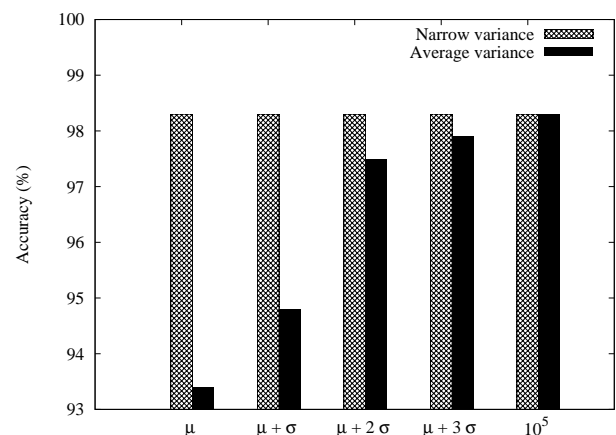


図 5 1 音符対 1 モーラの対応のみである曲の精度変化

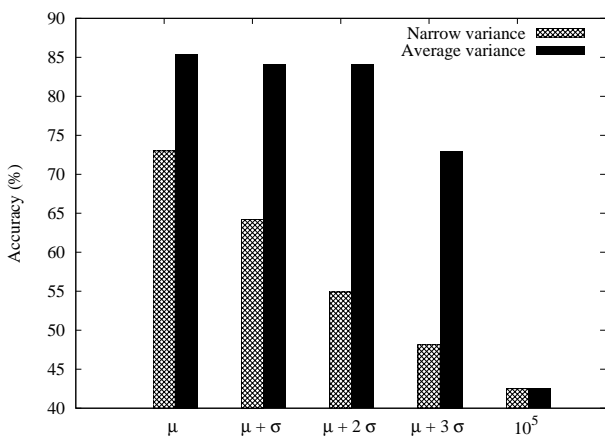


図 4 2 音符対 1 モーラの対応を含む曲の精度変化

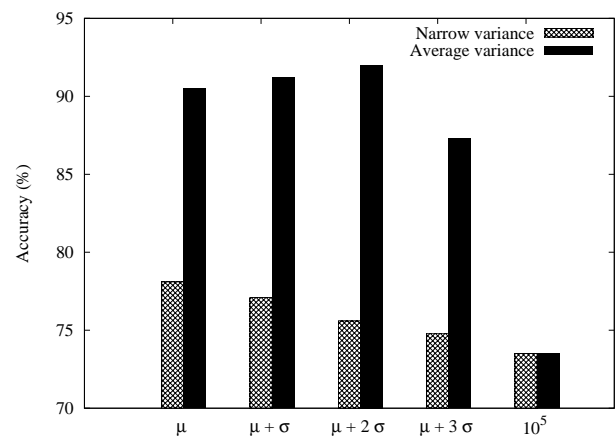


図 6 全曲まとめたときの精度変化

最終的に、全曲を用いた時の認識精度は、図6のようになつた。この結果は、評価データ中に多対1対応している部分がどの程度含まれるかによって変化すると思われるが、今回のデータベースの場合では、共分散行列を「平均的分散」とした上で、平均ベクトルを $\mu + 2\sigma$ とすることで、92.0%と最もよい認識精度が得られた。音符の区切り情報を用いない場合の認識精度が85.7%であったことから、適切なパラメータ設定を行うことで、6.3ポイントの認識精度の改善を得ることができた。

5. まとめ

本論文では、歌唱音声中の歌詞情報を高精度に認識するため、歌唱中で長音化している場所での挿入誤りを減らす方法を提案した。基本的に歌詞中の1モーラは1音符と対応している事に注目し、別途推定した音符の区切り時刻でのみモーラ間の遷移を許すように認識アルゴリズムを変更した。具体的には、通常の音声から得られる特徴量とはかけ離れた値を持つ「マーカーベクトル」を、入力音声から得られた特徴量ベクトル系列中の音符の区切り時刻の位置に挿入する。また、マーカーベクトルの値を出力確率分布の平均値に持つ1状態の「マーカーHMM」を定義し、発音辞書において、すべてのモーラ間にマーカーHMMを挿入する。こうすることで、強制的に1モーラが1音符と対応することとなり、長音化による挿入誤りを抑制することを可能とした。

また、曲によっては1音符が2モーラに対応したり、2音符が1モーラに対応したりする事もあることから、マーカーHMMの出力確率分布の平均ベクトル(=マーカーベクトルが持つ特徴量の値)と共分散行列の値を適切に設定することで、こうした多対1の対応にも対応する方法を提案した。

27名が歌唱した童謡歌唱データを用いて認識実験を行ったところ、マーカーHMMの出力確率分布に適切な値を設定することで、単語認識精度を85.7%→92.0%と大きく改善させることができた。

参考文献

- [1] 伊藤彰則, 鈴木基之, 牧野正三: この曲、何だっけ? 歌で音楽を探す「歌声検索」, *DTM MAGAZINE*, Vol. 183, pp. 102-103 (2009).
- [2] 尾関弘尚, 鎌田貴幸, 後藤真孝, 速水 悟: 歌声の歌詞認識における音高の影響について, 2003年秋季音講論集 1-1-1, 日本音響学会 (2003).
- [3] Suzuki, M., Hosoya, T., Ito, A. and Makino, S.: Music Information Retrieval from a Singing Voice Using Lyrics and Melody Information, *EURASIP Journal on Advances in Signal Processing*, Vol. 2007, pp. Article ID 38727, 8 pages (2007). doi:10.1155/2007/38727.
- [4] 川井大陸, 山本一公, 中川聖一: DNN-HMMを用いた歌声の自動歌詞認識の検討, 音楽情報科学研究会研究報告, Vol. 2015-MUS-107, No. 58, pp. 1-6 (2015).

- [5] Bello, J. P., Daudet, L., Abdallah, S., Duxbury, C., Davies, M. and Sandler, M. B.: A Tutorial on Onset Detection in Music Signals, *IEEE Trans. Speech and Audio Processing*, Vol. 13, No. 5, pp. 1035-1047 (2005).
- [6] 鈴木基之, 岡松竜徳, 任 福継: 音程に注目した歌唱音声の音符区間推定, 音楽情報科学研究会研究報告, Vol. 2010-MUS-85, No. 9, pp. 1-6 (2010).
- [7] Lee, A., Kawahara, T. and Shikano, K.: Julius — an open source real-time large vocabulary recognition engine, *Proc. EUROSPEECH*, pp. 1691-1694 (2001).