

音長を考慮した音響モデルによる歌声の自動歌詞認識

栗田 渉平† 酒向 慎司‡ 北村 正‡

†名古屋工業大学情報工学科

‡名古屋工業大学大学院

1 はじめに

近年、膨大な数の楽曲を提供する楽曲配信サービスが音楽販売の主流となり、音楽情報検索の需要が高まっている。楽曲を特定するための重要な情報として、曲名、歌手、ジャンル、旋律、歌詞が挙げられる。歌詞認識は、楽曲の歌唱パートの歌詞を書き起こすものであり、音楽情報検索の研究分野の一つである。歌詞認識は、第一段階として伴奏のない歌声のみの楽曲を対象に行われている。歌詞認識は音声認識技術を拡張して行われているが、大規模な歌声データベースの不足により、音響モデルを歌声から学習することは困難である。従来研究 [1] では、朗読音声で学習した音響モデルを少量の歌声で適応学習して歌声に対応させた。しかし、歌声は様々な長さの音韻の引き伸ばしにより挿入誤りが増加するため、音韻の引き伸ばしに頑健な音響モデルが必要である。

本研究では、音韻の引き伸ばしに対応するために、様々な音長の歌声データを歌声合成によって確保する。さらに、その歌声データを使用して、音長に応じて母音のモデルを分けた音響モデルを構築する。

2 音響モデル

2.1 歌声合成音声による音響モデル

歌声合成音声は、オンラインサービス HMM 歌声合成システム Sinsy[2] で生成する。Sinsy は MusicXML で記述された歌詞つきの楽譜から任意の歌声を合成することができる。本研究では様々なテンポで合成した歌声データを用いて音響モデルを学習する。

2.2 音長に応じて母音のモデルを分けた音響モデル

音韻の引き伸ばしは、メロディに従って母音が発せられることにより生じる。様々な長さの母音を捉えるために本研究では、音長に応じて母音のモデルを分けて音響モデルを学習する。図 1 のように、学習データの母音のラベルを 100 ミリ秒ごとに分ける。ラベルを分けた学習データで音響モデルを学習する。その後、音素ごとにクラスタリングを行い、類似したモデルの状態を共有する。

3 評価実験

3.1 実験条件

提案した音響モデルを用いて音節認識実験を行った。以下の 3 種類のトライフォンモデルを比較する。

- 同一テンポの歌声による音響モデル (M_1)
- 様々なテンポの歌声による音響モデル (M_2)

0ms ~ 100ms /a/ → a:00
 101ms ~ 200ms /a/ → a:01
 ⋮ ⋮
 1001ms ~ /a/ → a:10

図 1 音長に応じた母音のラベル付け

表 1 学習に使用した歌声データ

モデル	BPM	曲数	合計時間
M_1	100	60 曲	1.0 時間
M_2	50, 60, ..., 250	1260 曲	19.4 時間
M_3			

表 2 実験結果

モデル	正解率	挿入誤り率	認識精度
M_1	56.00 %	18.43 %	37.57 %
M_2	55.80 %	23.51 %	30.40 %
M_3	57.00 %	10.10 %	46.99 %

- 音長に応じて母音のモデルを分けた音響モデル (M_3)

音響モデルの学習には、日本語の童謡を合成した歌声データを用いた。使用した歌声データを表 1 に示す。デコーダには Julius[3] を使用した。言語モデルは使用せず、任意の音節列を出力するネットワーク文法を用いた。

評価データには様々なテンポで合成された日本語の童謡 32 曲を使用した。この歌声データは日本語の童謡 16 曲に対して、BPM を 50, 60, ..., 250 の中から各曲に 2 つ無作為に設定し、合成したものである。

3.2 実験結果

同じ母音が連続して出現する場合、それらを一つの母音として統合した。表 2 に実験結果を示す。 M_1 と M_2 を比較すると、 M_2 の方が挿入誤り率が高い。しかし、 M_3 の挿入誤り率は 10.10% であり、他のモデルよりも高い精度を得られた。これは、様々なテンポの歌声で学習した音響モデルは音長に応じて母音のモデルを分けることで有効であることを示す。

4 むすび

本研究では、音響モデルの学習に様々なテンポの歌声合成音声を使用した。また、その歌声合成音声を使用して音長に応じて母音のモデルを分けた音響モデルを構築し認識実験を行った。様々なテンポの歌声で学習し、音長に応じて母音を分けた音響モデルは、音韻の引き伸ばしに対応した音響モデルであるといえる。

今後の課題として、言語モデルを使用した単語単位の認識実験が挙げられる。

参考文献

- [1] A. Mesaros, T. Virtanen: "Automatic Recognition of Lyrics in Singing", EURASIP, 2010.
- [2] 大浦 圭一郎他: "Sinsy: 「あの人に歌ってほしい」をかなえる HMM 歌声合成システム", 情報処理学会研究報告, 2010-MUS-86, 2010.
- [3] 李 晃伸: "大語彙連続音声認識エンジン Julius ver.4", 情報処理学会研究報告, 2007-SLP-69-53, 2007.