

パラ言語情報によるラベリングに基づく 音声合成用の演技音声の検索システム

吉田 聡平 伊藤 克亘 (法政大学情報科学部)

1 はじめに

音声合成による自動音声読み上げが様々な場面で使われるようになってきている。本稿では、声の類似性の知覚におけるパラ言語内容の重要性に基づき、パラ言語情報による演技音声へのマルチラベルスコアリングの、演技音声の自動選択への応用可能性を検証する。今回は漫画を音声合成によってフルボイス化する際に、元となる音声の選択を自動化することを想定し実験を行う。

2 演技音声の知覚におけるパラ言語内容

Obin らの研究 [1] により、声のカテゴリ (話者の特性および状態) への抽象化が、演技音声の類似性の人間の知覚において重要な役割を果たすと述べられている。本稿では、音声間の認識の類似度を求めるのに用いられたマルチラベル分類を、音声をクエリとしない状況に応用できるかを検証する。検証に用いるシステムの概要は以下の通りである。

1. ユーザーが漫画を読んで演技音声をイメージする。今回はコマ単位での検索を想定している。
2. ユーザーがパラ言語のラベルを選択方式で入力する。
3. 演技音声のデータベースの中から尤度の高いものを検索する。このとき、ユーザーの入力はバイナリのクエリとして処理し、SD 法と分散によって算出したコマごとの印象に大きな差が出るものを重視する重みをつけた。

3 演技音声の収集とマルチラベリング

実験には、実際にアニメやゲームで使われている演技音声を利用する。1 発話 2 秒以上の音声を対象に 118 キャラクター 4487 個の演技音声を収集した。

3.1 ラベルの種類

熊本らの手順 [2] に従って、男女 4 名ずつの計 8 人に 15 個のコマを対象に形容表現 10 個の選定を行った。なお、今回は検索に用いるラベルのため、対義語としてグループ化されたものも形容表現として選出した。(表 1)

3.2 ラベルの学習

演技音声へのラベルの付与は、一部を手動にて行い、それをもとに学習をすることによって残りの演技音声にも付与を行う。各ラベルに該当する演技音声をデータベースの中か

表 1. 選定されたラベル

声が高い	声が低い
声が大き	声が小さい
優しい	
焦り	
怒り	
喜び	悲しみ
早口な	ゆっくりと
強い	弱い
テンションが高い	テンションが低い
呆れた	

ら抽出して教師データとし、GMM を用いてモデルを形成する。GMM に適用する音響特徴量には、メル周波数ケプストラム係数 (MFCC)、F0 平均、F0 範囲、パワー平均、ラウドネス、スペクトル重心、話速を使用した。混合数は実験的に 6 とした。

4 システムの実装と評価

4.1 再現率

演技音声全体の 5 % をチェックして正解の割合を推定し、全体の正解数だけシステムで演技音声を検索して再現率を求める、という手法を取った。画像 4 枚に対する結果を表 2 に示す。なお、画像 2 は「焦り」が必須なラベルとなるが、検索された音声に「焦り」のラベルが含まれるものがなかったため正解個数が 0 個となってしまった。

表 2. 再現率

画像番号	推定個数	正解音声数	再現率 (%)
1	80	50	62.5
2	40	0	0.00
3	60	42	70.0
4	80	45	56.3

4.2 適合率

主観評価実験によって評価する。ユーザーは各ラベルについて当てはまるかどうかを選択する。システムは各ラベルに重み付けを行って検索をし、予め用意しておいたマルチラベル付きの演技音声を尤度が高い順に表示する。今回は対照実験とするため、システムによって選出された音声 5 個と無作為に選出された音声 5 個を織り交ぜ、それぞれについてあっているかどうかを判断させた。その結果、システムによって尤度が高いと判断された上位 5 位の演技音声に適するものを含むのは 16 件中 11 件で適合率は 32.5 %、無作為に選出された演技音声に適するものを含むのは 16 件中 4 件で適合率は 5.0 % であった。

5 まとめ

複数の音響特徴による GMM により形容表現ごとのラベリングと検索が可能であった。しかし、「怒り」や「焦り」など必須なラベルが存在すると検索性能が著しく下がってしまう可能性がある。今後の課題として、ユーザーによる必須ラベルの選択を導入することで、反映されなければいけないラベルを漏らさないようにすることで精度の向上が見込まれる。

参考文献

- [1] Nicolas Obin ら “Similarity Search of Acted Voices for Automatic Voice Casting”, IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2016, Volume: 24, Issue: 9, pp1642-1651
- [2] 熊本ら “印象に基づく検索のための印象語選定法の提案”, 情報処理学会論文誌 44(7), 1808-1811, 2003-07-15