

発話の韻律変換による感情表現の検討

A Consideration on Expression of Feeling by Utterances Whose Prosody was Converted

谷口 聖人, 大島 千佳, 中山 功一 Taniguchi Masato, Oshima Chika, Nakayama Koichi

佐賀大学 Saga University

1. はじめに

音声は、感情を表現する方法として一般的なものである。人間は、音声の強弱や高低などの韻律を変化させ、自身の感情を表現している。しかし、音声での感情の表現や判別には、個人で大きく差が出てしまう。

先行研究[1]において、システムにより変換された自身の発話をリアルタイムに聞かせると、発話者を任意の感情に誘導できることが示唆されている。しかし、発話者の気分や印象を聞き手に決定づける発話要因が、自分の気分を誘導させる発話要因と同一であるかは明らかではない。筆者は、音声に含まれる感情を決定づける音響特徴について機械学習を用いて特定することで、特定した音響特徴を利用した音声変換を実現し、任意の感情を自然に表現できるのではないかと考えた。

以上の背景から、本研究では音声の韻律変化による自然な感情表現を実現するために、機械学習を用いた感情認識プログラムを開発する。また、開発したプログラムを利用し、あらかじめ感情がラベル付けされた音声データや、音声変換システムで任意の感情を付与した音声に対し、正しく感情の認識が行われるか実験し、自然な感情表現が可能であるか検証する。

2. 音声変換システム

2.1 EPROCs

EPROCs は、筆者らの研究室で開発された音声に含まれる音量と音高をリアルタイムに変換することのできるシステムである。実験により、本システムでリアルタイムに音声の音量・音高を変換することで、被験者に対し任意の感情への誘導が可能であるという結果が示唆されている[1]。これらの研究結果を参考に、そのEPROCs と変換アルゴリズムを利用する。また、音高と音量を用いた際の音響的特徴の変化と、その音響的特徴が機械学習における感情の認識にどのような影響を与えるか検証する。

2.2 DAVID

DAVID(Da Amazing Voice Inflection Device)[2]は、早稲田大学理工学術院渡辺克巳教授らの研究グループが開発した、音声に感情表現を与えることのできるデジタルプラットフォームである。DAVID は、オンラインで音声に感情フィルタをかけることができるために、音声知覚、感情知覚、自己知覚などに関する実験心理学・認知心理学・神経科学的研究を含めた幅広い研究分野での活用が見込まれている。本研究では、EPROCs による変換と DAVID による変換とを比較し、考察する。

3. ラベル付き音声データ

本研究で用いた音声データは、音声資源コンソーシアムで公開されている、感情評定値付きオンラインゲーム音声チャットコーパスである[3]。今回の実験では、ANG(怒)、JOY(楽)、SAD(悲)、SUR(驚)の4種類の感情を用いた。各感情で用いる音声ファイルの数は、ANG が 320 個、JOY が 336 個、SAD が 336 個、SUR が 384 個の計 1376 個である。

4. 機械学習を用いた感情分類

感情ラベル 4 種類 1376 個の音声データのうち、960 個の音声データをトレーニングデータとして用いて分類器を構築し、416 個の音声データをテストデータとして分類器を用いて分類することで実験を行う。検証方法は、2 種類とする。

1 つ目は、MFCC と呼ばれる音質データを用いた分類である。音声データそのものから音質データを抽出し、そのデータを特徴量として学習を行う。

2 つ目は、音量・音高を用いた分類である。2 章 2.1 節で述べた EPROCs で音声から音量・音高の値を抽出し、csv ファイルとして出力し、そのデータを特徴として学習を行う。

以上の 2 種類の学習結果から、分類結果を用いた音声の韻律変換による自然な感情表現が可能であるか検証する。本研究で用いる機会学習における評価指標は適合率と F 値を用いる。

5. 評価実験

5.1 実験方法

実験 1 から実験 13 までの 13 種類の実験を行った。実験 1 では、音声変換を行わずに音質データに基づいて分類する。実験 2 から実験 7 は、DAVID を用いて変換した音声の音質データを学習データとして用いて実験する。実験 8 では、音声変換を行わずに音量・音高データに基づいて分類する。実験 9 から実験 13 は、EPROC を用いて変換した音声の音量・音高データを学習データとして用いて実験する。本稿では、代表例として、音質データを用いた実験 1 と、音量・音高データを用いた実験 8 を比較した結果のみ報告する。

5.2 実験結果

実験 1 と実験 8 で、適合率を比較した結果を図 5.1 に示す。音質データを用いた実験 1 の方が、音量・音高データを用いた実験 8 に比べて、高い適合率を得た。

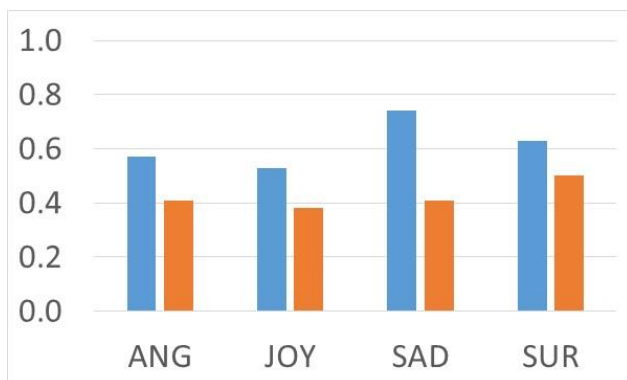


図 5.1 適合率による比較

実験 1 と実験 8 で、適合率と再現率の調和平均である F 値を比較した結果を図 5.2 に示す。適合率の結果と同様に、音質データを用いた実験 1 の方が、音量・音高データを用いた実験 8 に比べて、高い適合率を得た。

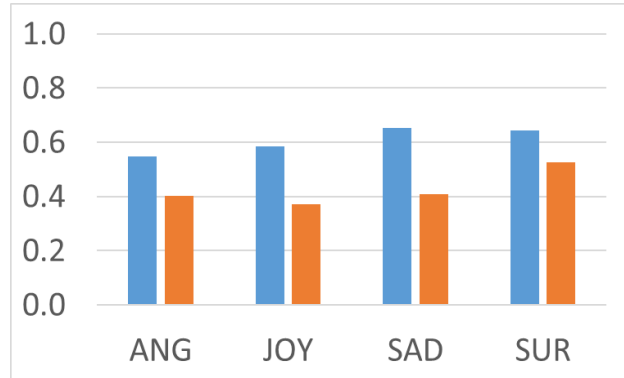


図 5.2 F 値による比較

5.3 考察

今回の実験で、音量・音高データを用いた感情の分類は、音質データを用いた分類よりも全体的に精度が低かった。特に、ANG や JOY は実験ごとに適合率の値が大きく異なっており、変換による影響を大きく受けているのではないかと考えられる。また SUR に関しては、それぞれの実験で適合率に大きな差が見られなかったため、SUR の感情に関しては変換による影響をほとんど受けていないのではないかと考えられる。

6. まとめと今後の展望

本研究では音声の韻律変化による自然な感情表現を実現するために、機械学習を用いた感情認識プログラムを開発した。また、開発したプログラムを利用し、あらかじめ感情がラベル付けされた音声データや、音声変換システムで任意の感情を付与した音声に対し、正しく感情の認識が行われるかの実験を行い、自然な感情表現が可能であるかの検証を行った。実験結果より、開発した感情認識プログラムが、ある特定の感情に対して分類を正しく行うことができ、音声変換で、正しく感情の付与を行い、自然な感情付与ができる可能性を検証できた。

参考文献

- [1] 東原凌平, 発話者の感情を誘導する音声変換アルゴリズムの検討, 2016 年度学士論文
- [2] Aucouturier, J. J., et al.:Covert digital manipulation of vocal emotion alter speakers' emotional states in a congruent direction, PROC NATL ACAD SCI,113(4), 948-953, 2016.
- [3] 感情評定値付きオンラインゲーム音声チャットコーパス(OGVC).
<http://research.nii.ac.jp/src/OGVC.html>
(2018-12-18 参照)