

## 感情を含んだ自発対話音声と演技音声間に生じる音響的特徴の比較分析

情報科学科 仲村 渠 遥

指導教員：入部 百合絵

## 1 はじめに

近年、音声を用いた様々な感情認識の研究が進められている[1]. 一般的に音声による感情認識の研究で用いられる音声データは、プロの声優や俳優が決められた感情や言葉を読み上げる演技音声である。しかし、実際に感情認識したい音声は、発話に込められる感情が指定されていない自然に表出される自発音声である。有本らの研究[2]では、オンラインゲームによって発話内容や感情を指定せず感情を誘発した対話を自発対話音声として、自発対話音声と演技音声間での音響的特徴量の違いを指摘している。本研究では、主成分分析によって自発対話音声と演技音声間での音響的特徴量の違いを感情毎に比較する。また、その結果を用いて自発対話音声と演技音声の感情毎の違いを考慮せずに感情認識するための一手法を提案する。

## 2 自発対話音声と演技音声間に生じる音響的特徴の比較分析

## 2.1 利用した感情音声データの概要

今回、本実験で使用したのは無償で提供されている「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」[3]である。このコーパスにはオンラインゲームをしている大学生(男性9名, 女性4名)の自発対話音声と、プロの俳優(男性2名, 女性2名)が指定された感情を含んで読み上げた、演技音声が収録されている。また、自発対話音声の各発話に対して三人の評価者が10種の感情ラベルを用いて感情評定を行っている。今回は、その中からANG(怒り), ANT(期待), DIS(嫌悪), FEA(恐れ), JOY(喜び), SAD(悲しみ), SUR(驚き)の8つの感情ラベルが付与された演技音声276発話(各感情平均40発話), 自発対話音声70発話(各感情平均10発話)を用いた。

## 2.2 分析方法

自発対話音声と演技音声を感情ごとに分け、フレームエネルギーの二乗平均平方根、メル周波数ケプストラム(以降、MFCCと呼ぶ)、ジッタ、シマー、ピッチ、ゼロクロス確率、ボイスブロープの平均と標準偏差および微分値の平均と標準偏差を求めた。また、ジッタとはピッチの揺らぎ、シマーとはピッチの振幅の揺らぎである。これらの特徴量を主成分分析にかけ、自発対話音声と演技音声での音響的特徴の違いを明らかにする。

また、自発対話音声と演技音声を感情において比較を行うこ

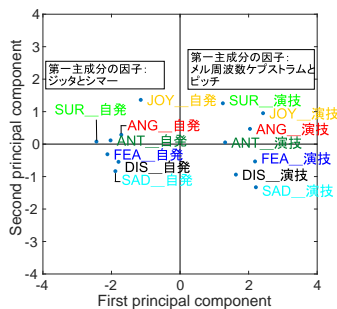


図1 特徴量の平均を変数とした主成分得点

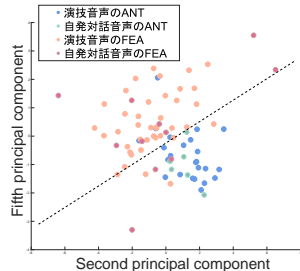


図2 ANT(期待)とFEA(恐れ)の主成分得点

表1 主成分を用いた感情識別の結果

感情	主成分	正解率
ANT-FEA	2,5	88.89 %
ANT-DIS	2,4	80.00 %
ANT-SUR	1,3,5	88.89 %
DIS-JOY	1,3	83.64 %
FEA-JOY	3,5	82.76 %
JOY-SAD	3,4	84.76 %

とで、二種類の音声間の違いをより詳細に分析する。

## 2.3 分析結果と考察

自発対話音声と演技音声に対する特徴量の平均を変数として算出した主成分得点を図1に示す。主成分分析の結果から因子を考えると、第1主成分はジッタとシマー(負の値)、MFCCとピッチ(正の値)が寄与していると解釈できる。図1より第1主成分方向の自発対話音声と演技音声は正負に分断されており、自発対話音声は感情を表出するのに基本周波数のゆらぎと周期毎の振幅のゆらぎが大きく、演技音声はどの感情においてもピッチが高めと言える。以上の結果より、演技と自発対話による感情音声の音響情報には差異が存在するため、演技音声と自発対話音声を混合して構築したモデルでは感情認識の精度に影響を及ぼすと考えられる。一方で、自発対話音声と演技音声の両音声から抽出した特定の感情を含む音声を混ぜて主成分得点を算出した結果例を図2に示す。軸をそれぞれ第2, 第5主成分とした場合、自発対話音声あるいは演技音声を問わずANTとFEAが分離していることが分かった。この結果より、ある特定の主成分は異なる感情を分離するのに有用であると考えられるため、主成分得点より特定した主成分を特徴量とした感情識別を行った。識別器はAdaBoost, RandomForest, SVMを用いて行った。特に正解率の高かったAdaBoostを用いた分類結果を表1に示す。表1から、主成分得点の図と同様に二種類の感情音声識別できていることが分かる。これにより、自発対話音声と演技音声の違いを考慮せずに感情認識を行うことのできる可能性を示した。また、表1には正解率が比較的高い結果を記載したが、記載していないANGに関しては他の感情との識別率が低い結果になった。従って、ANGを精度よく識別するための改善が必要である。また、今回は二種類の感情対での識別であったが、複数の感情に対しても高精度に識別するための手法を提案する予定である。

## 3 おわりに

本研究では、自発対話音声と演技音声における感情の違いを音響的特徴という側面から主成分分析を用いて比較した。また、その結果より自発対話音声と演技音声の感情毎の音響の違いに依存せずに感情認識するための手法を提案し、認識実験を行った。今後は8種類全ての感情を含んだ音声に対しての識別精度を向上させることが課題である。

## 参考文献

- [1] 岩瀬佳代子, 神田崇行, 石黒浩, 柳田益造, ”コミュニケーションロボットにおけるノンバーバル情報を用いた状況依存音声認識”, 人工知能学会 AI チャレンジ研究会, No20, pp33-38, 2004.
- [2] 有本泰子, 河津宏美, ”音声チャットを利用したオンラインゲーム感情音声コーパス”, 日本音響学会研究発表会講演論文集, Vol2013 秋, ROMBUNNO.1-P-46A, 2013.
- [3] 有本泰子, 河津宏美, ”感情評定位置付きオンラインゲーム音声チャットコーパス (OGCV) 使用説明書 Ver1.0”, 2013.