

機械学習を用いた人の状態認識システムの構築

機械学習を用いて、音声から人の感情を認識する技術があり、健康状態の推定や人と機械との対話の要素技術としての応用が期待される。一方、機械学習に必要な正解感情が定められた音声データベースの作成には非常にコストがかかるため、学習用の音声データが少量となることが、感情認識を困難とする一要因となっている。また、そのデータベースには自発音声と演技音声があり、自発音声の感情認識は難易度が高い。

本研究では、**音声による感情認識システム**を構築し、日本語の**自発音声**を対象とした**認識精度を高める**方法に関して検討する。

(1) 検討指針

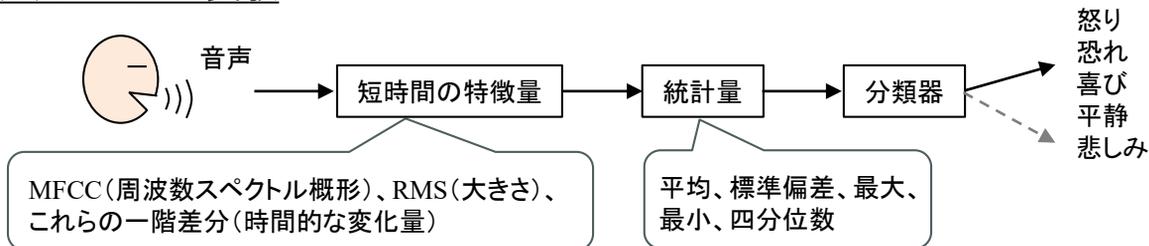
- ・少量学習データに対しては優位性がある**手法①**をベースにする。
- ・**手法②**のほうが精度が優れているが、通常は大量の学習データが必要
⇒対応策: **事前学習モデル**
画像分類等の大量のデータを用いて学習させた深層学習モデルの一部を感情認識モデルに転用

手法	精度	少量学習データ	特徴量抽出
①古典的手法	劣	優	手動
②深層学習	優	劣	自動

⇒ **手法①**の特徴量抽出部に、**手法②**の**事前学習モデル**を導入する方法の検討

本年度はベースとなる**手法①**のシステムを構築

(2) システム・実験



<データセット(OGVC*)>

感情	発話数
怒り	237
恐れ	142
喜び	595
平静	798
悲しみ	243
総発話数	2015

(話者11名)

<感情認識結果例>

平均**正解率**:39%

正解感情	認識感情(%)				
	怒り	恐れ	喜び	平静	悲しみ
怒り	19	6	30	43	3
恐れ	13	16	17	33	20
喜び	13	5	57	20	5
平静	7	5	14	65	9
悲しみ	4	12	8	35	40

・ **システムの基礎的な動作とともに、認識精度が概して低い課題を確認**

・ 次年度は、認識精度の向上のため、特徴量抽出への深層学習(事前学習モデル)の導入を検討予定

*国立情報学研究所 音声資源コンソーシアムから提供を受けた「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)」を利用した。