

温度画像を用いた母音発声に着目した表情認識

(知能情報システム学) 神田康成

1. はじめに

人とロボットが自然なコミュニケーションをとっていくためには、ロボットが人の感情を理解する必要がある。これを実現するため、さまざまな表情認識手法が研究されてきた。中でも、温度画像を用いた表情認識手法においては照明条件に左右されない表情認識が可能であり、5表情における平均が 90 %という認識率を得ている^[1]。既報^[2]の表情認識システムでは、被験者に、「タロー」と感情を込めて発声を行ってもらうよう、発声単語を指定した。既報では単語を指定し音声波形を基に「タ」と「ロー」の音素境界を定め、静止画像採取のタイミングを決定した。

本研究では汎用大語彙連続音声認識エンジン Julius^[3]を用い、音素境界を決定し母音発声に着目した表情認識を行った。

2. 開発環境

- ・ OS : Windows XP ・ 使用したプログラム言語 : Visual C++6.0
- ・ DELL Optiplex GX280 CPU : PentiumIV 3.6GHz メインメモリ : 2GB

3. 処理概要

① 入力画像を AVI ファイルに保存

温度動画像を入力し、「taro」30回発話分を AVI ファイルに保存する。

② 音声認識

保存された順に AVI ファイルを開き、その中の音声データを wave ファイルに変換、保存する。その wave ファイルを Julius を用いて音声認識する。そして発声開始までの時間、発声開始直後に出てくる母音(「taro」の場合「a」)、発声終了直前の母音(「taro」の場合「o」)、について発声開始及び終了までの時間を保存する。

③ 静止画像を採取

保存された AVI ファイルを開き、Julius の結果をもとに発声直前、発声開始直後の母音発声時、発声終了直前の母音発声時の 3 枚の静止温度画像を採取する。

④ 表情認識

1. 採取した静止温度画像の顔領域の位置、回転角度、サイズの補正処理を行ってから、表情の特徴が出やすい顔部品領域を抽出し、平均無表情画像との差分をとり、純粋な表情変化のみを抽出する。
2. 得られた差分画像に二次元離散的コサイン変換(2D-DCT)を施す。
3. DCT 特徴量から、特徴ベクトルを作成する。
4. 学習画像から得た特徴ベクトル空間に未知画像の特徴ベクトルを入力し、最小距離識別法による表情認識を行う。

4. 実験

4.1 方法

被験者に中立的な言葉(日本名「taro」)を5表情(怒り、喜び、無表情、悲しみ、驚き)で発声してもらった。このときの温度動画像を入力し、前述した処理によって表情認識実験を行った。1表情に対して30回の発声データを採取し、このうち20サンプルを学習用、10サンプルを認識用に用いた。1回の発声から得た静止画像は発声直前、「a」発声時、「o」発声時の3枚である。そして、既報の「ta」と「ro」の音素境界を波形から定めた方法での表情認識比較した。Juliusで誤認識が発生した場合、学習用画像は他の音素であっても学習データとして扱った。認識についてはJuliusで誤認識が発生した場合は認識用画像として取り扱わないようにした。

4.2 結果

既報の手法による被験者Aの表情認識結果^[2]を表1に、本法による、被験者Aの結果を表2に示す。

表1 既報の手法による認識結果^[2]

		入力表情				
		怒り	喜び	無表情	悲しみ	驚き
認識	怒り	100				
	喜び		100			
	無表情			100		
	悲しみ				70	
	驚き				30	100

表2 本法での認識結果

		入力表情				
		怒り	喜び	無表情	悲しみ	驚き
認識	怒り	100				10
	喜び		100			
	無表情			100		
	悲しみ				70	20
	驚き				30	70

音声認識を行った結果「taro」と以外の似た言葉に誤認識するパターンが見られた。発声開始までの時間、発声開始直後に出てくる母音、発声終了直前の母音、についてそれぞれの発声区間を認識することには100%成功した。

5. 結言

Juliusを用いて音声認識を行い、音素境界を決定し既報^[2]の手法を基にして表情認識を行った。その結果、認識率は既報の94%に対して本法では88%となった。音声の誤認識のため表情認識精度が悪くなった。表情認識に適応した音声データベースの構築を検討する必要がある。今後はさまざまな単語発声時に本法を適用して、実用化の課題を明らかにしていく必要がある。

参考文献

- [1] 池添史隆, 胡玲琴, 谷尻豊寿, 吉富康成, 「温度画像を用いた発声時の表情認識」, ヒューマンインターフェース学会論文誌, 6(2004), 19-27.
- [2] 中野真里, 「温度画像を用いた発声時の個人差にロバストな表情認識」平成19年度京都府立大学大学院人間環境科学研究科環境情報専攻 修士論文 2008.
- [3] 大語彙連続音声認識システム Julius Web ページ
<http://julius.sourceforge.jp/>