

ヒューマノイドエージェントにおける ユーザ識別と感情識別

A User Identification and Emotion Identification on a Humanoid Agent System

関根剛宏, 馬場朗, 日比谷新平, 大林史明, 寺澤章, 西山高史, 仲島了治
Takehiro SEKINE, Akira BABA, Shinpei HIBIYA, Fumiaki OBAYASHI, Akira TERASAWA,
Takashi NISHIYAMA, and Ryoji NAKAJIMA
松下電工(株) システム技術研究所

(〒571-8686 大阪府門真市大字門真 1048, sekine@trc.mew.co.jp, baba@trc.mew.co.jp, takashi@ai.mew.co.jp,
terasawa@icrl.mew.co.jp, hibiya@trc.mew.co.jp, obayashi@icrl.mew.co.jp, ryoji@ai.mew.co.jp)

1. はじめに

我々は現在, 総務省プロジェクト「ヒューマノイド・エージェントの構成技術と応用に関する研究」(平成 14 ~ 16 年度) の中で, ユーザが自然と親しみを持つ “人間型” を有し, 人工知能と人工感情を備えてユーザと対話するエージェントの研究を行っている [1]. 図 1 にヒューマノイド・エージェント (以下エージェント) の構成を示す. 本稿では図 1 のユーザ識別部及び感情推定部の表情推定部について述べる. エージェントは, マイクから取得された音声データ及びカメラから取得された顔画像データからユーザが誰であるかを識別するユーザ識別部と, 顔画像の時系列データからユーザの表情を識別する表情識別部を有する. 識別されたユーザ ID 及びユーザの感情に応じた対話制御を実現すれば, よりユーザフレンドリーな対話エージェントを構成できることになる. また, エージェントとの対話は家庭内での家族との対話シーンを想定している.

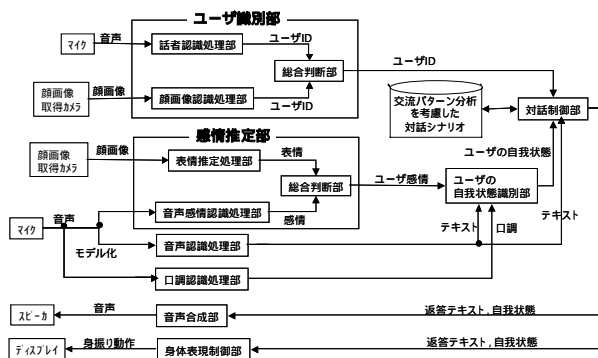


図1 ヒューマノイド・エージェントの構成

2. 話者識別部

これまでに GMM を用いた話者識別の研究としては [2] がある.

本プロジェクトでは話者識別の手法として HMM (Hidden Markov Model) を用い, これまでに基本的な識別の有効性を確認した. HMM の出力シンボルに話

者 ID を割り当て, 各話者の HMM を構成した. HMM の学習と別に用意した発話文章を用いてユーザ 10 名の場合と 200 名の場合で識別実験を行った. 表 1 に実験結果を示す.

音声の特徴パラメータは MFCC, MFCC, Power の計 25 次元, HMM の構成は left to right 型の GMM, 16 混合分布とした. HMM の学習および識別には HTK (Hidden Markov Toolkit) Ver.3.1 を用いた.

表 1 話者識別実験の諸元

	実験 1	実験 2
・特徴パラメータ	MFCC, MFCC, pow	計 25 次元
・HMM	3 状態, 16 混合 (性別非依存モノフォン)	
・使用データ	JNAS (ATR 音素/フランス文)	
・文法	START (f001 f002 ... fxxx) END	
・識別用データ数	10 文章/1 名	20 文章/1 名
・話者	10 名 (女声)	200 名 (男・女)
・学習回数	3 回	1 ~ 10 回
・学習文章数	1 ~ 30 文章	30 文章

実験 1 ではユーザ 10 名の場合に学習回数を固定とし, 学習文章数を可変にしたとき, どの程度の学習文章数で識別が可能であるかを調べた. 結果を図 2 に示す.

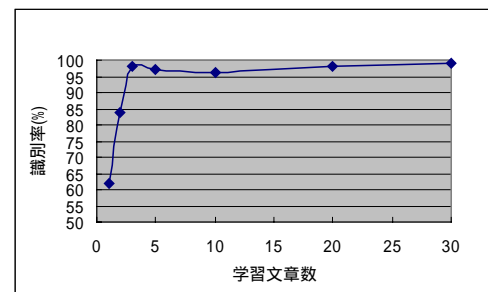


図 2 学習文章数と識別率の関係

実験から学習文章数は 3 文章以上の場合 (1 文章の発話時間は約 3 秒 ~ 6 秒), 識別率 96% 以上を確保でき, 3 文章未満の場合は識別性能が急激に低下することが分かった.

実験 2 ではユーザ 200 名 (女声 100 名, 男声 100 名) の

場合に学習文章数を固定とし、同一文章の学習回数を変化させ、識別率が変化するかを調べた。男女別の誤識別率の結果を図3に示す。

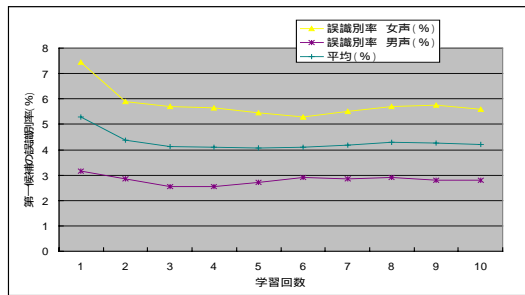


図3 学習回数と誤識別の関係

誤識別率は学習回数1回の場合を除き、平均で4.0%~4.4%となり学習回数にあまり依存しないことが分かった。また男女別に比較すると女声の方が男声よりやや誤りやすい傾向にあることが分かった。

2. 顔画像からのユーザ識別

顔画像からのユーザ識別部は、入力データとして顔の静止画を用いる。入力顔画像を、予め登録しておいたユーザ顔画像 D/B の特徴パラメータと照合し、距離の近いユーザ候補を出力する。

顔画像識別において、我々は特徴パラメータとして目、口などの顔器官の位置情報と顔の大きさをを用いる。図4に元画像から目の器官を抽出するフローを示す。

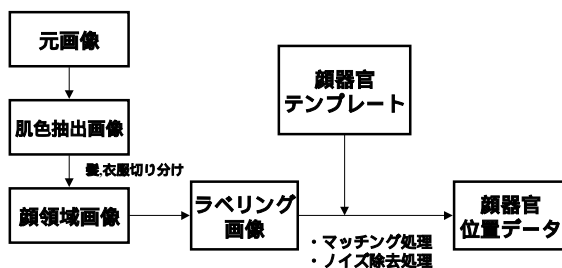


図4 顔器官抽出の流れ

顔器官の抽出は元画像を縮小処理した画像を基に肌候補の抽出処理を行い、エッジ画像を用いて顔と髪、衣服の切り分けを行い顔領域を求める。肌候補の抽出は照明による変動を考慮して HSV 表色系の色相成分を取得し、赤から黄色の比較的広範囲にあたるものを取得している。次に予め用意した顔器官テンプレートと入力画像とのテンプレートマッチングにより両目の位置を抽出し、両目の位置情報から推定される口位置領域のエッジ画像から口器官を抽出する。

開発中のエージェントは家族との対話を想定していることから[1]、現在の相手が家族のうちの誰かを識別する

必要がある。そこで、4人(1家族)×20世帯、計80名からなる顔画像データを収集し、これらについて両目と口の抽出実験を行い表2に示す7つの特徴パラメータの個人性を調べた。

両目と口が抽出できた抽出成功率は約96%であった。表2の特徴パラメータを両目間の距離で正規化したものについて、同一ユーザの2期間での変動と、全ユーザの分散値の比を調べた。表2に結果を示す。

表2 各特徴パラメータの個人性

	ユーザ内変動の平均と全ユーザ分散の比
両目の内側の距離	0.050
目と口の縦幅	0.467
左目の横幅	0.050
右目の横幅	0.050
口の両端の距離	6.597
顔の横幅	0.383
顔の縦幅	1.853

表2の比が大きい程、個人性があると考えられ、識別に有利なパラメータと考えることができる。

3. 顔画像からのユーザ感情推定部

ユーザの顔画像の時系列変化から現在のユーザの表情を推定するサブシステムであり、リアルタイムで顔画像、顔器官を抽出し、特徴パラメータを抽出する部分を開発した。2章と同様の手法で顔器官を抽出した。特徴パラメータは、両目と口の輪郭上側と下側における形状と相互の位置関係及びその時系列データである。現在この特徴パラメータを用いて、驚き、喜び、怒り、平静の4表情を推定するアルゴリズムの設計と実装ならびに実験を行っている。

4. おわりに

現在研究開発中の「ヒューマノイド・エージェントの構成技術と応用に関する研究」(平成14~16年度)におけるユーザ識別部及び顔画像からの表情識別について述べた。今後は音声と顔画像を統合して家族ユーザを識別する方式を検討していく予定である。家族内では、2章で挙げた顔識別のための特徴パラメータでは有効に識別されない可能性があり、話者識別とも上手く組み合わせることで総合的に判断する必要があると考えている。

参考文献

- [1]西山他：ヒューマノイド・エージェントの構成に関する研究，SICE SI 部門講演 会講演概要集，p.87，2003.
- [2]村井他：話者性と発話交代を考慮した複数話者対話音声の認識，電子情報通信学会論文誌，D-II Vol.J83-D-II，No.11，2000