

# ヒューマノイドエージェントへの口調識別の応用

馬場朗, 関根剛宏, 日比谷新平, 大林史明, 寺澤章, 西山高史, 仲島了治

Akira BABA, Takehiro SEKINE, 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, baba@trc.mew.co.jp, {sekine, hibiyaji}@trc.mew.co.jp,  
{obayashi, terasawa}@icrl.mew.co.jp, takashi@ai.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有する“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]。本稿ではヒューマノイド・エージェントがユーザの口調を識別することにより、ユーザの自我状態を推定する方法について報告する。

## 2. ユーザ口調の識別

本プロジェクトでは、ヒューマノイド・エージェントに人工知能と人工感情を与えるために、人間同士の対話における、双方の対話内容と身振りや感情表出のパターンに関する心理学知見である**交流パターン分析**[2]を導入する。交流パターン分析では、人間の心理状態を5つの自我状態(CP, NP, A, FC, AC)で捉え、個々の自我状態における人間の性質、言葉、声・声の調子、動作・表情・ジェスチャの特徴が分類されている。本報告では、ユーザの自我状態を推定するためのユーザ情報の一つとして、特にユーザの口調を識別する手法について考察する。

本システムにおいて、各自自我状態にあるユーザが取得する口調を表1のように、自我状態ごとに4種類の口調を想定した。以後、自我状態と口調種類の組み合わせ(例えばCP1「命令」など)を口調IDとする。

表1 口調の一覧

	1	2	3	4
CP	命令	脅す	叱る	吐き捨てる
NP	励ます	慰める	褒める	勤める
A	尋ねる	訊く	応える	答える
FC	喜ぶ	怒る	元気よく	甘える
AC	嘆く	ぼやく	ぼそっと	おそろおそろ

これまで、平静な読上げ音声を、怒り、喜び、悲しみなどの感情のこもった音声に GMM を利用して声質変換する

方法が提案されており[3]、感情という音声情報を GMM でモデル化することが可能であることがわかる。したがって口調の識別においても、同様に音声の韻律情報などを GMM でモデル化する方法が有効であると考えられるので、GMM による口調識別実験を行った。

## 3. 実験方法

実験方法について以下説明する。実験は次の3ステップからなる。

### 3.1 読上げ文章の収集

GMM の学習に用いる口調付き音声を収録するにあたり、読上げ文章について各口調を考慮した。読上げ文章は、口調をキーワードとして WWW 検索エンジンによりテキストを収集した。例えば口調 ID が CP1 「命令」の場合、検索クエリを  
「命令」と命令した。

のように閉じ括弧と活用形の組み合わせとし、検索結果であるテキスト

「前へ出る」と命令した。

から、括弧によって区切られた文章(前へ出る)を利用した。なお、表1に示した口調は、[2]を参考に各自自我状態において取得できると考えた100種類程度の口調について、WWW から文章を収集した結果、音声収録に必要と考える35文章以上を収集したものである。

### 3.2 音声の収録

読上げ文章を俳優が演技し、その音声を防音室で収録した。本システムは、例えば家庭におけるエージェントシステムなどを想定しているため、話者は、父親役の男性2名、母親役の女性2名、子供役の女性2名の合計6名とした。

収録は、話者ごとに表1に示す20口調、ならびに口調をつけない平坦な読上げで、前節で述べた収集文章(35文)について行った。収録文章数の構成を表2に示す。

表2 収録文章数

	学習用	評価用
口調あり	30	5
口調なし	5	5

### 3.3 GMM の学習

GMM の混合正規分布数は 64 であり, GMM の出力シンボルに口調 ID をあて, 口調識別 GMM を構成した. GMM は話者ごとに作成した. GMM の学習および識別には HTK (Hidden Markov Toolkit) Ver.3.1 を用いた. 音声信号からの特徴量抽出条件は以下の通りである.

- サンプル周波数 16(kHz)
- 量子化ビット数 16(bit)
- 分析窓長 25(ミリ秒), ハミング窓
- 分析間隔 10(ミリ秒)
- 特徴パラメータ MFCC+ MFCC  
+ Power(計 25 次元)

### 4. 実験結果

学習した口調識別 GMM を用いて, 評価用音声の識別実験を行った. 図 1 に実験結果を示す. 実験は, 6 名の話者それぞれについて, 各話者の口調識別 GMM により各話者の発声した評価用 5 発声の識別実験を行い, 正しく口調が識別された割合の全話者での平均を集計した. また, 実験は, 口調付き音声についてその口調を正しく識別できた割合(図中口調あり)と, 平坦な読上げを正しく識別できた場合(図中口調なし)それぞれについて集計してある. 全口調での平均は, 口調ありの場合で 49.5(%), 口調なしの場合で, 97.0(%)であった.

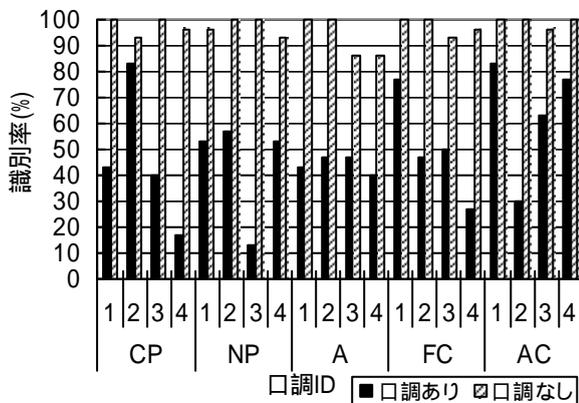


図 1 口調識別実験結果

### 5. 考察

識別実験結果における誤りを分析する. 口調付き音声の識別実験で頻度の高い誤りの組み合わせを表 2 に示す. 同表より, 同じ自我状態にマッピングされる口調 ID 同士での誤りが多いことがわかる. この原因は, ある自我状態内で取り得る口調同士が比較的類似していることによると考えられる.

また, 口調のない音声の入力に対して, 口調があると誤って識別された数は 18 個で, そのうち 13 個は A の自我状態に対応する口調 ID であった. これは A の自我状態における口調が平静に近いことを示しており, A の自我状態の特徴である, 「冷静で合理的な行動をとり, 単調で

落ち着いた声の調子になる」という性質に合致する.

表 2 識別誤りの上位 5 位

順位	入力音声	識別結果	誤り数
1	A1	A2	12
2	A2	A1	11
3	CP3	FC2	8
3	FC2	FC1	8
5	AC2	AC3	7
5	CP1	CP2	7
5	NP3	NP1	7

ユーザ音声の口調を識別する目的は, ユーザの自我状態を推定するためであるので, 自我状態の正解率での評価も必要である. 図 1 の実験結果の口調ありの場合を, 口調 ID に対応する自我状態ごとに集計した場合の結果を図 2 に示す. ただし, 口調がないと識別された場合には自我状態 A に相当するものとする.

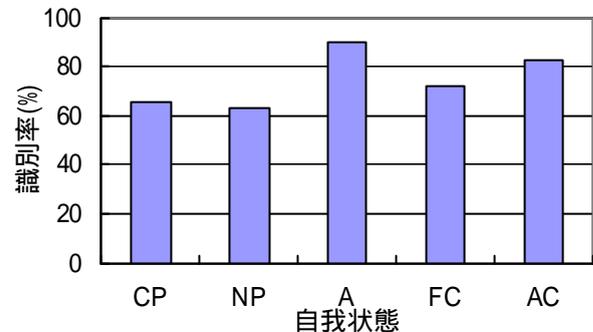


図 2 自我状態の識別率

図 2 のように, 全ての自我状態で 60(%)以上の識別性能が得られ, 平均では 74.7(%)であった. 以上より, GMM による口調識別が自我状態の推定に有効であることがわかった.

### 6. 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)において, ユーザの口調を識別することにより自我状態を推定する方法について述べた. 本研究の一部は, 総務省・戦略的情報通信研究開発推進制度からの委託を受けて推進中のものである.

### 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] 杉田: 交流分析, 日本文化科学社, 1985.
- [3] 岩見他: GMM に基づく声質変換を用いた感情表現の制御とその評価, 日本音響学会講演論文集, 1-20-24, Sept. 2003.