

## 日常的な対話における感情状態からの表情の生成

## Facial expression generation reflecting speaker's emotional state in daily conversation

大島 康十  
Koh Ohshima森 大毅十  
Hiroki Mori

## 1. はじめに

近年、アバターを用いたネットワークコミュニケーションツールが多く実用化されている。アバターは、ユーザがお互いに直接会ってコミュニケーションをしているような感覚を実現する。しかし、現在のところアバターの表情や動作がネットワークコミュニケーションにおいて果たす役割は小さく、ユーザの分身としては未だ不十分であるといえる。その要因として、現在のアバターは私たちが日常の対人場面において表出するような自然な表情を自動で生成することができないという点が挙げられる。

本研究では、話者の発話から表情を自動で生成するシステムの開発を目的としている。その処理は図1に示すように大きく2段階に分けられる。まず第1段階として、発話音声から知覚される感情状態の自動推定を行う。これは話者の感情状態と音響特徴量との関係を明らかにすることで実現できる。これについては文献[1-2]に委ねるものとしてここでは扱わない。次に第2段階として、感情状態から表情を決定し、顔画像合成を行う。

本稿ではこの第2段階となる、感情状態からの表情の生成について述べる。具体的には、様々な感情状態と、それに対する表情のデータを収集し、それらの対応データをもとに機械学習を行うことにより、任意の感情状態に対する表情の生成を試みた。

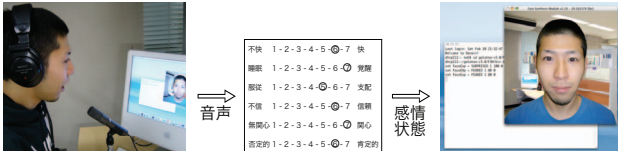


図1 音声入力から表情生成までの流れ

## 2. 表情合成の原理

## 2.1 感情状態から表情への変換の手法

本研究では、話者の感情状態の記述法として、後述する感情の次元説[3]にもとづく方法を用いる。日常的な対話において生じる感情状態から表情への制御規則を定めるためには、様々な感情状態に対する表情のデータを収集する必要がある。その手段として、本研究では以下の方法を提案する(図2)。

まず、ある発話から知覚される感情状態を評価する(A)。一方で、その発話にふさわしい表情を用意する(B)。そして、その表情と感情状態を対応させ、ひとつの発話に対する対応データとする。これを複数の発話に対して行うことで、様々な感情状態と表情との対応データを得ることが可能である。

本研究では、この方法により収集された対応データをもとにして、任意の感情状態に対する表情の生成を行う。

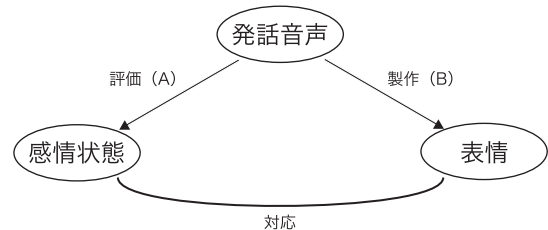


図2 対応データの収集法

## 2.2 使用する顔モデル

本研究では、顔画像の合成のための手段として、擬人化音声対話エージェントのソフトウェアツールキット、Galatea Toolkit[4]を使用している。このGalatea Toolkitの顔画像合成の基盤として用いられているエージェント生成ツールを使用することにより顔モデルを生成し、その眉や唇、眼球などの顔の各部位の動きを制御することが可能となっている。

## 2.3 表情の記述

表情の記述には、FACS (Facial Action Coding System)[5]を用いている。FACSは心理学者Ekmanらによって提案された表情の記述法であり、顔を解剖学的に考慮し、表情変化にともなう顔面の動きをアクション・ユニット(AU)と呼ばれる最小単位に分解している。そしてこのAUの組み合わせにより、人間が表出する様々な表情を記述することが可能であるとされている。

## 3. 感情状態と表情との対応データの収集

## 3.1 感情状態の記述

本研究では、感情の次元説にもとづく方法により感情状態を記述する。次元説とは、感情は離散的に特定できるものではなく、少数の次元で構成される空間上の座標として表されると考えるものである。ここでは、日常的な対話において知覚される話者の感情状態を表す手段として、以下に示す6次元の評価項目を用いる[6]。

- 1) 快-不快
- 2) 覚醒-睡眠
- 3) 支配-服従
- 4) 信頼-不信
- 5) 関心-無関心
- 6) 肯定的-否定的

それぞれの次元について、感情の度合いを7段階の値として記述する。具体的には、「快-不快」において、非常に快であれば7、かなり快であれば6、やや快であれば5、どちらでもなければ4と表す。これにより感情状態をパラメータとして扱うことが可能となる。

## 3.3 感情状態データ(A)

発話音声及びそれに対する感情状態のデータとしては、宇都宮大学パラ言語情報研究向け音声対話データベース[1]

を使用した。これは、前述の6項目による感情状態のラベルが与えられた音声対話データベースであり、日常的な対話において発せられた話者の発話に対して、評価者3名による感情状態の評価作業を行うことにより構築されたものである。

本研究では、このデータベースから対話音声3セッション分を抜粋した。ここでの対話は大学生の女性2名によるものである。このうち、片方の話者1名による219発話に対する評価者3名の評価値の平均値を感情状態データとして使用した。

### 3.4 表情データの製作(B)

感情状態に対応する表情のデータを用意するために、音声データの各発話に対して表情を製作していく作業を行った。この作業には Galatea Toolkit のエージェント生成ツールを使用した。これにより、表情の変化を確認しながら、それぞれのAUの動きの度合いを独立に操作することが可能となる。

表情製作の手順としては、発話を聴取し、顔の各部位をAU1から順に表情の変化を見ながら動かしていく作業を行う。そして、全体を見て調整を行い、最終的な表情を決定する。その表情に使用したAUの組み合わせとそれぞれの動きの大きさの度合い(0-100%)のパラメータを表情データとする。

この作業を全219発話に対して行い、それぞれの発話に対する感情状態と表情のデータを対応させた。これにより得られた対応データの例を図3に示す。

	評価値
快-不快	6.3
覚醒-睡眠	6.3
支配-服従	4.7
信頼-不信	6.7
関心-無関心	6.7
肯定的-否定的	6.7




図3 対応データの例

## 4. 感情状態からの表情の生成

収集した対応データをもとにして、感情状態から表情へのマッピングを行うための制御規則を定める。本研究ではその手法として、階層型3層ニューラルネットワークを用いた。ネットワークへの入力には感情状態を表す6次元の評価値、出力はAUパラメータである。また、中間層ユニット数は6個とした。

未学習の感情状態に対して、本手法により生成される表情の例を図4、図5に示す。このように、任意の感情状態に対して表情を生成することが可能となった。また、その表情は日常的に表出されるような自然なものであるといえる。しかし一方で、現段階で生成できる表情の幅は十分に広いとはいえない。これについては、今回の作業で得られなかった感情状態と表情との対応データを加えて学習させることで、より多くの感情状態を反映することができると考えている。

	評価値
快-不快	2.0
覚醒-睡眠	6.0
支配-服従	3.0
信頼-不信	3.0
関心-無関心	5.0
肯定的-否定的	2.0



図4 生成される表情の例(1)

	評価値
快-不快	4.0
覚醒-睡眠	7.0
支配-服従	5.0
信頼-不信	4.0
関心-無関心	6.0
肯定的-否定的	5.0



図5 生成される表情の例(2)

## 5. おわりに

本稿では、日常的な対話における感情状態から表情へのマッピングを実現する手法について述べた。また、実際に感情状態と表情との対応データを収集し、それらをもとにして表情の生成を試みた。その結果、任意の感情状態から表情を自動で生成することが可能となった。

今後の展望としては、生成される表情をより人間らしいものにするとともに、前述の処理の第1段階となる、音声入力からの感情状態の自動推定を目指す。そして、これらをリンクさせることにより、話者の発話からリアルタイムで表情を生成するシステムを実現する。

## 参考文献

- [1] H. Mori, H. Kasuya, "Voice source and vocal tract variations as cues to emotional states perceived from expressive conversational speech," Interspeech 2007, Antwerp, Belgium (to appear).
- [2] 佐竹智幸, "音声伝達するパラ言語情報の知覚および音響的性質に関する研究," 宇都宮大学卒業論文, 2007.
- [3] 濱治代, 鈴木直人, 濱保久, "感情心理学への招待 感情・情緒へのアプローチ," サイエンス社, 2001.
- [4] 川本真一, 下平博, 新田恒雄, 西本卓也, 中村哲, 伊藤克亘, 森島繁生, 四倉達夫他, "カスタマイズ性を考慮した擬人化音声対話ソフトウェアツールキットの設計," 情報処理学会論文誌, Vol. 43, No. 7, pp. 2249-2261, 2002.
- [5] P. Ekman, W. V. Friesen, 工藤力訳, "表情分析入門," 誠信書房, 1987.
- [6] 森大毅, 相澤宏, 粕谷英樹, "対話音声のパラ言語情報ラベリングの安定性," 日本音響学会誌, Vol. 61, No. 12, pp. 690-697, 2005.