

## 表情の変化速度がアバターの感情表出の自然性に与える影響

### The effect of display speed of facial expression on the naturalness of emotional expression by avatars

大島康<sup>1\*</sup> 森大毅<sup>1</sup> 中村真<sup>2</sup>

Ko Oshima<sup>1</sup>, Hiroki Mori<sup>1</sup>, Makoto Nakamura<sup>2</sup>

<sup>1</sup> 宇都宮大学大学院工学研究科

<sup>1</sup> Graduate School of Engineering, Utsunomiya University

<sup>2</sup> 宇都宮大学国際学部

<sup>2</sup> Faculty of International Studies, Utsunomiya University

**Abstract:** To make avatars generate natural facial expressions like human, it is important to control dynamic properties of facial expressions. The purpose of this study is to investigate relationships between emotional states in daily conversation and natural display speed of facial expressions. We conducted a psychological experiment, where subjects were presented an auditory stimulus and a visual stimulus simultaneously. Auditory stimuli are utterances in daily conversation, and visual stimuli are dynamic facial expressions. Six speed conditions were used in the experiment. The Subjects were asked to judge the naturalness of the presented dynamic facial expression with the 5-grade scale. Experimental results suggested that natural display speed of facial expressions relates to dimensions of pleasant-unpleasant and aroused-sleepy.

## 1 はじめに

人は日常のコミュニケーションにおいて様々な情報を伝達し合っている。その情報としては、文字で表される言語情報だけではなく、話しぶりや顔の表情、うなずきや首を傾げるといった動作、視線、身振り手振りといった様々な要素が存在する。人はそれらの情報を総合的に判断して瞬時に相手の感情や意図、態度などを認識している。これらのパラ言語情報が言語情報の不足を補うことで、発話者が真に伝えたい情報の正確な伝達が実現されているといえる。

近年では視線やジェスチャーなどの動作を行うアバターや擬人化エージェントの研究もされており、より人間らしく違和感の無いインターフェースを目指す試みがなされている。本研究では、パラ言語情報の中でも基本的な感情表出の手段であるアバターの顔の表情を対象として自然で人間らしい感情表出の実現を目指す。

表情においては、古くから静止画を用いた研究が行われてきたが、近年ではそれに加えて表情表出における変化速度や時間関数などの動的な情報が注目されている。代表的な研究としては、高速度カメラを用いて表情による感情表出を撮影し、顔の各部位の動きを定

量的に測定したもの [1] や、表情の変化速度が基本感情の認知に及ぼす影響を調べたもの [2] がある。特に [2] では、表情が表出する感情によって最も正しく認識される変化速度が異なることが明らかにされている。このように、表情による感情表出と動的な情報とは密接に関係していると考えられる。アバターを用いた表情生成においても、動的な情報を適切に制御することでより人間らしい感情表出が実現できるといえる。しかし、現段階では感情表出と動的な情報との関係は十分に解明されておらず、様々な感情表出における表情の動的な情報について調査する必要がある。

本研究は、「驚き」や「悲しみ」といった基本感情だけでなく、人対人の日常的な対話における感情状態を反映する表情を対象としている。人間が実際のコミュニケーションにおいて表出する感情や意図、態度などの情報をアバターに表出させることが可能になれば、より自然で人間らしいインターフェースが実現できる。筆者らは、これまでに自発音声から知覚される感情状態をパラメータとして表情生成を行うアバターを開発し、その有効性を確認した [3]。本稿では、アバターの表情による感情表出の自然性をさらに高めるための要素として、表情の変化速度に着目している。従来の研究により「驚き」や「悲しみ」などの基本感情と表情の変化速度との関係が示されていることから、アバターを

\*連絡先：宇都宮大学大学院工学研究科  
(〒 321-8585 栃木県宇都宮市陽東 7-1-2)  
E-mail: koh@speech-lab.org

用いた表情生成においても、人が自然であると感じる表情の変化速度には、そこに表出される感情状態が影響を及ぼしていることが予想される。本研究は、日常的な発話における感情状態と表情の自然な変化速度との関係を明らかにすることを目的とする。そこで、被験者に発話と同時に表出し始める表情の動画像を呈示し、その表情表出の自然性を評価させる心理学実験を行った。表情の呈示は、目標の表情への到達時間を変化させて行った。この実験の結果より、(1) 表情の変化速度が自然性の程度に与える影響、および (2) 表情の最適な変化速度と感情状態との関係、を分析した。

## 2 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 (UUDB)

本研究では、日常的な対話における感情状態を記述する手段として、感情の次元説 [4] にもとづく方法を用いる。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。

以下に本研究で用いる 6 次元の項目を示す [5]。

- 快-不快
- 覚醒-睡眠
- 支配-服従
- 信頼-不信
- 関心-無関心
- 肯定的-否定的

各次元について、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、のようにして表す。

本研究では、この 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る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 (UUDB) [6] を使用している。これは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。

ここに収録されている発話にラベリングされた感情状態をパラメータとして与えることでアバターの表情生成を行う。

## 3 表情合成の原理 [3]

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 [7] 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、FACS (Facial Action Coding System) [8] にもとづ

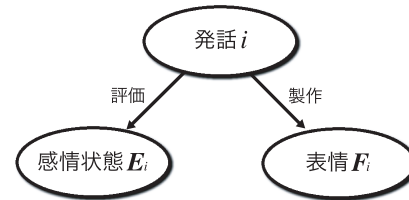


図 1: 対応データの収集.

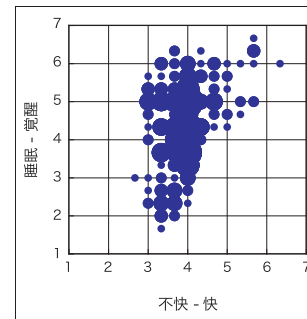


図 2: 話者 MKK による 249 発話の感情状態の分布.

き顔の各部位の動きを制御することが可能となっている。本研究では、感情状態に応じたアクション・ユニット (AU) をパラメータとして顔モデルを操作することで表情の生成を行っている。

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行う手法について述べる。具体的には、発話  $i$  に対する感情状態  $E_i$  と表情  $F_i$  の対応データを階層型 3 層ニューラルネットワーク (NN) に学習させることで、任意の感情状態に対して表情を決定する制御規則を構築する (図 1)。ここで、感情状態  $E_i$  は前述の 6 項目の次元の値、表情  $F_i$  は AU パラメータである。

対応データのうち、 $E_i$  については UUDB により与えられる感情状態の平均評価値を用いる。一方で、 $F_i$  については各発話に対して表情を製作する作業を行う。この作業には Galatea Toolkit のエージェント生成ツールを使用し、発話を聴取しながらその印象に合った表情を手作業で製作する。最終的に表情に使用した AU の組み合わせと動きの大きさの度合い (0 ~ 100%) のパラメータを表情データ  $F_i$  とする。

今回の実験では、話者 MKK による対話 C064 の 249 発話を対応データの収集に使用した。6 項目の中でも特に代表的な感情の次元である「快-不快」「覚醒-睡眠」の分布を図 2 に示す。ここで、円の面積はデータの個数を表している。図 2 より、使用した発話は幅広い感情状態の分布を持つことがわかる。

## 4 実験方法

前述の表情生成法を用いて、感情表出画像の呈示速度が自然性に与える影響を調べる心理学実験を行う。実

表 1: 各速度レベルにおける表情の表出時間.

	速度レベル	表出時間 $T(\text{ms})$
速	6	44
	5	87
	4	175
	3	349
	2	698
遅	1	1396

験では、発話音声と同時に変化し始める表情を被験者に呈示し、発話の印象に対する表情表出の自然性を5段階で評価させる。これを様々な変化速度に対して行うことで、表情の変化速度が自然性に与える影響を検討する。

実験で呈示刺激として使用する発話および表情について述べる。まず呈示する発話としては、「覚醒-睡眠」について幅広い分布を持つ話者 FSA を含む対話 C024 を UUDB から抜粋した。使用した発話数は 213 個で、うち話者 FSA が 113 発話、もう一方の話者 FTY が 100 発話である。実験では、これらを被験者に 1 発話ずつ発話順に呈示した。

呈示する表情は、各発話にラベリングされた感情状態の入力に対して NN により出力された AU パラメータによって操作した顔モデルである。これをターゲット表情とする。一方で、6 次元の評価値が全て中立 (4) の時の NN の出力をニュートラル表情とする。表情は発話の再生と同時にニュートラル表情から変化し始め、設定された時間後にターゲット表情に到達する。このときリップシンクは行わない。

表情の変化関数には式 (1) に示す 2 次系のステップ応答 (減衰定数=1) を使用した (図 3)。これは顔の動きは線形的な変化ではなく表出動作の完了近くになるにつれて変化量が減少していく [1] という知見を考慮したものである。

$$AU(t) = AU_{\text{target}} \{1 - e^{-\alpha t} (1 + \alpha t)\} \quad (1)$$

実験では AU の変化量がターゲット表情の 95% に到達したら 100% (ターゲット表情) に切り替えることとした。ここで、ターゲット表情の 95% に達する時間を表出時間  $T(\text{ms})$  とする。

表情変化における速度条件について述べる。本研究で扱う表情は従来の表情研究においてよく用いられる基本 6 感情のようなカテゴリ的なものではないため、過去の研究での表出時間と比較することは難しい。そこで、予備実験から 6 段階の速度レベルを選出した。各レベルの表出時間  $T(\text{ms})$  を表 1 に示す。実験では発話ごとに 6 速度のいずれかのレベルをランダムに呈示する。

表 2: 快グループと不快グループのデータ数.

	快グループ	不快グループ
話者 FSA	19	78
話者 FTY	31	42

表 3: 覚醒グループと睡眠グループのデータ数.

	覚醒グループ	睡眠グループ
話者 FSA	30	74
話者 FTY	59	34

評価作業の手順について述べる。被験者は心理学・社会学専攻の大学生の男 1 名女 9 名である。被験者は 1 名ずつ PC の前に座り、ヘッドフォンを装着して自身がコントローラ (図 3 左下) を操作しながら評価作業を進行する。

まず、画面にはニュートラル表情を表示させておく。被験者が再生ボタンを押すと、対話の 1 発話目が再生され、同時に表情変化が開始する。発話および表情の呈示が完了すると画面上に評価を促すメッセージが表示される。被験者は呈示された表情の自然性をスライダーを動かして評価し、次の 2 発話目へ進む。これを繰り返して評価作業を進める。

実験では、各発話に対して 6 速度がすべて評価されるように対話を 6 回繰り返す。すなわち被験者 1 名あたりの評価数は、213 発話  $\times$  6 速度 = 1278 である。これを 1 日 2 セッション (426 発話) ずつ、3 日に分けて評価実験を実施した。また、作業中の休憩は自由にとらせた。

## 5 実験結果

### 5.1 分析 1: 表情の変化速度が自然性の程度に与える影響

表情表出における変化速度の違いが自然性の程度に及ぼす影響を検討する。分析には、各表情に対する被験者 10 名による自然性評価の MOS (Mean Opinion Score) を用いる。以下に分析方法を述べる。まず実験に用いたデータを「快-不快」次元の値が中立である 4.0 より高いものは「快グループ」とし、一方で中立より低いものは「不快グループ」として分ける。そしてグループ内の全データに対して速度レベルごとの MOS の全体平均を求める。「覚醒-睡眠」次元に対しても同様に分析を行う。各グループのデータ数を表 2、表 3 に示す。

6 次元の項目のうち代表的な次元である「快-不快」「覚醒-睡眠」についての結果を図 4~7 に示す。

速度レベルを独立変数、自然性評価の MOS を従属変数として、各グループに対して一元配置の分散分析

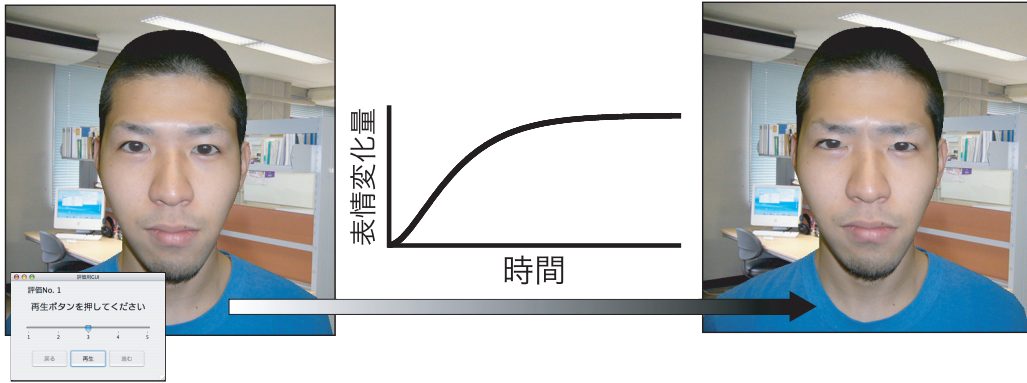


図 3: 表情変化の呈示方法.

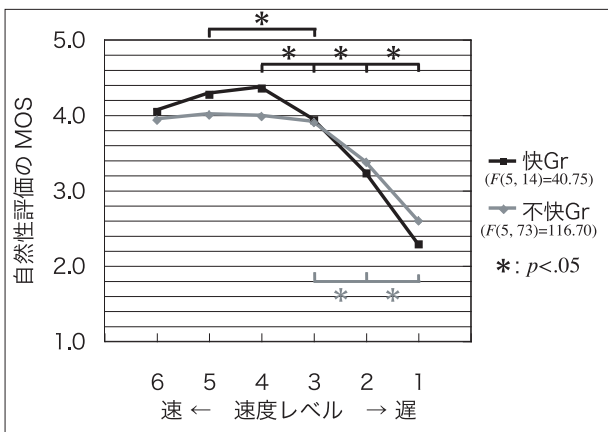


図 4: 表情の変化速度と自然性評価の MOS との関係 (FSA : 快 vs 不快).

を行ったところ、全てのグループにおいて速度の変化による MOS への有意な差が見られた ( $p < .05$ )。F 比を図 4~7 の図中に示す。また、下位検定 ( $p < .05$ ) の結果、速度レベル間で MOS の差が有意 ( $p < .05$ ) であったものは図中に\*を記す。

まず図 4 より、話者 FSA の快グループと不快グループどちらにおいても速度レベル 2、1 では MOS が大きく低下した。さらに速度レベル 6、5、4、3 において、快グループは速度レベル 4 を最大として MOS が山なりに変化していることがわかる。一方で、不快グループは速度レベル 6、5、4、3 間で MOS に有意な違いは見られない。

また、図 5 より、話者 FSA の覚醒グループと睡眠グループにおいても図 4 と似た傾向が見られた。すなわち、覚醒グループは速度レベルに対して MOS が山なりに変化するのに対して、睡眠グループは速度レベル 6、5、4、3 では有意な違いが見られない。図 6、図 7 の話者 FTY についても、快グループ、覚醒グループは速度レベル 6、5、4、3 において MOS に差が見られた。一方で不快グループ、睡眠グループには差が見られな

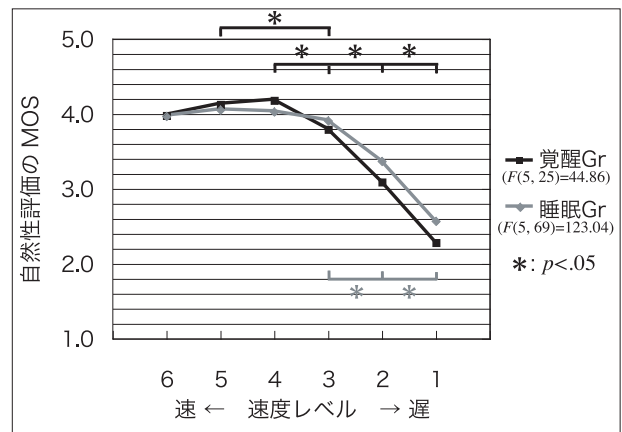


図 5: 表情の変化速度と自然性評価の MOS との関係 (FSA : 覚醒 vs 睡眠).

かった。

この結果より、表情変化における速度の違いが自然性の程度に与える影響は、表情が表出する感情状態によって異なることがわかる。特に「快」や「覚醒」といった感情状態については、表情の変化速度が自然性に大きく影響する。すなわち、表情の変化速度が自然な表情表出のためのより重要な要素となるといえる。一方で「不快」や「睡眠」などの感情状態については表情の自然性は変化速度に対して鈍感であるといえる。

## 5.2 分析 2: 表情の最適な変化速度と感情状態との関係

本研究は、日常的な発話における感情状態と表情の自然な表出速度との関係を明らかにし、アバターに感情表出を行わせる際に表情変化速度を適切に制御することを目指している。その実現のため、分析 2 では様々な感情状態に対する最適な表情変化速度を選出し、さらにそれらの間を関係を検討する分析を行う。

分析方法としては、各表情について 6 段階の変化速度のうち最も自然性評価の MOS が高いものを最適変



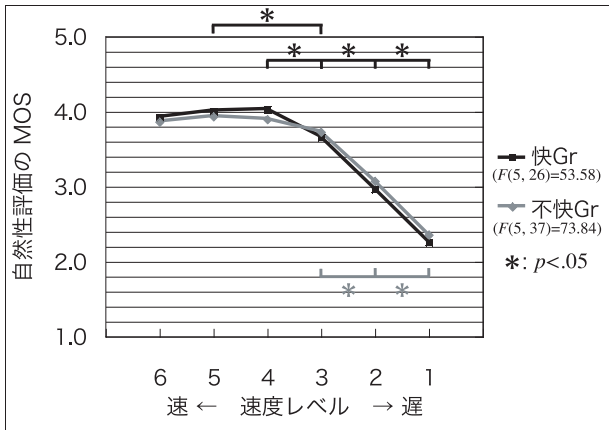


図 6: 表情の変化速度と自然性評価の MOS との関係 (FTY: 快 vs 不快).

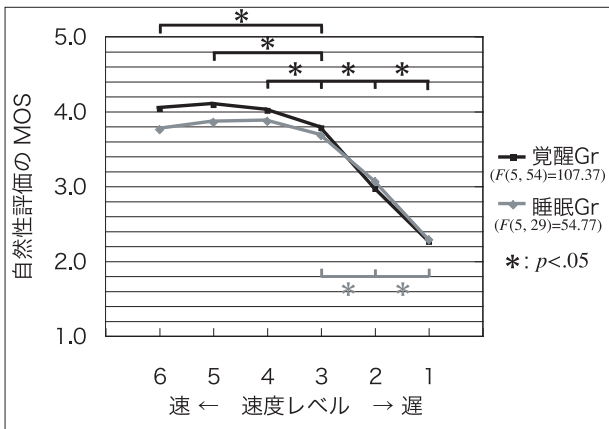


図 7: 表情の変化速度と自然性評価の MOS との関係 (FTY: 覚醒 vs 睡眠).

化速度として選出する。最高値が複数存在する場合にはどちらも選出することとする。話者 FSA、FTY それぞれについて、「快-不快」および「覚醒-睡眠」次元と表情の最適変化速度との関係を図 8、図 9 に示す。

分析の結果、図 8、図 9 において各感情次元と最適速度レベルとの間に目立った関係は見られないといえる。また、各感情次元について、中立である 4.0 より値が大きいデータと 4.0 未満のデータからなる 2 グループの感情状態を独立変数、最適速度レベルを従属変数として一元配置の分散分析を行ったが、感情状態の違いによる最適速度レベルの有意な違いは見られなかった。

この理由として発話時間が短い刺激に対する自然性評価の問題が考えられる。本研究では日常的な対話音声を使用しているため、実験に用いた発話には相槌などの短いものも含まれる。これらの発話時間が短い刺激に関しては、図 10 に示すように MOS が比較的良かった。また、短い発話では速い速度レベルが最適とされる傾向が見られた。この原因として、発話刺激の呈示

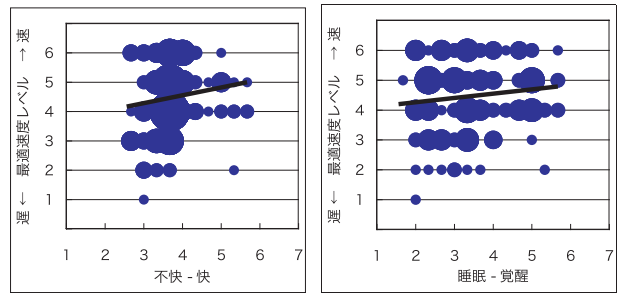


図 8: 感情次元と最適速度の関係 (FSA: 全データ).

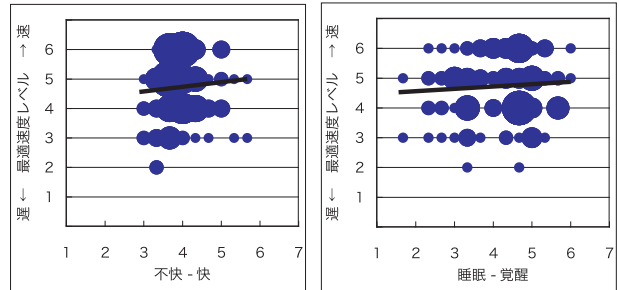


図 9: 感情次元と最適速度の関係 (FTY: 全データ).

時間が短いために発話の印象と表情表出との比較が困難であり、呈示タイミングが近い短い変化時間を最適とした可能性が考えられる。よって、これらは感情表出の印象によらず決定されたものとして、発話時間が短いものを除いた場合についても別途分析を行った。発話時間による影響を統制するため、しきい値は「覚醒-睡眠」次元にデータが同程度にばらつく 800ms 以上とした。話者 FTY については十分な「睡眠」の発話確保できなかったため分析対象としない。話者 FSA の発話時間が 800ms 以上の場合を図 11 に示す。

各感情次元について、中立である 4.0 より値が大きいデータと 4.0 未満のデータからなる 2 グループの感情状態を独立変数、最適速度レベルを従属変数として一元配置の分散分析を行ったところ、「快」と「不快」

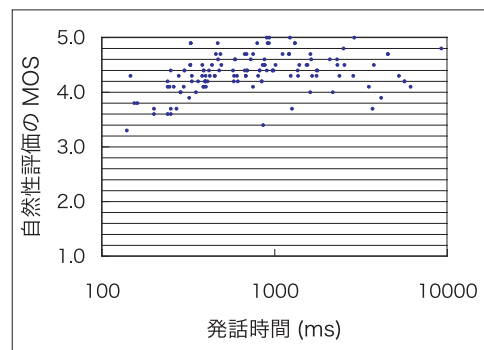


図 10: 最適速度として選出されたデータにおける発話時間と自然性評価の MOS との関係 (話者 FSA).

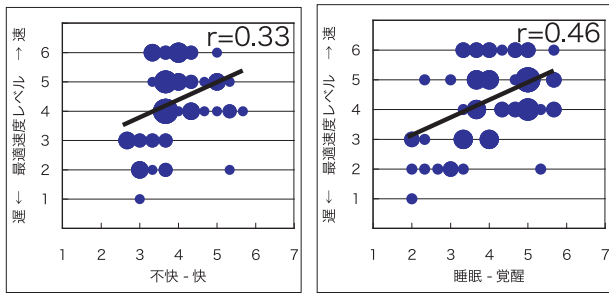


図 11: 感情次元と最適速度の関係 (FSA : 800ms 以上).

(図 11 左)の間に有意な差が見られた ( $F(1, 50) = 4.92, p < .05$ ). さらに「覚醒」と「睡眠」(図 11 右)の間にも有意な差が見られた ( $F(1, 50) = 8.02, p < .05$ ). さらに、相関係数は「快-不快」と最適速度レベルとの間で 0.33、「覚醒-睡眠」と最適速度レベルとの間で 0.46 であり、中程度の相関が見られた。

これより、発話時間がある程度長い刺激においては、「快」および「覚醒」の表情は「不快」および「睡眠」に比べて最も自然な表情変化速度が速くなるといえる。すなわち、「快」や「覚醒」の感情状態を表出する表情については比較的速く変化の方が自然であると知覚されている。一方で「不快」や「睡眠」の感情状態を表出する表情については比較的遅く変化の方が自然であると知覚されている。

感情次元のうち、より強い相関関係が見られた「覚醒-睡眠」は話者の心理活動の活発さを表す。実験の結果より、日常的な対話における話者の心理活動の活発さが、顔の部位の動きの速さによって表出されることで、感情表出がより自然なものとなると考えられる。また、「快-不快」は話者の気持ちの良さ悪しを表す。特に本研究で対象としている日常的な対話における「不快」とは話者が困っているような感情状態などを表すものであり、基本感情における「怒り」や「恐怖」のような感情とは異なる。今回の実験によって、そのような日常的な対話における「不快」に関しては表情が比較的遅く変化の方が自然であると知覚されることが明らかになったといえる。アバターを用いた人対人のコミュニケーションにおいても、これらの知見をもとに「快-不快」「覚醒-睡眠」の感情次元を独立変数として表情の変化速度を制御することで、ユーザにとってより違和感の無いインターフェースを実現できると考えられる。

## 6 まとめ

本研究では、アバターによる表情変化の動的な情報を制御するパラメータとして変化速度を取り入れることを目指し、日常的な感情状態を反映する自然な表情の変化速度を調査する心理学実験を行った。

実験の結果、表情変化における速度の違いが自然性の程度に与える影響は、表出する感情状態によって異なることが明らかになった。「快」や「覚醒」といった感情状態については、その変化速度が表情の自然性により大きく影響することがわかった。したがって、これらの感情状態をアバターに自然に表出させるためには変化速度を適切に制御することが重要であるといえる。一方で「不快」や「睡眠」などの感情状態については表情の自然性は変化速度に対して鈍感であった。

また、表情の自然性が最も高くなる変化速度と感情状態との関係を検討したところ、発話時間がある程度長い刺激において、「快」および「覚醒」の表情は「不快」および「睡眠」に比べて最も自然な表情変化速度が速くなるという結果が得られた。このことから、「快-不快」「覚醒-睡眠」の感情次元を表情変化速度の制御のための独立変数として用いることで、アバターの表情による自然な感情表出が実現できると考えられる。

今回行った心理学実験の結果、従来の研究で示されていた「驚き」や「悲しみ」などの基本感情に限らず、人対人の日常的な対話においても、そこで表出される感情状態が表情の自然な変化速度に影響を及ぼすことが明らかとなった。これはアバターを用いたネットワークコミュニケーションなどにおいて特に有用な知見となるといえる。本研究では今後アバターへの実装にむけて、今回得られたデータをもとに感情状態に対して変化速度を決定する制御規則を検討していく。

## 参考文献

- [1] 四倉, 内田, 山田, 赤松, 鉄谷, 森島: 高速度カメラを用いた顔面動作の分析および表情合成, 信学技報, IE2001-67, pp. 15-22 (2001).
- [2] 蒲池みゆき, 吉川左紀子, 赤松茂: 変化速度は表情認知に影響するか? — 動画刺激を用いた顔表情認知の時間特性の解明 —, 信学技報, HCS98-34, pp. 17-24 (1998).
- [3] H. Mori and K. Ohshima: Facial Expression Generation from Speaker's Emotional States in Daily Conversation, *IEICE Trans. Inf. & Syst.*, Vol. E91-D, No. 6, pp. 1628-1633 (2008).
- [4] 濱治世, 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).
- [5] 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).
- [6] H. Mori, T. Satake, M. Nakamura and H. Kasuya: UU Database: A spoken dialogue corpus for studies on paralinguistic information in expressive conversation, *Text, Speech and Dialogue*, LNCS 5246, Springer, pp. 427-434 (2008).
- [7] 川本, 下平, 新田, 西本, 中村, 伊藤, 森島, 四倉他: 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の設計, 情報処理学会論文誌, Vol. 43, No. 7, pp. 2249-2261 (2002).
- [8] P. Ekman and W. V. Friesen, 工藤力訳: 表情分析入門, 誠信書房 (1987).