

アバター間コミュニケーションのための 感情に基づく表情生成システム

服部 嵩[†] 長尾 確[‡]
名古屋大学 情報学部[†] 名古屋大学 大学院情報学研究科[‡]

1. はじめに

近年、Virtual Reality (以下 VR) を用いた VR ミーティングや VR SNS といった、VR 上におけるアバター間でのコミュニケーションが増加している。アバター間でのコミュニケーションは、現実とは異なり表情を読み取ることができず、コミュニケーションが阻害される課題がある。自動で表情生成する手法には、HMD を装着した人の顔の動きを読み取るフェイシャルトラッカーを利用し、現実の顔の動きをアバターの顔で再現する手法が存在する。しかし、違和感のない表情生成には詳細なアバターチューニングが必要であり、多大な労力が必要である。

本研究では、フェイシャルトラッカーと音声データから感情推定を行い、推定結果に基づく表情をアバターに反映するシステムを提案する。フェイシャルトラッカーのデータを直接利用しないことによりアバターの制限を緩和し、より柔軟な表情生成を可能とする。

2. 提案システム

本節では、提案システムの構成手法について述べる。図 1 は提案システムの構成図を示しており、感情を機械学習によって推定する推定段階と推定された感情をアバターの表情に反映する反映段階に 2 つの段階から構成される。次項からは、表情と感情の対応の説明のため反映段階、推定段階の順に各段階について述べる。

2.1 表情反映

2.1.1 感情の定量化

感情を分類し定量化するための手法には様々なものが存在するが、本研究では感情の定量化にラッセルの円環モデル[1]を利用する。ラッセルの円環モデルはすべての感情を Valence と Arousal の二軸で表現可能とする感情モデルであり、すべての感情状態を推定できるという点で優れているが、2 軸の値の絶対的な基準は存在しない。また、感情の分類法は厳密には定められていない。

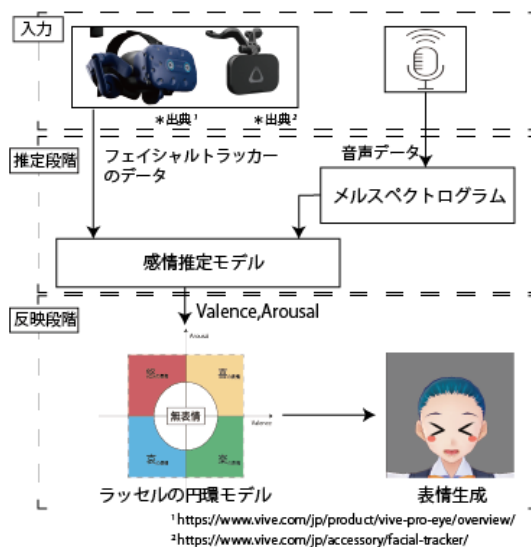


図 1. システム構成図

2.1.2 感情の分類と表情への表出

前項で述べたように、ラッセルの円環モデルでは感情の分類法を定めていないが、G.Valenza らの研究[2]では、第一象限 (Valence と Arousal が共に高い) を Happy、第二象限 (Valence が低く Arousal が高い) を Angry、第三象限 (Valence が高く Arousal が低い) を Relaxed、第四象限 (Valence と Arousal が共に低い) を Sad として分類している。これは日本語で言う喜怒哀楽に対応する感情分類であり一般的に広く浸透している。ラッセルの円環モデルは原点からの距離が感情の強さを示すため、原点近くの場合は感情がアバターに表出しないものとし、Valence と Arousal の値から図 2 のように表情生成を行うものとした。

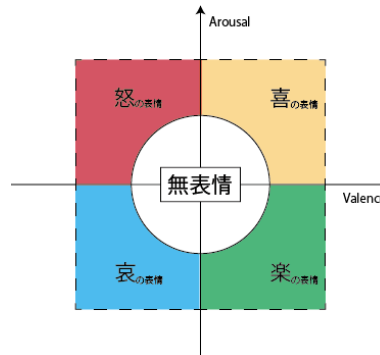


図 2. ラッセルの円環と表情表出の対応

Emotion-Based Facial Expression Generation System
for Inter-Avatar Communication in Virtual Space
[†]HATTORI, Takashi (thattori@nagao.nuie.nagoya-u.ac.jp)
[‡]School of Informatics, Nagoya University
^{††}NAGAO, Katashi (nagao@i.nagoya-u.ac.jp)
^{‡‡}Graduate School of Informatics, Nagoya University

2.1.3 アバター作成

本システムに利用するアバターには Pixiv が提供する 3D アバターモデリングソフト VRoid Studio¹ を用いる。VRoid Studio は簡単な操作で人型のアバターを作成することができ、それらは図3のように喜怒哀楽に対応する表情をデフォルトで表現できる。デフォルトの表情は現実の表情変化に似せたものだが、アバターの表情には感情が伝わりやすいようデフォルメ化された表情も存在する。そこで、デフォルメ表情の作成を行い、デフォルメ化された表情とそうでない表情どちらが優れているかの比較を行う。



図2. デフォルトで表現できる表情



図3. 作成したデフォルメ表情

2.2 感情推定

2.2.1 データセット作成

感情推定を機械学習により行うには、フェイシャルトラッカーデータ、音声データと Valence、Arousal の注釈を持つデータセットが必要である。本研究では、池田らの研究[3]を参考に、生体情報から得られる情報は他モダリティよりも正確であるとし、脈拍と脳波を用いて Valence、Arousal の注釈を付けデータセットの作成を行う。

2.2.2 データセット作成のための被験者実験

本システムは VR 上でコミュニケーションをとる人に対する感情推定を想定しているため、データセットも同様の条件下で作成する必要がある。そこで、フェイシャルトラッカーにより表情生成された VRoid アバター同士で対話可能な記録システムを作成し、実験手順に沿ってデータの記録を行った。なお、実験には 14 人(男性:13 名, 女性:1 名, 年齢:20~24 歳)に協力をして頂いた。

実験手順は以下の通りである。

1. 事前アンケートによる趣味嗜好の確認
2. 趣味嗜好の合う 2 名の選択、VR 上で自然な対話をするように説明
3. フェイシャルトラッカーを装備した HMD・脳波計・脈拍計を装着
4. 記録の後、事後アンケートの実施

¹ <https://vroid.com/studio>

2.2.3 事後アンケート結果

手順5で行ったアンケート(“1:全く, 2:少し, 3:かなり”とした程度の三段階評価)の結果を図4に示す。本アンケートは、記録用のシステムに対してのものであり、フェイシャルトラッカーを利用し、現実の顔の動きをアバターの顔で再現する手法に対しての意見である。結果としては、VR 上のコミュニケーションにアバターの表情は必要だと感じる人は多いにも関わらず、フェイシャルトラッカーを用いた手法ではうまく感情を伝達できていないことが確認できた。

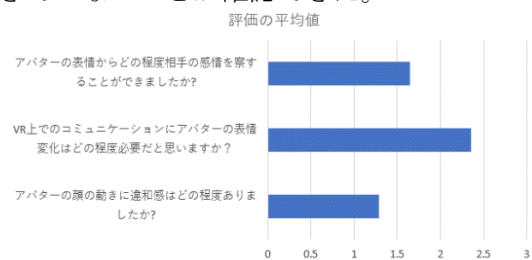


図4. アンケート結果

2.2.4 感情推定モデルの構築

被験者実験により作成したデータセットを用いて感情推定モデルの構築を行った。実現方法はまず、音声データをメルスペクトログラムに変換後 CNN にて特徴量抽出を行う。次に、顔の特徴量をフェイシャルトラッカーデータとし、音声データの特徴量と統合する。その後、統合した特徴量を P. Tzirakis らの研究[4]を参考に 2 層の LSTM モデルに入力し、出力に対して線形回帰を行い Valence、Arousal の値を推定する。

3. おわりに

本稿では、フェイシャルトラッカーと音声データから感情推定を基に表情をアバターに反映するシステムを提案した。また、機械学習データ収集のための実験を行った。その際のアンケートにより提案システムの必要性を確認した。今後は、表情生成システムを実装・評価した後、感情推定モデルの改善を行い、より適切な表情を表示可能にする予定である。

参考文献

- [1] J.A. Russell, “A Circumplex Model of Affect”, Journal of Personality and Social Psychology, 1980.
- [2] G. Valenza, L. Citi, A. Lanatá, E.P. Scilingo & R. Barbieri, “Revealing Real-Time Emotional Responses: A Personalized Assessment based on Heartbeat Dynamics”, Sci. Rep. 4, 2014.
- [3] 池田悠平, 岡田桂子, 堀江亮太, 菅谷みどり, “表情と生体情報を用いた感情の推測方法の検討”, マルチメディア, 分散, 協調とモバイル(DICIMO2017)シンポジウム, 2016.
- [4] P. Tzirakis, G. Trigeorgis, M.A. Nicolaou, B. Schuller, & S. Zafeiriou, “End-to-End Multimodal Emotion Recognition using Deep Neural Networks”, IEEE Journal of Selected Topics in Signal Processing, Vol. 11, No. 8, 2017.