

感情モデルへの自在なマッピングによるバーチャルエージェント 顔表情作成ツールの実現

A Tool for Creating Virtual Agent Facial Expressions by Flexible Mapping to Emotion Models

川添 和希^{1*} 坂本 大介¹ 小野 哲雄¹
Kazuki Kawazoe¹ Daisuke Sakamoto¹ Tetsuo Ono¹

¹ 北海道大学

¹ Hokkaido University

Abstract: In this research, we developed the tool for creating the facial expressions of virtual agents(VA) which correspond to valence and arousal of Russell's circumplex model. The tool uses specific facial parts of VA to create more expressive facial expression. Also, valence and arousal of Russell's circumplex model is able to be extracted from various things such as conversational texts and voices, so we expect we can use the tool in various fields. In this research, as an application of the tool, we create videos in which VA change their facial expressions more expressively by extracting valence and arousal from videos of a human.

1 はじめに

人間の感情を研究で利用しやすくするために、人間の感情を快-不快と覚醒-眠気の2次元情報にマッピングすることができるラッセルの円環モデル [1] が用いられている。ラッセルの円環モデルを用いて感情を数値に変換することで、ニューラルネットワークの入力などの数値として使用したい場合に感情を用いることができるようになる。また、ラッセルの円環モデルの値は幅広い感情を表現することができるという点も利点の1つであり、文章や音声、人間の顔表情画像などが表す感情をラッセルの円環モデル上の値として推定する研究 [2][3][4] も行われている。本研究では以降、ラッセルの円環モデル上の値のことを感情の値と記述する。

また近年、仮想のCGキャラクターを用いた技術が社会で利用される機会が増えている。仙台駅、盛岡駅などの4駅の案内をするバーチャルエージェント (VA) 「陸奥めぐ」や、オンライン動画プラットフォームで2Dや3Dアバターを用いて配信活動を行うVTuberなどが例として挙げられる。VA自身の感情を表すための手段の1つとしてVAの顔表情を用いることができる。しかし、VAの顔表情の種類は事前に作成した顔表情の数に依存し、限られた少ない種類の顔を用いるため、VAで表現できる感情の種類が限られてしまう。また、

VA自身の感情を強調して表現するために、黒目のサイズを小さくする、頬の色を赤くするなどをして、人間の顔では再現できない顔表情を用いることが多い。そのため、人間の顔表情をVAの顔に反映させる際に人間の顔のトラッキングを行うだけでは、VAの感情表現を強調した顔を作成することができない。

そこで本研究では、特定の感情にとどまらない幅広いVAの顔表情を作成できるようにすることを目的としたVA顔表情作成ツールを提案する。提案するVA顔表情作成ツールでは、黒目の大きさや頬の色など、VA特有の顔のパーツも用いた顔表情を作成することができるようにし、感情表現をより強調したVAの顔を作成できるようにする。また、感情の値をVA顔表情作成ツールの入力とすることで、ラッセルの円環モデル上の自由な感情の顔を作成できるようにするとともに、感情の値を取り出すことができる文章や音声、人間の顔表情画像などのさまざまなものを入力として使用できるようにする。また本研究では、人間の顔表情を、感情表現が強調されたVAの顔表情に反映させることを目的とし、人間の顔表情変化動画から感情表現が強調されたVAの顔表情変化動画の作成を行う。VAの顔表情変化動画の作成には、元となる人間の顔表情変化動画内の人間の顔表情が表す感情の値からVAの顔表情を作成する。人間の顔表情から感情表現が強調されたVAの顔表情への変化及び、VA表情変化動画作成により、提案するVA顔表情作成ツールの汎用性を示す。

*連絡先: ヒューマンコンピュータインタラクション研究室
〒060-0814 北海道札幌市北区北14条西9丁目
E-mail: kawazoe.kazuki.b8@elms.hokudai.ac.jp

2 関連研究

感情の値から VA の顔表情を作成する研究として、Gobron ら [5] や Ahn ら [6] の研究があるが、各研究において使用する VA の顔が実際の人間に近い顔を使うことを想定しており、頬の色や黒目の大きさなどを使用した感情表現を強調した顔表情を作成することができない。

また、顔表情変化についての研究として、Recio ら [7] や大島ら [8]、中谷ら [9] の研究がある。Recio ら [7] と大島ら [8] の研究では、特定の顔表情から特定の顔表情へのモーフィング速度に対する検討が行われていたが、モーフィングが終わった後にさらに他の顔へモーフィングを行うといった連続したモーフィング処理に関しては議論されていなかった。中谷ら [9] の研究では、リアルタイムで人間の顔の特徴点から VA の顔表情に変更するという、連続的なモーフィング処理を行っていた。しかし、変更対象の顔は準備された決まった顔のみしか使用できないようになっていた。

本研究では、目の大きさや口の大きさなど、実際の人間とは異なる顔のパーツを持つ VA も使用するとともに、人間にはできない頬の色や黒目の大きさも表情の 1 つとして考慮した顔を感情の値から作成できるツールを開発する。また、開発する VA 顔表情作成ツールを利用し、モーフィング処理における変更後の顔を定期的に作成することで、元とする人間顔表情変化動画内の人間の顔に合わせたさまざまな顔へ変化する VA 顔表情変化動画の作成を行う。

3 提案する VA 顔表情作成ツール

3.1 使用する VA



図 1: 使用する VA

本研究では、顔表情を変更する VA として図 1 の Live2D の女性キャラクターと男性キャラクターを用いる。顔表情の表現に用いる顔のパラメータとして、女性キャラクターでは目の開閉度、目尻の下がり具合、黒目の大きさ、口の開閉度、口角の上がり具合、眉の寄

り具合、眉の上がり具合、頬の赤さの計 8 個を使用した。また、男性キャラクターにおいては頬の赤さを抜いた計 7 個のパラメータを使用した。

3.2 ツールの作成方法



図 2: ツールに使用する各感情の顔表情

VA の顔表情作成ツールでは、入力とする感情の値から、事前に作成した特定の感情の顔のパラメータを何割使用するかを求め、各表情のパラメータを求めた割合のみ使用する顔を作成する。VA の顔表情作成ツールの作成にあたって、最初にツールで使用する angry, happy, calm, sad, sleepy, astonished の計 6 つの感情の顔表情を図 2 のように作成する。使用する 6 つの感情のうち angry, happy, calm, sad は各象限における代表感情として取り上げ、各表情は Yamada[10] の基本感情の顔のイラストや Tottenham ら [11] の NimStim Set に含まれる特定の感情を表している人間の顔表情画像をもとに作成した。また、sleepy, astonished においては、顔表情作成ツールの表現の幅を広げるために使用した。sleepy, astonished の顔は図 2 のようになって

おり、sleepy の顔は各象限の代表感情の顔よりも目の開閉度が小さく、astonished の顔は口の開閉度が大きい、眉の y 座標が大きい、口の開閉度が大きいという特徴がある。VA 顔表情作成ツールは表情作成のために用いる顔のパラメータを使用するため、sleepy、astonished の顔表情を VA 顔表情作成に使用することにより、VA 顔表情作成ツールにて sleepy と astonished の顔の持つ特徴を表現することができるようになる。

VA 顔表情作成に使用する感情 k の使用割合 r_k は式 (1) のように求め、式 (1) で使用する r_{top_k} は式 (2) のようにして求める。式 (2) ではラッセルの円環モデル上における入力感情の値の座標から、VA 顔表情作成に使用する 6 つの感情の座標までの各距離 d_k を用いる。ラッセルの円環モデル上における各感情の座標は Paltoglou ら [12] のラッセルの円環モデルをもとに設定する。また、 w は使用割合を求める際の距離 d_k の重要度を表している。 w が大きいとき、入力感情の値の座標から距離が近い感情の使用割合が大きくなり、 w が小さいとき、各感情の使用割合は入力感情の値の座標に依存しない平均的な値となる。本研究では、女性キャラクターと男性キャラクターの顔表情作成において $w = 2$ とする。

$$r_k = \frac{r_{top_k}}{\sum_{i=1}^6 r_{top_i}} \quad (1)$$

$$r_{top_k} = \frac{1}{(100d_k)^w + \epsilon} \quad (2)$$

式 (1) から求めた各感情の割合 r_k と、各感情の表情における顔の各パーツのパラメータ p_k を用いて VA の作成する顔の各パーツのパラメータ p を式 (3) のように求める。

$$p = \sum_{k=1}^6 p_k r_k \quad (3)$$

4 VA 顔表情作成ツールの評価実験

4.1 実験参加者

VA 顔表情作成ツールの評価実験では、クラウドソーシングサービスである Amazon Mechanical Turk を用いたアンケート調査を行った。アンケート調査は男性 55 名、女性 45 名の計 100 名（年齢: Mean=36.45, SD=10.12）に行った。

4.2 実験概要



図 3: VA 顔表情作成ツール評価実験のアンケートページ

実験では、3 節で提案した VA 顔表情作成ツールにおける各感情の顔表情の使用割合をナイーブな計算式で求めるものと、提案する計算式で求めるものの比較実験を行った。ナイーブな計算式では式 (1) で使用する r_{top_k} を式 (4) のようにして求める。式 (4) のナイーブな計算式は、単純に距離の割合を使用割合とする考え方であり、 $range_{xy}$ はラッセルの円環モデル上における 2 軸の範囲の大きさである。本研究ではラッセルの円環モデル上における 2 軸の範囲を $[-3, 3]$ としたので、 $range_{xy} = 6$ とした。

$$r_k = range_{xy} - d_k \quad (4)$$

実験では、図 3 のように特定の感情の値から作成した顔を参加者に見せ、入力で使用した感情の値に対応する感情にどれくらい見えるかということ調査した。実験の入力で使用した感情は Paltoglou ら [12] のラッセルの円環モデル上における 98 個の感情のうち、単語として使用頻度が高い 27 個の感情を使用した。使用頻度の高さは Word frequency data[13] を参考にした。また、VA 顔表情作成ツールで作成した顔を参加者に評価してもらう際に、比較対象とする Neutral な顔表情を隣に配置した。各 VA における Neutral な顔表情は Yamada[10] の Neutral の顔のイラストや Tottenham ら [11] の NimStim Set に含まれる人間の Neutral な顔表情画像を参考にして作成し、図 1 の顔表情にした。

アンケートでは、使用した感情に見えるかどうかを、以下のようなアンケート文にどれくらい賛成しているか、賛成していないかということ 7 段階形式で回答してもらった。

(アンケート文) She/He on the left looks (emotion).
(左側の彼女/彼は (emotion) に見える。)

また、質問は男性キャラクターと女性キャラクターのセクションに分け、各セクションの順番とそれぞれの

セクションにおける質問の順番をランダムに表示するようにした。

4.3 実験結果

表 1: VA 顔表情作成ツールのアンケート結果（女性キャラクター）

感情	ナイーブ		提案		t 値	p 値
	平均	SD	平均	SD		
afraid	4.73	1.54	5.12	1.47	2.184	0.031*
ambitious	4.68	1.55	4.17	1.99	-2.824	0.006**
angry	3.86	1.85	5.02	1.74	6.157	<.001**
apathetic	4.48	1.53	4.55	1.49	0.445	0.657
astonished	5.45	1.26	6.22	1.17	6.180	<.001**
attentive	5.04	1.41	5.36	1.35	2.209	0.029*
bitter	4.27	1.69	4.54	1.73	2.015	0.047*
bored	5.01	1.44	5.24	1.37	1.472	0.144
confident	4.98	1.55	5.35	1.35	2.304	0.023*
convinced	4.70	1.51	4.59	1.64	-0.690	0.492
disappointed	5.12	1.58	5.56	1.24	2.921	0.004**
disgusted	4.28	1.81	4.33	1.80	0.279	0.781
dissatisfied	4.57	1.56	5.55	1.29	5.891	<.001**
embarrassed	4.52	1.60	4.62	1.75	0.552	0.582
excited	5.01	1.63	5.15	1.81	0.769	0.444
friendly	5.45	1.28	5.53	1.53	0.502	0.617
glad	5.24	1.43	5.73	1.27	3.510	<.001**
happy	4.88	1.75	6.01	1.23	6.801	<.001**
interested	5.19	1.37	5.62	1.37	3.202	0.002**
jealous	4.18	1.77	4.76	1.66	3.870	<.001**
passionate	4.60	1.55	5.05	1.43	3.349	0.001**
peaceful	5.47	1.35	5.71	1.34	1.830	0.070
sad	5.39	1.34	5.97	1.23	4.017	<.001**
serious	4.24	1.70	4.36	1.82	0.811	0.419
suspicious	4.32	1.72	4.51	1.55	1.183	0.240
tired	5.19	1.51	6.00	1.19	5.517	<.001**
worried	4.70	1.66	4.78	1.63	0.542	0.589

* $p < .05$ ** $p < .01$

VA 顔表情作成ツールから作成した女性キャラクターの顔表情のアンケート結果は表 1 のようになった。ナイーブな計算式を用いて作成した VA 顔表情の評価と、提案する計算式を用いて作成した VA 顔表情の評価の平均に差があるのかを調べるために t 検定を行ったところ、27 個の感情のうち、16 個の感情において有意差がみられた。有意差が見られた 16 個の感情のうち ambitious を除く 15 個の感情において、提案する計算式を用いて作成した VA 顔表情の方が有意に大きい評価を得られた。

VA 顔表情作成ツールから作成した男性キャラクターの顔表情のアンケート結果は表 2 のようになった。ナイーブな計算式を用いて作成した VA 顔表情の評価と、

表 2: VA 顔表情作成ツールのアンケート結果（男性キャラクター）

感情	ナイーブ		提案		t 値	p 値
	平均	SD	平均	SD		
afraid	4.65	1.51	4.39	1.70	-1.588	0.115
ambitious	4.85	1.59	4.03	1.96	-4.307	<.001**
angry	3.66	1.90	6.33	1.18	11.495	<.001**
apathetic	4.80	1.48	4.81	1.38	0.066	0.947
astonished	5.59	1.25	6.19	1.21	5.405	<.001**
attentive	4.99	1.47	5.14	1.39	0.917	0.361
bitter	4.44	1.87	4.71	1.58	1.642	0.104
bored	5.47	1.32	5.74	1.18	2.139	0.035*
confident	4.99	1.48	5.36	1.45	2.595	0.011*
convinced	4.88	1.45	4.79	1.56	-0.682	0.497
disappointed	5.48	1.31	5.89	1.29	2.762	0.007**
disgusted	4.12	1.78	4.93	1.50	4.312	<.001**
dissatisfied	5.14	1.37	5.84	1.14	4.772	<.001**
embarrassed	4.45	1.76	4.84	1.68	2.256	0.026*
excited	5.68	1.18	6.03	1.27	2.836	0.006**
friendly	5.55	1.47	5.85	1.24	2.398	0.018*
glad	5.45	1.28	5.86	1.21	3.397	<.001**
happy	5.61	1.23	6.29	1.24	4.370	<.001**
interested	5.46	1.36	5.84	1.20	2.732	0.007**
jealous	3.91	1.72	4.62	1.53	4.376	<.001**
passionate	4.63	1.57	5.11	1.40	3.550	<.001**
peaceful	5.56	1.27	5.79	1.24	1.662	0.100
sad	5.70	1.24	6.29	1.25	5.542	<.001**
serious	4.14	1.85	4.10	1.92	-0.218	0.828
suspicious	4.38	1.83	4.57	1.63	1.192	0.236
tired	5.49	1.36	6.10	1.21	4.502	<.001**
worried	4.57	1.65	4.76	1.77	1.243	0.217

* $p < .05$ ** $p < .01$

提案する計算式を用いて作成した VA 顔表情の評価の平均に差があるのかを調べるために t 検定を行ったところ、27 個の感情のうち、17 個の感情において有意差がみられた。有意差が見られた 17 個の感情のうち ambitious を除く 16 個の感情において、提案する計算式を用いて作成した VA 顔表情の方が有意に大きい評価を得られた。

4.4 考察

女性キャラクターにおいても、男性キャラクターにおいても、提案する計算式を用いて作成した VA 顔表情の方が多くの感情において評価が有意に大きかったため、VA 顔表情作成ツールは有用であるといえると考えられる。しかし、有意に評価が大きい感情が女性キャラクターには 11 個、男性キャラクターには 10 個あった。

ambitious は両方のキャラクターにおいて評価が有意

に小さくなっていた。ambitiousはPaltoglouら[12]のラッセルの円環モデル上において、VA顔表情作成に使用する感情であるastonishedの座標に近い。提案する計算式を用いた場合、ナイーブな計算式を用いた場合よりも距離の近い感情の使用割合が大きくなるので提案する計算式を用いて作成したambitiousのVA顔表情はastonishedのVA顔表情に近くなっていた。そのことから、ambitiousの評価が有意に小さくなっていたのは、astonishedのVA顔表情はambitiousの顔表情に見えないと考えた人が多くいたからではないかと考える。

apathetic, convinced, peaceful, serious, suspicious, worriedは両方のキャラクターにおいて評価に有意差がみられなかった。apatheticとworriedはPaltoglouら[12]のラッセルの円環モデル上において、VA顔表情作成に使用する6感情の重心の近くの座標にあった。VA顔表情作成に使用する6感情の重心に近いほど、ナイーブな計算式を用いて作成したVA顔表情と提案する計算式を用いて作成したVA顔表情は同じになるため、apatheticとworriedのVA顔表情における評価に差が出なかったのだと考える。また、convinced, serious, suspiciousはナイーブな計算式を用いて作成したVA顔表情と提案する計算式を用いて作成したVA顔表情の両方の評価が低く、かつ作成したVA顔表情はナイーブな計算式のものと同様で異なっている。そのため、2つの異なるVA顔表情が同じくらいにconvinced, worried, suspiciousのVA顔表情に見えなかったということであり、convinced, serious, suspiciousのVA顔表情がどのような顔であれば評価が高くなるかを検討する必要があると考える。

angry, sad, dissatisfied, disgustedなどのPaltoglouら[12]のラッセルの円環モデル上における快-不快の軸において負の座標にある感情の顔の評価が女性キャラクターの方が男性キャラクターよりも低いことから、女性キャラクターは快-不快の軸における負の座標にある感情を男性キャラクターより表現できていなかったということがわかる。女性キャラクターは男性キャラクターよりも実際の人間の顔から離れた顔であることから、女性キャラクターの顔表情が見慣れないものとなっており、快-不快の軸の負の表現を弱めていた可能性があると考えられる。

5 VA顔表情変化動画の生成

VAの顔表情変化動画作成では、元動画とする人間の表情変化動画の各フレームにおける人間の顔表情の情報を利用する。元動画における顔検出にはOpenCVを使用し、特徴点の検出にはToisoulら[14]のEmoFANを使用する。本研究では、作成するVA顔表情変化動画

におけるVAの目と口は、開いていると検出したときに開け、閉じていると検出したときに閉じるようにする。目の開閉検出にはEye Aspect Ratio (EAR)、口の開閉検出にはMouth Aspect Ratio (MAR)[15]を使用した。本研究ではEARとMARの閾値はそれぞれ0.2とした。

VAの顔更新は、更新前の顔から更新先の顔に0.3秒間で変更するようにする。0.3秒毎に更新先の顔表情を変更することで、定常的に顔表情が変化するようにする。顔表情変化にかかる時間を0.3秒としたのは、Recioら[7]の研究において、顔表情変化に適した時間が0.2~0.5秒だと示しているからである。

5.1 ナイーブなVA顔表情変化動画生成方法

ナイーブなVA顔表情変化動画作成方法として、元動画において検出した感情の顔表情に変化する方法を用いる。人間の顔表情が表す感情の推定にはToisoulら[14]のEmoFANを使用する。また、EmoFANで推定できる感情であるSurprise, Happy, Fear, Anger, Contempt, Disgust, Sad, Neutralに対応する顔表情をVA顔表情作成ツールにて作成し、VA顔表情変化動画におけるVAの顔表情として使用する。使用する各感情のラッセルの円環モデルにおける座標はPaltoglouら[12]のラッセルの円環モデルを参考にした。Paltoglouら[12]のラッセルの円環モデル上において存在しなかった感情であるSurprise, Fear, Neutralは、本研究ではSurpriseはPaltoglouら[12]のラッセルの円環モデル上におけるastonishedの座標、Fearはafraidの座標、Neutralは(0,0)とする。

VA顔表情変化動画において、0.3秒ごとに更新先の顔表情を変更する必要があるが、ナイーブなVA顔表情変化動画作成方法では、更新先の顔表情として、元動画における0.3秒毎の代表感情の顔表情を使用する。0.3秒毎の代表感情とは、元動画における各0.3秒間に含まれるフレームにおいて、最も検出した数が多い感情である。

5.2 提案するVA顔表情変化動画生成方法

本研究では、VA顔表情変化動画作成方法として2つの方法を提案する。

1つ目は、元動画において検出した感情の値の顔表情に変化する方法である。0.3秒ごとの更新先の顔表情として、元動画における0.3秒毎の感情の値を使用して作成した顔表情を用いる。0.3秒毎の感情の値とは、元動画における各0.3秒間に含まれるフレームにおいて、検出した感情の値の平均の値である。

2つ目は、更新先の顔表情として、元動画における1.5秒間分の感情の値を利用して作成した顔を用いる方法である。0.0~0.30秒における0.3秒間、0.31~0.60秒における0.3秒間、0.61~0.90秒における0.3秒間、0.91~1.20秒における0.3秒間、1.21~1.50秒における0.3秒間における感情の値を利用し、更新先の顔表情作成に使用する感情の値を求め、それぞれの0.3秒間における感情の値を $ev_1, ev_2, ev_3, ev_4, ev_5$ とする。更新先の顔表情作成に使用する感情の値 $emov$ は、式(5)のようにして求める。式(5)のように求めることで、1つ目の提案方法よりも更新先の感情の値の振動を抑えつつ、更新先の感情の値を強調するため、表情変化がより滑らかになる。また、 s は更新先の顔の強調度を表している。 s が大きいとき、各0.3秒間における感情の値の強調度合いの差が大きくなるため、瞬間的に推定した感情の値の顔になるとともに、表情変化の滑らかさが減少する。 s が小さいとき、各0.3秒間における感情の値の強調度合いの差が小さくなるため、1.5秒間における平均的な顔になるとともに、表情の滑らかさが増加する。本研究では、女性キャラクターと男性キャラクターの顔表情作成において $s = 4$ とする。

$$emov = \frac{ev_1 + (1+s)ev_2 + (1+2s)ev_3 + (1+s)ev_4 + ev_5}{1 + (1+s) + (1+2s) + (1+s) + 1} \quad (5)$$

$$= \frac{ev_1 + (1+s)ev_2 + (1+2s)ev_3 + (1+s)ev_4 + ev_5}{5 + 4s}$$

本研究では以降、ナイーブな方法を感情の方法、1つ目の提案方法を感情の値の方法、2つ目の提案方法を5段階の方法と呼ぶ。

6 VA 顔表情変化動画の評価実験

6.1 実験参加者

VA 顔表情変化動画の評価実験では、クラウドソーシングサービスである Amazon Mechanical Turk を用いたアンケート調査を行った。アンケート調査は男性59名、女性41名の計100名(年齢: Mean=34.99, SD=9.39)に行った。

6.2 実験概要

実験では、4節で述べた感情の方法で作成したVA顔表情変化動画と提案する2つの方法で作成したVA顔表情動画、計3種類の動画の比較実験を行った。それぞれのVA顔表情変化動画を作成する際に入力として使用する人間の顔表情変化動画には図4の2つの15秒

間の動画を用いた。2つの元動画では、女性が常に顔表情を変化させ続けていた。また、VAは女性キャラクターと男性キャラクターの2種類を使用した。3種類のVA表情変化動画生成方法、2つの元動画、2種類のVAを使用したため、計12個のVA顔表情変化動画を使用した。



(a) 元動画 1

(b) 元動画 2

図 4: 実験で使用した人間の顔表情変化動画

実験では、最初に元動画を見せたのちに、その元動画から作成したVA顔表情変化動画を見せ、以下のアンケート(Q1),(Q2)を回答してもらった。アンケート(Q1)においてはアンケート文にどれくらい賛成しているか、賛成していないかということを7段階形式で回答してもらい、(Q2)においては自由記述で回答してもらった。

(Q1) The character changes her/his facial expression naturally in the video above.

(上の動画において、キャラクターは表情を自然に変化させていた。)

(Q2) When the character's facial expression is changing, if there are any parts of the character's face that you think is unnatural, please enter the parts' names. If you think all of the face parts are natural, please enter "nothing".

(キャラクターが表情を変化させているとき、不自然な顔のパーツがあったら教えてください。もしなければ"nothing"と入力してください。)

アンケート(Q1),(Q2)を回答してもらった後、図5のようにVA顔表情変化動画の横に元動画を配置した動画を見せ、以下のアンケート(Q3),(Q4)を回答してもらった。元動画の大きさは元動画内の人間の顔の大きさとVAの顔の大きさが等しくなる程度の大きさにした。また、アンケート(Q3)においてはアンケート文にどれくらい賛成しているか、賛成していないかということを7段階形式で回答してもらい、(Q4)においては自由記述で回答してもらった。

(Q3) The character expresses the feelings of the woman on the left in the video above.

(上の動画において、キャラクターは左の女性の感情を表現していた。)

(Q4) Why did you answer (Q3) in that way? Please describe the reason briefly.

(なぜ(Q3)をそのように答えたのですか?理由を簡

潔に教えてください。)

実験では女性キャラクターを用いて作成した VA 顔表情変化動画の評価を行うセクションと、男性キャラクターを用いて作成した VA 顔表情変化動画の評価を行うセクションに分けた。そのため、1つの元動画を見せたのち、その元動画から作成した VA 顔表情変化動画（感情の方法、感情の値の方法、5段階の方法）を見せ、アンケートに回答してもらおうということを各 VA のセクションにおいて2つの元動画の分、行った。各セクションの順番と3つの VA 顔表情変化動画の順番はランダムに表示するようにした。



図 5: 実験で使用した VA 顔表情変化動画

6.3 実験結果

表 3: VA 顔表情変化動画の Q1 と Q3 のアンケート結果（女性キャラクター）

質問	元動画 1			元動画 2		
	感情	感情の値	5段階	感情	感情の値	5段階
Q1	5.05 ± 1.78	5.03 ± 1.66	5.07 ± 1.64	4.86 ± 1.78	5.19 ± 1.59	5.09 ± 1.64
Q3	5.13 ± 1.72	5.08 ± 1.63	4.87 ± 1.84	4.97 ± 1.87	4.60 ± 1.97	4.40 ± 2.10

女性キャラクターの VA 顔表情変化動画の Q1 と Q3 のアンケート結果は表 3 のようになった。感情の方法、感情の値の方法、5段階の方法を用いて作成した VA 顔表情変化動画の評価の平均に差があるのかを調べるために繰り返しなしの二元配置分散分析を行ったところ、元動画 1 の Q1 ($p = .970$) と元動画 2 の Q1 ($p = .100$)、元動画 1 の Q3 ($p = .169$) は有意差がみられなかったが、元動画 2 の Q3 ($p = .000$) は有意差がみられた。

元動画 2 の Q3 において Bonferroni 法による多重比較検定を行った結果、感情の値の方法-5段階の方法間 ($p = .589$) には有意差がみられなかったが、感情の方法-感情の値の方法間 ($p = .009$) と感情の方法-5段階の方法間 ($p = .001$) には有意差がみられた。感情の方法-感情の値の方法間、感情の方法-5段階の方法間ともに、感情の方法の方が有意に大きい評価を得られた。

Q2 の VA の顔の不自然なパーツを聞いた質問では、元動画 1 において”nothing”と答えた数が一番多かったのは感情の値の方法 (57人) であり、元動画 2 において

も感情の値の方法 (60人) であった。また、”nothing”以外では、目や口と答える人が多かった。元動画 1 においては、すべての方法において目と答えた人が一番多かった (感情: 24人, 感情の値: 15人, 5段階: 18人)。元動画 2 では、感情の方法では眉と答えた人が一番多く (20人)、感情の値の方法と5段階の方法では口と答えた人が一番多かった (感情の値: 17人, 5段階: 20人)。

Q4 の Q3 の回答理由を聞いた質問では、元動画 1, 2 のすべての方法における肯定的な意見は「happy や sad の顔表情が全体的に一致していたから」というような意見が多かった。元動画 1, 2 の各方法における否定的な意見は、感情の方法では「表情変化が速かったから」「happy の顔でないときに happy の顔になっていたから」「感情表現が強すぎるように見えたから」というような意見が多く、感情の値の方法では「sad や happy の表現が (人間の) 女性よりも弱く見えたから」という意見が多く、5段階の方法では「sad や happy の表現が (人間の) 女性よりも弱く見えたから」「顔の動きが遅かったから」という意見が多かった。また、元動画 1, 2 のすべての方法において「目と口の動きが一致していないから」「頭や眉の動きが不十分であったから」という否定的な意見が多くあった。

表 4: VA 顔表情変化動画の Q1 と Q3 のアンケート結果（男性キャラクター）

質問	元動画 1			元動画 2		
	感情	感情の値	5段階	感情	感情の値	5段階
Q1	5.05 ± 1.80	5.02 ± 1.64	5.18 ± 1.60	4.70 ± 1.86	5.16 ± 1.68	5.07 ± 1.58
Q3	4.87 ± 1.85	5.35 ± 1.62	5.14 ± 1.66	5.11 ± 1.78	4.63 ± 1.97	4.51 ± 2.02

男性キャラクターの VA 顔表情変化動画の Q1 と Q3 のアンケート結果は表 2 のようになった。感情の方法、感情の値の方法、5段階の方法を用いて作成した VA 顔表情変化動画の評価の平均に差があるのかを調べるために繰り返しなしの二元配置分散分析を行ったところ、元動画 1 の Q1 ($p = .551$) は有意差がみられなかったが、元動画 2 の Q1 ($p = .020$) と元動画 1 の Q3 ($p = .012$)、元動画 2 の Q3 ($p = .001$) は有意差がみられた。

元動画 2 の Q1 と元動画 2 の Q3、元動画 1 の Q3 においてそれぞれ Bonferroni 法による多重比較検定を行った。元動画 2 の Q1 においては、感情の方法-5段階の方法間 ($p = .145$) と感情の値の方法-5段階の方法間 ($p = 1.597$) には有意差がみられなかったが、感情の方法-感情の値の方法間 ($p = .005$) には有意差がみられた。元動画 1 の Q3 においては、感情の方法-5段階の方法間 ($p = .401$) と感情の値の方法-5段階の方法間 ($p = .313$) には有意差がみられなかったが、感情の方法-感情の値の方法間 ($p = .018$) には有意差がみ

られた。元動画2のQ3においては、感情の値の方法-5段階の方法間 ($p = 1.357$) には有意差がみられなかったが、感情の方法-感情の値の方法間 ($p = .019$) と感情の方法-5段階の方法間 ($p = .004$) には有意差がみられた。元動画2のQ1の感情の方法-感情の値の方法間においては、感情の値の方法の方が有意に大きい評価を得ることができ、元動画1のQ3の感情の方法-感情の値の方法間と、元動画2のQ3の感情の方法-感情の値の方法間および感情の方法-5段階の方法間においては、感情の方法の方が有意に大きい評価を得ることができた。

Q2のVAの顔の不自然なパーツを聞いた質問では、元動画1において”nothing”と答えた数が一番多かったのは5段階の方法(51人)であり、元動画2においては感情の値の方法(57人)であった。また、”nothing”以外では、眉や口と答える人が多かった。元動画1においては、すべての方法において口と答えた人が一番多かった(感情: 18人, 感情の値: 19人, 5段階: 17人)。元動画2においても、すべての方法において口と答えた人が一番多かった(感情: 22人, 感情の値: 21人, 5段階: 17人)。

Q4のQ3の回答理由を聞いた質問では、元動画1, 2のすべての方法における肯定的な意見は「顔表情がhappyを主として全体的に一致していたから」というような意見が多かった。元動画1, 2の各方法における否定的な意見は、感情の方法では「表情変化が速かったから」「顔の動きが激しく機械的に見えたから」「感情表現が強すぎるように見えたから」というような意見が多く、感情の値の方法では「sadやhappyの表現が(人間の)女性よりも少し弱く見えたから」という意見が多く、5段階の方法では「sadやhappyなどの表現が(人間の)女性よりも弱く、落ち着いて見えたから」「顔の動きが遅かったから」という意見が多かった。また、元動画1, 2のすべての方法において「目と口の動きが一致していないから」「頭や眉の動きが不十分であったから」という否定的な意見が多くあった。

6.4 考察

Q1の質問において、元動画2の男性キャラクターを用いて作成したVA顔表情変化動画以外は各方法間で有意差が見られなかった。しかし、各手法間によって作成したVA顔表情変化動画におけるVAの顔表情変化の動きは大きく異なっていたため、Q1の質問において有意差がみられなかったのはQ2やQ4の意見として多くあった目や口の動きが原因の1つとして考えられる。目や口の動きの不自然さによって全体的に等しく悪い印象を与えていた可能性がある。6節の実験ではVAの目と口は実際の人間の口の開き具合を考慮し

ているのではなく、口が閉じているか開いているかのみを反映させていたため、人間の口と目の開閉具合を利用してVAの口と目のパラメータ調整を行うことで評価が変化すると考える。

Q3の質問において、元動画1の女性キャラクターを用いて作成したVA顔表情変化動画以外は感情の方法の評価が有意に大きくなっていった。Q4の質問において感情の値の方法と5段階の方法で共通して「感情表現が弱い」というような否定的な意見が多かったことから、元動画から適切な感情の値が取り出せていなかった可能性がある。元動画から適切な感情を取り出せているかどうか調べるために、元動画の人間が表している感情の値がPaltoglouら[12]のラッセルの円環モデル上においてどのような感情に該当しているかを具体的に調査し、検証する必要があると考える。

7 おわりに

7.1 本研究のまとめ

本研究では、特定の感情にとどまらない幅広いVAの顔表情を作成できるツールを作成することを目的とした。入力に汎用性の高い感情の値を利用することで離散的でないさまざまな感情を表現できるようにするとともに、先行研究では利用されていない黒目の大きさや頬の色といったVA特有のパーツを利用した感情表現を強調したVAの顔を作成するツールを実現できた。

また、もう1つの目的であるVA顔表情作成ツールの応用性を示すためのVA顔表情変化動画の作成においては、提案する方法で作成したVA顔表情変化動画とナイーブな方法で作成したVA顔表情変化動画よりも高い評価を得ることができなかった。全体的に評価を下げていた目や口の動きの反映方法や元動画から推定した感情の値の誤認識などが原因である可能性があり、目や口の開閉度をパラメータ調整に利用するといったことや、元動画から読み取った感情の値の検証が必要であると考えられる。

7.2 今後の展望

本研究では、VA顔表情作成ツールを作成するとともに、VA顔表情作成ツールを利用したVA顔表情変化動画生成を行ったが、VA顔表情変化動画への応用ができていなかったため、原因と考えられる目や口の開閉度を利用したパラメータ調整方法を検討する必要がある。また、リアルタイムの表情変化動画をVA表情変化動画に反映させることも今後の展望として考えているため、動きの連動性を上げるために姿勢推定による頭の動きの反映なども実装したいと考えている。

VA 顔表情作成ツールでは感情の値を入力とすることで、幅広い感情の VA 顔表情を作成できるようにした。感情の値は研究分野において様々なことに利用されているが、人間自身が自身の表したい感情を感情の値に直感的に変換することができないため、本研究で提案した VA 顔表情作成ツールを人間が直接扱うことは困難である。人間が直感的に VA 顔表情作成ツールを利用できるようになることで、自身の代わりとするアバターの顔を状況に適した顔に簡単に換えられるようになり、コミュニケーションの際の負担を減らすことが期待できる。そのため、人間自身が表現したい感情を感情の値に変換できるようにする手法を検討することも今後の課題である。

謝辞

本研究は、JST CREST (JPMJCR21D4) の支援を受けたものである。

参考文献

- [1] Russell, J. A.: A Circumplex Model of Affect, *Journal of Personality and Social Psychology*, 39(6), pp. 1161-1178, 1980
- [2] Wei-Yi Chang, Shih-Huan Hsu, Jen-Hsien Chien: FATAUVA-Net : An Integrated Deep Learning Framework for Facial Attribute Recognition, Action Unit Detection, and Valence-Arousal Estimation, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 17-25, 2017
- [3] Seliz Gulsen Karadogan, Jan Larsen: Combining Semantic and Acoustic Features for Valence and Arousal Recognition in Speech, *Proceedings of the 3rd International Workshop on Cognitive Information Processing (CIP)*, 2012
- [4] Jin Wang, Liang-Chin Yu, K.Robert Lai, Xuejie Zhang: Dimensional Sentiment Analysis Using a Regional CNN-LSTM Model, *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pp. 225-230, 2016
- [5] Stephane Gobron, Junghyun Ahn, Georgios Paltoglou, Michael Thelwall, Daniel Thalmann: From sentence to emotion: a real-time three-dimensional graphics metaphor of emotions extracted from text, *Journal of The Visual Computer*, 26, pp. 505-519, 2010
- [6] Junghyun Ahn, Stephane Gobron, Quentin Silvestre, Daniel Thalmann: Asymmetrical Facial Expressions based on an Advanced Interpretation of Two-dimensional Russell's Emotional Model, *Proceedings of ENGAGE 2010*, 2010
- [7] Guillermo Recio, Annekathrin Schacht, Werner Sommer: Classification of dynamic facial expressions of emotion presented briefly, *Journal of Cognition and Emotion*, 27(8), pp. 1486-1494, 2013
- [8] 大島康, 森大毅, 中村真: 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, *HAI シンポジウム 2008*, 2008. (online), <https://hai-conference.net/proceedings/HAI2008/pdf/1e-3.pdf>
- [9] 中谷将大, 佐伯幸郎, 中村匡秀, 安田清: 認知症者のための音声対話システムにおける個人向けヴァーチャルエージェントの生成, *信学技報*, 117(464), pp. 93-98, 2018
- [10] Hiroshi Yamada: Visual Information for Categorizing Facial Expression of Emotions, *Journal of Applied Cognitive Psychology*, 7(3), pp. 257-270, 1993
- [11] Nim Tottenham, James W. Tanaka, Andrew C. Leon, Thomas McCarry, Marcella Nurse, Todd A. Hare, David J. Marcus, Alissa Westerlund, BJ Casey, Charles Nelson: The NimStim set of facial expressions: Judgments from untrained research participants, *Journal of Psychiatry Res.*, 168(3), pp. 242-249, 2009
- [12] Georgios Paltoglou, Mike Thelwall: Seeing Stars of Valence and Arousal in Blog Posts, *Journal of IEEE Transactions on Affective Computing*, 4(1), pp. 116-123, 2013
- [13] Mark Davies. (2021). *Word frequency: based on one billion word COCA corpus*. Retrieved from <https://www.wordfrequency.info>
- [14] Antoine Toisoul, Jean Kossaiifi, Adrian Bulat, Georgios Tzimiropoulos, Maja Pantic: Estimation of continuous valence and arousal levels from

faces in naturalistic conditions, *Journal of Nature Machine Intelligence*, 3, pp. 42-50, 2021

- [15] Alessio Bellino: SEQUENCE: a Remote Control Technique to Select Objects by Matching their Rhythm, *Journal of Personal and Ubiquitous Computing*, 22(4), pp. 751-770, 2018