

自動表情認識を用いた遠隔感情共有システムの開発

平島 健大* 大久保 雅史*

Development of Emotion Sharing System Using Automatic Facial Expression Recognition

Kenta Hirashima* and Masashi Okubo*

Abstract - In recent years, information and emotions are shared in a comment form among the remote viewers in live streaming video sharing service. However, it takes time for a person to input comments, and temporal dissociation occurs between video and comments. On the other hand, "facial expression" is an important factor as nonverbal information expressing emotion, and recently a technology to recognize it automatically is being developed. In this research, we aim to improve the sense of unity among video viewers and the sense of pleasure on video by sharing their feelings among the remote viewers through avatars with expression in real time. We have developed the emotion sharing system in which the facial expression of avatar changes based on the viewers' facial expression automatically.

Keywords: facial expression, automatic recognition, remote communication, entertainment and non-verbal information

1. はじめに

近年、インターネット環境の発展に伴い、「ニコニコ生放送」や「Ustream」などのライブストリーミング動画共有サービスが普及している。これらの特徴として、遠隔地のユーザと同じ動画をリアルタイムで視聴し、動画に対する感想をコメント入力によって共有できる点が挙げられる。しかし、ユーザ同士で感想や感情を共有するためには、動画を視聴しながら、コメントを読み、書き込むといった作業を同時に行う必要がある。また、コメントというバーバル情報での共有方法が大部分を占めており、かつ、ユーザの能動的な操作に依存している。これらのことから、動画を視聴しながら遠隔地のユーザと動画に対する感情を十分に共有しているとは言い難い。

一方、感情を他者に伝達する手法としてノンバーバル情報の利用がある。多くの専門家が対面コミュニケーションにおいてノンバーバル情報が大きな役割を担っていると説いている^[1]。しかし、遠隔コミュニケーションではノンバーバル情報が十分に伝達されないため、コミュニケーションに支障をきたす可能性がある^[2]。

そこで、本研究では遠隔地の視聴者との動画視聴において、お互いのノンバーバル情報を共有することで、より感情の共有を可能とするシステムの構築を目指す。感情を表現するノンバーバル情報として代表的なものには「表情」がある。対面コミュニケーションにおいて表情が重要な役割を果たすように、遠隔コミュニケーションにおいても表情は円滑なコミュニケーションに寄与する可能性がある^[3]。Ekman は 6 つの基本感情（喜び、怒り、嫌悪、恐れ、悲しみ、驚き）を表す人の表情は文化や環

境に依存せず、人類に普遍的な特徴であると述べている^[3]。この提唱に基づいて、市川らは、他者の表情に対して、観察者の同調的反応の有無を検証している^[4]。検証の結果、6 つの基本感情のうち、喜びと驚きを表す表情に対して、観察者の同調的反応が有意に生じることが明らかになっている。さらに、この結果に着目し、大森らは、友人 1 人と隣り合って笑いを誘発する動画を視聴する他者有り条件と、1 人で視聴する他者無し条件で実験を行い、他者の存在が、動画に対する主観的面白さと、笑いの表情の表出に及ぼす影響について検証している^[5]。検証の結果、他者有り条件のほうが動画は面白いと評定され、笑いの表情の表出も多かったことが明らかにされている。しかし、この結果は友人との対面環境における結果であり、遠隔環境においては検証されていない。

本研究では、動画視聴時の視聴者の笑いの表情に着目し、遠隔地の 2 人の視聴者の感情をリアルタイムに共有させることで、動画視聴者同士の一体感の向上や動画に対する楽しさを助長することを目的としている。具体的には、各視聴者の笑いの表情を、アバタを介してリアルタイムで共有できるシステムを提案する。さらに、提案システムにおけるアバタの表情を、自動表情認識を用いて切り替える方法と手動で切り替える方法を比較し、自動的に表情を切り替えて感情を共有することが、視聴者の笑いや視聴動画に対する評価に及ぼす影響について検証している。

2. 遠隔感情共有システム

2.1 提案システムの概要

本研究では、動画視聴時に 2 人の視聴者の笑いの表情をリアルタイムで共有できるシステムを提案する。図 1

*: 同志社大学大学院 理工学研究科 情報工学専攻
*: Graduate School of Science and Engineering, Doshisha University

にシステムの概要を示す。遠隔環境における 2 人の視聴者の PC の画面上にはそれぞれ自身の表情アバタと他視聴者の表情アバタを表示する。動画を視聴しながら表情アバタを変化させることで、2 人の視聴者で表情を共有することができる。

また、表情の伝達手段として使用するアバタは自動または手動で切り替えることができる。

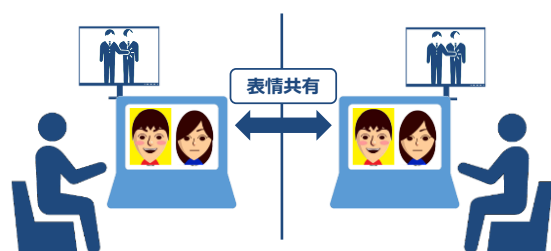


図1 提案システムの概要

Fig.1 System configuration

2.2 表情アバタ

図2に提案システムで用いる表情アバタを示す。アバタの表情はEkmanのFACS(Facial Action Coding System)^[3]を参考に男性と女性でそれぞれ「無表情」と「笑いの表情」の2パターンを用意する。また、視聴者が表情アバタに注意を向けすぎると、動画の内容に集中できなくなる可能性がある。そこで、笑いの表情ではアバタの背景を黄色にすることによって、視聴者が表情アバタに注目せずとも表情の変化に気づきやすくしている。

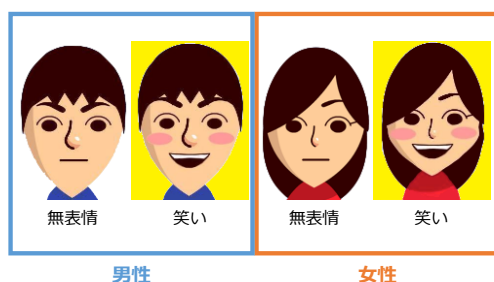


図2 表情アバタ

Fig.2 Avatar with facial expression

2.3 表情切替方法

提案システムには、視聴者が笑いの表情を共有する方法として、「自動表情変化モード」と「手動表情変化モード」の2種類のモードを搭載する。

自動表情変化モードでは、視聴者の表情を自動的に推定するために、affectiva社の感情認識技術Affdexを使用する。Affdexは、PCのWebカメラを使用して撮影したストリーミング動画から人の顔を検出し、感情をリアルタイムで推定する。Affdexを用いて推定できる感情は、喜び、悲しみ、怒り、軽蔑、嫌悪、恐怖、驚きの7種類

であり、各感情の推定結果が0~100の数値で表される^[6]。提案システムでは、喜びの感情のみ自動推定を行い、推定結果が70以上であれば「笑い」、70未満であれば「無表情」とする。推定結果のサンプリング周期は0.1秒であり、15分間の動画視聴でサンプリング数は9000個である。図3に自動表情認識による表情アバタの切り替えを示す。AffdexのSDKを組み込み、Webカメラから取得した視聴者の表情を自動推定し、アバタの表情を自動的に切り替える。

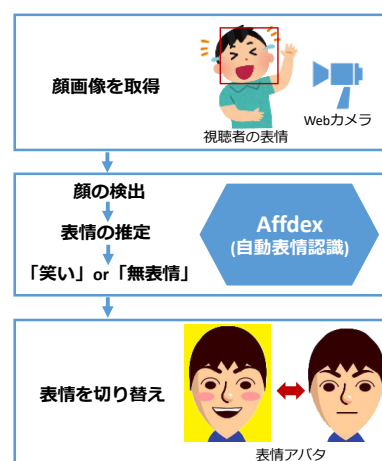


図3 自動表情認識による表情アバタの切り替え

Fig.3 Switching the avatar's expression using automatic facial expression recognition

手動表情変化モードでは、視聴者が面白いと感じたタイミングで自身のアバタをクリックさせる。クリック操作により、提案システムはアバタを笑いの表情に切り替え、1秒後に無表情に切り替える。また、手動表情変化モードでもバックグラウンドでAffdexによる表情の推定を行い、視聴者の表情データとして記録している。

2.4 表情データの共有

図4に表情データの送受信の流れを示す。提案システムにおいて、クライアントはサーバから発行されたユーザIDを保持し、リアルタイムでアバタの表情データを更新する。表情データは自動表情認識による推定値やクリック操作状況から、笑いであるか無表情であるかを0か1の数値で表す。

サーバは各ユーザID、登録された各ユーザの名前や性別、表情データ、アバタが切り替えられる時刻を保持する。また、クライアントとサーバの間で以下のように情報をやりとりする。

- ① クライアントはアバタの表情データが変化したタイミングで表情データをサーバに送信する。
- ② サーバはクライアントから表情データを受信した時刻を記録する。

- ③ サーバは各クライアントへアバタの切り替え命令を送信する。

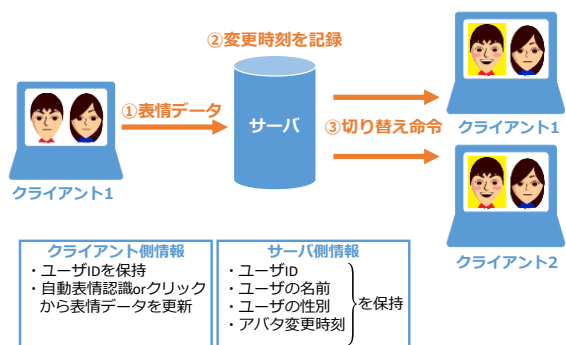


図4 表情データの送受信の流れ
Fig.4 Flow of data of facial expression

3. 評価実験

3.1 実験目的

本実験では、提案システムのアバタの表情を自動で切り替える方法と手動で切り替える方法を用いて実験協力者に笑いを誘発する動画を視聴させる。2つの方法の比較により、自動的に表情を切り替えて感情を共有することが、実験協力者の笑いや動画に対する評価に及ぼす影響を検証することを目的とする。

3.2 実験方法

実験は2人1組で行う。同志社大学、同志社女子大学の学生、男性同士(9組)、女性同士(5組)、男女ペア(2組)の、合計16組32名を対象に実施する。また、実験協力者はすべて友人同士である。

図5に実験風景を示す。各実験協力者には「視聴者」として遠隔環境を再現した2つの個室で、提案システムを使用しながら、15分間の笑いを誘発する動画を同時に視聴させる。動画はユーザの前面の壁にプロジェクタで投影し、手前のPCの画面に2人のアバタを表示する。



図5 実験風景
Fig.5 Experimental scene

図6に実験手順を示す。実験は順序効果を考慮し、動画1、動画2の2種類の動画と、自動表情変化モードと手動表情変化モードの2種類のモードを組み合わせ、4パターンの組み合わせで実施する。

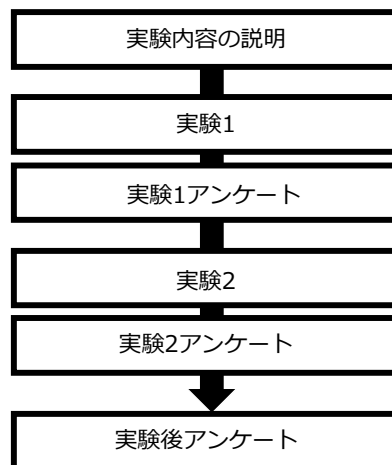


図6 実験手順
Fig.6 Experimental process

実験前の説明において、提案システムの自動表情変化モードと手動表情変化モードのどちらを使用する場合においても、Webカメラによって実験協力者の表情を撮影し、表情データを記録することを伝える。

定量的評価指標として、各視聴者の表情の推定値の時系列データと表示されるアバタの表情変化の時系列データを記録する。

また、定性的評価指標として、1つの動画を視聴し終えるごとにアンケートに回答させる。視聴者の動画に対する集中度、アバタの表情変化のタイミング、他視聴者との笑いの共有感などについて、「1:全くあてはまらない」「2:あてはまらない」「3:あまりあてはまらない」「4:どちらでもない」「5:ややあてはまる」「6:あてはまる」「7:とてもあてはまる」の7段階評価で回答させる。

また、本実験では、視聴者の笑いを誘発させるための題材として、プロのお笑い芸人の漫才やコン트의動画を使用する。視聴者の好みやお笑い芸人の能力による「面白さ」の偏りを無くすために、動画1、動画2ともに同じ3組のお笑い芸人のパフォーマンスを採用する。1組の芸人につきパフォーマンスを2つずつ選定し、1つの動画が合計15分となるように動画1と動画2を構成する。

4. 実験結果と考察

表1にアンケート内容、図7にアンケート結果を示す。なお、アンケート結果の解析にはWilcoxonの符号順位検定を用いる。アンケート結果より、Q1~Q7のすべての質問に対して自動表情変化モードを使用して動画を視聴し

た方が高く評価された。とくに、Q1の「動画の内容に集中できたか」や、Q5の「他視聴者の感情が伝わってきたか」、Q7「笑いが共有されていると感じたか」については有意水準 1%で自動表情変化モードの方が高く評価された。提案システムの自動表情変化モードは、動画視聴中の操作が一切必要なく、自動表情認識による表情の推定結果がリアルタイムで共有される。このことが動画に対する集中度や、他視聴者との感情の伝達及び笑いの共有感を高めた要因であると考えられる。

表 1 アンケート内容

Table 1 Contents of the questionnaire

番号	アンケート内容
Q1	動画の内容に集中できたか
Q2	他視聴者に感情や表情をタイミングよく伝えることができたと感じたか
Q3	他視聴者から感情や表情をタイミングよく受け取れると感じたか
Q4	動画視聴中楽しいと感じたか
Q5	他視聴者の感情が伝わってきたか
Q6	他視聴者との一体感を感じたか
Q7	他視聴者と笑いが共有されていると感じたか

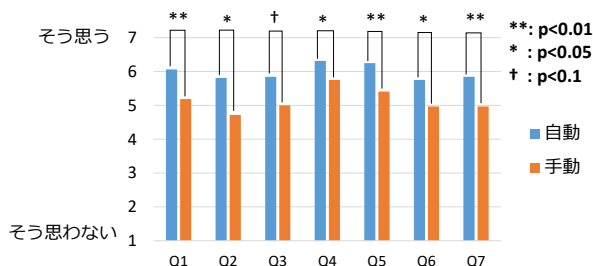


図 7 アンケート結果

Fig.7 Result of questionnaires

本実験では、定量的評価指標として 3 種類の時系列データを用いて比較する。まず、自動表情変化モードでは、自動表情認識による 2 人のアバタの表情変化の時系列データ（以下、「自動__表情認識」）を得る。また、手動表情変化モードでは、クリック操作による 2 人のアバタの表情変化の時系列データ（以下、「手動__クリック」）と、そのバックグラウンドでの自動表情認識による 2 人の視聴者の表情変化の時系列データ（以下、「手動__表情認識」）を得る。なお、検定には t 検定を用いる。

図 8、図 9 に表情変化モード別の自動表情認識による視聴者の表情の笑いの回数と累計時間の平均を示す。動画 15 分間で、視聴者の表情の笑いの回数は自動表情変化モードでは 63.3 回、手動表情変化モードでは 59.6 回であり、有意な差はなかった。一方、笑いの累計時間を比較すると動画 15 分間で、自動表情変化モードでは 242.7 秒、手動表情変化モードでは 288.2 秒であり、統計的な有意差は見られなかったが、手動表情変化モードのほうが笑いの時間が長い傾向にあった。

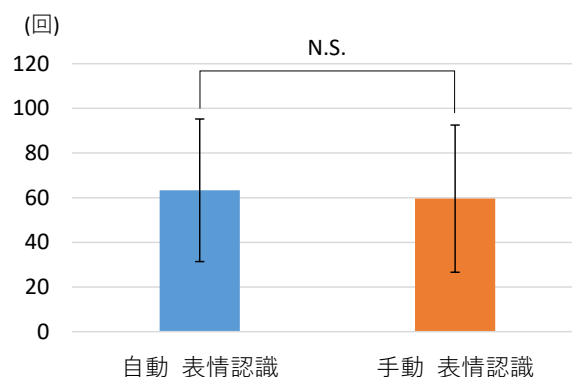


図 8 視聴者の笑いの回数の平均

Fig.8 Average of number of viewers' laughing

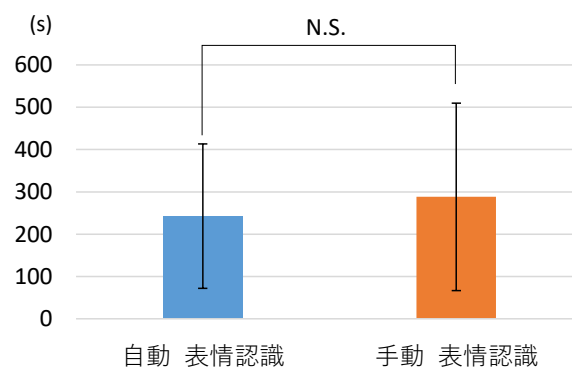


図 9 視聴者の笑いの累計時間の平均

Fig.9 Average of viewers' laughing time

図 10 に表情変化モード別の自動表情認識による視聴者の 1 回あたりの笑いの時間の平均を示す。この平均値は、全実験協力者 32 人の笑いの回数と笑いの時間のデータから、1 人分の 1 回あたりの笑いの時間を求め、それらを 32 人で平均したものである。1 回あたりの笑いの時間は、自動表情変化モードでは 3.65 秒、手動表情変化モードでは 4.73 秒であり、手動表情変化モードのほうが長く、統計的に有意傾向が見られた。実験中、自動と手動のどちらの場合においても、視聴者は、面白いと感じたポイントからしばらく時間が経過しても笑いの表情のままであることが多く見られた。自動表情変化モードでは、視聴者の表情がリアルタイムに表情アバタに反映されるため、視聴者が面白いと感じたポイントから時間が経過しても自身のアバタが笑いの表情であることに気づき、自身の表情を意識的に無表情に抑制した可能性がある。

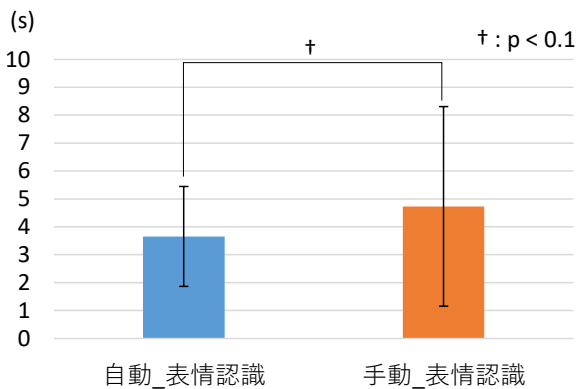


図 10 1 回あたりの笑いの時間の平均
Fig.10 Average of time per laughing

つぎに、動画視聴中の視聴者の笑いの同調現象について検証する。図 11 に 2 人の表情の時系列データの相関値の求め方を示す。視聴者の笑いの同調現象を定量的に評価するために、15 分間の動画視聴における表情変化の時系列データから 1 分ごと（データ数：600）に 2 人のアバタの表情変化の相関値を式(1)に基づいて求める。相関値が高いことは 2 人の表情変化の推移の類似度が高いことを示す。

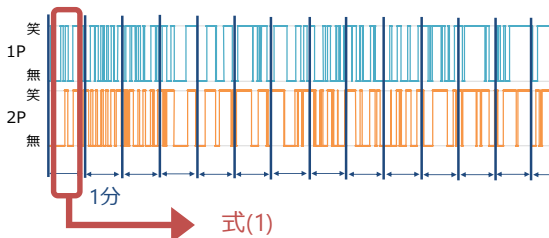


図 11 表情の時系列データからの相関値の求め方
Fig.11 How to calculate correlation value from time series data of facial expression

$$r = \frac{\sum_{t=1}^{600} (x(t) - \bar{x})(y(t) - \bar{y})}{\sqrt{\sum_{t=1}^{600} (x(t) - \bar{x})^2 \sum_{t=1}^{600} (y(t) - \bar{y})^2}} \quad (1)$$

\bar{x}, \bar{y} は $x(t), y(t)$ の平均

図 12 に手動表情変化モードにおける、自動表情認識による 2 人の視聴者の表情変化の相関値の平均とクリック操作による 2 人のアバタの表情変化の相関値の平均を示す。視聴者の表情のほうが高い相関値が得られ、統計的に有意な差が見られた。すなわち、視聴者が手動でアバタの表情を笑いに変化させる場合において、2 人のアバタの笑いの相関値よりも、2 人の視聴者の笑いの相関値

の方が高い結果となった。この結果から、視聴者のクリック操作よりも、自動表情認識によって視聴者の表情をアバタに反映させる方法の方が、視聴者の笑いの同調現象が起こる可能性が高いことが明らかになった。

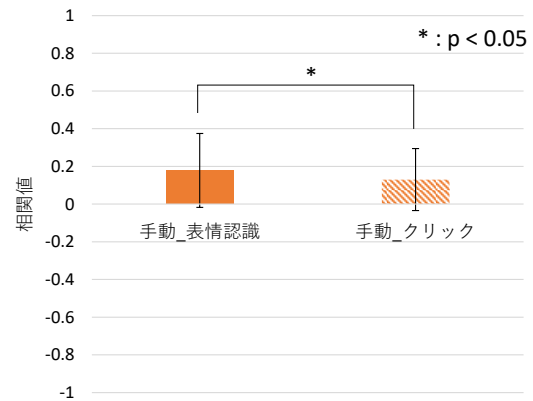


図 12 手動表情変化モードにおける視聴者の表情とアバタの表情の相関値の平均
Fig.12 Average of correlation between value viewers' facial expression changes and between avatars' expression changes in the manual expression change mode

図 13 に自動表情変化モードでの自動表情認識による 2 人のアバタの表情変化（すなわち、2 人の視聴者の表情変化）の相関値の平均と、手動表情変化モードでの自動表情認識による 2 人の視聴者の表情変化の相関値の平均を示す。自動表情変化モードのほうが高い相関値が得られ、統計的に有意な差が見られた。アバタの表情を変化させるためのクリック操作が 2 人の視聴者の笑いの同調現象に影響している可能性が考えられる。

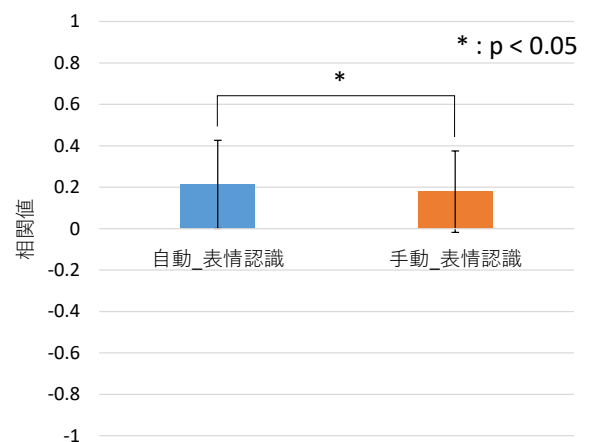


図 13 表情変化モード別での 2 人の視聴者の表情変化の相関値の平均
Fig.13 Average of correlation value between viewers' facial expression changes by expression change modes

図 12, 13 の結果を踏まえ, 図 14 に, 自動表情変化モードでの自動表情認識による 2 人のアバタの表情の相関値の平均と, 手動表情変化モードでのクリック操作による 2 人のアバタの表情の相関値の平均を示す. 自動表情変化モードのほうが高い相関値が得られ, 統計的に有意な差が見られた. このことから, 自動表情変化モードを用いて動画を視聴する方が笑いの同調現象が多く見られることが示された. この結果が, アンケート結果 Q4 の「動画視聴中楽しいと感じたか」や, Q6 の「他視聴者との一体感を感じたか」において高く評価された要因であることが考えられる.

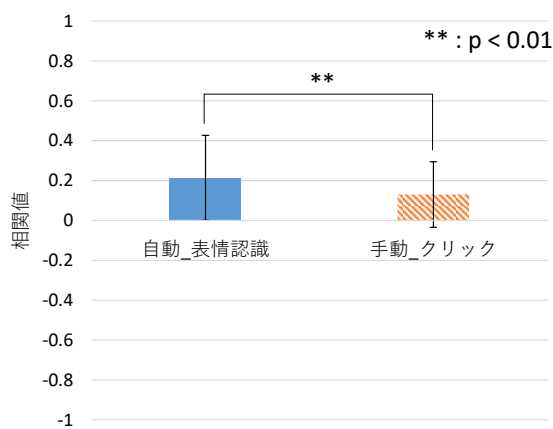


図 14 表情変化モード別の 2 人のアバタの表情変化の相関値の平均

Fig.14 Average of correlation value between avatars' facial expression changes by expression change modes

5. おわりに

本研究では, 動画視聴時の視聴者の笑いの表情に着目し, 遠隔地の 2 人の視聴者の笑いをリアルタイムで共有させるシステムを提案した. また, 提案システムにおけるアバタの表情を自動で切り替える方法と手動で切り替える方法を比較し, 自動的に表情を切り替えて感情を共有することが, 視聴者の笑いや視聴動画に対する評価に及ぼす影響について, 評価実験を行い検証した.

実験の結果より, 自動表情変化モードを用いた動画視聴の方が手動表情変化モードを用いて視聴するよりも, 2 人の視聴者のアバタの表情変化の相関値が高くなり, 笑いの同調現象が多く見られた. この結果が, アンケートの結果から得られた, 2 人の視聴者の一体感の向上や, 動画視聴中の楽しさを助長する要因であると考えられる.

動画に対する集中度や, 他視聴者との感情の伝達についても自動表情認識を用いた笑いの共有方法のほうが高く評価された. 提案システムの自動表情変化モードは, 動画視聴中の操作が一切必要なく, 自動表情認識による表情の推定結果がリアルタイムで共有されることが要因であると考えられる.

一方, 笑いの回数に差は見られなかったが, 自動表情変化モードを用いた場合, 視聴者の笑いの時間の減少が見られた. これは自動表情認識によるアバタの表情が視聴者に笑いを意識的に抑制させた可能性がある. これは, 遠隔地の 2 人の視聴者の感情共有方法において, 自身の表情アバタの有無が視聴者の表情や行動に影響を及ぼす可能性が考えられ, 今後検証する予定である.

また, 視聴者間ではさまざまな感情の共有の可能性が挙げられる. 本研究では, 喜びの感情を表す笑いの表情のみに着目して検証した. 表情の同調的反応は, 喜び以外の感情を表す表情に対しても見られ, とくに驚きの感情を表す表情に対して多く見られることが明らかになっている^[4]. このことから, 驚きの表情やその他の表情に対しても, 自動表情認識を用いて動画視聴者の感情を共有する方法が有用である可能性があり, システムの改良を行う. また, 本研究では感情を共有する対象を友人と一対一と設定したが, 不特定多数の視聴者に対しても, 自動表情認識を用いて動画の視聴者同士の感情を共有する方法が有用である可能性があり, 今後はこれについても検証する.

参考文献

- [1] 楠見孝, 米田英嗣, 小島隆次: 3 次元マルチユーザ仮想空間におけるコミュニケーション -表情アバタによる感情表出と理解-; 情報処理学会第 65 回全国大会論文集, pp439-442 (2003).
- [2] 黒川隆夫: ノンバーバルインタフェース; オーム社 (1994).
- [3] P.Ekman, W.V.Friesen: 表情分析入門-表情に隠された意味を探る; 誠信書房 (1987)
- [4] 市川寛子, 牧野順四郎: 刺激表情に対する観察者の同調的表情; 心理学研究, 75 巻, 2 号, pp.142-147 (2004).
- [5] 大森慈子, 千秋紀子: 他者の存在が映像に対する面白さと笑い表情の表出に与える影響; 仁愛大学研究紀要 人間学部篇, 10 号, pp.25-31 (2011).
- [6] affectiva : 感情認識ソフトウェア Affdex; <http://www.affectiva.com/>