デジタルカメン：組込型光センサアレイを用いた近接表情認識機能をもつデジタルマスクの設計と実装

竹川 佳成[1]　徳田 雄嵩[3]　梅澤 章乃[1]　鈴木 克洋[2]　杉浦 裕太[2]　正井 克俊[2]　杉本 麻樹[2]
Diego Martinez Plasencia[4]　Sriram Subramanian[4]　平田 圭二[1]

概要：本論文では，実空間での対面コミュニケーションを支援するために，ユーザの顔の表情全体をアバタの表情に反映させることができる薄型のデジタルフルフェイスマスクディスプレイ「デジタルカメン」の設計と実装について述べる．カメラを用いた表情認識技術により，アバタによる顔の拡張が可能になったが，その応用は仮想空間での対面コミュニケーションに限られていた．本研究では，実空間でのアバタによる顔の拡張を可能にするために，軽量でフレキシブルなディスプレイと表情認識器を統合したデジタルカメンを提案する．顔全体に分散した 40 個の光センサから構成される光センサアレイを用いて表情を推定するため，カメラによる表情認識と異なり近接で表情認識できる．機械学習アルゴリズムの 1 つである SVM (Support Vector Machine) を用いた 10 クラスの表情認識モデルを構築する．提案する表情認識モデルの平均正答率は 79%であった．被験者にデジタルカメンを装着してもらい，鏡越しにデジタルカメン上に表示されるアバタを見てもらった．装着者自身の表情に同期して変化するアバタの表情変化は，アバタ再現性やアバタの表情変化の応答性という評価指標において，カメラベースの表情認識結果を表情付アバタとして生成および表示する機能をもつ従来手法（e2-Mask）と比較して同程度の結果が得られた．

1. 背景

顔はその人自身の印象を決める重要なパーツの 1 つである．初対面の人の第一印象を決定づける要因の 55%は，視覚情報が占めると心理学者のメラビアン [17] は述べている．また，Secord[28] は，人は相手の顔の特徴に基づいて性格を推測すると述べている．川西 [11] は，顔の特徴が好ましいと思う人と個人的親しみやすさとの間に，正の相関関係があることを示している．このように，顔の表情は対面コミュニケーションの重要な要素であるとともに，感情状態を表し，他者が認知する対話者のペルソナを暗黙のうちに形成する．

近年，仮想空間で人の顔の物理的な制約を超えたペルソナを創造および表現できるアバタが新たなオンラインコミュニケーション手段として積極的に活用されている．例えば，zoom などのオンラインビデオ会議システムにおいて，自身の顔をアバタに変えるユーザがいる．また，そのアバタは，カメラベースの表情認識技術を用いることで，話者の表情をリアルタイムに再現している．一方で，実空間でのアバタによるデジタル自己表現の機会を拡張するために，デジタルな顔拡張技術の研究も多数実施されている [1], [5], [8], [12], [24], [33]．例えば，Akaike らは，装着者が見ている対話者の顔に，アバタの静止画を重ね合わせられる AR ベースの顔面拡張ヘッドマウントディスプレイ（HMD）を提案している．しかし，このようなカメラを介したビデオシースルー型手法では会話に参加する全員が HMD を装着する必要があり，実空間での対面コミュニケーションへの利用は難しい．

そこで本研究では，実空間におけるアバタを用いた対面コミュニケーションの機会を拡張するため，近接表情認識機能をもつ薄型デジタルフルフェイスマスクディスプレイ「デジタルカメン」を提案する．デジタルカメンは，マスク装着者の表情をリアルタイムに認識し，その認識結果を反映させたアバタを，軽量薄型の曲面有機 EL ディスプレイの画面に表示する．また，顔全体に分散した 40 個の光センサから構成される光センサアレイを用いて表情を推定するため，カメラによる表情認識と異なり近接で表情認識できる．デジタルカメンは図 2 に示すように，従来のアナログな仮面では困難であった喜怒哀楽などの様々な表情を動的に表現・制御できる表情拡張型対面コミュニケーションを可能にする．例えば，筋萎縮性側索硬化症（ALS）患者

1 公立はこだて未来大学
2 慶應義塾大学
3 City University of Hong Kong
4 University College London

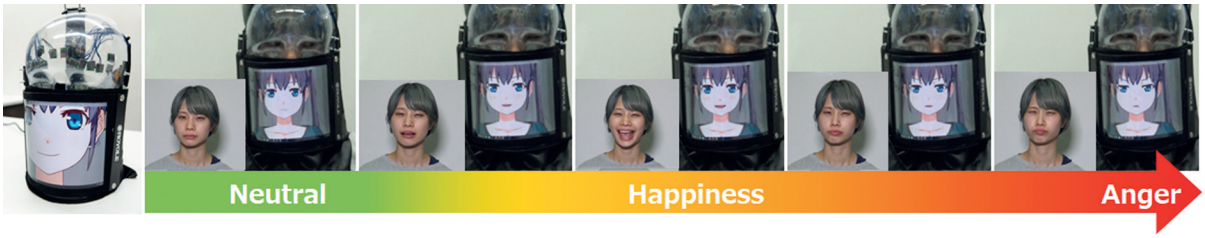

図 1 装着者の表情によって変化するデジタルカメン上のアバタアニメーション

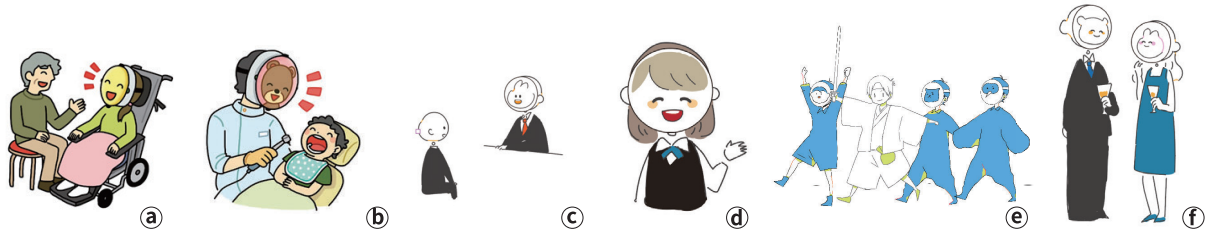

図 2 デジタルカメンの応用例

の多様な表情生成支援（図 2(a)），小児患者の歯科治療中の不安感やストレスの軽減（図 2(b)），面接官の表情制御による受験者の緊張緩和（図 2(c)）などが考えられる．また，最適なペルソナを柔軟に表現するメディアとして，接客業における感情労働の軽減（図 2(d)），演劇におけるパフォーマンスと表現力の自由度の向上（図 2(e)），オンラインゲームやフォーラムのユーザがオフラインのミーティングでもオンラインのデジタルペルソナを維持できるようにする（図 2(f)）などの応用が期待できる．

本論文では，顔面拡張のためのデジタルカメンの概念実証（Proof-of-concept）について議論する．デジタルカメンは，マスク表側にアバタ可視化用フレキシブルディスプレイを搭載し，マスク裏側に光センサアレイを内蔵する．機械学習アルゴリズムの 1 つである SVM (Support Vector Machine) を用いて表情認識モデルを構築した．また，複数の被験者による 10 種類の表情データセットを構築し，提案モデルの妥当性について議論する．さらに，装着者自身の表情に同期するアバタを見てもらうことで，アバタの再現性およびアバタの表情変化の応答性という観点でユーザスタディを実施する．最後に，現行のデジタルカメンの技術的な限界を議論し，円滑な対面コミュニケーションのための展望について議論する．

本研究の貢献は以下のとおりである．

- 実空間におけるデジタルな顔拡張を実現する表情認識機能付き薄型デジタルフルフェイスマスクディスプレイの概念実証
- 光センサアレイを用いた近接表情認識手法の検証
- 装着者自身の表情に同期するアバタに対する有用性の検証

2. 関連研究

2.1 顔の拡張

これまでに，顔の全体や一部を仮想的に変える事例がいくつかある．メディアアートとして，山田太郎プロジェクト [34] や，TABLETMAN[7] といったタブレットを顔に装着するといった作品がいくつか存在する．山田太郎プロジェクトとは，iPad を用いて街中で人の顔を撮影し，それを自分の顔に投影するという一時的かつ匿名性のある演出である．TABLETMAN は，株式会社東芝が宣伝のために作り出したキャラクタで，光るラインの入った特撮ヒーローのようなスーツに，いくつものタブレット PC を頭部や腕や腹部に搭載している．これらのように，自身の顔の代替や，商品などの宣伝として体にタブレットを貼り付けることは演出として用いられている．しかしながら，これらの事例は対面でのコミュニケーションに関しては考慮されておらず，本研究で提案する表情認識機能を持ち合わせていない．

テレプレゼンスシステムとして，タブレットを顔に装着するシステムがある．Skype などのビデオチャットや遠隔ユーザの顔を表示する ChameleonMask [19] がある．テレプレゼンスとは，遠隔地のメンバーとその場で対面しているかのような臨場感や存在感を提供する技術である．ChameleonMask では，テレプレゼンスにおいて遠隔ユーザの明確化や身体的存在感を出すために，遠隔ユーザの顔が表示されたディスプレイを代理人が着用する．これにより，代理人は遠隔ユーザへの成り代わりが可能となり，遠隔ユーザとその対話者の会話に親近感や臨場感をもたらす．Sakashita ら [27] は，ユーザの身体と顔の動きを人形に反映させるシステムを提案した．人形の目にカメラを搭載し，ユーザが装着している HMD にカメラの映像を提示

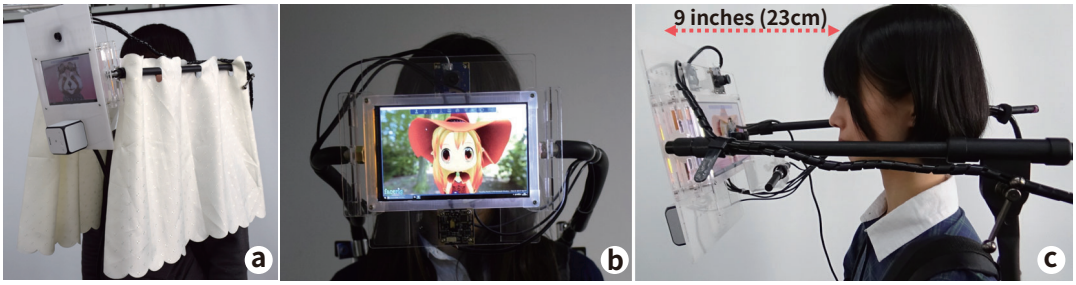

図 3 (a) ウェブカメラベースの表情認識機能をもつ e2-Mask[39] の外観 (b) 正面から見た e2-Mask (c) 側面から見た e2-Mask

する．これにより，ユーザは人形の視界の確保だけでなく，人形の周囲の雰囲気を視覚的に理解できる．他にも，遠隔対話者の見た目を変えることで，遠隔共同作業における創造力を向上させるシステム [20] や，遠隔対話者の非言語情報を拡張し，会話に臨場感を与えるシステム [31] がある．これらの顔を拡張する研究はいずれも遠隔でのコミュニケーションを前提とする一方，本研究は実世界における対面コミュニケーションを前提としている点で異なる．そのために，光センサアレイによる近接表情認識機能を備えたマスク型ディスプレイを提案し，対面コミュニケーションにおけるアバターを用いた表情拡張をめざす．

大澤の AgencyGlass[24] は，サングラスの形をしており，目と同じ大きさの液晶ディスプレイをサングラスのレンズに配置し，あらかじめ撮影した装着者の目の動きを液晶に表示する．例えば，接客業において，店員は悲しい感情を抱えている場合でも自らの感情を制御し，笑顔で顧客に接客しなければならない．そこで，店員が笑った時の目の動きを AgencyGlass に表示することで，店員は自らの感情を制御せずにその場に適した目の動きを表出できる．このように，AgencyGlass は，装着者の感情と異なる目の動きを表出し，感情労働の負担を軽減できる．石井らの MouthOver [37] や HappyMouth[38] は，サージカルマスクにディスプレイを搭載したマスクデバイスである．ディスプレイには本来マスクで隠されている装着者の口を表示している．AgencyGlass と同様，感情労働の負担軽減などに応用できる．HYPERFACE[4] は帽子のつばの部分に小型プロジェクタを内蔵する帽子型ウェアラブルデバイスで，人の顔に表情を拡張する映像効果を顔にプロジェクションマッピングする．例えば，笑顔の人の頬に赤みを帯びた映像を投影することで，バーチャルなメイクを実現している．また，ChromoSkin[9], [10] は，従来の化粧品にサーモクロミックパウダーを加え，熱を加えることでアイシャドウの色を制御するデジタルメイクを実現している．これらの事例のように，目や口といった顔の一部を変えたり，メイクアップだけでは，装着者がもつ顔の特徴（年齢，性別，顔つきなど）が残ってしまう．本研究では，装着者がもつ顔全体を変えるという点で異なる．

山本らは聴衆に肯定的な反応を重ねる方法を提案し，シースルー HMD[36] を使用して各オーディエンスメンバーに笑顔のカボチャの画像を重ねるシステムを実装した．このシステムはプレゼンテーションで使用される．しかし，デジタルカメンは会話などの対面コミュニケーションの利用を想定しており，笑顔だけでなく，さまざまな種類の表情も必要である．

2.2 表情認識インタフェース

表情認識機能をもつウェアラブルインターフェースが多数研究されている [2]．Fukumoto ら [6] は，フォトインタラプタを用いた笑顔や笑い声の認識方法を提案している．Masai ら [16] は，表情認識のための光反射センサを組み込んだスマートアイウェアを提案している．このスマートアイウェアは 8 種類の表情を認識することができる．Suzuki ら [31] は，HMD に埋め込まれた光センサを用いた顔の表情認識手法を提案している．これらの研究では，感情を表す顔の表情を認識することに重点が置かれているため，話し手の口の動きを認識することは困難である．Le ら [13] は， 8 個のストレインゲージセンサを内蔵し，顔の上部の表情を認識するとともに，奥行きカメラを用いて顔の下部の表情を認識することで，顔の性能をセンシングする HMD を提案している．坂下ら [27] は，口の開閉を検出するために光センサアレイを使用している．これらの研究と同様に，デジタルカメンでは顔の表情認識に光センサを利用している．光センサをマスクの顔全体に配置することで，顔の表情と口の動きを正しく認識するだけでなく，マスクに表示されたアバタのアニメーションにも反映させることができる．

Mimicat[29] は，着ぐるみ装着者の顔形状の変化を赤外線カメラを用いて認識する手法を提案している．また，別のアプローチとして，8 個の光センサにより着ぐるみ装着者の顔形状の変化を認識する手法も提案している．さらに，認識した顔形状をもとにアクチュエータを使って着ぐるみの表情を変化させており，装着者の顔の認識および再現という点で本研究と類似している．本研究との相違点は，本研究では表示にディスプレイを使っていること，光センサ

を増やすことによって笑顔や驚きなど多様な表情を認識できること，表情認識結果を確率値として出力し中間表情を再現できていることである．

2.3 カメラベースのデジタルマスクディスプレイ

筆者らの研究グループは，カメラを使ったデジタルマスクディスプレイ e2-Mask[39] を開発した．これは，2 台のディスプレイと 2 台のウェブカメラを使って，装着者の顔を撮影し，アバタの表情に反映させる（図 3）．2 台のディスプレイとカメラは，マスクデバイスの装着者側と対話者側のそれぞれに設置されている．装着者側のディスプレイには，装着者から見て外側に設置されたウェブカメラの映像が表示され，一人称視点の映像を見ることができる．Facerig[30] は，装着者側のカメラで認識した装着者の顔の表情や向きをもとに，アバタの表情を制御する．アバタは，ディスプレイの前面に表示されている（図 3(b)）．e2-Mask を面接という状況で利用してもらう認知心理学実験を実施し，面接官が e2-Mask を装着することで，受験者の緊張が緩和するという結果が得られた．一方，カメラを用いたデジタルマスクの限界として，顔全体の表情を認識するためには，カメラと装着者の間に最低 23cm の距離が必要となり，マスクと呼ぶには不自然なほどディスプレイが前面に飛び出してしまう（図 3(c)）．さらに，装着者の顔を隠すためのカーテンが必要となる（図 3(a)）．これらの問題を解決するために，本研究では，フレキシブルなディスプレイをベースにした薄型デジタルマスクディスプレイの設計と実装，および光センサアレイをベースにした近接表情認識モデルを構築する．

3. デジタルカメン

3.1 ワークフロー

デジタルカメンのワークフローを図 4 に示す．また，アバタの表情が変化している様子を図 1 に示す．デジタルカメンは，顔ディスプレイマスクと顔キャプチャマスクから構成されている．顔ディスプレイマスクとは，アバタを出力するディスプレイである．一方で，顔キャプチャマスクとは装着者の表情を認識するマスクである．顔ディスプレイマスクは，顔を人の顔と同様に立体的に再現するために，Flexible Top Hat [26] を使用する．Flexible Top Hat は ROYOLE 社が販売している有機 EL ディスプレイを搭載した帽子である．これを，顔キャプチャマスクを覆うように加工する．

顔キャプチャマスクの裏面に配置される光センサは Arduino Pro mini で一括管理され，シリアル通信で PC に送られる．PC は，光センサデータをもとに表情を識別し，表情付きアバタを生成する．生成したアバタの映像を顔ディスプレイマスクに表示させる．顔ディスプレイマスクとして使用する有機 EL ディスプレイは Apple 社が提供している AirPlay を利用して PC の映像をミラーリングできる．これを利用して PC の映像を顔ディスプレイマスクに表示させる．

装着者の視界を確保するために，顔キャプチャマスクの目の位置に穴をあけ，顔ディスプレイマスクを目より下の位置に設置した．

顔キャプチャマスクおよび顔ディスプレイマスクの総重量は 485g である．なお，この重量には，PC，有機 EL ディスプレイを駆動するためのモバイルバッテリ，デジタルカメンと PC をつなぐケーブルは含まれていない．

3.2 表情認識モデル

ハードウェア　光センサの大きさは小型（W × L × H= 2.7 × 3.2mm × 1.4(mm)）である．40 個の光センサを図 5 に示すように顔キャプチャマスクに配置した．図 5 に実装した顔キャプチャマスクおよび光センサを示す．先行研究 [21], [23], [32] における表情認識手法を参考として，デジタルカメンを装着したときに，マスク装着者の目・鼻・口が光センサに当たらないように眉・目元・頬・口元に光センサを配置した．Arduino Pro mini を利用して，各光センサデータを集積し，Arduino Pro mini から PC にセンサデータを送信する．Arduino Pro mini および PC 間のデータの送受信はシリアル通信を利用した．

ソフトウェア　表情を変えるたびに顔の皮膚の盛り上がり方が変わり，同時に表情ごとに光センサと顔の皮膚の距離感も変わる．この特性を利用し，各光センサデータを特徴量とし，Support Vector Machine (SVM) を使用して表情認識モデルを構築する．具体的には，Arduino 環境下で，各光センサからの入力を 10 ビットデータに変換し，PC 上のソフトウェアで各センサデータを 0 から 1 の範囲に正規化した．その後，rbf カーネルの SVM (C = 10, gamma = 1.0) アルゴリズムを適用し，表情認識モデルを構築した．ハイパーパラメータは，非線形識別および線形識別におけるいくつかの方式を実装し予備実験を行った結果のもとで決定した．機械学習における表情認識アルゴリズムは多数存在するが，ウェアラブルでリアルタイムなデジタルフェイスマスクの表示を考慮して，計算コストが低いという利点がある SVM を選択した．また，光センサを用いた表情認識の実現性については，先行研究 [16] で報告されており実績がある．

表情認識モデルの構築と同様，正規化した 40 個の各光センサデータを特徴量とし，これらを表情認識モデルに適用することで表情クラスを推定する．アバタの中間表情を生成するために，入力した特徴量に対して，各表情クラスに分類される確率を最終的に出力する．なお，SVM は Python 上で Scikit-learn ライブラリを用いて構築した．

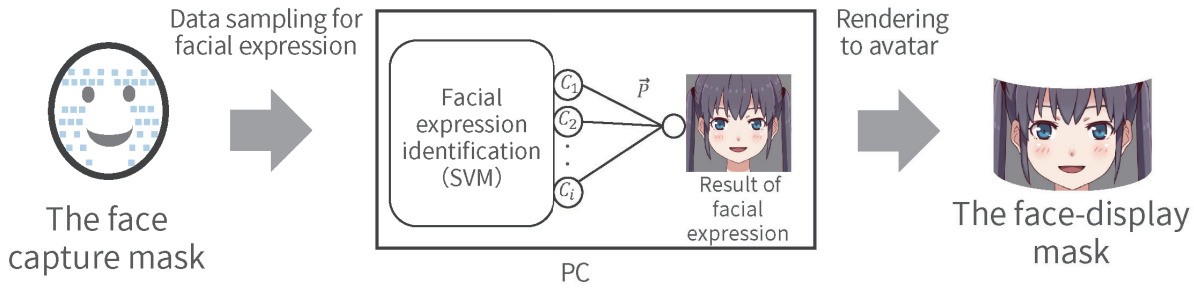

図 4 デジタルカメンのワークフロー

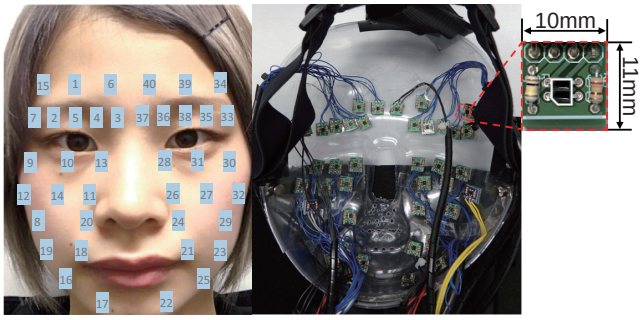

図 5 表情をサンプリングするための 40 個の光センサアレイの配置（左），顔キャプチャマスク（中央），光センサモジュールの寸法（右上）

3.3 アバタの出力

前節で説明した表情認識結果をもとに装着者の表情をアバタに反映する．本手法は鈴木ら [32] を参考とした．アバタに反映させる表情を任意の n 次元の表情パラメータで表現できる場合，左目，右目の開き具合や口の幅のパラメータなどを格納したアバタの表情パラメータ $\vec{P}$ は，真顔の表パラメータを $\vec{P_B}$，真顔の表情以外の m 個の表情パラメータを $\vec{P_i}(i = 1, 2, ..., m)$，表情の多クラス分類の確率値を $\vec{C} = (c_1, c_2, ..., c_m)$ とすると下記の式で合成できる．

$$\vec{P} = \vec{P_B} + \sum_{k=1}^{n} c_i \times (\vec{P_i} - \vec{P_B})$$

上記の手法でアバタの表情を変化させることで，基本表情だけでなく微小な表情の表現や表情が変化する際の中間表情の出力が可能となる．なお，アバタのモデルの生成には Live2D[14] を使用し，アバタの表示には物理シミュレーションを搭載したゲームエンジンである Unity を使用した．PC 上のプログラム間の通信は UDP を利用し，表情認識結果，キーボード入力などをそれぞれ伝送している．

4. 表情認識モデルの性能評価

表情認識モデルの性能を評価するための実験を実施した．実験に用いた表情は図 6 に示す 10 種類である．1 番から 5 番までの表情は Ekman によって定義された基本感情 [25]

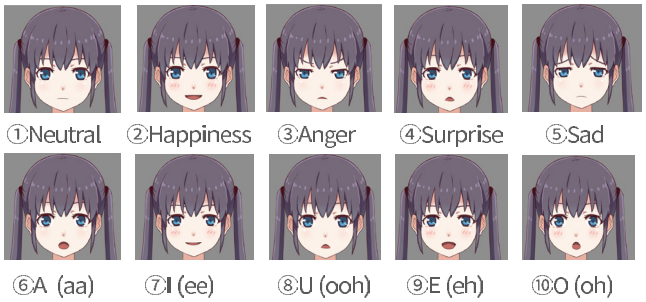

図 6 実験に使用した表情

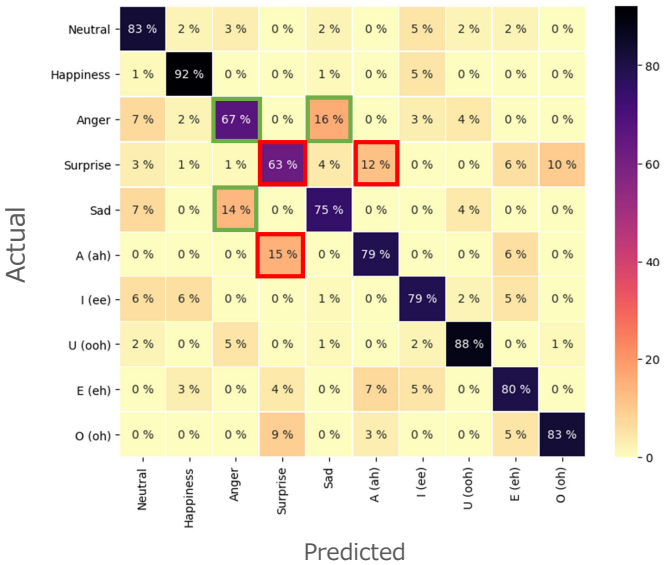

図 7 表情認識結果

である．6 番から 10 番は母音を発音したときの表情である．デジタルカメンを装着した状態で会話するため，感情を表した表情だけでなく，話している最中の口の動きを表現させる必要がある．そのため，感情を表した表情と母音を発音した際の口の動きを表現している 10 種類の表情を用いた．なお，Ekman によって定義された基本感情は，幸せ，嫌悪，怒り，驚き，恐怖，悲しみの 6 表情であるが，事前実験により，嫌悪や恐怖は表情として再現しづらい [35] ことが知られており除外した．

被験者は公立はこだて未来大学に所属する 20 代の大学生および大学院生 7 名で女性 4 名，男性 3 名である．本実験の内容いずれの被験者も日本人で，欧米人と比較して，彫が浅く，鼻が低いという特徴をもつ．また，本実験への参加およびデジタルカメンのプロトタイプの利用は初めてであった．実験の手順として，被験者にプロトタイプを装着してもらう．次に，図 6 に示す表情を 1 番から 10 番まで順に真似してもらうよう指示した．これを 1 セットとし，6 セット繰り返し実施した．1 つの表情に 10 サンプルのデータを取得し，10 種類の表情を 6 セット採取したため，1 人の被験者に対して合計 600 サンプルのデータを取得した．なお，顔キャプチャマスクのプロトタイプは 1 機しか存在せず，被験者は同じサイズの顔キャプチャマスクを使用した．

4.1 結果

600 サンプルのデータに対し 100 サンプルのデータをテストデータ，500 サンプルのデータを学習データとして，6 分割交差検証（leave-one-set-out）を適用した．その結果，正答率は 79%であった．

4.2 考察

4.2.1 表情認識器の正答率

図 7 に 600 サンプルのデータを用いた場合の全被験者を合わせた混合行列を示す．デジタルカメンは本人の表情を学習した条件で 79%の正答率を示しており，表情認識結果は全体的に高い正答率となった．また，キャリブレーションにかかった時間は，モデル生成を含めて平均 3 分であり，短時間でキャリブレーションを完了できる．また，1 サンプルに対して表情の認識にかかる平均時間は 1.38(msec)，標準 0.54(msec) であり，リアルタイムに表情を認識できる．

喜びの表情の正答率が最も高く，驚きの表情の正答率が最も低い結果となった．最も認識正答率の悪かった「驚き」の表情についてその原因を分析する．驚きの正解データの内，驚きと予測されたデータは 63%であった．驚きの正解データの内「あ」と予測されたデータは 12%であった．一方，「あ」の正解データの内，驚きと予測されたデータは 15%であった．これらの結果，驚きの表情は「あ」の表情に誤認識しているといえる．

また，2 番目に正答率が低い表情は怒りの表情であった．怒りの正解データの内，怒りと予測されたデータは 67%であった．怒りの正解データの内，悲しみと予測されたデータは 16%であった．一方，悲しみの正解データの内，怒りと予測されたデータは 14%であった．被験者から「悲しみと怒りの表情の分別が難しい」という指摘があった．被験者によっては怒りと悲しみの表情の区別がつけにくく，誤認識が生じたと考えられる．

怒りの表情と悲しみの表情を個別に表現しづらいといったように，デジタルカメン利用者自身が表現しづらい表情があるのであれば，デジタルカメンを利用する事前設定において，怒りと悲しみを 1 つの表情として登録することで，認識正答率の低下を回避できる．また，「あ」の表情および驚きの表情をそれぞれ反映したアバタは似ており，これらを誤認識したとしても対話者はそのことに気づかない可能性も考えられる．本実験では表情認識モデルの正答率を評価するために 10 種類の表情を利用したが，実際にデジタルカメンを運用する場合には，状況を考慮した上で一般的に利用する表情をデジタルカメン装着者および対話者の両面から精査する必要がある．

4.2.2 光センサデータの変化量

図 5 に示すように，デジタルカメンは顔キャプチャマスクに 40 個の光センサを配置している．そのため，顔の形や大きさが異なっていても表情認識に寄与するセンサが存在し，汎用性を高めている．

600 サンプルのデータを用いて，各表情の光センサデータの平均値を算出し，真顔を基準とし，真顔と各表情のフォトリレクタセンサデータの差分の絶対値を算出した．その絶対値を累積した値に対して正規化した結果を変化量と定義する．その結果，被験者によって表情ごとの光センサの変化量が大きく異なった．本論文では特徴的であった 3 名の被験者の結果を例にあげる．3 名の被験者の変化量の結果を図 8 に示す．グラフの横軸は光センサの ID を表し，縦軸は変化量を表している．積み上げた変化量が大きい光センサほど表情変化に対して敏感に反応しているといえる．また，図 8 の下部に示す顔のダイアグラム上の矩形は，マスク上に配置された各光センサの変化量をヒートマップで可視化したものである．色が濃いほど変化量が大きいことを示す．

図 8 に示す被験者 1 では 9 番，13 番から 26 番，32 番の光センサデータにおける表情の変化量が，被験者 2 では 1 番から 13 番，28 番から 40 番の光センサデータにおける変化量がそれぞれ大きい．被験者 3 では全体的に光センサデータにおける変化量はほぼ同じである．これより，被験者によって表情変化に寄与する光センサは異なるといえる．なお，被験者 3 の変化量は，被験者 1 や被験者 2 の変化量と比較して全体的に小さい．しかし，3 名の被験者の正答率は 73%以上であるため，積み上げた光センサデータの変化量の大小に関係なく表情認識は可能である．

デジタルカメンは様々な人が使用することを想定している．したがって，顔の形やサイズに関わらず高い正答率で表情を認識できる必要がある．デジタルカメンは顔キャプチャマスク一面に光センサを配置している．これにより，顔の形やサイズが異なっていても平均正答率が 79%と高い正答率で表情を認識できている．

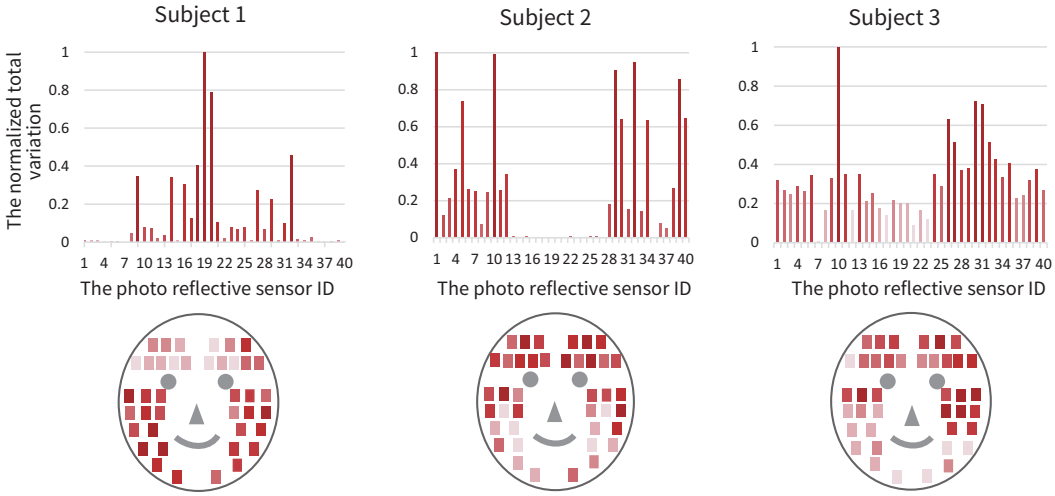

図 8 3 名の被験者の光センサアレイデータの変化量

表 1 ユーザスタディの結果

質問番号	質問内容	回答	提案手法 M	提案手法 SD	比較手法 M	比較手法 SD	P
1	どれくらいあなたの表情をアバタは再現していましたか？	1: 再現していた - 5: 再現していなかった	2.8	1.1	3.0	1.0	0.50
2	アバタの表情変化はどれくらいスムーズに変化できていましたか？	1: スムーズであった - 5: スムーズでなかった	2.0	0.5	2.6	1.1	0.28

5. ユーザスタディ

デジタルカメンが生成および表示するアバタの有用性を調査するために，カメラベースで認識した表情を表情付アバタとして生成および表示する機能をもつ e2-Mask を比較手法 [39] としてユーザスタディを実施した．

5.1 実験内容

被験者は，公立はこだて未来大学に所属する 20 代の大学生および大学院生 8 名で女性 4 名，男性 4 名である．4 章の実験と同様，被験者は日本人で，欧米人と比較して，彫が浅く，鼻が低いという特徴をもつ．被験者には，e2-Mask（比較手法）およびデジタルカメン（提案手法）それぞれを装着してもらい，実験課題として，2 種類のデジタルマスクディスプレイ（e2-Mask とデジタルカメン）上で，図 6 に示す 5 つ（真顔，幸福，怒り，驚き，悲しみ）の表情を，各被験者に表現してもらった．いずれの被験者もデジタルカメンや e2-Mask の利用は初めてであった．また，リアルタイムに表示される表情付アバタが投影されたディスプレイを視聴してもらった．提案手法を利用してもらう前に，各被験者から 300 サンプルのデータ（5 表情× 10 サンプル × 6 セット）を収集し，表情認識モデルを構築するためのキャリブレーションを実施した．一方，カメラベースの表情認識 [30] である比較手法は，特にキャリブレーションを必要としなかった．被験者には，e2-Mask とデジタルカメンの両方のデジタルマスクディスプレイについて，最大 20 分間，被験者が満足するまで，練習してもらった．練習後，被験者にいずれかのデジタルマスクディスプレイを装着してもらい，鏡から 30cm 離れたワイドミラー（2000mm × 830mm）に面した席に案内し，鏡に映った被験者自身の姿を見てもらった．カウンターバランスをとるために，8 人の被験者のうち 4 人には提案手法を装着して実験してもらい，もう 1 つの被験者グループには比較手法を装着してもらった．また，疲労による表情変化の影響を軽減するために，実験の前半に 10 分間の休憩時間を設け，その後，デジタルマスクディスプレイを入れ替え，後半の実験に取り組んでもらった．図 6 に示す 5 つ（真顔，幸福，怒り，驚き，悲しみ）の画像を模範画像として記憶してもらった．実験中は模範画像を使わずにランダムに指示された表情を再現してもらった．1 つの表情につき 3 回，計 15 回表情を再現してもらった．実験終了後，2 種類のデジタルマスクディスプレイについて，表 1 に示す 2 つの質問に理由を添えて 1〜5 で回答してもらった．

5.2 結果

表 1 は，アンケートの各質問項目について，提案手法と比較手法における平均値（M），標準偏差（SD），両側有意確率（P）をそれぞれ示している．平均値については，値が小さいほど質問に対して肯定的な回答が多いことを示している．有意確率（P）は，提案手法と比較手法の両方について，各質問の回答結果を Wilcoxon signed-rank 検定で算出した．質問 1 では，提案手法（デジタルカメン）の平均値が 2.8 であった一方，比較手法（e2-Mask）の平均値

は 3.0 であった．質問 2 では，提案手法の平均値が 2.0 であった一方，比較手法の平均値は 2.6 であった．両質問において提案手法の平均値は，比較手法の平均値よりも小さくなったが，有意差は見られなかった．

5.3 考察

各質問項目において，有意差は観測されず，提案手法および比較手法いずれも再現性および応答性において同程度の結果が得られた．比較手法で利用されているカメラベースの表情認識は商用のアプリケーションであり，vTuber などで一般に広く使われている．本アプリケーションと同程度の結果が得られたという結果は，デジタルカメンの有用性において一定の信頼性があるといえる．

質問項目 1 について，両手法とも被験者が期待した表情を再現できない場合があった．例えば，比較手法では，真顔の閉じた口をトラッキングできず，アバタには半開きの口が表示されていた．また，提案手法ではアイトラッキングが適用されなかったため，視線やまばたきを正しく再現できなかった．これらの理由から，各手法において質問項目 1 の再現性という点でスコアが低くなった．質問項目 2 について，両手法とも 5 段階尺度の中央値である 3 を下回っており，表情制御の反応速度に肯定的な回答が得られた．有意差は観測されなかったものの，被験者は提案手法の方が比較手法よりも反応速度が速いという結果が得られた．1 名の被験者は，「表情の変化に遅れがあると，アバタに対して自分が操作している感覚が薄れる」と比較手法に対してコメントしている．また，別の 1 名の被験者は「自分の表情とアバタの表情がスムーズに連動していることがわかり，自分が操作しているように思えた」と提案手法に対してコメントしている．

これらの結果から，提案手法はカメラベースの比較手法と同程度にスムーズにアバタを再現できているといえる．

6. リミテーションおよび展望

6.1 表情認識モデル

提案する表情認識モデルは，現時点では，10 種類の個別の表情しか認識できず，視線の方向やまばたきなどを識別することはできない．この問題を解決するために，今後は，目の周りにさらに光反射型センサアレイを配置し，機械学習 [15] を用いた目の状態の推定モデルの構築に取り組む．また，今後はより高密度な光センサアレイを用いて，3 次元の物理的な顔モデル [13] を構築し，マイクロエクスプレッション [18] を推定することで，拡張アバタの顔の表情をよりスムーズに変化させることが可能になると考えられる．表情認識の正答率を向上させるためには，SVM アルゴリズム以外の機械学習アルゴリズムが考えられる．例えば，ランダムフォレストを用いた多クラス分類 [3] を適用することで，表情間の表情遷移の確率を導出し，現在の表情から将来遷移する可能性のある表情を絞り込むことで，表情認識正答率の向上が見込まれる．

また，本論文では被験者に意図的に作ってもらった表情に対して表情認識モデルを構築し，その正答率を評価した．一方，意図的に作らない自然な表情における表情認識正答率に関しては十分に評価できていない．また，図 2(a) で示したように ALS の患者のための表情認識は，患者自身がどこを動かせるかというチューニングしたうえで表情を認識する必要がある．これらは表情認識の方式自体から検討する必要があるように考えられる．これらの実現に関しては今後の課題である．

6.2 顔キャプチャマスクの筐体

顔キャプチャマスクと装着者の顔の大きさの不一致も実用上の問題として残っている．マスクが装着者の顔よりもはるかに小さい場合，一部の光センサが顔に触れてしまい，対応する顔の特徴を正しく取得できなかった．装着者の肌とセンサの擦れにより装着者が不快感をもってしまう．単純な解決策としては，マスクの端に鼻パッドやパッドを追加することで，顔とセンサアレイの位置の間の不一致を克服できる．逆に，マスクが装着者の顔よりも大きい場合，センサと肌の距離を変更できるスペーサを挿入できるセンサホルダを実装することで解決できる．

6.3 デジタルカメンを装着する意義

本論文では，表情認識モデルの性能（4 章）や，意図的に作った表情に対してデジタルカメン上に生成されるアバタの有用性について主に議論しており，図 2 で示した用途での評価や代替手段との比較議論ができていない．本人がいるにもかかわらずあえて仮面を通して会話するという状況は，一般的には奇異な状況で，社会的に受容されるかは大きな課題といえる．筆者らの研究グループは採用面接という状況において，受験者がデジタルカメンを用いた面接官の印象を調査するという認知心理学実験を実施した [22]．デジタルカメンのアバタの種類によって受験者の緊張度の減衰あるいは増幅するという結果が得られた．この認知心理学実験も緊張度という側面でしか評価できておらず，受験者がもつ面接官に対する信頼の度合いなど多面的に調査する必要がある．デジタルカメンを装着する意義の調査に関しては今後の課題である．

6.4 デジタルカメンの副作用

宗教的理由で顔を出せない人々が顔を表現する手段としてデジタルカメンを利用する可能性はある一方，実世界において顔を隠したい人々への支援となるかについて，長期的な影響を含めた議論が必要である．例えば，デジタルカメンで一時的に顔を隠せたとしても，生涯顔を隠し通すことは難しい．デジタルカメンが顔を晒すという機会を奪っ

てしまう可能性はある．デジタルカメンはデジタルであるがゆえにアバタを表示させることもできれば，デジタル的に本人の顔を表示することもできる．例えば，初対面や見知らぬ人にはアバタを表示し何度も会うにつれ自分自身の顔を少しずつ表示する，笑顔のときは本人の顔を表示するなど，本人の顔を晒すという行為に多様な選択肢を提供できる．

6.5 デジタルカメン装着時の行動制限

図 1 に示すようにデジタルカメンは目の下にディスプレイを配置し，装着者の視界を確保している．また，ディスプレイは顎部分まで覆いかぶさる．このため，デジタルカメンを装着しながら，食べたり，飲んだり，鼻をかんだりするという行為は難しい．バイクヘルメットにおけるバイザー部分のフリップアップ機構のように，デジタルカメンにフリップアップ機構を導入し，デジタルカメンの脱着を容易にできるようにすることでこの問題を解決できると考えられる．

また，歩行に関して，首を動かさずに目だけ動かして足元を確認することは難しく，足元に注意を払わないといけないときは，首を動かして足元を確認する必要がある．透過性の高い有機 EL ディスプレイを利用することで視野を広げられると考えられる．

7. まとめ

本論文の目的は，人間の顔をアバタに置き換えることで，デジタルな自己表現を可能にし，実空間での対面コミュニケーションを支援するデジタルカメンの設計と実装である．デジタルカメンは，個人が実世界の中で他者から認識される理想的なペルソナを柔軟に表現するための新しいメディアやインターフェースとなり，外見や確立された個人的関係などの制約を超えて，対面コミュニケーションの体験を拡張できる．従来のカメラによる表情認識技術は，顔とカメラの間に一定の距離が必要であった．本論文では，この技術的課題に取り組んだ．本研究では，フレキシブルな有機 EL ディスプレイと，40 個の光センサから構成される光センサアレイデータを SVM に学習させることで構築した表情認識モデルを統合したデジタルカメンを開発した．評価実験の結果，近接表情認識モデルは，デジタルカメンを装着したユーザの表情を平均 79 %の正答率で分類できることが明らかになった．また，デジタルカメンで表示されるアバタの有用性を検証するためにユーザスタディを実施し，その結果，デジタルカメンは比較手法である e2-Mask と比較して，有意差は観測されなかったが，表情の再現性および表情変化の応答性という点で同程度の結果が得られた．

謝辞 本研究は JSPS 科研費 19H04157 の助成を受けたものである．

参考文献

[1] Akaike, Y., Komeda, J., Kume, Y., Kanamaru, S. and Arakawa, Y.: AR Go-Kon: A System for Facilitating a Smooth Communication in the First Meeting, *2014 IEEE 11th Intl Conf on Ubiquitous Intelligence and Computing and 2014 IEEE 11th Intl Conf on Autonomic and Trusted Computing and 2014 IEEE 14th Intl Conf on Scalable Computing and Communications and Its Associated Workshops*, pp. 120–126 (2014).

[2] Bernal, G., Yang, T., Jain, A. and Maes, P.: PhysioHMD: A Conformable, Modular Toolkit for Collecting Physiological Data from Head-Mounted Displays, *Proceedings of the 2018 ACM International Symposium on Wearable Computers*, New York, NY, USA, Association for Computing Machinery, p. 160–167 (online), DOI: 10.1145/3267242.3267268 (2018).

[3] Dapogny, A., Bailly, K. and Dubuisson, S.: Pairwise Conditional Random Forests for Facial Expression Recognition, *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 3783–3791 (2015).

[4] designboom: eun kyung shin's social mask is an AI gadget that displays human emotion in real time, https://www.designboom.com/technology/eun-kyung-shin-hyperface-artificial-intelligence-social-mask-08-04-2017/ (2017). Date last accessed July 20, 2018.

[5] Frueh, C., Sud, A. and Kwatra, V.: Headset Removal for Virtual and Mixed Reality, *ACM SIGGRAPH 2017 Talks*, New York, NY, USA, Association for Computing Machinery, (online), DOI: 10.1145/3084363.3085083 (2017).

[6] Fukumoto, K., Terada, T. and Tsukamoto, M.: A Smile/Laughter Recognition Mechanism for Smile-Based Life Logging, *Proceedings of the 4th Augmented Human International Conference*, New York, NY, USA, Association for Computing Machinery, p. 213–220 (online), DOI: 10.1145/2459236.2459273 (2013).

[7] GREATWORKS: TABLET MAN, http://www.greatworks.co.jp/works/tablet-man.html (2013). Date last accessed April 24, 2017.

[8] Hagiwara, S. and Kurihara, K.: Development and Evaluation of a "Gaze Phobic KOMYUSHO" Support System using See-through HMD based on Social Welfare Approach, *Computer Software*, Vol. 33, No. 1, pp. 52–62 (2016).

[9] Kao, C. H.-L., Mohan, M., Schmandt, C., Paradiso, J. and Vega, K.: ChromoSkin: Towards Interactive Cosmetics Using Thermochromic Pigments, pp. 3703–3706 (online), DOI: 10.1145/2851581.2890270 (2016).

[10] Kao, C. H.-L., Nguyen, B., Roseway, A. and Dickey, M.: EarthTones: Chemical Sensing Powders to Detect and Display Environmental Hazards through Color Variation, pp. 872–883 (online), DOI: 10.1145/3027063.3052754 (2017).

[11] 川西千弘：正確さへの動機づけが対人認知における顔の機能に及ぼす効果，心理学研究，Vol. 68, No. 6, pp. 465–470 (1998).

[12] Koulieris, G., Akşit, K., Stengel, M., Mantiuk, R., Mania, K. and Richardt, C.: Near-Eye Display and Tracking Technologies for Virtual and Augmented Reality, *Computer Graphics Forum*, Vol. 38, pp. 493–519 (online), DOI: 10.1111/cgf.13654 (2019).

[13] Li, H., Trutoiu, L., Olszewski, K., Wei, L., Trutna, T., Hsieh, P.-L., Nicholls, A. and Ma, C.: Facial Performance Sensing Head-Mounted Display, *ACM Trans. Graph.*, Vol. 34, No. 4 (online), DOI: 10.1145/2766939

(2015).
[14] Live2D: Live2D Cubism, https://www.live2d.com/en/ (2020). Date last accessed February 24, 2020.
[15] Masai, K., Kunze, K. and Sugimoto, M.: Eye-Based Interaction Using Embedded Optical Sensors on an Eyewear Device for Facial Expression Recognition, *Proceedings of the Augmented Humans International Conference*, New York, NY, USA, Association for Computing Machinery, (online), DOI: 10.1145/3384657.3384787 (2020).
[16] Masai, K., Sugiura, Y., Ogata, M., Kunze, K., Inami, M. and Sugimoto, M.: Facial Expression Recognition in Daily Life by Embedded Photo Reflective Sensors on Smart Eyewear, *Proceedings of the 21st International Conference on Intelligent User Interfaces*, New York, NY, USA, Association for Computing Machinery, p. 317–326 (online), DOI: 10.1145/2856767.2856770 (2016).
[17] Mehrabian, A. et al.: *Silent messages*, Vol. 8, No. 152, Wadsworth Belmont, CA (1971).
[18] Merghani, W., Davison, A. K. and Yap, M. H.: A Review on Facial Micro-Expressions Analysis: Datasets, Features and Metrics (2018).
[19] Misawa, K. and Rekimoto, J.: ChameleonMask: Embodied Physical and Social Telepresence using Human Surrogates, *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, Vol. 14, No. 2, pp. 115–128 (online), available from ⟨https://doi.org/10.1145/2702613.2732506⟩ (2017).
[20] Nakazato, N., Yoshida, S., Sakurai, S., Narumi, T., Tanikawa, T. and Hirose, M.: Smart Face: Enhancing Creativity during Video Conferences Using Real-Time Facial Deformation, *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work Social Computing*, New York, NY, USA, Association for Computing Machinery, p. 75–83 (online), DOI: 10.1145/2531602.2531637 (2014).
[21] Nishime, T., Endo, S., Yamada, K., Toma, N. and Akamine, Y.: Feature Acquisition From Facial Expression Image Using Convolutional Neural Networks, *Journal of Robotics, Networking and Artificial Life*, Vol. 3, p. 9 (online), DOI: 10.2991/jrnal.2016.3.1.3 (2016).
[22] Noguchi, K., Takegawa, Y., Tokuda, Y., Sugiura, Y., Masai, K. and Hirata, K.: Study of Interviewee's ImpressionMade by Interviewer Wearing Digital Full-Face Mask DisplayDuring Recruitment Interview, p. 323–327（オンライン）, 入手先 ⟨https://doi.org/10.1145/3472307.3484662⟩, Association for Computing Machinery (2021).
[23] Nomiya, H., Sakaue, S. and Hochin, T.: Recognition and Intensity Estimation of Facial Expression Using Ensemble Classifiers, *International Journal of Networked and Distributed Computing*, Vol. 4, pp. 203–211 (online), DOI: https://doi.org/10.2991/ijndc.2016.4.4.1 (2016).
[24] Osawa, H.: Emotional Cyborg: Complementing Emotional Labor with Human-Agent Interaction Technology, *Proceedings of the Second International Conference on Human-Agent Interaction*, New York, NY, USA, Association for Computing Machinery, pp. 51–57 (online), DOI: 10.1145/2658861.2658880 (2014).
[25] Paul Ekman, F. W. V.: *Facial action coding system*, Stanford University, Palo Alto (1977).
[26] Royole: Flexible Top Hat, https://global.royole.com/jp/flexible-top-hat.
[27] Sakashita, M., Minagawa, T., Koike, A., Suzuki, I., Kawahara, K. and Ochiai, Y.: You as a puppet: evaluation of telepresence user interface for puppetry, *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, pp. 217–228 (2017).
[28] Secord, P.: Facial features and inference processes in interpersonal perception, *Person Perception and Interpersonal Behavior*, pp. 300–315 (1958).
[29] Shoji, R., Yoshiike, T., Kikukawa, Y., Nishikawa, T., Saori, T., Ayaka, S., Baba, T. and Kushiyama, K.: Mimicat: Face Input Interface Supporting Animatronics Costume Performer's Facial Expression, *ACM SIGGRAPH 2012 Posters*, New York, NY, USA, Association for Computing Machinery, （オンライン）, DOI: 10.1145/2342896.2342983 (2012).
[30] Steam: Facerig, http://store.steampowered.com/app/274920/FaceRig/ (2017). Date last accessed April 24, 2017.
[31] Suzuki, K., Nakamura, F., Otsuka, J., Masai, K., Itoh, Y., Sugiura, Y. and Sugimoto, M.: AffectiveHMD: Facial Expression Recognition and Mappingto Virtual Avatar Using Embedded Photo Sensors, *Transactions of the Virtual Reality Society of Japan*, Vol. 22, No. 3, pp. 379–389 (2017).
[32] Suzuki, K., Kionshita, Y., Sakurai, S., Narumi, T., Tanikawa, T. and Hirose, M.: Gender-Impression Modification Enhances the Effect of Mediated Social Touch between Persons of the Same Gender, *Journal of Augmented Human Research*, Vol. 1, No. 2, pp. 379–389 (online), available from ⟨https://doi.org/10.1007/s41133-016-0002-y⟩ (2016).
[33] Thies, J., Zollhöfer, M., Stamminger, M., Theobalt, C. and Nießner, M.: FaceVR: Real-Time Gaze-Aware Facial Reenactment in Virtual Reality, *ACM Trans. Graph.*, Vol. 37, No. 2 (online), DOI: 10.1145/3182644 (2018).
[34] vimeo: Yamada Taro Project, https://vimeo.com/82250584 (2017). Date last accessed April 24, 2017.
[35] Wataru, S., Sylwia, H., Kazusa, M. and Sakiko, Y.: Facial Expressions of Basic Emotions in Japanese Laypeople, *Frontiers in Psychology*, Vol. 12, pp. 1–11 (online), DOI: 10.3389/fpsyg.2019.00259 (2019).
[36] Yamamoto, K., Kassai, K., Kuramoto, I. and Tsujino, Y.: Presenter Supporting System with Visual-Overlapped Positive Response on Audiences, *Advances in Affective and Pleasurable Design*, Springer, pp. 87–93 (2017).
[37] 石井綾郁, 小松孝徳, 橋本直ほか: MouthOver: マスク型デバイスによる対面コミュニケーション能力の拡張, 情報処理学会インタラクション 2017 論文集, pp. 844–845 (2017).
[38] 石井綾郁, 小松孝徳, 橋本直ほか: HappyMouth: マスク型デバイスによる対面コミュニケーション能力の拡張, 研究報告ヒューマンコンピュータインタラクション (HCI), Vol. 2018, No. 7, pp. 1–7 (2018).
[39] 梅澤章乃, 竹川佳成, 平田圭二, 杉浦裕太: e2-Mask:顔の外見を変える仮面型ディスプレイを用いた緊張緩和に関する効果の検証, 映像情報メディア学会誌, Vol. 75, No. 5, pp. 682–690（オンライン）, DOI: 10.3169/itej.75.682 (2021).