

組み込み型光センサを搭載した表情認識機能をもつデジタルカメンの軽量化および日光に対する頑健性の向上

鈴木 あゆみ¹ 竹川 佳成¹ 平田 圭二¹

概要: 近年、バーチャルアバタを通じた交流やコミュニケーションが普及してきた。それに伴い、実環境でバーチャルアバタを用いることで、対話者の不安やストレスの軽減が期待される。そこで、筆者らの研究グループは、実環境の対面コミュニケーションでも同様にバーチャルアバタを使用できるデジタルカメン e2-MaskZ を開発してきた。e2-MaskZ は組み込み型光センサであるフォトリフレクティブセンサを搭載し、組み込まれたセンサから得られたデータをもとに表情を推定する。ユーザの表情に合わせてリアルタイムでアバタの表情を更新し、マスクに搭載されている曲面の有機 EL ディスプレイにマッピングさせることで、実環境でのアバタの使用を可能にした。しかし、既存の e2-MaskZ は重く、装着者への負担が大きい。そのため、マスクを 30 分以上の連続使用が困難である。また、屋外で使用すると日光の影響が強くなり認識精度が著しく低下する。これらの問題を解決するために、ハードウェア面から改良した。

1. 背景

初対面において顔は、その人自身の印象を決める重要なパーツの 1 つである。心理学者の Mehrabian[4] は、初対面における第一印象を決める要因の内訳が視覚情報 (55%)、聴覚情報 (38%)、言語情報 (7%) から形成されると述べている。また、Secord[9] は対面コミュニケーションにおいて、対話者の顔の特徴から性格を推測すると主張している。川西 [15] は、対話者の顔の印象が好ましいほど、親近感を抱きやすい傾向にあるとした。

また、Skype や Zoom などのオンライン会議ツールを利用した遠隔のコミュニケーションにおいて、仮想的に顔の印象を操作することが可能である。近年、バーチャルアバタを使用したコンテンツが注目を集めており [12]、生身の自身ではなくバーチャルアバタを通して、インターネット上にて動画投稿やライブ配信を活動することが身近な技術になりつつある。アバタ作成から動画配信まで簡単にできるアプリなどもあり、スマートフォン 1 つで簡単にしたい自分になれる [11]。同様に、バーチャルアバタを通して交流できるゲームも普及している。このような交流が可能な仮想空間やゲームはメタバースと呼ばれ、さまざまな企業が事業参入を図っている。このようにバーチャルアバタを通じたコンテンツや交流が普及している中で、実環境でバーチャルアバタを用いることで、対話者の不安やストレスの軽減が考えられる。例えば、赤池 [1] らは対話者の顔

に静止画のアバタを HMD (Head Mounted Display : ヘッドマウントディスプレイ) 上に映し出すことで「外見が第一印象に及ぼす影響」を緩和できることを示した。このことから、面接官の表情をアバタに置き換えることで受験者の緊張緩和や、小児患者の歯科治療中の不安感やストレスの軽減などが期待される。

顔の印象をコントロールすることは、対面コミュニケーションに不安を抱える人への支援にもつながる。そこで、新田ら [16] は実環境の対面コミュニケーションでも同様にバーチャルアバタを使用できるインタフェースとして、図 1 に示す e2-MaskZ を提案してきた。e2-MaskZ は組み込み型光センサ (以降、センサと表記する) を搭載したマスクである。組み込まれたセンサから得られたデータをもとに表情を推定し、ユーザの表情に合わせてリアルタイムでアバタの表情を更新している。このアバタを、マスクに搭載されている曲面の有機 EL ディスプレイにマッピングさせることで、実環境で顔のみをアバタに変換させることを可能にした。

しかし、e2-MaskZ (以降は、従来マスクと表記する) は重く、装着者への負担が大きい。従来マスクの重さは約 1.0kg ある。マスクそのものの重量が重いことに加え、センサが肌に触れないようにするために鼻当てを導入しており、鼻根部にマスクの重みが一点に集中する。このため、鼻根部が圧迫され、鼻根部に痛みが生じる。また、従来マスクは人によって精度が低くなる。これは、人によって顔の形状は異なる一方、センサの位置が固定されており、個人々に

¹ 公立はこだて未来大学



図1 e2-MaskZ

合わせることができないことが原因である。装着者の顔の大きさや、凹凸に違いがあるため、センサと肌の距離が大きくなってしまい、値が変化しづらいセンサがあると考えられる。さらに、光センサを使用しているため、日光が差し込む場所で使用するとき、表情認識精度（以降、精度と表記する）が著しく低下する。そこで、本研究では従来マスクのこれらの課題を解決するために、ハードウェア面で改良を行った。具体的には、従来マスク上にある40個のセンサの中で表情認識に寄与する重要なセンサを選定した。また、装着者の顔の形状にフィットするようにセンサを固定している筐体を改良した。さらに、電子回路全体の設計を見直すことで、ケーブルを最適化し、ケーブルの本数や長さを削減した。これらの工夫により軽量化と精度の維持の両立に成功した。また、日光を遮断するために3種類の異なる筐体を実装した。

2. 関連研究

2.1 顔の拡張

これまで、顔の一部や全体をタブレットなどを活用して仮想的に置き換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^{*1}や、TABLETMAN^{*2}がある。山田太郎プロジェクトは、街中で出会った人の顔を撮影し、撮影した顔を自分の顔に装着したタブレットに投影する。一時的かつ匿名性のあるメディアアートである。TABLETMANは、光るラインの入ったスーツを纏い、タブレットを複数搭載した特撮ヒーローのような演出である。タブレットは頭部や腕、腹部と複数搭載し、株式会社東芝の商品などを宣伝している。これらのように、自分の顔などをタブレットで代用したり宣伝に用いる演出は一般的なものである。しかしながら、これらの事例は顔の表情の再現はなくコミュニケーションに関して考慮されていない。

遠隔プレゼンシステムとして、代理人の顔にタブレットを装着させ表情と声を代替するシステムがある。このような事例として、ChameleonMask[5]がある。ChameleonMaskは、遠隔にしながら代理人を立てることでその場に本人がいるように再現するシステムであり、代理人の顔と声はプレゼン本人の顔に置き換えることで遠隔ユーザ本人だと感じさせる。Sakashitaら[7]は、人間のパフォーマーの体

と顔の動き通りにパペットが動くシステムを提案した。このシステムは、人形による視聴覚フィードバックをHMDを用いてパフォーマーに送信することにより、人形目線の情報が得られる。ほかにも、遠隔者の表情を適切なものに変えることで、遠隔者との共同作業における創造性をシステム[8]などがある。これらは、遠隔でのコミュニケーションを想定しているが、本研究では実世界での対面コミュニケーションを目指している点で異なる。そのため組み込み型光センサを用いてユーザの表情を認識する機能を備えたマスクを提案し、実環境においてアバタを用いた表情拡張を目指す。

大澤のAgencyglass[6]は、メガネ型のデバイスでありレンズ部分にディスプレイが設置されている。このディスプレイに事前に収録した目を映し出すことで、装着者の目の動きを表現する。例えば、接客業において店員は自分の感情を抑制し愛想良く笑顔にすることが強制される。そこで、笑ったときの目の動きをAgencyglassに表示させることで、感情の制御にかかる負担を軽減させることができる。このように、Agencyglassは装着者の感情と切り離れた目を出し、反対の感情などを表すためにかかる負担を軽減できる。

石井らのMouthOver[13]やHappyMouth[14]は、サージカルマスクにディスプレイを搭載したマスク型デバイスである。装着者の精神や肉体の状態に関わらず、状況に応じた発話と表情を相手に伝えることができる。AgencyGlassと同様に、感情表現に関わる負担を軽減させ、マスクに隠された表情を伝えることができる。

HYPERFACE^{*3}はサンバイザー型デバイスで、つばの部分に小型プロジェクトが内蔵されている。目の部分に透明のスクリーンを搭載し、そこにハートなど映し出すことができる。例えば、アイシャドウやチークの代わりにバーチャルでメイクができる。また、ChromoSkin[2], [3]は、従来のアイシャドウにサーモクロミックパウダーを加えることで、熱で色が変わるアイシャドウである。これらの事例のように、顔の一部を演出したり、変えたりする化粧だけではユーザ本来の顔の特徴（輪郭、年齢、顔つきなど）が残ってしまう。本研究では装着者の顔全体を変えるという点で異なる。

山本らは、観客にポジティブな反応を重ねるシステムを提案した。このシステムは、シースルーHMD[10]を使用して観客に笑顔のカボチャを重ねて、ユーザの不安や緊張を和らげる。しかし、提案マスクは対面コミュニケーションの利用を想定しているため、笑顔だけではなく、真顔などの表情も必要である。

^{*1} Yamada Taro Project,
<https://vimeo.com/82250584>

^{*2} TABLETMAN,
<http://www.greatworks.co.jp/works/tablet-man.html>

^{*3} 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/>

3. 設計と実装

1章で述べたように、従来マスクは約1.0kgと重いことや日光による精度の低下などの課題がある。これらの課題を解決するために、本研究ではマスクの軽量化（3.1節）や、日光に対する頑健性の向上（3.2節）に取り組んだ。以下、それぞれについて詳細に説明する。

3.1 軽量化

図2と図3にそれぞれ従来マスクと提案マスクを示す。

従来マスクの裏側には図2のようにセンサが配置されている。これらのセンサは筐体に1つずつ埋め込まれているため調節不可能である。センサデータの集積には Arduino Pro Mini を利用しており、有線ケーブルによるシリアル通信を用いてセンサデータを Arduino Pro Mini から PC へ送信している。提案マスクでは、従来マスクでのセンサの特徴量をもとに使用するセンサやセンサの位置を検討した。具体的には、図4および図5に示すように、従来マスクでの特徴量の重要度を算出した。図4の横軸はセンサの通し番号であり、縦軸は重要度を示す。特徴量の重要度は、学習アルゴリズムに使用している Catboost ライブラリに備わっている関数 `get.feature_importance` を用いた。図5では特に重要なセンサの位置を赤（重要度が10-40）、次に重要なものを黄緑（重要度が5-10）、あまり重要ではないものを青（重要度が0-5）で表現している。重要度が0に近いものが多く、口周りのセンサが重要であることがわかった。そこで、表情の識別に重要ではないセンサを減らし、図3のように配置した。センサは、4個で1セットとし、土台を動かせるようにした。センサと肌の距離を調節することができるようになるため、顔の違いによる精度への影響を減らすことが期待される。また、センサデータの集積を Arduino Pro Mini から M5 Atom Lite に変更し、ワイヤレス (Bluetooth) 通信での送受信を可能にした。また、従来マスクではセンサから出ている配線が多く、メンテナンスが困難であった。そこで、センサからマイクロコンピュータに繋がる配線を図6のように変更した。従来マスクでは、1つのセンサから3本の配線が出ていたため、3本*センサ40個の計120本の配線が使われていた。提案マスクでは、センサ4個を1セットとし、GNDやVCCの端子を共通化することで4個当たり6本で、6本*センサ16個/4の計24本にまで減らした。配線の減少による軽量化以外に、不具合があるセンサと繋がっている配線を見つけることが容易になり、修理や交換ができるようになった。さらに、配線を少なくすることで配線がセンサに重なってしまうことを防ぐことができる。従来マスクと提案マスクの重さの変化を表1に示す。これらの改良の結果、0.6kg減らすことができた。また、顔を覆っていた部分を図7のようなアクリルテープに変更した。アクリルテープで額の両端2点と後頭



図2 従来マスク

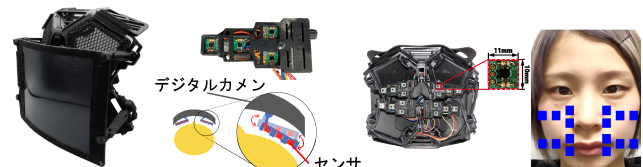


図3 提案マスク

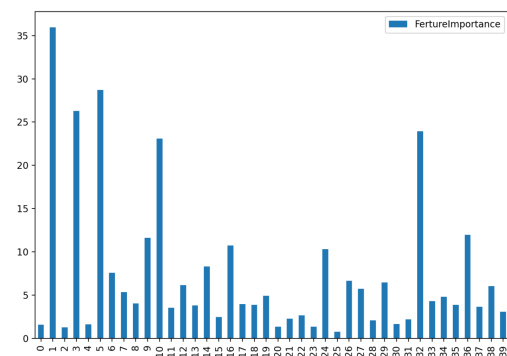


図4 各センサの重要度

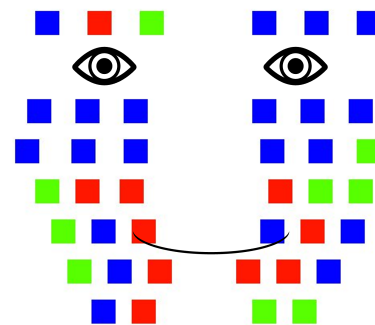


図5 重要なセンサの位置

表1 筐体の変化

	従来マスク	提案マスク
センサ数	40	16
筐体の重さ (kg)	1.0	0.4

部の3点で固定するため、鼻根部に集中していた重さが分散する。これにより、従来マスクにあった鼻根部の痛みが解消された。従来マスクと提案マスクでは全体的に小さくなり、顔からディスプレイの距離が短くなったため、ディスプレイが前面に出すぎる違和感が軽減された。

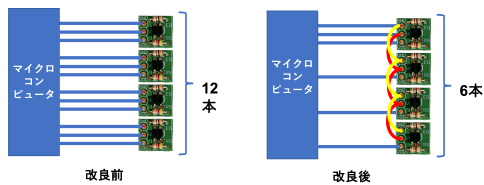


図 6 配線の最適化



図 8 各モードの外観



図 7 提案マスクを着用した様子

3.2 日光に対する頑健性の向上

提案マスクは表情認識のためにフォトリフレクティブセンサを使用している。フォトリフレクティブセンサは、赤外線 LED から放出された光を物体に照射し、反射光をフォトトランジスタで受け取ることでセンサと物体までの距離を計測する。このため日光に対して強く影響を受け、例えば、屋外では正しく表情認識できなくなってしまう。この問題を解決するために、以下に示す 3 つのアプローチを提案する。図 8 に各モードの外観を示す。

パーカーモード

パーカーのフードを深く被り提案マスクに固定したモードである。どの角度からも光が入りにくいが、服装が制限されてしまう。

遮光板モード

画面と顔の隙間に入る光を防ぐために、画面横に 3D プリンタで作成した日よけを付けたモードである。顔側面から入る光を防ぐことが期待され服装も制限されないが、角度によって光が入り込んでしまう。

縦画面モード

縦画面モードは有機 EL ディスプレイの裏側の一部を切り抜くことで視界を確保しつつ遮光板を備え付けたマスクである。この文章以降はこのマスクを縦画面マスクと表記する。有機 EL 自体で遮光されており、側面にも遮光板が備わっているが遮光板モードの半分ほどの大きさである。

4. 表情認識精度の検証

4.1 屋内での表情認識精度

提案マスクでの表情認識精度を評価するため屋内で実験を実施した。提案マスクは従来マスクに比べて、センサを 40 個から 16 個に減らしたことが表情認識精度に及ぼす影響の変化を調査する。

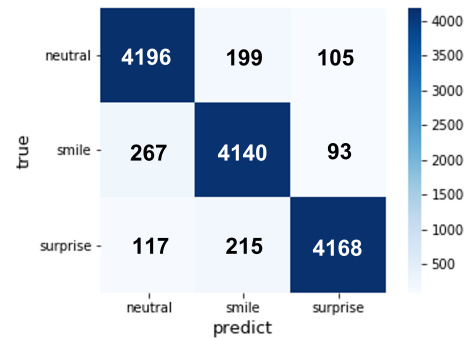


図 9 混合行列

表 2 3 表情の認識精度

	正解率	適合率	再現率	f 値
従来マスク	0.93	0.93	0.93	0.93
提案マスク	0.94	0.94	0.94	0.94

実験方法

初めに、被験者に学習させる表情を説明し、表情の学習を行った。実験者がパソコンのキーボードを操作し、画面に映る表情を口頭で伝えながら、順番に画面に映る表情と同じ表情をしてもらった。実験に用いた表情は真顔、笑顔、驚きの 3 種類で、各表情あたり約 2200 データ*15 セット分収集した。1 人当たり 3 表情*2200*15 セットの約 100000 データをサンプリングした。サンプリングしたデータから、各表情ごとに 1500 個使用し、計 4500 データを学習に用いて、6 分割交差検証を行った。指標として、平均の認識精度（正解率、適合率、再現率、f 値）を算出した。

被験者

従来マスクでの被験者は、大学生 6 名と大人 2 名の男女 8 名である。提案マスクでの実験では、大学生の男 2 名と女 6 名の計 8 名である。

結果

従来マスクと提案マスクでの認識精度を表 2 に示す。提案マスクでサンプリングしたデータでの混合行列を図 9 に示す。提案マスクでの平均正解率は 0.94、平均適合率は 0.94、平均再現率は 0.94、平均 f 値は 0.94 であった。従来マスクでの平均正解率は 0.93、平均適合率は 0.93、平均再現率は 0.93、平均 f 値は 0.93 であることから、全ての指標において精度に大きな変化がないことがわかった。

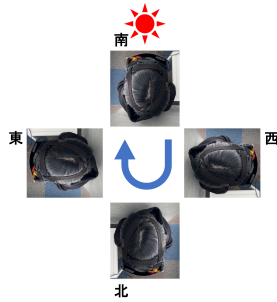


図 10 実験のイメージ

考察

センサ数が 40 個の従来マスクと、センサ数が 16 個の提案マスクでは、どちらも精度に大きな違いがみられないことから、センサ数を減らしても精度に影響しないといえる。また、センサと肌の距離を調節することで値が反応しやすい。これら結果から、センサの位置を反応しやすい位置に配置し、センサを可動式にして個人の顔の形状に合うように筐体を改良したことで、センサの数を減らしても精度は変わらないため運用に問題ないと言える。

4.2 屋外での表情認識精度

3.2 節で作成した日光対策を施した 3 つのモード（パーカモード、遮光板モード、縦画面モード）の表情認識精度を評価するため、快晴かつ日陰がない屋外で実験を実施した。

実験方法

図 10 に実験状況を示す。データセットを構築した時間帯には、太陽が南にあった。東西南北の 4 方向でそれぞれデータを取得した。4.1 節と同様の 3 種類の表情に対して、同数のデータをサンプリングした。表 3 から表 6 に示すように、学習データおよびテストデータに対して、方向ごとに組み合わせを適用し、正解率を算出した。

3.2 節で作成した日光対策を施した 3 つのモードに加え、比較手法として図 7 に示す日光対策を施していないモード（以降、ノーマルモードと呼ぶ）の 4 種類のモードに対して正解率を算出した。

被験者

被験者は本論文の第一著者で 1 名である。

結果

ノーマルモードの正解率を表 3 に示す。パーカモードの正解率を表 4 に示す。同様に、遮光板モードの正解率を表 5 に、縦画面モードの正解率を表 6 に示す。平均正解率は学習データと、テストデータが異なる正解率を標本とする。ノーマルモードの平均正解率は 0.32 に対して、パーカモードの平均正解率は 0.45、遮光板モードの平均正解率は 0.46、縦画面モードの平均正解率は 0.30 であった。パーカモードと遮光板モードのときはノーマルモードよりも約 0.13 ほど正解率が高いが、縦画面モードの正解率は、ノーマルモードの正解率よりも低くなった。

表 3 ノーマルモードの正解率

学習データ	テストデータ			
	南	西	北	東
南	0.68	0.35	0.36	0.32
西	0.34	0.73	0.29	0.33
北	0.34	0.24	0.73	0.19
東	0.42	0.23	0.45	0.68

表 4 パーカモードの正解率

学習データ	テストデータ			
	南	西	北	東
南	0.69	0.40	0.38	0.43
西	0.37	0.77	0.42	0.62
北	0.37	0.49	0.80	0.50
東	0.22	0.46	0.72	0.86

表 5 遮光板モードの正解率

学習データ	テストデータ			
	南	西	北	東
南	0.84	0.34	0.23	0.22
西	0.27	0.89	0.80	0.77
北	0.20	0.62	0.87	0.82
東	0.17	0.40	0.68	0.92

表 6 縦画面モードの正解率

学習データ	テストデータ			
	南	西	北	東
南	0.60	0.36	0.30	0.16
西	0.21	0.77	0.35	0.33
北	0.27	0.28	0.38	0.38
東	0.22	0.36	0.34	0.83

考察

パーカモードと遮光板モードの正解率は、ノーマルモードの正解率と比較して高くなった。しかし、平均正解率は約 0.45 と低く、改善の余地は大いにある。また、被験者 1 名分のデータでしか検証できていないため信頼性に欠ける。筐体の改良以外に、例えば、日光によって過度に反応しているセンサを無効化するなど、認識アルゴリズムを改良することで、精度の向上を実現する。一方、縦画面モードの正解率は、ノーマルモードの正解率と比較して、低くなってしまった。これは、縦画面モードの側面にある遮光板は遮光板モードの半分ほどの大きさしかなかったため、遮光効果が薄かったことが原因であると考えられる。縦画面モードにおける遮光板を大きくし、遮光性を高くすることで、正解率の向上の実現をめざす。

5. まとめ

本研究では、従来マスクの軽量化と屋外での日光に対する頑健性の向上を目的とし、センサ位置の検討によるセンサ数の削減、電子回路全体の見直しによるケーブル本数および長さの削減、データ送受信の無線化などを実現した。

センサの位置は重要度が高い口周りに配置し数を40個から16個に減らし、センサの土台を可動式にした。従来マスクの平均正解率が0.93に対して提案マスクの平均正解率は0.94と精度を維持しつつセンサを減らすことができた。筐体は1.0kgから0.4kgまで軽くなり、無線化によりマスク着用時の負担を軽減した。次に屋外での表情認識精度の向上を目指し、環境光の影響を軽減させるためにパーカモードと遮光板モードと縦画面モードの3つのモードを検証した。ノーマルモードよりもパーカモードと遮光板モードの平均正解率は約0.13ほど向上したが、平均正解率はそれぞれ0.45, 0.46と依然低い。この結果から、筐体の改良だけで日光に対する精度低下を解決することは難しく、認識アルゴリズムの改良が必要であることがわかった。

今後の課題としては、屋外での表情認識精度の向上のための認識アルゴリズムの改良、被験者を増やした屋外での表情認識精度の検証などが挙げられる。

謝辞 本研究はJSPS科研費21K18518の助成を受けたものです。

参考文献

- [1] 赤池勇磨, 金丸智史, 米田純, 久米由花, 荒川豊: 拡張現実技術によるコミュニケーション能力への影響, 研究報告コンシューマ・デバイス&システム(CDS), Vol. 2014, No. 5, pp. 1–8 (2014).
- [2] 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).
- [3] 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).
- [4] Mehrabian, A.: Nonverbal betrayal of feeling., *Journal of Experimental Research in Personality* (1971).
- [5] 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*, pp. 401–411 (2015).
- [6] Osawa, H.: Emotional cyborg: Complementing emotional labor with human-agent interaction technology, *Proceedings of the second international conference on Human-agent interaction*, pp. 51–57 (2014).
- [7] 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).
- [8] 櫻井翔, 中里直人, 吉田成朗, 鳴海拓志, 谷川智洋, 廣瀬通孝: 表情変形フィードバックによる遠隔協調作業における創造力向上支援, 日本バーチャルリアリティ学会論文誌, Vol. 20, No. 4, pp. 323–332 (2015).
- [9] Secord, P. F.: Facial features and inference processes in interpersonal perception, *Person perception and interpersonal behavior*, pp. 300–315 (1958).
- [10] Yamamoto, K., Kassai, K., Kuramoto, I. and Tsujino, Y.: Presenter Supporting System with Visual-Overlapped Positive Response on Audiences, *Advances in Affective and Pleasur-*

able Design, Springer, pp. 87–93 (2017).

- [11] 工藤義礎: REALITY 株式会社, 日本バーチャルリアリティ学会誌, Vol. 26, No. 1, pp. 28–29 (オンライン), DOI: 10.18974/jvrsj.26.1_28 (2021).
- [12] 原田伸一郎: バーチャル YouTuber の肖像権, 情報通信学会誌, Vol. 39, No. 1, pp. 1–11 (オンライン), DOI: 10.11430/jsicr.39.1_1 (2021).
- [13] 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).
- [14] 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).
- [15] 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).
- [16] 新田野乃華, 竹川佳成, 鈴木あゆみ, 徳田雄嵩, 鈴木克洋, 杉浦裕太, 正井克俊, 杉本麻樹, 平田圭二: 組み込み型光センサアレイを用いたフルフェイスマスクの表情認識精度の向上, 研究報告エンタテインメントコンピューティング(EC), Vol. 59, No. 45, pp. 1–7 (2018).