

# 反射型光センサアレイを用いた眼鏡型装置による 作り笑いと自然な笑いの識別

齊藤 千紗<sup>1,a)</sup> 正井 克俊<sup>2,1,b)</sup> 杉本 麻樹<sup>1,c)</sup>

受付日 2020年11月30日, 採録日 2021年7月7日

**概要:** 笑いは日常生活で頻繁に観察される表情の1つであり, 非言語コミュニケーションに不可欠な役割を果たす. 笑いには可笑しさや喜びといった快感情から生じる自然な笑いとは異なる作り笑いがある. この2つの笑いをコンピュータが適切に推測することができれば, ユーザへの理解が深まり, また, インタラクティブシステムに応用可能である. 本稿では, 日常の使用に適した形状である眼鏡型の装置に搭載した反射型光センサアレイを用いて, 2種類の笑いの識別可能性を検証する. 実験では, 12人の参加者が動画を視聴して生じた自然な笑いとは異なる作り笑いの2種類の笑いのデータを収集した. センサから得られた反射強度の分布である幾何学的特徴と時間軸の時間的特徴に対してサポートベクタマシンを適用した結果, ユーザ依存の学習の場合, 12人の実験参加者で平均精度が94.6%であった. これはデータを収集した際の表情変化の動画から人間が判定した場合(90.2%)よりも高い精度であった. さらに, 畳み込みニューラルネットワークを用いた個人間の学習においても82.9%であった.

**キーワード:** 笑い識別, ウェアラブル, 光センサ

## Classification of Genuine and Posed Smiles by Photo-reflective Sensors Embedded with Smart Eyewear

CHISA SAITO<sup>1,a)</sup> KATSUTOSHI MASAI<sup>2,1,b)</sup> MAKI SUGIMOTO<sup>1,c)</sup>

Received: November 30, 2020, Accepted: July 7, 2021

**Abstract:** Smile is one of the representative facial expressions which is observed frequently in daily life and essential for various non-verbal communications. People make genuine smiles and posed ones. If computers can recognize properly whether a person's smile is genuine or posed, an interactive system can understand the inner feelings of the user from their smile or we can use it as an input system by smile. In this study, we propose a smile classification system with a smart eyewear that equips photo-reflective sensors and examine whether we can distinguish two types of smiles; genuine smiles caused by funny videos and posed smiles evoked by instructions. We extract geometric features: reflection intensity distribution of sensors and temporal features in the time axis. By applying Support Vector Machine, we observed 94.6% as the mean accuracy among 12 participants with user-dependent training. The accuracy was higher than human judgement with videos (90.2%). Also, the accuracy when using Convolutional Neural Network with user-independent training was 82.9%.

**Keywords:** smile classification, wearable device, optical sensor

<sup>1</sup> 慶應義塾大学大学院理工学研究科  
Graduate School of Science and Technology, Keio University  
Yokohama, Kanagawa 223-8522, Japan

<sup>2</sup> NTTコミュニケーション科学基礎研究所  
NTT Communication Science Laboratories  
Atsugi, Kanagawa 243-0198, Japan

a) c.saito@imlab.ics.keio.ac.jp

b) katsutoshi.masai.fa@hco.ntt.co.jp

c) sugimoto@ics.keio.ac.jp

### 1. 背景

対人コミュニケーションにおいて, 表情, 声のトーンや身振りなどの非言語情報は重要な役割を果たす [1]. 表情は, 自己の内面の表出や他者への表現として作用し, 感情や意図などの非言語情報を無意識的または意識的に伝達する [2]. 特に笑いは日常で頻繁に観察される表情の1つで

あり、一般的に喜び、楽しさ、面白さなどの快感情にともなって表出する。一方、写真撮影の際に作る笑いや、社会的文脈における他の感情の隠蔽、雰囲気操作などの調整の手段としての笑い [3] など、快感情にかかわらず人は笑う場面がある。本稿では、前者の快感情がともなう笑いを自然な笑いとして定義し、後者の快感情をともなわない笑いを自然な笑いとは異なる笑い、つまり作り笑いとして定義する。先行研究 [4] によると、人は自然な笑いを作り笑いを知覚できることが示されている。

Krumhuber ら [5] は模擬採用面接実験において実験参加者の自然な笑い、作り笑い、無表情の3つの表情が面接官に与える印象や評価、意思決定に与える影響を調査した。その結果、自然な笑いは作り笑いとは無表情に比べて面接官からより高い評価を得た。自然な笑いは相手に好意的にとらえられるため、自分にとって有利な判断や決断を相手に促すことを指摘している。

コンピュータが快感情がともなわない作り笑いとは自然な笑いを検出できれば、このような人間の笑いの種別をふまえたシステムの振舞いが設計できると考えられる。また、自然な笑いとは作り笑いの識別技術は、表情を入力ジェスチャとしたインタラクティブシステムに活用できる。たとえば、車椅子の制御を行うために笑顔を入力の一つとして扱うシステムが提案されているが [6]、ユーザの作り笑いとは快感情をともなう自然な笑いをシステム上で識別することで、コマンドとしての笑顔と自然な笑いを区分して取り扱うことができると考えられる。

自然な笑いとは作り笑いの識別技術は環境設置型のカメラを用いたコンピュータビジョンや装着型の筋電センサを用いて実現されている [7], [8]。しかし、これらの手法では、計測範囲が装置の撮像範囲に限定されてしまったり、センサの肌への直接的な接触による不快感があったりするなどの制約をともなう。装着型かつ非接触のセンサを用いることで、こうした制約を緩和することが可能であると考えられる。

本稿では、装着型かつ非接触のセンサとして反射型光センサアレイを組み込んだ眼鏡型装置 [9] を用い、自然な笑いとは作り笑いの識別が可能かの初期的な検討として、快感情をともなう自然な笑いとしての動画刺激によって誘発した笑いとは、快感情をともなわない作り笑いとしてのコンピュータからの指示によって意図的に作らせた作り笑いとは識別できるかについて解析を行う (図 1)。

使用する装置は装着型であり、非侵襲な計測が可能である。フレームに配置した反射型光センサアレイは、表情筋の動きを目の周りの皮膚変形から計測することができる。それぞれの条件におけるセンサデータを12人の参加者から取得し、このデータセットを用いて、サポートベクタマシン (Support Vector Machine, 以下 SVM) による識別可能性を検証する。また、実験中の笑顔を動画によって記録

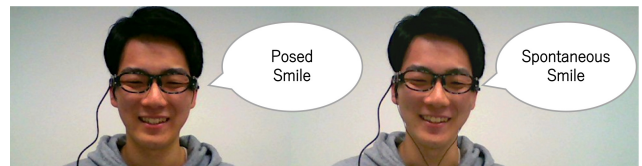


図 1 本稿の概要図。ユーザが装着している眼鏡型の装置を用いて自然な笑いとは作り笑いの識別を行う

Fig. 1 A user wears a glasses-shaped facial expression classification device with photo-reflective sensors. We tried to classify genuine and posed smiles using this device.

し、人間による識別精度と比較検討する。さらに畳み込みニューラルネットワーク (Convolutional Neural Network, 以下 CNN) を用いて個人間学習での識別を試みることで汎化性能を検証する。

## 2. 関連研究

### 2.1 画像を用いた笑いの識別手法

表情を分析する研究において最も一般的な手法はコンピュータビジョン技術を用いたものである。Dibeklioglu ら [7], [10] は、画像から自動で作り笑いとは自然な笑いを識別するシステムを開発した。識別には目、頬、口の端の動きの振幅、持続時間、速度、加速度が使用された。さらに、同時に400人のユーザの自然な笑いとは作り笑いを含むデータベース [11] を提示した。このデータベースはコンピュータビジョン技術を使用した笑いに関する研究で幅広く使用されている。Yang と Wang [12] は既存の笑いのデータベースを使用して、作り笑いとは自然な笑いの時間的パターンを表現したモデルを提案した。

画像情報から笑いを分析する研究 [13], [14], [15] によると、笑いは開始期、頂点期、収束期の3つの段階に分割することができ、特に開始期では表情変化における振幅と速度が異なることが示されている。Schmidt ら [13] は、自然な笑いの開始期が特徴的な顔の信号を構成することを示した。また作り笑いとは自然な笑いの開始期よりも短い期間で大きく速く変化することを指摘した [16]。Cohn ら [14] は、顔面上の特徴を追跡する技術を使用して自然な笑いおよび作り笑いの開始期の振幅と持続時間の関係を調査し、笑いの開始タイミングの違いから2種類の笑いを区別できることを示した。

このようにカメラを用いた手法は、笑顔の種類を分析する研究で多く利用されている。しかし、カメラはつねに顔をフレーム内に収めている必要があるが、遮蔽やユーザの頭部運動により、顔を捉えられず困難な場合がある。そのため、日常で生じる笑顔を識別するシステムには不向きである。またつねにカメラで撮影を行うことはプライバシーの観点からリスクがある。

## 2.2 筋電センサを用いた笑いの識別

カメラを用いた手法以外では、顔の筋肉の活動を計測できる筋電センサ (electromyograph, 以下 EMG) を利用した自然な笑いと作り笑いの識別が試みられている [8]. EMG を用いた装置はコンピュータビジョンの手法より高い精度で作成笑いと自然な笑いを区別できること [17] が示されている. EMG はウェアラブルであり, ユーザの動きやプライバシーに関する問題は発生しない. しかし, 電極を顔の肌表面に直接接触させる必要があり, 不快感がともなう可能性がある. また, センサの値は接触状態と皮膚の状態による導電性の変化に影響されてしまう.

## 2.3 反射型光センサを用いた表情識別

反射型光センサによる, 笑顔を含む表情識別の手法も提案されており, 眼鏡型装置にセンサを埋め込む形で検討されている. 眼鏡型装置は装着性に優れており, 日常の使用に適していると考えられる. Fukumoto らは微笑みと笑いを反射型光センサによって検出し, 日常で生じる笑いを活用したライフログシステムを提案した [18]. この手法では, ユーザや装着状態の変化に応じて閾値を定める必要がある. 一方, Masai らが開発した手法 [9] も反射型光センサアレイを搭載したウェアラブルな眼鏡型装置であるが, 複数のセンサデータに対して機械学習を適用することで, 笑顔を含む 8 つの基本的な表情を識別した. 教師あり学習を用いることで笑顔以外にも多様な表情を識別できることが示されている. Suzuki らは, 頭部装着型ディスプレイに反射型光センサを組み込むことによって, 装着者の表出強度を含めた表情をバーチャル環境のアバタに反映するシステム [19] を構築している. Asano らは, Masai らと同様の眼鏡型装置に加えて, 耳掛けの部分に取り付けた反射型光センサから取得したデータを用いて, 顔に貼り付けたマーカの 3 次元位置を推定した [20], [21]. 顔の形状が高精度で推定できたことから, 反射型光センサによって基本表情以外の複雑な表情を再構成できる可能性を示した.

これらの研究では, 笑いが自然な笑いか作り笑いかを識別することは対象としていない. 本研究では Masai らと同様の反射型光センサを搭載した眼鏡型装置を使用して, 自然な笑いと作り笑いの識別が可能かを検討し, その精度を評価する.

## 3. データ収集実験

反射型光センサを搭載した眼鏡型装置による笑いの識別可能性は Masai ら [9] によって確かめられている. そこで, 本実験では, 笑いが識別された際のその種類の識別可能性の検証を目的として, 自然な笑いと作り笑いのデータセットを作成する. 実験には 21~25 歳の 12 人の日本人学生が参加した. 実験は小さな会議室で参加者ごとに実施した. 実験に先立ち, 実験者は実験方法と実験で得られる顔画像な

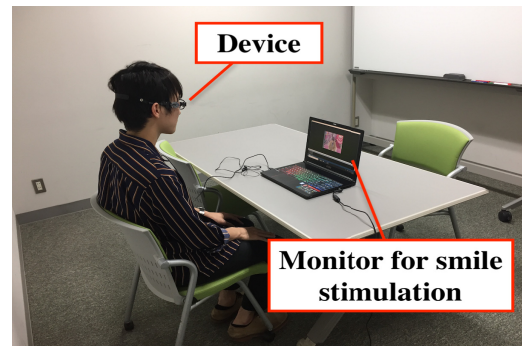


図 2 実験の外観

Fig. 2 Experiment environment.

どの個人情報の使用について説明し, 参加者に書面による同意書へ署名してもらった. この研究プロトコルは慶應義塾大学理工学部生命倫理委員会によって承認済みである.

実験は作り笑いのデータを集めるセッション, 自然な笑いのデータを集めるセッションの順番を 1 セットとし, 2 セット繰り返した. また, 実験全体の最初と最後に, これらのデータを補正するためのデータを収集するセッションを実施した. 図 2 に実験環境を示す. 実験中, 参加者は装置をつねに着用し, 外さないように指示された. 参加者の前にパソコンを設置し, 実験中の参加者への指示および動画の提示はこの画面上に表示した. 参加者の表情はパソコンの内蔵カメラによって撮影した. 画像とセンサ値は約 30 Hz で取得し, タイムスタンプによって同期させた.

### 3.1 笑いの計測に使用した眼鏡型装置

本研究では眼鏡型装置 [9] を使用する. この装置には IR (Infra Red) LED と IR フォトトランジスタで構成される 16 個の反射型光センサが取り付けられている. 表情筋の動きによって皮膚変形が生じると, 眼鏡フレームに取り付けたセンサと皮膚表面の間の距離が変化する. 各センサが計測する反射強度はこの距離によって変わり, このセンサデータから表情を識別する. センサは小型であるため, ユーザの外見を大きく損ねることはなく, 通常の眼鏡のように簡単に装着することが可能である.

図 3 は, 本研究で用いる眼鏡型装置と眼鏡のフレームに取り付けられた 16 個の反射型光センサの位置を示している. センサデータの A/D 変換後, 装置は取得したセンサ値をワイヤレス通信の ESP-WROOM-32 Bluetooth モジュールを介してコンピュータに送信する. シリアル通信を使用して, 毎秒 115,200 ビットで実験用のアプリケーションに接続される. アプリケーションでは約 30 Hz でセンサデータを取得する.

### 3.2 補正データ収集セッション

眼鏡フレーム上のセンサと顔の皮膚までの距離が近いほど, 反射型光センサで受光する反射光が増加し解像度は高

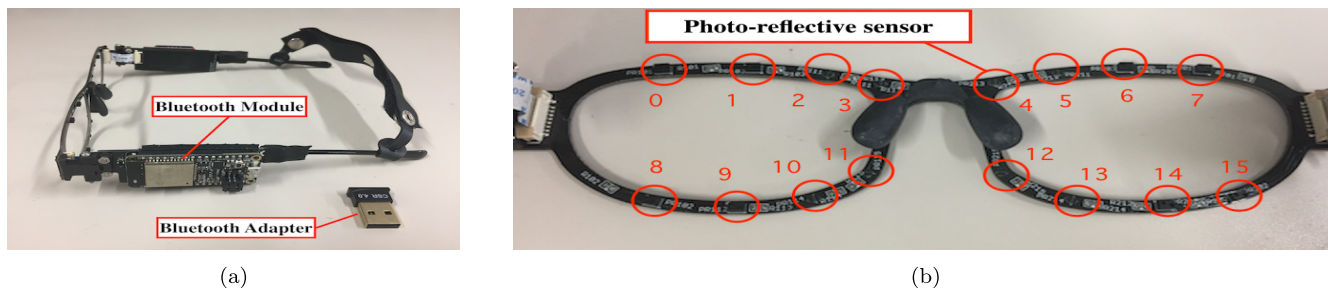


図 3 (a) 反射型光センサを搭載した装置, (b) フレーム上の 16 個のセンサの配置

Fig. 3 (a) Eyewear device (b) Sensor layout of the device. 16 photo-reflective sensors are placed on the eyewear flame.

くなる。センサ値は、対象の反射係数が一定であれば計測最小距離以上の範囲においては肌表面とセンサ間の距離の指数関数に反比例すると考えられ、ユーザの顔の形と顔のどこで測定されるかに応じて変化する。このセッションでは各参加者の各センサ値の変化の範囲を把握するために、参加者にできるだけ自由に顔を動かすように指示した。最初と最後の 2 回のセッションで得たデータの中で、各センサの最大値と最小値を取得した。

### 3.3 作り笑いとは自然な笑い

先行研究 [8], [11], [22] をふまえて、本研究においても作り笑いとは自然な笑いを誘発する手法を用いた。作り笑いは、快感情をとまなわぬ笑いであり、何かしらの意図をもって作られる。作り笑いを誘発するための指示には、楽しそうに笑うことを指示するもの [22] や自然な笑いを演じさせるようなもの [8] など自然な笑いに近づけようとさせる指示がある。本実験では、実験参加者が普段から作り慣れている作り笑いとして、写真撮影で見せるような笑顔を作るように指示した。写真撮影における笑顔は、一般の人々が日常において作る機会が多く、写真を見た人への印象をよく見せるという意図を持つ笑いであり、快感情をとまなわぬ笑いの一種である。

自然な笑いは快感情から自然と生じる笑いである。先行研究で使用されている手法 [8], [11], [22] にならぬ、本実験でも面白いと考えられる動画を視聴した際に表出した笑いを自然な笑いとした。

### 3.4 作り笑いのセッション

このセッションでは、参加者の目の前のパソコンのディスプレイに「Smile」という単語が表示されたときに、写真撮影で見せるような笑顔を作るように指示した。参加者には 10 秒の時間が与えられ、その時間内に笑顔を作り、任意のタイミングで通常表情に戻る。1 つのセッションにおいて、この試行を 10 回繰り返した。

### 3.5 自然な笑いのセッション

実験参加者は、セッションごとに 1 つの動画を視聴し、

2 セッションで日本のお笑い番組 (7 分) とコメディ番組 (13 分) の映像の 2 つを視聴した。参加者は映像に集中できるように、小さな会議室に 1 人となり、周りには誰もいない状況下で視聴した。

### 3.6 笑いタイミングの切り出し

すべての実験が終了した後、実験者はそれぞれの笑いのセッションにおける笑いのタイミングに対応するセンサ値をまとめ、データセットを作成した。

作り笑いについては、コンピュータ上の表示による指示で参加者が笑い始めてから通常表情に戻ったタイミングまでのセンサ値を切り出した。自然な笑いについては、実験者が実験中に撮影された顔の画像を観察し、その期間を特定した。参加者の笑い始める直前 (通常表情) から笑いが終わるまでの期間を 1 つの笑いとして抽出した。用意したビデオごとに最初の 10 回分の笑いを切り出し、参加者ごとに 20 回分の自然な笑いのデータを収集した。

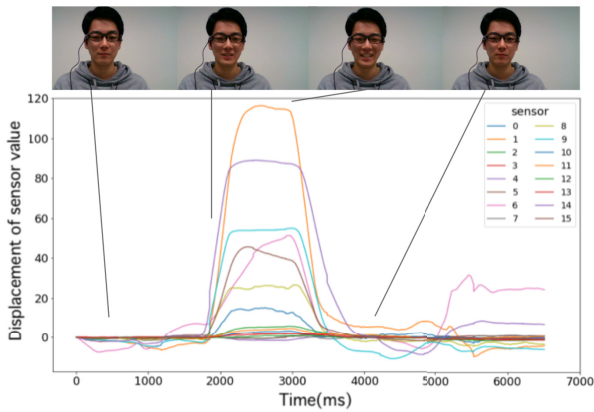
作り笑いのセッションと自然な笑いのセッションはそれぞれ 2 回ずつ繰り返したため、このデータセットには、12 人の参加者の 20 回分 (2 セッション × 1 セッションあたり 10 回) の作り笑いとは 20 回分の自然な笑いが含まれている。

## 4. 自然な笑いとは作り笑いの識別

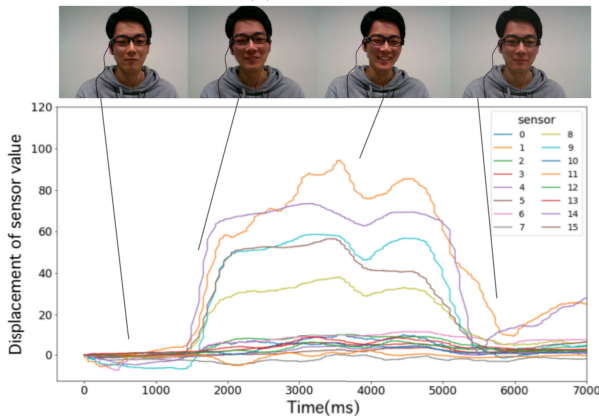
自然な笑いとは作り笑いの識別には、幾何学的特徴と時間的特徴の 2 種類の特徴量を使用した。それぞれの特徴量は、笑いが生じた際のセンサ値の変位とその変位にかかった時間から抽出したものである。これらの特徴量を使い、機械学習による検証 (SVM を用いたユーザ依存の学習による検証、個人間の学習による検証) と人間の視覚的判定による検証を行った。

### 4.1 開始期の特定

図 4 は実験で得た作り笑いとは自然な笑いにおけるセンサ値の一例を示している。笑いの時系列には開始期、頂点期および収束期の 3 つの段階が存在する。開始期とは、笑いが生じ始めてから最も大きく笑うまでである。作り笑いのセッションでは参加者が自分の任意のタイミングで笑いを



(a) 作り笑い



(b) 自然な笑い

図 4 実験で取得したセンサ値の例. 分析で用いる特徴量はセクション 4 で記述した手順により抽出した

Fig. 4 This figure shows how 16 sensor values change in (a) posed smile and (b) genuine smile in one participant and pictures of facial expressions captured at each time frame. The features were extracted by the process described in section 4.

やめることができるが、それぞれの教示の長さが 10 秒であることがそのタイミングに影響を与えていたことが観測された。開始期以降である、頂点期と収束期は教示の長さの影響を受ける可能性があったため、本研究ではその影響を除外するために開始期のデータのみを分析対象とした。

まず、補正データ収集セッションで取得した各センサの最大値と最小値を用いて開始期の各センサデータが 0 から 1 の範囲になるように正規化した (セクション 3.2)。正規化後、反射型光センサへのノイズの影響を減らすため、ローパスフィルタを適用した。具体的には、前 10 サンプルのセンサ値の移動平均を使用した。

次に、主成分分析を 16 個のセンサデータに対して実行し、時系列方向の第 1 主成分となるデータを求めた。そして、このデータを最小二乗法により 20 次の多項式に近似し、得られた近似曲線を微分して、極大値と極小値を計算した。近似曲線の決定係数はすべての参加者で平均 0.9 以上であった。第 1 主成分の各極小値からそれに隣り合う

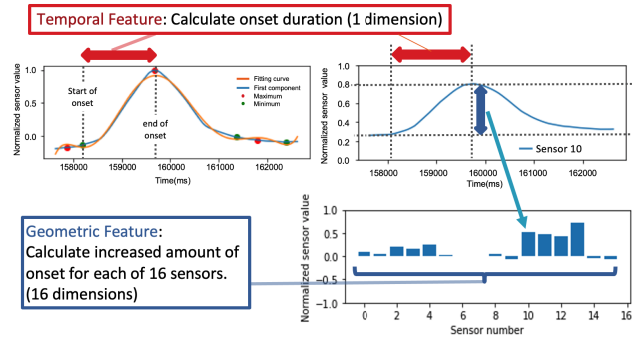


図 5 開始期の特定. 16 個のセンサ値の第 1 主成分の時系列データに対する近似式の極小値と極大値から定義した. 時間的特徴は開始期の長さを示す 1 次元の特徴量であり, 幾何学的特徴はその開始期間における笑い開始からのそれぞれのセンサ値の変動を示す 16 次元の特徴量である

Fig. 5 Onset phase detection. This is an example of a graph representing the first principal component of 16 sensor values, its fitting curve, and the local minimum and local maximum point that were automatically detected. We defined the largest change in the graph as the onset. The one-dimensional temporal feature indicates the length of the onset period, and the 16-dimensional geometric feature indicates the displacement of each sensor value from the start of smiles in that onset period.

極大値までの増加量を計算し、増加量が最大になる期間を開始期とした。この開始期は、笑い始め（センサ値が増加し始めるタイミングとしての極小値）から最も大きく笑う瞬間（センサ値の立ち上がり終了するタイミングとしての極大値）までの期間を反映していると考えられる。図 5 に、16 個のセンサ値の第 1 主成分とその近似グラフ、極値の検出結果とその結果から特定した開始期の位置および各特徴量の概要の例を示す。

## 4.2 特徴抽出

4.1 節で求めた開始期の 16 個のセンサの各値の変位を幾何学的特徴として定義した (図 5 右下)。各センサについて立ち上がり開始のセンサ値を立ち上がり終了のセンサ値から差し引いた。したがって幾何学的特徴は 16 次元のデータとして構成される。図 6(a) は、16 次元データを 3 次元に次元圧縮した後の各参加者の幾何学的特徴データの分布を示している。

開始期の時間的特徴を取得するために、4.1 節の手順から指定された開始期の期間を時間的特徴として定義した (図 5 左上)。図 6(b) は、幾何学的特徴と時間的特徴を組み合わせた合計 17 次元のデータを 3 次元に次元圧縮したデータの分布を示している。それぞれの参加者について、すべての笑いの時間間隔を収集したデータセットは、全体で平均が 0、分散が 1 になるように正規化した。図 7 は参加者ごとの作り笑いとは自然な笑いの時間的特徴量の平均と標準偏差を示している。

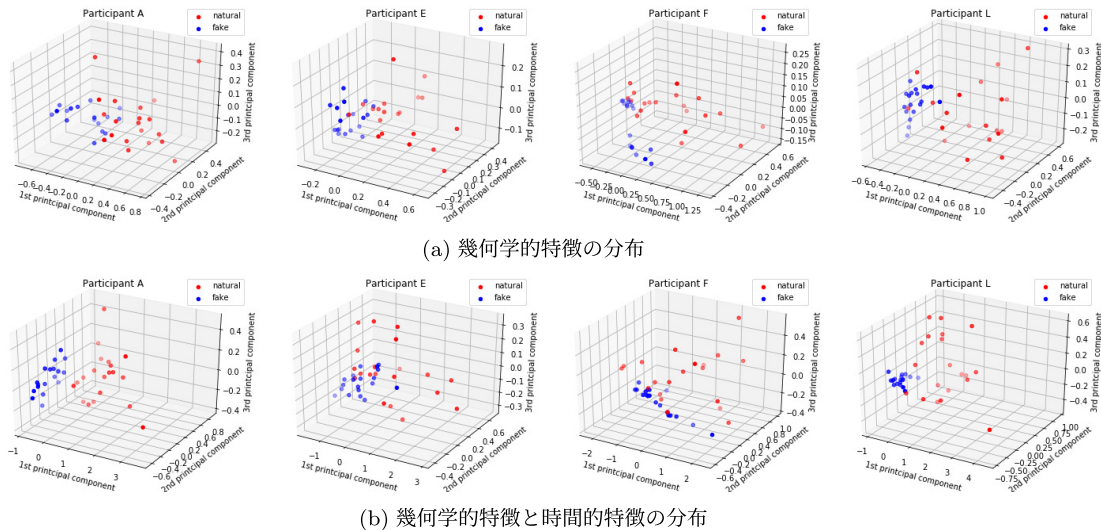


図 6 特徴量の三次元分布

Fig. 6 The results of a 3D plot using up to the third principal component of (a) the data of displacement of 16 sensors in onset phase and (b) the 17 dimensional data, which includes displacement of 16 sensors and the duration in onset phase.

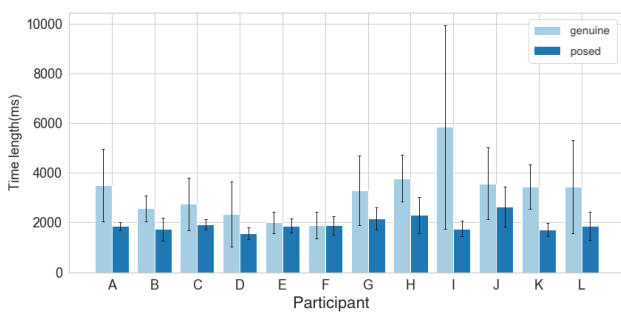


図 7 参加者ごとの時間的特徴量. グラフは計 20 回分の笑いから取得した値の平均値を示している

Fig. 7 The graph shows the average value of 20 (a) genuine and (b) posed smiles. The value of the first principal component was used for classification.

### 4.3 検証方法

#### 4.3.1 機械学習（ユーザ依存の学習）による検証

これらの幾何学的特徴と時間的特徴を用いて、教師あり機械学習法である SVM による識別を試みた。放射基底関数カーネル (Radial basis function kernel, 以下 RBF カーネル) を使用し、参加者ごとに層化 5 分割交差検証を用いて学習および評価をした。テストデータは作り笑いと自然な笑い、それぞれ数が一致するようにランダムに選択した。5 分割交差検証におけるハイパーパラメータはそれぞれの訓練データに合わせて最適化した。具体的には、訓練データのうち 3 分の 2 をグリットサーチ (Python の scikit-learn) 用の訓練データとして使用し、残りの 3 分の 1 を識別した際に最も精度の高かったハイパーパラメータに決定した。 ( $C, \gamma = 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^2, 10^3$ )。このパラメータを用いて、訓練データ全体を学習した識別器によってそれぞれのテストデータを評価した。

#### 4.3.2 機械学習（個人間の学習）による検証

続いて訓練データとして別の実験参加者のデータを使用した識別を行った。11 人分のデータを訓練データとし、残り 1 人分のデータをテストデータとする試行をすべての実験参加者に対して行った。カーネルは RBF カーネルを用いて、ハイパーパラメータはユーザ依存の識別と同様に、グリットサーチによって決定した。ただし、グリットサーチのハイパーパラメータ検証用データは訓練データの 4 分の 1 とした。SVM を用いて、幾何学的特徴量のみ、または幾何学的特徴量と時間的特徴量の両方を入力として作り笑いと自然な笑いそれぞれ 20 回分のテストデータを識別した。

#### 4.3.3 人間の視覚的判定による検証

反射型光センサを用いた識別に使用した笑いの開始期における人の笑いが、人から見て識別ができるのかを検証した。実験では、12 人から作り笑いと自然な笑いを 20 回分ずつ顔表情の画像を取得しており静止画データと動画データをそれぞれ 480 個作成した。静止画データには、センサ値から特定した開始期の終了時点に取得されたものを利用した。動画データには、開始期全体で取得した連続する顔写真データを約 30 fps で再構成したものを利用した。動画データには音声は含まれていない。これらはデータ収集実験において得られたものを利用して、機械学習のテストで用いたシーンにおける動画データである。

13 人 (男性 11 人) が判定者となり、480 個のデータが作り笑いか自然な笑いかを判定した。全判定者は静止画の判定実験を先に行い、静止画の実験の影響を回避するため 1 週間以上の期間を開けて動画の判定実験を行った。

表 1 自然な笑いと作り笑いの識別結果

Table 1 Classification accuracy of genuine smiles and posed ones.

参加者	ユーザ依存の学習による識別率		個人間学習による識別率		人間による判定の識別率		性別
	幾何学的特徴	幾何学的・時間的特徴	幾何学的特徴	幾何学的・時間的特徴	静止画	動画	
A	0.975	1.000	0.050	0.475	0.663	0.825	Male
B	0.900	0.925	0.500	0.800	0.746	0.873	Female
C	1.000	1.000	0.550	0.675	0.602	0.937	Male
D	0.950	0.975	0.500	0.575	0.848	0.965	Male
E	0.950	0.925	0.375	0.425	0.696	0.885	Male
F	0.900	0.850	0.700	0.650	0.683	0.935	Female
G	0.925	0.925	0.375	0.400	0.627	0.954	Male
H	0.950	0.950	0.575	0.750	0.685	0.896	Male
I	0.950	0.975	0.800	0.825	0.777	0.940	Female
J	0.900	0.900	0.650	0.650	0.738	0.858	Male
K	0.975	0.975	0.500	0.850	0.704	0.860	Female
L	0.875	0.950	0.500	0.675	0.813	0.908	Male
平均	0.938	0.946	0.506	0.646	0.715	0.903	-

4.4 結果

表 1 に自然な笑いと作り笑いの識別結果について前セクションの 3 種の検証方法による結果を示した。

ユーザ依存の学習において、幾何学的特徴のみを使用した場合、参加者全体の平均精度は 93.8% (標準偏差 = 3.8%) であった。一方、幾何学のおよび時間的特徴の両方を入力とした場合は 94.6% (標準偏差 = 4.4%) であった。16 次元の幾何学的特徴のみを考慮する場合と比較して、時間的特徴の追加により、識別精度がわずかに改善された。E と F を除くすべての参加者において、幾何学的特徴と時間的特徴の両方を使用した場合の識別精度は、幾何学的特徴のみを使用した場合よりも高い結果が得られた。参加者 L では時間的特徴を追加すると、精度が 7.5% 向上した。図 6 より、時間的特徴を追加することで参加者 L の笑いの分布の密度が高くなっていることが確認できる。また精度が 2.5% 向上した参加者 A では、幾何学的な特徴のみを使用する場合と比較して、両方の特徴を使用する場合は、自然な笑いと作り笑いの 2 つの分布を明確に区別することができる。例外として、参加者 E および F では時間的特徴を使用すると識別精度がそれぞれ 2.5% および 5.0% 低下した。識別率が低下した参加者 F の識別結果を分析すると、20 回分の自然な笑いのうち 5 つが作り笑いとして識別されていた。また図 6 では、時間的特徴が追加されたことによって自然な笑いの分布が作り笑いとは重なっていることが分かる。E と F では開始期の期間が自然な笑いと作り笑いで近い値であった。このため時間的特徴を加えた条件において、自然な笑いと作り笑いの識別精度の低下が生じた可能性がある。

個人間学習では、参加者全体の平均精度は幾何学的特徴のみを用いた場合、50.6% とほぼランダムチャンスと同等の結果になった。また、幾何学的特徴と時間的特徴の両方を使用した場合 64.6% であった。実験参加者 B, I, K で

は 80% 以上の識別率であったが、多くの参加者の識別率は 70% 以下となった。最も識別率の低かった実験参加者 G では 20 個中 19 個の作り笑いが自然な笑いと判別された。反対に、識別率が 47.5% の A の場合は 14 個の自然な笑いが作り笑いに見なされた。

人間による判定率は、静止画を用いた場合、実験参加者全体で平均 71.5% (標準偏差 = 6.97%)、動画を用いた場合 90.2% (標準偏差 = 4.21%) となった。静止画を用いた判定では、実験参加者 C と G が最も判定率が低く、それぞれ 60.2%, 62.7% であった。どちらも作り笑いが自然な笑いと見なされ、判定率が低くなった。動画を用いた場合ではそれぞれ 93.6%, 95.4% と高くなった。一方、実験参加者 A の判定率は静止画の場合 66.3%、動画の場合 82.5% と動画の判定率が参加者全体で最も低くなった。A の場合は自然な笑いが作り笑いに見なされたことが要因である。

4.5 考察

すべての参加者のユーザ依存の学習による平均精度は、幾何学的特徴を用いた場合と幾何学のおよび時間的特徴を含むデータセットを用いた場合の両方で 90% 以上であった。これらの結果は人間が動画を見て判定した場合の精度とほぼ同等かそれ以上であった。このことから、提案システムの反射型光センサを用いた手法は表情の幾何学的な違いと笑い表出時の時間的な違いにより、自然な笑いと作り笑いを高い精度で識別できると推測できる。各笑いでの時間的特徴量を示した図 7 にみられるように、開始期の期間 (時間的特徴) は自然な笑いよりも作り笑いの方が短くなる傾向があった。つまり作り笑いを表出する際の表情変化の速度が自然な笑いよりも速いということが分かる。この傾向が個人間学習において、重要な特徴量となっていた。また、センサデータの変化量の大きさよりも変化量の分散の大きさが自然な笑いと作り笑いで異なり、自然な笑いの

方がより分散が大きくなっていったことから変化量の分散を特徴量として扱える手法が個人間学習での精度を向上させるうえで必要になることが分かった。

図 6 が示すように、すべての参加者の作り笑いは自然な笑いよりも 1 点に集中して分布していた。自然な笑いの場合、参加者の笑いの大きさと持続時間は笑いを引き起こす刺激の程度と長さに依存するが、作り笑いは単一の指示によって作られたため表出の仕方にはばらつきが少なかった。この作り笑いは、単純な指示のみによるもので再現性が高く、自然に笑った場合と判別可能であったと考えられる。

実験中のカメラ画像を確認すると、作り笑いはカメラ目線になっている場合があり、自然な笑いはカメラ目線にならないことが一般的であった。指示によって、条件間で定常的な視線方向に一定の差異が生じたことが考えられる。一方で、本研究では、幾何学的特徴、時間的特徴の算出において、センサ値の笑い始めた時点からの変化量を使用しているため、定常的な視線方向の影響を除外したデータにおいて、笑いの識別精度を検証できていると考えられる。

### 5. 深層学習を用いた個人間学習での識別

SVM での個人間学習では、参加者全体の平均精度は 64.6%であった。識別精度が良好ではなかった理由として、幾何学的特徴、時間的特徴では、個人間での識別に必要な情報量が不足していた可能性が考えられる。そこで、4.5 節で言及したように、時系列情報（時系列で観測したセンサの変化の様子や速度情報）を考慮した手法として CNN による識別を追加検討した。開始期の最初から各センサ値を任意のフレーム幅で切り取り、二次元配列に整形する (input1)。また、30 Hz で取得した各センサデータの前フレームとの差分を計算したデータ（速度情報）も同様にして切り取る (input2)。この 2 つの入力を用いて識別を行った。図 8 に使用した CNN の構造を示す。それぞれの入力データは (センサ个数 : 16) × (フレーム幅 N : 70, 80, 90, 100) の次元を持った二次元配列のデータである。畳み込み層のフィルタ数は順に 32, 64, 128, 256 で畳み込みは input1, input2 それぞれ各センサごとに行い、フィルタサイズは時系列方向に 3, スライドは 1 である。活性化関数には ReLU を用いた。最大プーリング層のプーリングサイズは 2, スライドは 2 である。畳み込み層ではバッチ正規化を行った。また過学習を抑制するため EarlyStopping を使用した。Epoch 数は最大 200 に設定した。CNN の実装には python の keras を用いた。

これまでの実験と同様に、11 人分のデータを訓練データとし、残り 1 人分のデータをテストデータとする精度検証をすべての実験参加者のデータに対して行った。個人間で学習した CNN の訓練データの時系列の範囲フレームごとの平均識別率を表 2 に示す。また、実験参加者ごとの識別率を図 9 に示す。情報量が最も多い 100 フレームの時系

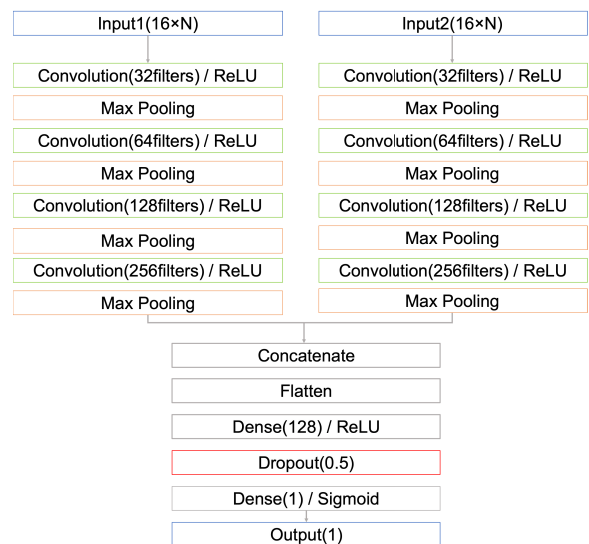


図 8 使用した CNN のネットワーク構造  
Fig. 8 CNN architecture.

表 2 フレーム数ごとの CNN による個人間学習の平均識別率  
Table 2 Classification accuracy of user-dependent training with CNN by frames.

フレーム数	平均識別率 [%]	標準偏差 [%]
70	81.8	13.5
80	82.7	13.4
90	82.7	13.9
100	82.9	15.4

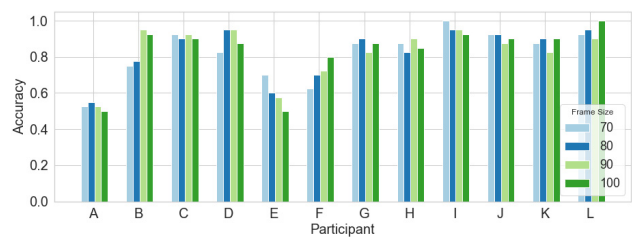


図 9 CNN を用いた個人間識別の結果  
Fig. 9 Classification accuracy of user-dependent training with CNN.

列データを用いたときに最も識別率が高くなり、実験参加者 12 人中 10 人は 80%以上の識別率であった。実験参加者 L ではすべての笑いが正しく識別できた。

SVM で用いた幾何学的特徴であるセンサ値の変化量では個人間の学習がうまくいかなかったのに対し CNN を用いることで大きく精度が改善された。このことから、それぞれの笑いにおけるセンサ値の変化の仕方に個人差を吸収する特徴量があったと考えられる。先行研究においては、作り笑いと自然な笑いには顔の各パーツが動き出す順番や変化の速度などの動的な情報に違いが確認されており [13], [14], 時系列のセンサデータがこれらの情報を含んでいたため識別できたと考えられる。

一方、SVM を用いたときと同様に、実験参加者 A と E



の識別率はおよそ 50% と他人のデータを用いた識別はできなかった。実験参加者 A と E はどちらも大半の作り笑いが自然な笑いと判別されており (A は 20 個すべて, E は 18 個), この原因の 1 つとして A と E の笑い時の表情変化が他の実験参加者に比べて小さかったことが考えられる。そのためセンサ値の変化量も小さくなり, 変化の仕方の特徴量の違いが CNN でとらえにくくなったと考えられる。こうした実験参加者の笑いを高い精度で識別するためには, ユーザごとに学習した識別器を用いるか, 類似した表情特徴を持つデータによって学習した識別器を用いることが必要になることが示唆される。

## 6. 今後の課題

本研究で用いた眼鏡型装置はウェアラブルであるため, カメラのように人の顔の動きに合わせて装置の位置をつねに調整する必要はなく, 日常生活のコミュニケーションで生じる笑いを分析するのに適している。その最初のステップとして実験室環境において作り笑いと自然な笑いを識別対象として実験した。本研究で識別対象とした作り笑いは, 意図的な作り笑いの 1 つである写真撮影での笑いを指定した。この笑いは, 多くの実験参加者が作り慣れている笑顔であり作り笑いと自然な笑いが提案手法によって区別できるかを検証する手始めとしてこのような取り組みをした。人が自然に笑っているかのように演技した作り笑いと, 自然な笑いの識別が可能であるかについての検証は今後の課題である。また, 笑いは多様なニュアンスを含み, 他人とのコミュニケーション中に発生する社交的な笑いなども考えられる。これらの笑いの識別において, 本装置を用いた手法の日常での有効性を検証するのは重要な課題である。

実験の順序として, 自然な笑いを誘発する動画刺激より先に, 作り笑いを作る指示から始めた。この順序は, 動画視聴中に参加者が自分の笑いを測定されていること, そして笑うことを期待されていることを認識させた可能性がある。つまり, 動画視聴中にも作り笑いを誘発させた懸念がある。この場合, 自然な笑いが作り笑いに近くなることで識別が困難になることが予測される。ユーザ依存の学習による識別, 人による動画での判定の識別いずれも 90% 以上となったことからその影響は少ないと考えられるが, 順序効果に対する影響を検討する必要がある。

本研究で識別精度が高かったユーザ依存型のモデルでは, 事前に各ユーザの訓練データを収集する必要がある。また, 学習には, 自然な笑いと作り笑いの両方のセンサデータが必要である。実験ではビデオを視聴させた後, 実験者による笑いタイミングを切り出したが, 実際には自然な笑いのデータおよびそのアノテーションを事前に収集することは難しい。この問題の解決策として個人間で識別モデルを共有する方法がある。CNN を用いた個人間学習のモデルでも 80% 以上の識別率であった。現状の識別率では, 特

定のコンテキストのみのアプリケーション, たとえば, 電話がかかってきた際に作り笑いによって電話を受けるなどが考えられるが, より実用的なモデル, アプリケーションを検討することも課題である。

ウェアラブル装置に取り付けられた反射型光センサは眼鏡のフレームと顔の皮膚表面との間の反射強度を測定しているが, そのセンサ値は, 距離と皮膚表面の反射係数に依存し, ユーザの顔の形状や眼鏡型装置の着用位置によって異なる。したがってユーザが装置を 1 度外した場合, センサ値が変化する。本研究で使用した特徴量はセンサ直の変位を利用しているため, 再装着の影響は Masai ら [9] と比較すると受けにくいと考えられるが, その影響を評価することも今後の課題である。

実験中, 一部の参加者は動画を視聴しながら笑っている最中に上半身と頭を大きく動かす傾向があった。大きな動きで眼鏡の位置がずれると, センサの値に影響し, 実験における識別率低下の要因の 1 つとなった可能性がある。一方で, この頭部の動きを加速度センサなどで検出し, 特徴量に加味することで, 識別率が向上するか検討したい。

## 7. 結論

本研究では, 16 個の反射型光センサを搭載した眼鏡型の表情識別装置を使用して自然な笑いと作り笑いの識別を試みた。2 種類の笑いに対応するセンサデータの収集実験を行い, 開始期のセンサ値の変位を幾何学的特徴, 開始期の期間を時間的特徴として定義し特徴抽出を行った。SVM による識別の結果, ユーザ依存の学習の場合, 12 人の参加者の平均精度は幾何学的特徴のみを使用した場合は 93.8%, 幾何学的特徴と時間的特徴の両方を使用した場合は 94.6% であった。これらの結果は人間が動画を見て判定した場合の精度 (90.2%) 以上であった。また, CNN を用いた個人間の学習では平均 82.9% の精度であった。

謝辞 本研究は JST CREST JPMJCR14E1, 科学研究費補助金 (16H05870) の支援を受けたものである。

## 参考文献

- [1] Mehrabian, A. et al.: *Silent messages*, Vol.8, No.152, Wadsworth Belmont, CA (1971).
- [2] Keltner, D., Ekman, P., Gonzaga, G.C. and Beer, J.: *Facial expression of emotion* (2003).
- [3] Scott, S.K., Lavan, N., Chen, S. and McGettigan, C.: *The social life of laughter*, *Trends in Cognitive Sciences*, Vol.18, No.12, pp.618–620 (2014).
- [4] Miles, L. and Johnston, L.: *Detecting happiness: Perceiver sensitivity to enjoyment and non-enjoyment smiles*, *Journal of Nonverbal Behavior*, Vol.31, No.4, pp.259–275 (2007).
- [5] Krumhuber, E., Manstead, A.S., Cosker, D., Marshall, D. and Rosin, P.L.: *Effects of dynamic attributes of smiles in human and synthetic faces: A simulated job interview setting*, *Journal of Nonverbal Behavior*, Vol.33, No.1, pp.1–15 (2009).

- [6] Pinheiro, P.G., Pinheiro, C.G. and Cardozo, E.: The Wheelie—A facial expression controlled wheelchair using 3D technology, *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pp.271–276, IEEE (2017).
- [7] Dibeklioglu, H., Salah, A.A. and Gevers, T.: Recognition of genuine smiles, *IEEE Trans. Multimedia*, Vol.17, No.3, pp.279–294 (2015).
- [8] Perusquía-Hernández, M., Hirokawa, M. and Suzuki, K.: A wearable device for fast and subtle spontaneous smile recognition, *IEEE Trans. Affective Computing*, Vol.8, No.4, pp.522–533 (2017).
- [9] 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, *Proc. 21st International Conference on Intelligent User Interfaces*, pp.317–326, ACM (2016).
- [10] Dibeklioglu, H., Valenti, R., Salah, A.A. and Gevers, T.: Eyes do not lie: Spontaneous versus posed smiles, *Proc. 18th ACM International Conference on Multimedia*, pp.703–706, ACM (2010).
- [11] Dibeklioglu, H., Salah, A.A. and Gevers, T.: Are you really smiling at me? Spontaneous versus posed enjoyment smiles, *European Conference on Computer Vision*, pp.525–538, Springer (2012).
- [12] Yang, J. and Wang, S.: Capturing Spatial and Temporal Patterns for Distinguishing between Posed and Spontaneous Expressions, *Proc. 25th ACM International Conference on Multimedia, MM '17*, pp.469–477, Association for Computing Machinery (online), DOI: 10.1145/3123266.3123350 (2017).
- [13] Schmidt, K.L., Cohn, J.F. and Tian, Y.: Signal characteristics of spontaneous facial expressions: Automatic movement in solitary and social smiles, *Biological Psychology*, Vol.65, No.1, pp.49–66 (2003).
- [14] Cohn, J.F. and Schmidt, K.: The timing of facial motion in posed and spontaneous smiles, *Active Media Technology*, pp.57–69, World Scientific (2003).
- [15] Guo, H., Zhang, X.-H., Liang, J. and Yan, W.-J.: The dynamic features of lip corners in genuine and posed smiles, *Frontiers in Psychology*, Vol.9, p.202 (2018).
- [16] Schmidt, K.L., Bhattacharya, S. and Denlinger, R.: Comparison of deliberate and spontaneous facial movement in smiles and eyebrow raises, *Journal of Nonverbal Behavior*, Vol.33, No.1, pp.35–45 (2009).
- [17] Perusquia-Hernández, M., Ayabe-Kanamura, S., Suzuki, K. and Kumano, S.: The Invisible Potential of Facial Electromyography: A Comparison of EMG and Computer Vision when Distinguishing Posed from Spontaneous Smiles, *Proc. 2019 CHI Conference on Human Factors in Computing Systems*, p.149, ACM (2019).
- [18] Fukumoto, K., Terada, T. and Tsukamoto, M.: A Smile/Laughter Recognition Mechanism for Smile-Based Life Logging, *Proc. 4th Augmented Human International Conference, AH '13*, pp.213–220, Association for Computing Machinery (online), DOI: 10.1145/2459236.2459273 (2013).
- [19] Suzuki, K., Nakamura, F., Otsuka, J., Masai, K., Itoh, Y., Sugiura, Y. and Sugimoto, M.: Recognition and mapping of facial expressions to avatar by embedded photo reflective sensors in head mounted display, *2017 IEEE Virtual Reality (VR)*, pp.177–185 (online), DOI: 10.1109/VR.2017.7892245 (2017).
- [20] Asano, N., Masai, K., Sugiura, Y. and Sugimoto, M.: Facial performance capture by embedded photo reflec-

ive sensors on a smart eyewear, *Proc. 27th International Conference on Artificial Reality and Telexistence and 22nd Eurographics Symposium on Virtual Environments*, pp.21–28, Eurographics Association (2017).

- [21] Asano, N., Masai, K., Sugiura, Y. and Sugimoto, M.: 3D facial geometry analysis and estimation using embedded optical sensors on smart eyewear, *ACM SIGGRAPH 2018 Posters*, p.45, ACM (2018).

- [22] Petridis, S., Martinez, B. and Pantic, M.: The MAH-NOB laughter database, *Image and Vision Computing*, Vol.31, No.2, pp.186–202 (2013).



齊藤 千紗

2018年慶應義塾大学理工学部情報工学科卒業。2020年同大学院理工学研究科開放環境科学専攻前期博士課程修了。光センシングと機械学習を用いた笑いの識別に関する研究に従事。



正井 克俊

2018年慶應義塾大学大学院理工学研究科後期博士課程修了。博士(工学)。日本学術振興会特別研究員(DC2)、慶應義塾大学理工学部情報工学科助教等を経て、現在、日本電信電話株式会社NTTコミュニケーション科学基礎研究所リサーチアソシエイト。人の非言語動作をセンシングするウェアラブルインタフェースに関する研究に従事。



杉本 麻樹 (正会員)

2006年電気通信大学大学院電気通信学研究科博士後期課程機械制御工学専攻修了。博士(工学)。日本電信電話株式会社NTTコミュニケーション科学基礎研究所客員研究員、日本学術振興会特別研究員(DC1/PD)、慶應義塾大学大学院メディアデザイン研究科特別研究講師、慶應義塾大学理工学部情報工学科専任講師、准教授等を経て、現在、慶應義塾大学理工学部情報工学科教授。サイバーフィジカル空間におけるユビキタス光センシングに関する研究に従事。