

多人数遠隔コミュニケーションにおける 肯定的感情表出支援の効果と支援適用タイミングの決定

二瓶 芙巳雄^{1,a)} 田口 和佳奈^{1,b)} 中野 有紀子^{1,c)} 深澤 伸一^{2,d)} 赤津 裕子^{2,e)}

受付日 2020年4月17日, 採録日 2020年11月5日

概要: 本研究では, 遠隔ビデオコミュニケーションにおいて, 肯定的な表情や音声の表出を補助することによる議論活性化の効果に着目し, 次の3つの課題に取り組んだ; 1) 肯定的な感情を適切なタイミングで人工的に表出することによる議論の活性化効果の検証, 2) 肯定的な感情を表出すべきタイミングの推定手法の提案, 3) 推定されたタイミングの有用性の検証. はじめに肯定的な感情を人工的に表出するツールを使用する状況での議論を収録し, 収録したデータを分析することで, 利用者が適切なタイミングでツールを使用することによりコミュニケーションが活性化することを確認した. 次に音声・表情情報に基づき肯定的な感情を表出すべきタイミングを推定するモデルを提案し, 提案モデルが肯定的な感情を表出すべきタイミングを予測することにより会話の活性化が期待できることを議論した. 最後に肯定的な感情を表出すべきタイミングを可視化する推定機構を実装した.

キーワード: 肯定的な感情の表出, コミュニケーション支援, ビデオ会議, 深層学習

Effect of Computer-supported Positive Emotion Expression and Prediction of System Support Timings in Video-mediate Communication

FUMIO NIHEI^{1,a)} WAKANA TAGUCHI^{1,b)} YUKIKO I. NAKANO^{1,c)} SHINICHI FUKASAWA^{2,d)}
HIROKO AKATSU^{2,e)}

Received: April 17, 2020, Accepted: November 5, 2020

Abstract: Focusing on the effect of computer-supported positive facial and vocal expressions to activate discussions in remote video communication, this study addressed the following three issues; 1) investigating whether artificially exaggerating positive emotional expressions displayed at appropriate timing contribute to activating the discussion, 2) proposing a method for estimating the timing when positive emotions should be expressed, and 3) examining the usefulness of the estimated timing. First, we recorded discussions in which conversation participants used a tool that exaggerates positive emotions, and by analyzing the recorded data, we confirmed that the tool is useful for activating the communication if the tool is used at appropriate timing. Next, we proposed a model that estimates the timing that positive emotions need to be expressed based on speech and facial expression information, and discussed that the proposed model is expected to activate discussion by suggesting the timing to express positive emotions. Finally, we implemented an estimation mechanism that visualizes when positive emotions should be expressed.

Keywords: positive emotion expression, communication supporting, video mediated meeting, deep learning

¹ 成蹊大学
Seikei University, Musashino, Tokyo 180-8633, Japan

² 沖電気工業株式会社
Oki Electric Industry Co., Ltd., Warabi, Saitama 335-8510,
Japan

a) nihei.fumio@st.seikei.ac.jp

b) dm186205@cc.seikei.ac.jp

c) y.nakano@st.seikei.ac.jp

d) fukasawa856@oki.com

e) akatsu232@oki.com

1. はじめに

ビデオ通話やビデオ会議の推進と普及により、映像をとともなう遠隔コミュニケーションは私生活やビジネスを問わず利用されている。しかし遠隔コミュニケーションには自身や相手の様子を共有しにくいという課題がある。この課題の解決のため、視線の方向を補正し遠隔地の他者に提示する支援 [10] や、VR 空間上に対面状況を構築するような支援 [5] が提案されている。

遠隔コミュニケーション支援の多くは遠隔環境のため欠落した情報を補うことを目的とするが、コミュニケーション信号を強調することでコミュニケーションを円滑に調整する支援が必要な場合もある。たとえば、合意を得ることが難しい交渉場面や会話が停滞している場面では、肯定的な感情を表出させにくい。このような状況を解消するためには、肯定的な感情の表出をシステムにより強調するような補助が有効であり、このようなシステムの支援により膠着してしまった状況を活性化させることが期待できる。

肯定的な感情の表出方法として、声のトーンや表情を明るくすることが有効であると期待できる。いくつかの先行研究でも、システムによる非言語シグナルの強調（デフォメーション）が議論を活性化させる効果を有することを報告している。しかし多くのシステムは会話全体 [4], [21], [25], あるいは表情模倣のタイミング [23] でのみ非言語シグナルを強調するという、固定的あるいは限定的なタイミングで強調を適用している。したがって、強調を適用すべき適切なタイミングについての議論はいまだ十分にされておらず、また、非言語シグナルを適切なタイミングで強調したときに議論を活性化させることができるかについても十分に検証されていない。加えて実際の支援のためには、肯定的な感情を強調すべきタイミングを推定する手法も必要である。そこで本研究では、肯定的な感情を人工的に表出することにより遠隔コミュニケーションを活性化させることを狙い、次の3つの課題に取り組む；1) 利用者にとって適切なタイミングで肯定的な感情を人工的に表出することによる、コミュニケーションにおける効果の検証、2) 肯定的な感情を人工的に表出すべきタイミングの推定手法の提案、3) 推定されたタイミングの有用性の検証。

以降、本論文を次のように構成する。まず2章で本研究に関連する取り組みを概観し、本研究の位置づけを議論する。3章では肯定的な感情を人工的に表出するツールを使用したうえでの遠隔会話を収録した実験について説明し、4章において、利用者にとって適切なタイミングでそのツールを使用したときの遠隔会話における効果を分析する。5章では肯定的な感情を表出すべきタイミングを推定するモデルを機械学習により作成し、6章では推定モデルの特性や、推定モデルが議論を実際に支援する場合に期待できる効果を議論する。提案モデルの有用性を議論した

後、将来実施するユーザスタディに向け、支援適用タイミング推定機構のプロトタイプを7章で実装する。最後に8章において本研究を総括する。

2. 関連研究

2.1 コミュニケーション支援

遠隔コミュニケーション支援の研究では相手の様子が伝わりにくいという課題から、自身と相手の様子を正しく理解する、あるいは共有することを目的とすることが多い。映像をとともなわないテキストベースの遠隔コミュニケーションの支援では、感情を表現するアイコンであるエモティコン (emoticon) や利用者の感情を表出するキャラクターがテキストチャットと併用されることで、利用者はテキストのみのチャットよりも満足度が高く、また会話が盛り上がったと感じることが報告されている [9], [17]。一方で、ビデオ会議のような遠隔コミュニケーションの支援では、遠隔コミュニケーション特有の問題である、視線が分かりにくい状況を解決するために、視線方向を示すディスプレイ [14] や、画像処理により修正された視線方向を表出するシステム [10]、また VR 空間上に対面状況での会議を構築することで、遠隔でも対面会議と同様に議論できる環境を提案する取り組み [5] もある。

コミュニケーション支援には、以上の取り組みのような、自身や他者の様子を正確に伝えることに焦点を当てたもののほかに、コミュニケーション自体にポジティブな影響を与えることを目的とするものもある。対面でのグループコミュニケーションにおいて Schiavo らは、各人のディスプレイに注目すべき参加者を示すことによる微かなファシリテーションが、グループの活動をバランスさせることを明らかにした [19]。テーブル上に発話時間や視線行動といったグループの social dynamics を投影するシステムは、発話しすぎる参加者の発話を抑え、発話の少ない参加者に発話を促す効果がある [24]。Shimizu らは指向性スピーカによる音声により受動的な参加者の発話量が増えたことを報告した [20]。協調学習の研究では、沈黙を検知すると発光するアンビエントなインタフェースにより参加者の発話量が増加する効果が報告されている [8]。

以上の研究が提案したシステムは、対面コミュニケーションをポジティブに変化させることを報告している。ここで我々の研究に最も関連する研究として、櫻井らは、遠隔コミュニケーションにおける創造力に対して、表情が与える影響を検証した [21]。表情を笑顔や悲しい顔に変形するシステムを使用したユーザスタディの結果、笑顔での議論では通常表情や悲しい顔での議論と比較し、より多くのアイデアが創出されることを報告した。このシステムでは遠隔会議中、他者の表情を考慮せず、利用者の表情をつねに笑顔や悲しい顔に変形させている。その一方で人対人、また人対ヒューマノイドのコミュニケーションでは、

感情の模倣 (Emotion mimicry) のようなタイミングを考慮したコミュニケーションが重要であることが報告されている [3], [12], [15], [22]. Suzuki らはシステムによる笑顔の模倣の効果を調査するため、会話相手の笑顔を自動で検出し、対象者の表情を笑顔に変形させる FaceShare を提案した [23]. 主観評価の結果、FaceShare を使用した参加者と変形した表情を見た参加者の両者が、会話が盛り上がったと感じていたと報告している。

文献 [21] と [23] では二者間対話を分析対象としたが、我々は三者間対話を分析対象とする。多人数会話における肯定的な感情を表出すべきタイミングは二者間対話と異なり、Suzuki らが扱った表情模倣の場面だけでなく、次に誰が会話のターンをとるのが決まらない場面や、他参加者の様子を見て同意すべきかどうか決めかねている場面なども考えられる。我々が定義する肯定的な感情を表出すべきタイミングは、多人数会話特有の場面を網羅する点が Suzuki らの研究とは異なる。また文献 [21] と [23] では発話音声の変形については焦点を当てていない。参加者の声のトーンを変化させることが対立する状況でのコミュニケーションにおいて不安の解消に貢献する [4] ことや、支配性が偏った二者間会話における説得力に影響を与える [25] ことが報告されていることから、表情だけでなく、音声モダリティの変形も会議の活性化に対して有効であることが期待できる。さらに我々は、音声を変形させるタイミングを考慮したうえでのコミュニケーションへの影響を調査する点が文献 [4], [25] とは異なる。

以上の議論より本研究では、肯定的な感情の表出を支援するシステムによる遠隔コミュニケーションの活性化のため、まず利用者の所望のタイミングで表情と音声を肯定的に変容させるシステムを開発する。そしてそのシステムを利用者が使用することで、遠隔コミュニケーションが活性化することを分析により明らかにする。

2.2 感情認識

感情に基づくコミュニケーション支援技術を構築するうえで、感情認識技術は重要な基盤となる。

感情認識には機械学習アプローチがよく使用されるが、従来の機械学習アプローチでは音声情報や映像情報などの非言語特徴量が使用されている [27]. 具体的な音声特徴量はピッチやエネルギー、MFCC であり、映像特徴量では顔ランドマークの座標や頭部動作があげられる。深層学習アプローチによる感情認識においても従来の推定手法と同様に音声情報や映像情報 [11], [26] が入力データとして使用されている。加えてモダリティごとのネットワークの統合や、CNN や LSTM などの入力データの特徴を顕著にするレイヤの組合せが柔軟である深層学習ネットワークを構築することで、高性能な感情認識モデルが実現されている [11], [28].

我々が対象とする肯定的な感情を表出すべきタイミングの推定には、参加者の感情状態が大きく関与することから、感情認識と同様に音声情報や映像情報などの非言語情報が強く寄与することが予想できる。

2.3 マルチパーティインタラクション

我々が分析対象とする 3 名以上が参加する遠隔会議は、マルチパーティインタラクションとしての性質もあわせ持つ。マルチパーティインタラクションのモデル化としては、創発的リーダーシップ [18] や会話における優位性やリーダーシップ [1], [16], 会話参加者の機能的役割 [2], [13], また参加者の高い関心が集まった時間 [7] の推定があげられる。また深層学習を用いた会議における重要発言推定モデル [6] は深層学習のネットワーク形状の参考にできる。我々が対象とする遠隔コミュニケーションはマルチパーティな会議であることから、マルチパーティインタラクションの性質をとらえることができるネットワーク構造の深層学習モデルを検討する。

以上より本研究では、マルチパーティな遠隔会議において、肯定的な感情を表出すべきタイミングを非言語情報に基づき推定する深層学習モデルを提案する。

3. 遠隔会話収集実験

利用者が所望のタイミングで、肯定的な感情の表出を強調することによる効果を検証するためまず、肯定的な感情の表出を支援するツールを使用する状況での遠隔会話を収録する実験を実施した。この章では、我々が開発した会話参加者の顔映像と発話音声を肯定的にデフォルメーションさせるツールについて、またそのツールを使用した条件下での遠隔会話収集実験のデザインについて説明する。

3.1 感情表出調節ツール

会話参加者の表情と音声を明るく変容させることができる、感情表出調節ツールを作成した。会話参加者が任意のタイミングでキーボードの任意のキーを押すとツールがオンの状態となり、自身の表情と音声はデフォルメーションされる。デフォルメーションは 10 秒間持続する。デフォルメーションに気づいた他者がコミュニケーションを意識的に変化させることを避けるため、感情表出調整ツールのデフォルメーションは他者から気づかれにくい程度に調整^{*1}した。顔映像と音声のデフォルメーションの具体的な処理を以下に説明する。

顔映像 顔映像をデフォルメーションするソフトウェア^{*2}を基に開発したアプリケーションを使用し、実際の映像

^{*1} 予備実験においてデフォルメーションが他者から気づかれにくいように調整した。また本実験実施後の参加者へのインタビューにおいて、ツールによるデフォルメーションと参加者自身の感情表出を明確に区別できた参加者は稀であることも確認している。

^{*2} <https://www.auduno.com/clmtrackr/examples/facedeform>



図 1 感情表出調節ツール使用による表情の変化と実験環境
Fig. 1 Deforming facial image using the emotional deformation tool and experimental settings.

よりも口角を上げた笑顔の表情を作成した(図 1)。会話参加者が任意のキーを押すと、デフォルメーションが施された映像が他参加者のディスプレイに表示される。
音声 会話参加者が任意のキーを押すと、これをトリガに実験者がオーディオエフェクタ (Roland: AIRA VT-3 Voice Transformer) を用いて手でピッチを 20% 上昇させる操作を行った。他参加者にはデフォルメーションされた音声が表示される。

会話参加者の表情と音声を参加者自身が変化させるのではなく、感情表出調整ツールを使用し変化させる理由として、会話をしながら表情と音声を意識的に変化させることは、無意識的に変化させることよりも認知的な負荷が高いことがあげられる。また不自然な変化は相手に気づかれ、会話の雰囲気悪くする恐れもある。そこで我々は、他者から気づかれない程度の肯定的な感情の表出を支援することで会話の雰囲気の悪化を避けながらも、利用者自身が肯定的な表情を表出することの認知的な負荷を抑えるシステムを提案した。

3.2 タスクと被験者

遠隔会話を再現した環境下において、3名の参加者からなるグループが次の3つの議題について会話した；(1) コンビニの良いと思うメニュー、(2) 新宿駅の問題点、(3) アメリカ人をもてなすディナーメニュー。各議題についての会話時間は約 15 分間である。各議論後に、グループの意思決定として上位 3 つの観点について報告してもらった(たとえば、(1)であれば、コンビニメニューのベスト 3 を議論して決めてもらった)。

実験参加者は計 33 名(女性 18 名、男性 15 名)の社会人や大学生である。平均年齢は 28.7 歳(標準偏差: 7.22)であった。3 名で 1 グループを構成したため、計 11 グループの実験グループとなった。

3.3 手続き

実験環境 遠隔コミュニケーション環境を構築するため、実験室内に 3 名分の小部屋を設営した。各部屋に参加

者を 1 名ずつ割り当て、参加者が互いの声や表情を直接観測できないようにした。各参加者の前には、28 インチと 11.6 インチの大小 2 つのディスプレイが設置された(図 1)。大きいディスプレイには、自分以外の 2 名の参加者の様子が映され、デフォルメーションされた映像がここに提示される。各実験参加者の前に置かれたキーボードの任意のキーを押すと感情表出調節ツールがオンになり、顔映像や音声のデフォルメーションが適用される。

収集データ 顔映像、音声、感情表出調節ツール操作ログデータを行動データとして収集した。顔映像と音声についてはデフォルメーションが適用される前と後の両方のデータを収録した。また、会話中の印象を尋ねるアンケートを会話ごとに実施し、実験後にはインタビューを行った。

教示 参加者には任意のキーを押すことにより感情表出調整ツールを適用することができることを教示した。また、明るい表情に変更したいときにはツールを積極的に使用するよう指示した。

実験実施後、収録したデータを整備し、不備があるデータを除外した。以降から 10 グループのデータを分析対象にする。

4. 音声・映像デフォルメーションの効果

適切なタイミングで音声・映像を肯定的にデフォルメーションすることによる影響を調べるために、感情表出調節ツールが有効である時間とそうでない時間で、参加者の快表情や発話の量が異なるか否かを検証した。

ツールを使用するとツール使用者の音声と表情が 10 秒間デフォルメーションされるが、デフォルメーションが影響を与える範囲は利用者だけでなくデフォルメーションを見聞きした参加者にも及ぶことが想定される。すなわち、参加者の誰かがツールを使用した時間区間には、会話の参加者全員がデフォルメーションの影響を受ける。加えて、デフォルメーションが終了してから 10 秒後程度もデフォルメーションの影響が持続すると仮定し、ツールの使用開始から 20 秒後までをツールの効果がある時間と定義した。

ツールの効果がある時間とない時間における統計的な有意差を、次の 2 つの指標において検証する；1) 発話割合、2) 表情の Valence。発話割合は任意の時間区間において参加者のいずれかが発話している時間の割合であり、 $((\text{時間長} - \text{無音時間長}) / \text{時間長})$ により計算される。Valence は Affectiva^{*3}により推定された、デフォルメーションが適用される前の顔映像の快・不快の強度である。ここで、ツールの効果がある時間とない時間の平均値はそれぞれ、5 分 50 秒と 9 分 23 秒だった。

*3 <https://www.affectiva.com>

t 検定の結果を図 2 に示す. ツールの効果がある時間のほうが, 発話割合, Valence の値ともに高くなり, 有意差が見られた ($t(98) = 2.286, p < .05, d = 0.23$), ($t(98) = 2.987, p < .01, d = 0.30$). さらにいずれの d 値も 0.2 を超えることから, 小程度の効果量も認められた. 被験者は適切なタイミングでツールを起動したと考えると, この結果は, 適切なタイミングで表情と音声デフォルメーションすることにより, 発話量と Valence の値が有意に増加することを示しているといえる. また被験者はつねにデフォルメーションを利用するわけではないことも明らかになった. したがって, 従来研究ではタイミングについては議論していないが, デフォルメーションのタイミングを適切に決めることが重要であることが明らかになった.

ここで先行研究では, 会話中常時表情を笑顔に変形させるシステムによりアイデアの量が増える [21] こと, また相手の笑顔を自動で検出して笑顔を模倣するように表情を変形させるシステムにより会話が盛り上がったと感じさせる [23] ことが報告されている. 先行研究における発見はいずれも会話中での発言や意見に関する分析での統計的に有意な影響を報告している. 我々も発言や意見におけるツールの有効性, すなわち適切なタイミングでデフォルメーションすることにより発話割合が統計的に有意に増加することを明らかにした. これは先行研究の報告と一致する結果である. 加えて本研究では適切なタイミングでデフォルメーションすることにより, 快表情の強度も統計的に有意に増加することを明らかにしている. 本研究と先行研究 [21], [23] は, 利用者にとって適切なタイミングでデフォルメーションすること, また音声モダリティもデフォルメーションする点が異なる. すなわち以上の分析は, 利用者にとって適切なタイミングで表情や音声をデフォルメーションすることは, 発言や意見に関する側面に影響することを先行研究と同様に明らかにしただけでなく, 表情にも有意な影響を与えることも明らかにしたものと見える.

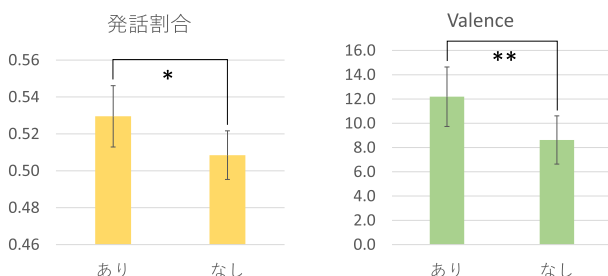


図 2 感情表出調節ツールの効果がある時間とない時間における発話割合と Valence (*: $p < .05$, **: $p < .01$). エラーバーは標準誤差を示す

Fig. 2 Speech time ratio and the average valence value in the tool affected/not affected time. Error bars show standard error.

5. 支援適用タイミング推定モデル

4 章において, 感情表出調節ツールを適切なタイミングで使用することでコミュニケーションが活性化することが確認できた. 実験では, 参加者自身がツール適用開始を決定していたが, これを自動化することができれば会議参加者自身がツールを起動する負荷なく議論を活性化することが期待できる. しかしそのためには, デフォルメーション機能が適用されるべきタイミング, すなわち支援適用タイミングを自動推定する必要がある. そこで本章では, 音声, 表情モダリティから支援適用タイミングを機械学習アプローチにより推定するモデルを作成する.

5.1 教師データの作成

支援適用タイミングの教師データを作成するために, アノテータに 4 章で収集した議論のビデオを観察してもらい, 支援適用タイミングを判断してもらった. アノテータが作業を行った実験ビデオのスナップショットを図 3 に示す. 画面下部に 1 名の会話参加者を表示し, 画面上部にそれ以外の 2 名の会話参加者の映像を表示している. 画面下部に表示した 1 名の会話参加者を評価対象者とする. アノテータには自分が評価対象者の立場であった場合に, 肯定的な表情や音声を表出すべきかどうかを判断するよう指示した. 作業用のビデオクリップは会話全体を 15 秒ずつオーバーラップなく切り出したものである. アノテータは合計 12 名で, 全員女性である (平均年齢: 40.8 歳, 標準偏差: 10.8). 合計 5442 クリップのデータを収集した. これ以降, ビデオクリップを分析単位とする.

次に, 得られた 5442 クリップを正例・負例に割り振るために, 各クリップにおける合意率を算出した (表 1). ここで合意率とは, あるクリップに対して肯定的な表情や音

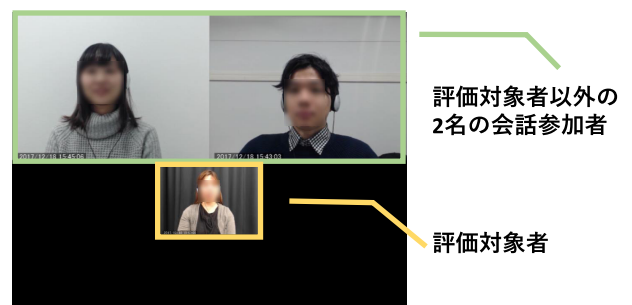


図 3 実験ビデオの一例

Fig. 3 A video sample for annotating support timing.

表 1 合意率の分布. 合意率 0.75 以上のデータを正例とした
Table 1 Distribution of consensus ratio. Cases whose consensus ratio was higher than 0.75 were used as positive cases.

合意率	1.00	0.75	0.50	0.25	0.00
割合	0.30	0.26	0.21	0.18	0.04

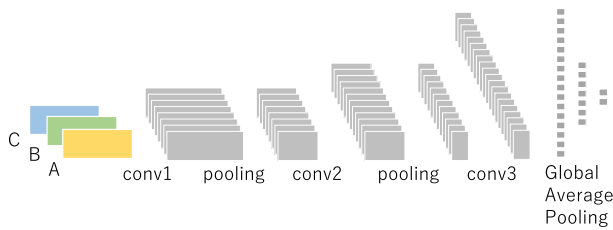


図 4 2D CNN モデルのネットワーク構造
Fig. 4 Network structure for 2D CNN model.

表 2 カーネルサイズとフィルタ数
Table 2 Kernels sizes and pooling filter sizes.

	2D モデル	3D モデル
Conv	(5, 56)	(5, 56, 3)
Pooling	(2, 1)	(2, 1, 1)
Conv1 フィルタ数	32	32
Conv2 フィルタ数	64	64
Conv3 フィルタ数	128	128

声を表出すべきと合意した人の割合と定義する。結果として、正例、負例がほぼ等しくなるように合意率が 0.75 以上を正例、それ未満を負例とした。

5.2 音声モデル

支援適用タイミングの推定に音声貢献すると考え、音声データのみを使用したモデルを作成する。

5.2.1 音声入力データ

音声の学習データを作成するために、56 段階の周波数分解能を持つ音声のスペクトログラムを会話参加者ごとに作成した。サンプリング周波数は 44.1kHz、窓関数として、窓幅 1024 点のハニング窓を用いた。音声スペクトログラムは想定される最大値で除することで正規化した。

5.2.2 音声モデルのネットワーク

音声データを畳み込むモデルが有効であると考え、会話参加者の情報をチャンネルで与える 2D CNN モデルを作成した。しかし、チャンネルで与えてしまうと、畳み込み後には会話参加者の情報を明示的に扱うことができない。そこで、入力の特徴数、時間、会話参加者の 3 次元とする 3D CNN モデルを作成した。

2D CNN ネットワーク 会話参加者の情報をチャンネルで与える 2D CNN モデルのネットワークの構造を図 4 に示す。チャンネルとして与えるスペクトログラムの順序は、評価対象者を基準に実験時の席で反時計回りの順序で与えた。中間層では、3 回の畳み込み層、2 回のプーリング層の後、Global Average Pooling 層、全結合層を通り、最終的に softmax 層で支援適用タイミングかどうかの予測を行う。カーネルサイズは表 2 に示す。

3D CNN ネットワーク 会話参加者の情報を明示的に扱うために、会話参加者の情報を次元とする 3D CNN

モデルを作成した。スペクトログラムの順序は、会話参加者の情報をチャンネルとして扱うときと同様に、評価対象者を基準に実験時の席で反時計回りの順序で与えた。ネットワーク構造としては 2D CNN における 2 次元畳み込み演算を 3 次元畳み込み演算に、2 次元プーリングを 3 次元プーリングに置き換えたものと同じである。中間層では、3 回の畳み込み層、2 回のプーリング層の後、Global Average Pooling 層、全結合層を通り、softmax 層で支援適用タイミングかどうかの予測を行う。カーネルサイズは、表 2 に示す。

5.3 表情モデル

音声データと同様に表情データでも深層学習による訓練を実施したが、最適化に失敗し、作成されたモデルはすべてのインスタンスを正例か負例のどちらか一方として推定するものになった。そこで従来手法である SVM を用いてモデルを作成する。

5.3.1 表情入力データ

顔特徴抽出ソフトである OpenFace^{*4}を使用し、顔から推定されるデータを取得した後、各データを正規化することで特徴量を定義した。定義した特徴量を以下に述べる。

視線方向 x 軸、 y 軸についての視線の回転角度。正面方向を向くと 0° をとる。左右方向への変化は x 軸に影響し、上下方向の変化は y 軸に影響する。 x 、 y 各軸がとる値を 60° で除した値を特徴量として定義した。 60° は観測された値の最大値。

頭部回転 x 、 y 、 z 軸についての頭部回転角。視線方向と同様に、 x 、 y 、 z 各軸における角度を 60° で除した値を特徴量として定義。

Action Unit (強度) 頬を上げる、目を伏せるなどの、表情を構成する顔動作である Action Unit (AU) のうち 17 種類の強度についての推定値。使用した AU は 1, 2, 4, 5, 6, 7, 9, 10, 12, 14, 15, 17, 20, 23, 25, 26, 45 番。各 AU につき 0~1 の数値をとるように正規化した。

Action Unit (有無) 18 種類の AU が発生したかどうかの推定値。使用した AU は 1, 2, 4, 5, 6, 7, 9, 10, 12, 14, 15, 17, 20, 23, 25, 26, 28, 45 番。各 AU につき 0 か 1 をとる。

2D 顔ランドマーク距離 まず顔の特徴点である 68 個の顔ランドマークの二次元座標を取得した。そして映像中での顔の位置と顔の大きさを正規化するため、次の 2 段階の演算を実施した；1) 各ランドマークと鼻の先端との間 (計 67 組) の距離を計算する、2) 67 組の距離を、鼻の先端と鼻の付け根の間の 2 倍の距離で除算する。最終的な特徴量数は 66 (= 67 組 - 鼻の先端と付

*4 <https://github.com/TadasBaltrusaitis/OpenFace>

表 3 表情特徴量の t 検定の結果

Table 3 Results of t-test for facial features.

特徴量	有意差が見られた特徴量数	すべての特徴量数
視線方向	1	2
頭部回転	1	3
Action Unit (強度)	12	17
Action Unit (有無)	17	18
2D 顔ランドマーク距離	54	66
3D 顔ランドマーク距離	62	66
2D 目ランドマーク距離	43	56
3D 目ランドマーク距離	0	56

け根の組) 個である。

3D 顔ランドマーク距離 3次元座標での68個の顔ランドマークを、2D 顔ランドマーク距離と同様の方法で特徴量化した。特徴量数は2D 顔ランドマーク距離と同様に66個である。

2D 目ランドマーク距離 目の輪郭に点在する56個の目ランドマークの二次元座標を計算した後、2D 顔ランドマーク距離と同様の方法で特徴量化した。2D 顔ランドマーク距離と同一の正規化定数を使用。特徴量数は56個である。

3D 目ランドマーク距離 目の輪郭に点在する56個の目ランドマークの三次元座標を計算した後、2D 顔ランドマーク距離と同様の方法で特徴量化した。3D 顔ランドマーク距離と同一の正規化定数を使用。特徴量数は56個である。

定義した特徴量について正例、負例それぞれの区間における平均値を算出し t 検定を行った (表 3)。ここで有意差が見られた特徴量である視線方向、頭部回転、AU (強度)、AU (有無)、3D 顔ランドマーク距離、2D 目ランドマーク距離をモデル作成に使用する。2D 顔ランドマーク距離にも有意差は確認できたが、訓練時のデータサイズを加味し、使用しなかった。よって合計 162 の特徴量を入力データとして使用した。

5.3.2 表情 (SVM) モデルの概要

3名分の表情データを入力とする SVM モデルを作成した。1クリップ中の各特徴量の平均値を使用したため、入力サイズは486次元 (162次元 × 3名) である。

5.4 フュージョンモデル

音声モダリティと表情モダリティの両方を使用するマルチモーダルなモデルを作成した (図 5)。2D CNN モデルの最終全結合層の出力と、表情データを連結したデータを入力とする SVM を作成した。入力データは3名分の表情データ (162次元 × 3名) と音声モデルの中間層から得られた3名分のベクトル (64次元) で合計550次元である。

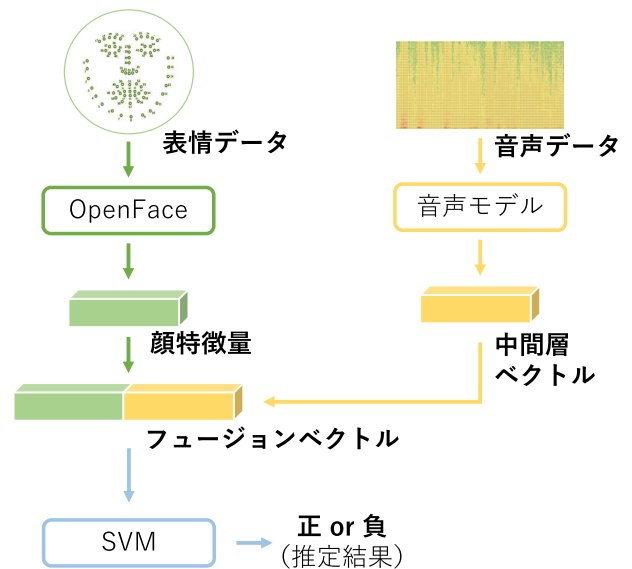


図 5 音声、表情モダリティのフュージョン手法
Fig. 5 A method for fusing speech/facial modality.

表 4 モデルの性能

Table 4 Performances of the models.

モデル	Acc.	F1
音声 (2D CNN)	0.67	0.65
音声 (3D CNN)	0.64	0.51
表情 (SVM; $C = 2^{-6}$, $\gamma = 2^{-7}$)	0.69	0.64
音声・表情 (SVM; $C = 2^{-6}$, $\gamma = 2^{-2}$)	0.71	0.72

5.5 モデルの訓練と評価

音声モダリティを使用する 2D CNN と 3D CNN モデル、表情モダリティを使用する SVM モデル、音声と表情の情報をフュージョンする SVM モデル、以上 4 種類のモデルを作成した。

音声 (2D/3D CNN) モデルの最適化アルゴリズムには Adadelta を使用し、エポック数は 50、バッチサイズは 32、目的関数には交差エントロピーを使用し、訓練を行った。SVM を使用する表情モデル、音声・表情フュージョンモデルのカーネルには RBF を使用した。ハイパーパラメタである C と γ はそれぞれ、 $2^{-10} \sim 2^{16}$ 、 $2^{-20} \sim 2^{11}$ の範囲でグリッドサーチを行った。合計 806 通りのモデルを作成し、Accuracy が最も優れていたモデルを採用した。

訓練、検証、テストのデータ数はそれぞれ 2,722, 1,101, そして 1,439 ビデオクリップであり、各会話グループは訓練、検証、テストのどれか 1 つに振り分けた。訓練データはアンダーサンプリングを行った。評価尺度は Accuracy と F-measure である。

4つのモデルの性能比較の結果を表 4 に示す。Accuracy と F-measure それぞれにおける最良のモデル性能は太字で修飾されている。

ユニモーダルモデルである音声モデル (2D/3D CNN) と表情モデルを比較すると、Accuracy においては表情モ

デルが最良 (0.69) であり, F-measure においては音声モデル (2D CNN) が最良 (0.65) であった.

フュージョンモデルである音声・表情モデルはすべての尺度において最良の性能を達成し, その性能は Accuracy で 0.71, F-measure で 0.72 である. Accuracy では, ユニモーダルモデルの中で最も高性能であった表情モデルよりも 2%, F-measure では, ユニモーダルモデルの中で最も高性能であった音声モデル (2D CNN) よりも 7%, それぞれ高い性能が得られた. 以上から, フュージョンモデルは, 音声, 表情データのどちらかだけを使用したユニモーダルモデルよりも高い性能を持つといえる.

6. 議論

本章ではまず, 5章で提案したモデルの特性を議論するため, モデルの推定結果を考察する. そして推定機構に対して期待できるコミュニケーション支援の効果を議論する.

6.1 提案モデルの特性

提案モデルが予測した肯定的な感情を表出すべきタイミングの特性を議論するため, モデルの推定結果を議論の映像とともに観察した. その結果, 支援適用タイミングとして推定された状況における参加者の発話と笑顔の様子には, 次の2種類の特徴が確認できた; 1) 参加者のうち少なくとも1名が積極的に発話する一方, 参加者間での笑顔が同期しない, 2) 参加者のうち少なくとも1名が積極的に発話しかつ参加者間での笑顔が同期する一方, 笑顔の程度が弱い.

以上より, 作成したモデルはある程度会話が円滑であるグループに対して, より会話が円滑になるように支援することが可能なモデルであることが示唆される.

6.2 提案機構により期待できる会話活動の改善

肯定的な感情の表出は発話量と快感情の増加に貢献することを4章において述べた. ここで注意として, 4章で分析対象とした感情表出調整ツールの使用タイミングは, 会話参加者自身により判断されたものである. 一方, 5章のモデルは, 会話参加者自身ではない外部評価者が評定した支援適用タイミングを推定する. そこで提案モデルに基づき肯定的な感情を表出する場合においても, 4章と同様に, 発話量と快感情の増加が期待できるかどうかを検証する.

そのために, 以下の2種類の尺度による評価を行った.

S-P 会話参加者自身が判断した感情表出調整ツールの使用タイミングの, 提案モデルによる予測性能

S-O 会話参加者自身が判断した感情表出調整ツールの使用タイミングと外部評価者が評定した支援適用タイミングの一致の程度

上記の2つの尺度に対して, Precision と Recall を計算した. その結果を表5に示す. まずS-Oでは, 外部評価

表5 参加者判断と客観評価の一致度 (S-O), モデルによる参加者判断の推定性能 (S-P)

Table 5 Degree of agreement between subjective judgment and objective evaluation, and model performance for predicting subjective judgment.

尺度	S-O	S-P
Precision	0.185	0.194
Recall	0.603	0.680

者が判断した支援適用タイミングが会話参加者自身の判断にどの程度適合しているか (Precision), またどの程度再現できているか (Recall) を算出した. S-Oにおいて, Precision は Recall に比べて大きく低下することから, 外部評価者は会話参加者自身よりも頻繁に肯定的な感情を表出すべきと評価したといえる. これは, 会話に参加しながら感情表出調整ツールを操作することは, 会話参加者にとって負荷の高い課題であった, つまり会話参加者はツールをオンにしてもよい状況すべてに反応したわけではないことが原因であると考えられる.

S-P の Precision, Recall はいずれも S-O よりも高く, 特に S-P の Recall は 0.68 であり, モデルは会話参加者自身が肯定的な感情を表出すべきと判断したタイミングの7割程度を網羅できていたことが分かった. モデルが肯定的な感情を表出すべきと推定した場合, その推定結果は外部評価者の判断よりも会話参加者自身の判断に近いといえる.

以上の分析に基づき, 提案モデルを用いて遠隔コミュニケーションを支援する方法を検討する. Precision よりも Recall が良好であることを考慮すると, 提案モデルの推定結果に基づき会話参加者の表情や音声を自動的にデフォメーションさせることよりも, モデルが推定した支援適用タイミングを会話参加者へ提示し, ツールの使用を促す方式が適当であると考えられる. 会話参加者は, システムからの促しを参考に感情表出調整ツールを使用するか否かを決定することで, 判断への認知的負荷を抑えながらも, ツールを積極的に利用することができ, 会話での発話量や快感情の向上が期待できる.

以上の議論は提案モデルを使用した遠隔会話支援のデザインについて, またそれをシステム化することにより期待できる会話の改善についての意味のある議論である一方, 遠隔会話支援インタフェースとしての評価は, 実際にシステムを実装したうえでユーザスタディを行う必要がある. これについては将来の課題である.

7. 支援適用タイミング推定機構

6章において, 提案モデルに基づく支援により議論の活性化が期待できることを示唆した. そこで本章では, システム化の可能性を示すために, 支援適用タイミング推定機構のプロトタイプを開発する. 本推定機構は5章で作成し

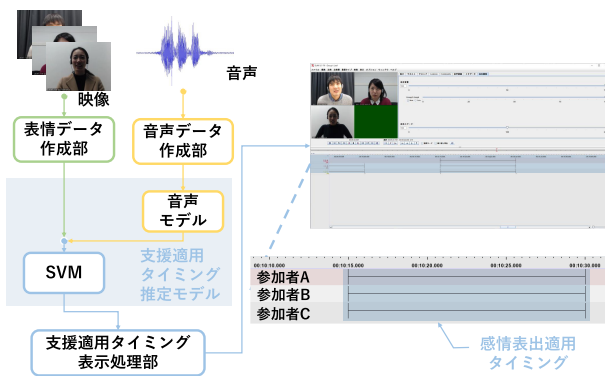


図 6 システム適用タイミング推定機構の構成図

Fig. 6 System components of the support timing estimation mechanism.

たモデルを搭載したオフラインシステムであり、会話参加者3名分の映像と音声を入力とすることで、時間経過とともなう支援適用タイミングを可視化することができる。

図 6 に推定機構のシステム構成を示す。まず、表情データ作成部と音声データ作成部が、映像データと音声データを推定モデルに入力できる形式に加工する。加工されたデータはモデルに供給され、支援適用タイミングを推定する。支援適用タイミングは15秒で区切られた会話セグメントごとに推定される。会話全体の推定結果が蓄積され次第、可視化のためのソフトウェアが会話全体の支援適用タイミングを可視化する。

以降では推定機構を構成する各モジュールについて詳細に説明する。

表情データ作成部 まず顔が記録された映像に対して OpenFace を適用し、視線方向、頭部回転、Action Unit、3次元の顔のランドマーク、2次元の目のランドマークを取得した。次に5.3.1項で説明した方法で各種データを特徴量化し、クリップ単位で各特徴量の平均値を算出した。最後に3名の会話参加者のデータを連結することで、表情データを作成した。

音声データ作成部 まず5.2.1項と同様の方法でFFTを使用して、音声スペクトログラムを作成した。音声スペクトログラムを正規化したのち、3名の会話参加者の1クリップ分のデータを1つにまとめたファイルを作成した。

支援適用タイミング推定モデル 表情、音声データから、各クリップに対して支援適用タイミングかどうかを判定する。推定モデルには最も Accuracy が高かったフュージョンモデルを採用した。

支援適用タイミング表示処理部 各クリップに対する推定結果をアノテーションソフトウェア ELAN^{*5}で表示するために、注釈層名、開始時間、終了時間、注釈内容、対応するビデオをxmlで記述したeafファイルを作成

した。作成したeafファイルを実行することで、支援適用タイミングが示されているELANの画面が起動する。

支援適用タイミングの表示 ELANによって支援適用タイミングを表示する。ELANには会話参加者3名のタイムラインが表示されており、支援適用タイミングがアノテーション区間として示される。

今後、リアルタイム化したツールを作成することにより、可視化された推定結果、あるいはそれに基づくシステムからの示唆を参考に、会話参加者がデフォルトタイミングを決定することにより、表情と音声をデフォルトタイミングに対する認知的負荷を抑えながらも、会話での発話量や快感情の向上が期待できる。

8. まとめと今後の課題

本研究では多人数での遠隔コミュニケーションを活性化させるため、はじめに、我々が作成した肯定的な感情を人工的に表出するツールを使用する状況でのコミュニケーションを収録した。収録されたコーパスを分析することで、適切なタイミングでツールを使用することによりコミュニケーションが活性化することを確認した。次に肯定的な感情を表出するタイミングを示唆するシステムの実装をめざし、音声情報と表情情報に基づき肯定的な感情を表出すべきタイミングを自動推定する学習モデルを提案した。提案モデルの性能はF値で0.72であり、さらにモデルが会話参加者に対して感情表出調整ツールを使用することを促すことで、会話の活性化が期待できることが示唆された。最後にコミュニケーション支援システムの実現に向け、支援適用タイミング推定機構のプロトタイプを実装した。

今後の課題として、現在オフライン処理のみが可能な支援適用タイミング推定機構を改良し、リアルタイムでの推定を可能にする。さらにリアルタイム化された推定機構を使用した評価実験を実施し、肯定的な感情を表出すべきタイミングを提案モデルが参加者に対して示唆する場合のコミュニケーションの活性化について調査する。

参考文献

- [1] Andolina, S., Orso, V., Schneider, H., Klouche, K., Ruotsalo, T., Gamberini, L. and Jacucci, G.: Investigating Proactive Search Support in Conversations, *Proc. 2018 Designing Interactive Systems Conference, DIS '18*, pp.1295-1307, Association for Computing Machinery (online), DOI: 10.1145/3196709.3196734 (2018).
- [2] Basu, S., Choudhury, T., Clarkson, B. and Pentland, A.: Towards measuring human interactions in conversational settings (2001).
- [3] Bickmore, T.W. and Picard, R.W.: Establishing and Maintaining Long-Term Human-Computer Relationships, *ACM Trans. Comput.-Hum. Interact.*, Vol.12, No.2, pp.293-327 (online), DOI: 10.1145/1067860.1067867 (2005).

*5 <https://tla.mpi.nl/tools/tla-tools/elan>

- [4] Costa, J., Jung, M.F., Czerwinski, M., Guimbretière, F., Le, T. and Choudhury, T.: Regulating Feelings During Interpersonal Conflicts by Changing Voice Self-Perception, *Proc. 2018 CHI Conference on Human Factors in Computing Systems, CHI '18*, Association for Computing Machinery (online), DOI: 10.1145/3173574.3174205 (2018).
- [5] Dijkstra-Soudarissanane, S., Assal, K.E., Gunkel, S., ter Haar, F., Hindriks, R., Kleinrouweler, J.W. and Niamut, O.: Multi-Sensor Capture and Network Processing for Virtual Reality Conferencing, *Proc. 10th ACM Multimedia Systems Conference, MMSys '19*, pp.316–319, Association for Computing Machinery (online), DOI: 10.1145/3304109.3323838 (2019).
- [6] 二瓶芙巳雄, 中野有紀子, 高瀬 裕: 言語・非言語情報に基づく議論要約のための重要発言推定, *電子情報通信学会論文誌 A*, Vol.J102-A, No.2, pp.35–47 (2019).
- [7] Gatica-Perez, D., McCowan, I.A., Zhang, D. and Bengio, S.: Detecting Group Interest-level in Meetings, *IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP)* (2005).
- [8] Hirai, Y. and Kaneko, K.: Ambient Conversation Support in Small Face-to-Face Group Meetings, *Proc. 6th International Symposium on Information and Communication Technology, SoICT 2015*, pp.239–246, Association for Computing Machinery (online), DOI: 10.1145/2833258.2833274 (2015).
- [9] 服部憲治, 渡辺富夫, 山本倫也: タイピング駆動型身体引き込みキャラクタチャットシステム InterChat, *ヒューマンインタフェース学会論文誌*, Vol.15, No.4, pp.389–398 (2013).
- [10] Kuster, C., Popa, T., Bazin, J.-C., Gotsman, C. and Gross, M.: Gaze Correction for Home Video Conferencing, *ACM Trans. Graph.*, Vol.31, No.6 (online), DOI: 10.1145/2366145.2366193 (2012).
- [11] Li, S., Zheng, W., Zong, Y., Lu, C., Tang, C., Jiang, X., Liu, J. and Xia, W.: Bi-Modality Fusion for Emotion Recognition in the Wild, *2019 International Conference on Multimodal Interaction, ICMI '19*, pp.589–594, Association for Computing Machinery (online), DOI: 10.1145/3340555.3355719 (2019).
- [12] Lisetti, C., Amini, R., Yasavur, U. and Rishe, N.: I Can Help You Change! An Empathic Virtual Agent Delivers Behavior Change Health Interventions, *ACM Trans. Manage. Inf. Syst.*, Vol.4, No.4 (online), DOI: 10.1145/2544103 (2013).
- [13] Otsuka, K., Yamato, J., Takemae, Y. and Murase, H.: Quantifying Interpersonal Influence in Face-to-face Conversations Based on Visual Attention Patterns, *CHI '06 Extended Abstracts on Human Factors in Computing Systems, CHI EA '06*, pp.1175–1180, ACM (online), DOI: 10.1145/1125451.1125672 (2006).
- [14] Otsuki, M., Maruyama, K., Kuzuoka, H. and Suzuki, Y.: Effects of Enhanced Gaze Presentation on Gaze Leading in Remote Collaborative Physical Tasks, *Proc. 2018 CHI Conference on Human Factors in Computing Systems, CHI '18*, Association for Computing Machinery (online), DOI: 10.1145/3173574.3173942 (2018).
- [15] Riek, L.D., Paul, P.C. and Robinson, P.: When my robot smiles at me: Enabling human-robot rapport via real-time head gesture mimicry, *Journal on Multimodal User Interfaces*, Vol.3, No.1, pp.99–108 (online), DOI: 10.1007/s12193-009-0028-2 (2010).
- [16] Rienks, R., Zhang, D., Gatica-Perez, D. and Post, W.: Detection and Application of Influence Rankings in Small Group Meetings, *Proc. 8th International Conference on Multimodal Interfaces, ICMI '06*, pp.257–264, ACM (online), DOI: 10.1145/1180995.1181047 (2006).
- [17] Rivera, K., Cooke, N.J. and Bauhs, J.A.: The Effects of Emotional Icons on Remote Communication, *Conference Companion on Human Factors in Computing Systems, CHI '96*, pp.99–100, Association for Computing Machinery (online), DOI: 10.1145/257089.257180 (1996).
- [18] Sanchez-Cortes, D., Aran, O., Jayagopi, D.B., Schmid Mast, M. and Gatica-Perez, D.: Emergent leaders through looking and speaking: From audio-visual data to multimodal recognition, *Journal on Multimodal User Interfaces*, Vol.7, No.1-2, pp.39–53 (online), DOI: 10.1007/s12193-012-0101-0 (2013).
- [19] Schiavo, G., Cappelletti, A., Mencarini, E., Stock, O. and Zancanaro, M.: Overt or Subtle? Supporting Group Conversations with Automatically Targeted Directives, *Proc. 19th International Conference on Intelligent User Interfaces, IUI '14*, pp.225–234, Association for Computing Machinery (online), DOI: 10.1145/2557500.2557507 (2014).
- [20] Shimizu, R., Takase, Y. and Nakano, Y.I.: Audio Based Group Conversation Support System, *Proc. 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proc. 2017 ACM International Symposium on Wearable Computers, UbiComp '17*, pp.197–200, Association for Computing Machinery (online), DOI: 10.1145/3123024.3123079 (2017).
- [21] 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 (特集) 教育・訓練・支援, *日本パーソナルリアリティ学会論文誌*, Vol.20, No.4, pp.323–332 (2015).
- [22] Sonnby-Borgström, M.: Emotional mimicry: underlying mechanisms and individual differences, *Studies in Emotion and Social Interaction*, pp.125–161 (online), DOI: 10.1017/CBO9781107587595.008, Cambridge University Press (2016).
- [23] Suzuki, K., Yokoyama, M., Yoshida, S., Mochizuki, T., Yamada, T., Narumi, T., Tanikawa, T. and Hirose, M.: FaceShare: Mirroring with Pseudo-Smile Enriches Video Chat Communications, *Proc. 2017 CHI Conference on Human Factors in Computing Systems, CHI '17*, pp.5313–5317, Association for Computing Machinery (online), DOI: 10.1145/3025453.3025574 (2017).
- [24] Terken, J. and Sturm, J.: Multimodal Support for Social Dynamics in Co-Located Meetings, *Personal Ubiquitous Comput.*, Vol.14, No.8, pp.703–714 (online), DOI: 10.1007/s00779-010-0284-x (2010).
- [25] Wang, T.-Y., Kawaguchi, I., Kuzuoka, H. and Otsuki, M.: Effect of Manipulated Amplitude and Frequency of Human Voice on Dominance and Persuasiveness in Audio Conferences, *Proc. ACM Hum.-Comput. Interact.*, Vol.2, No.CSCW (online), DOI: 10.1145/3274446 (2018).
- [26] Wang, Y., Wu, J. and Hoashi, K.: Multi-Attention Fusion Network for Video-Based Emotion Recognition, *2019 International Conference on Multimodal Interaction, ICMI '19*, pp.595–601, Association for Computing Machinery (online), DOI: 10.1145/3340555.3355720 (2019).
- [27] Wu, C.-H., Lin, J.-C. and Wei, W.-L.: Survey on audiovisual emotion recognition: databases, features, and data fusion strategies, *APSIPA Trans. Signal and Information Processing*, Vol.3, p.e12 (online), DOI: 10.1017/ATSIP.2014.11 (2014).
- [28] Zhou, H., Meng, D., Zhang, Y., Peng, X., Du, J., Wang, K. and Qiao, Y.: Exploring Emotion Features and Fu-

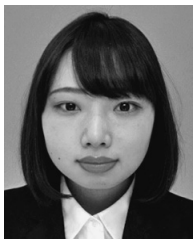
sion Strategies for Audio-Video Emotion Recognition, *2019 International Conference on Multimodal Interaction, ICMI '19*, pp.562-566, Association for Computing Machinery (online), DOI: 10.1145/3340555.3355713 (2019).



二瓶 美巳雄 (正会員)

2016年成蹊大学大学院理工学研究科修士課程修了。2019年同大学院博士課程修了。現在、成蹊大学特別共同研究員。マルチモーダル情報に基づくグループ会議の分析とモデル化の研究に従事。ACM, 人工知能学会各会員。

博士 (理工学)。



田口 和佳奈

2020年成蹊大学大学院理工学研究科修士課程修了。マルチモーダル情報に基づく遠隔グループ会議の分析とモデル化の研究に従事。修士 (理工学)。



中野 有紀子 (正会員)

1990年東京大学大学院教育学研究科修士課程修了。同年日本電信電話(株)入社。2002年MIT Media Arts & Sciences 修士課程修了。JST 社会技術研究開発センター専門研究員, 東京農工大学大学院工学府特任准教授, 成蹊大

学理工学部情報科学科准教授を経て, 現在, 成蹊大学理工学部情報科学科教授。2019年カーネギーメロン大学客員研究員。人との言語・非言語コミュニケーションが可能な会話エージェントの研究に従事。博士 (情報理工学)。



深澤 伸一

2008年千葉大学大学院自然科学研究科修士課程修了。同年沖電気工業株式会社入社, 研究開発本部配属。以来, ユーザインタフェース, 感情推定技術, 遠隔コミュニケーション支援システムの研究開発に従事。日本人間工学会, HI学会各会員。

会, HI学会各会員。



赤津 裕子 (正会員)

玩具メーカーを経て, 2002年沖電気工業株式会社に入社。ATM, 券売機, 複合機, コミュニケーションシステム等のインタフェースに関する研究開発に従事。早稲田大学大学院理工学研究科博士課程修了。工学博士。ヒューマン

インタフェース学会, 日本人間工学会, 日本認知科学会各会員。