

3D アバタを利用した遠隔授業での 頭部動作の認識を用いた意思表示インタフェース

長尾優花^{†1} 北野圭介^{†1} 大島登志一^{†1} 望月茂徳^{†1}

概要: ライブ配信型遠隔授業におけるオンラインミーティングツールの活用が, 教員-学生間での双方向コミュニケーションの容易性から急速に拡大している. 本研究では, 教育用途でのオンラインミーティングツールの活用において, 教員-学生間のコミュニケーション増強のため, 3D アバタと顔の動作による意思表示に着目したアプリケーションの提案を行う. 具体的には, 首振りや頷きによって理解状況を直感的に示すための実装を行う.

A Communication Interface using Head Movement Recognition in Online Classes with 3D Avatars

YUKA NAGAO^{†1} KEISUKE KITANO^{†1}
TOSHIKAZU OHSHIMA^{†1} SHIGENORI MOCHIZUKI^{†1}

1. はじめに

2019年度から世界的に流行した新型コロナウイルス感染症の蔓延防止により, 対面コミュニケーションから CMC (Computer-Mediated-Communication) への転換が企業や高等教育機関などで行われた. この転換により, 通信機器の整備や ICT の利活用が推進されている. 遠隔授業の暫定的な評価の一例としては, 学生を対象とした神奈川大学の調査において, 対面授業と遠隔授業を比較したとき「とても向上した」「向上した」「変わらない」と答えた割合が半数を超えていることが挙げられる[1]. 感染症の収束後も, テレワークや遠隔授業を始めとする CMC の積極的な利用は継続されると予想されている.

高等教育における遠隔授業の導入方式は, (1) ライブ配信型 (2) 動画教材提示型 (3) 資料・課題提示型に大きく区分される. 最も多く利用され満足度も高いと報告されているのは, ライブ配信型である[2]. Web 会議システムを利用したライブ配信型遠隔授業は最も対面形式に近い. このため, 他の遠隔授業の形式と比べて, 質問や発言などの双方向コミュニケーションを容易に行うことができる.

しかし, 授業アンケートでは学生向け[1], 教員向け[3]共に教員と学生のコミュニケーションの情報量や質の低下が指摘されている. 現在のライブ配信型遠隔授業では, ビデオ機能をオンにした場合, 学生の顔が教員や他の学生に公開される. そのため, 自身の真正面を向いた顔が常に公開されることを嫌い, ビデオ機能をオフにしてしまう学生が多く見られる. ビデオ機能をオフにすることによる視覚情報の欠落は, コミュニケーションの情報量や質の低下の原因の一つである. そこで, 学生が公開する動きの情報に着目して教員と学生のコミュニケーションを増強する必要が

ある.

本研究では遠隔授業における双方向コミュニケーションの支援のために, 3D アバタと顔認識技術を利用した頭部動作と口や目の動きの再現によって, 使用者の動作の伝達と動作による意思表示の強調を行うアプリケーションを提案する. 特に強調する動作は, 首を縦に振る「頷き」と横に振る「首振り」を対象とする.

2. 関連研究

3D アバタを用いた遠隔コミュニケーションは, 一般向けのサービスとして提供され始めている. Apple のビデオ通話アプリ Face Time において利用できる「ミー文字」[4]や, コミュニケーションアプリ LINE のビデオ通話機能で利用できる「LINE アバタ」[5]などで, 3D アバタに使用者の動きを反映させることが可能である. しかしながら, 教育機関において多く活用されている「Zoom」や「Microsoft Teams」などのオンラインミーティングツールにおいては, 現時点において使用者に連動する 3D アバタは標準機能として実装されていない. また, バーチャル 3D 空間での遠隔コミュニケーションにおいては, 「cluster」や「Mozilla Hubs」のような Web VR サービスがポスター発表のような双方向での交流が求められる場面で利用されている[6].

ビデオ通話における 3D アバタの利用に関する調査では, 山本らはアバタ・写真・ビデオを提示した音声通話と, 何も提示しない音声通話を比較して, 二者間の音声通話においては対話者の外見ではなく動きを伝えることが重要であり, アバタがビデオの代替になりうる可能性を示している[7]. また, 3D アバタに反映される使用者の振る舞いに重畳する試みとしては, 瀬島らの研究が挙げられる. 発生音声

^{†1} 立命館大学大学院映像研究科
Graduate School of Image Arts, Ritsumeikan University

から自動生成したうなずきモデルを重畳させることで、身体的インタラクションの促進効果が認められたと明らかにしている[8].

3. アプリケーション概要

3.1 頭部動作の表示と強調のためのデザイン

本研究で開発するアプリケーションでは、学生が教員や他の学生に対して、頭部動作と口や目の動きによって強調された意思表示を行うことを可能とする。

基本的な機能では、認識した顔の特徴点の動きを 3D アバタに反映させることにより、学生の頭部動作と口や目の動きを表示させる。この機能によって、参加している学生の動きを教員や他の学生が確認できる。お互いに動作を確認することにより、場の共有感が生まれる。遠隔授業に関する探索的調査において、技術に対する理解と個人学習の効率に肯定的な回答をしなかった学生は孤独感を抱くと指摘されているが[9], 他の学生の頭部動作の確認による場の共有感によって孤独感の緩和が期待できる。

意思表示の強調機能では (1) 小さな傾き・首振りを検出した場合に表情の変更 (小) (2) 大きな傾き・首振りを検出した場合に表情の変更 (大) と背景画像の変更, という二つの状態に遷移する。

(1) 小さな傾き・首振りの場合

小さな傾き・首振りが検出されるとき、小さな傾きでは口角を少し上げ目を瞑る動作を行う(図 1)。小さな首振りにおいては、眉毛が下がり口を窄める動作を行う(図 2)。この機能では、明示的に意思表示を行う必要がない場面において、学生側は肯定的、または否定的な意思表示が可能である。



図 1 小さな傾きを検知した場合



図 2 小さな首振りを検知した場合

(2) 大きな傾き・首振りの場合

(1)と同様に大きな傾き・首振りが検出されたとき、(1)よりも大きな動作の変化と背景画像の変更を行う。通常状態の背景画像は青であるが、大きな傾きの判定時には緑に変更する(図 3)。大きな首振りの判定時では赤に変更する(図 4)。この機能では、教員が理解度の確認をしたときなど学生が反応をする必要がある場面において、肯定的・否定的な意思表示をすることが可能である。教員側にとっては、一目で複数の学生の意思を確認することができる。

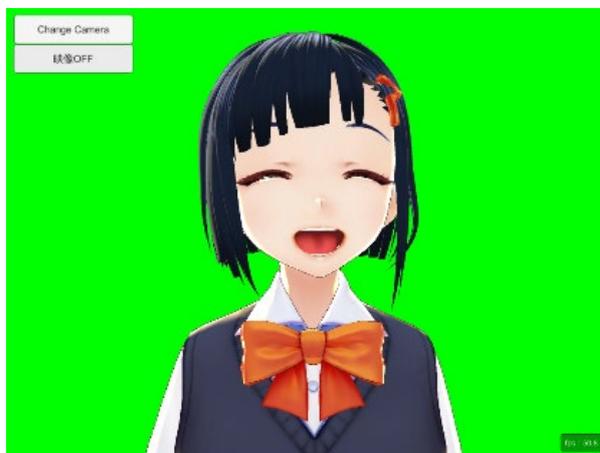


図 3 大きな傾きを検知した場合



図 4 大きな首振りを検知した場合

3.2 コミュニケーション動作の検出方法

上記に示されている各状態への遷移は、アプリケーションが検出した傾き・首振りの強弱によって行う。状態遷移の流れは図 5 に示す。

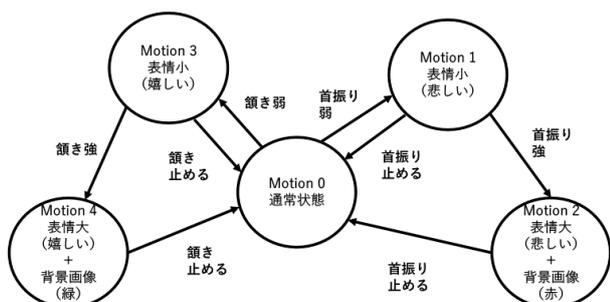


図 5 傾きと首振りの状態遷移図

傾きと首振りの判定には、図 6 に示される特徴点の一つである特徴点 30 のフレーム毎の変量を用いる。検出された特徴点座標のノイズを除去し、ユーザの姿勢の変更や急激な顔位置の移動を傾き・首振りとは判定しないよう除外するため、特徴点の x 座標、y 座標それぞれに指数移動平均によるフィルタを適応する。これらの値 (ave_sum_x, ave_sum_y) を動作の検出に用いる。これらの値の推移例を図 7、図 8 に示す。横軸はフレーム番号であり、縦軸は ave_sum_x, ave_sum_y である。

傾きの場合は、ave_sum_y が一定値以上で推移し、かつ ave_sum_x が相対的に低い値で推移し続ける状態となる (図 7)。対して、首振りの場合は ave_sum_x が一定値以上で推移し、かつ ave_sum_y が相対的に低い値で推移し続ける状態となる (図 8)。この検出方法では、起動時の値が x 座標、y 座標ともに高くなるため、どちらも高い値の場合は判定から除外する。傾きおよび首振りに対して、各変量をもとにした大きな動作と小さな動作に分ける判定を行っているが、その閾値については現時点では実験的に結果を見ながら調整している。

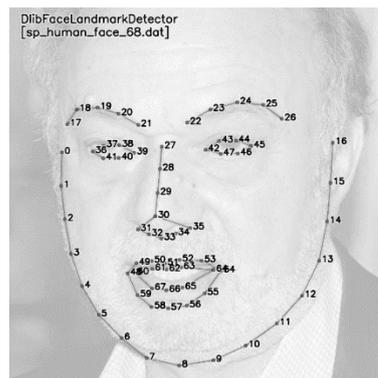


図 6 本アプリケーションにて検出する特徴点一覧[10]

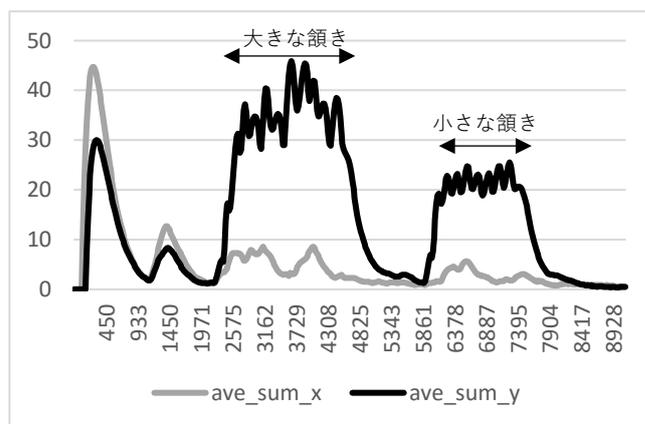


図 7 傾きの検出における変量の推移

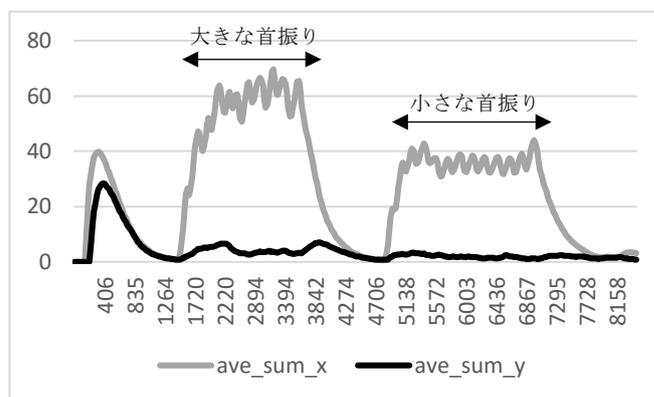


図 8 首振りの検出における変量の推移

3.3 システム構成

前述のアプリケーションデザインに基づいて構築したプロトタイプシステムの構成を示す。

本研究では、既存のオンラインミーティングツールを基盤としたアプリケーション開発を行う。Web カメラを用いてユーザの上半身を含む映像 (横 640pixel×縦 480pixel) を取得する。フレームレートは 60fps である。

使用する Web 会議システムはクラウド型ビデオチャットサービスである Zoom とした。開発ツールはゲームエン

ジン Unity を使用する。顔の特徴点の検出のため OpenCV for Unity と Dlib FaceLandmark Detector という Asset を用いた。顔の特徴点から 3D アバタの頭部動作と口や目の動きを制御するために CV Vtuber Example という Asset を利用し、3.2 で記述したコミュニケーション動作の識別を行った。使用する 3D アバタは VRM モデルを利用した。

ユーザが利用する本アプリケーション画面は、OBS Studio の仮想カメラ機能を用いて Zoom のカメラ入力として利用することとした。

4. アプリケーションデモンストレーション

アプリケーションの実現可能性を検証するため、制作したアプリケーションのデモンストレーションを行った。デモンストレーションでは、簡易な模擬授業と質問紙調査、インタビューを行った。

4.1 デモンストレーション設定、および手順

デモンストレーションは立命館大学院生 6 名を対象に行った。各参加者は一人一台の端末を用いて、Zoom を通じてのみ参加した。

今回のデモでは、被験者 6 名全員が制作したアプリケーションを使用して参加し、教員役として参加する実験者のみビデオ機能をオンにした通常のビデオ通話を利用した。アプリケーション内で使用する 3D アバタについては、3D キャラメイカー VRoid Studio にて提供されているサンプルモデルを改変して制作した。男性と女性の二種類を用意し、被験者がいずれかを選択した。

まず、首振り・顔きの機能を実験者の指示に合わせて被験者全員が使用した。その後、本アプリケーションを用いた意思表示の練習として簡単なクローズ質問（「朝食を食べましたか」）を行った。簡易的な模擬授業の再現として、本研究の説明を 5 分間行い、この間にもクローズ質問（「遠隔授業においてライブ配信型を最も多く利用しますか」、「遠隔授業で普段ビデオ機能をオンにしますか」）を 2 回実施した。その後、質問紙調査とインタビューを行った。

4.2 デモンストレーション結果

図 9・図 10 はデモンストレーション内にて、被験者が顔きや首振りの機能を利用した画像である。



図 9 顔きの機能を利用したとき



図 10 首振りの機能を利用したとき

質問紙調査では、「教員役が学生の状態を把握していた」「自分が授業に参加している意識を持てた」「自分の意思表示を行うことができた」「他の学生の状況を確認できた」という 4 つの項目についての回答を図 11 に示す。これらの質問項目については、五段階のリッカート尺度（全くそう思わない/あまりそう思わない/どちらともいえない/まあまあそう思う/正にそう思う）により回答を求めた。現時点ではデモンストレーション段階に留まるものの、どの質問においても肯定的な意見（まあまあそう思う/正にそう思う）が 3 人以上であり、少なくとも過半数は双方向コミュニケーションを促進する効果について肯定的であると読み取れる。

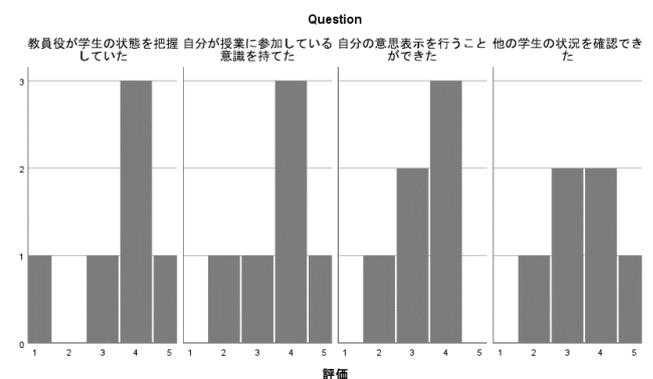


図 11 質問紙調査結果

インタビュー・自由記述においては、ビデオ機能をオンにすることに対して億劫に感じるという理由から、3D アバ

タを利用することに肯定的な意見も見られた。意思表示の機能に対しては、現時点ではかなり自覚的に動きを行う必要があるものの、意思が相手に伝わることを通常のライブ配信型遠隔授業よりも意識したという回答を得られた。

対して、遅延や検出の精度については複数の参加者から問題として指摘された。例えば、図 9・図 10 に見られるように、全員が頷きや首振りを行うよう指示した場合、被験者 6 名中 4 名は頷きの状態に遷移しており、残り 2 名は通常状態のままである場面もあった。動作検出の精度については、アプリケーションを実行させるための PC 環境や頭部動作の検出方法について引き続き精査が必要である。3D アバタについては、アニメ調の 3D モデル男女 2 種類しか用意していなかったことにより、自己と同一視できないという回答があった。

5. おわりに

本論文では、ライブ配信型遠隔授業で利用するオンラインミーティングツールの活用において、3D アバタと顔認識技術を利用した、使用者の動作の伝達と意思表示の強調を行うアプリケーションの提案を行った。検出の精度や使用する 3D アバタなど問題も見られるものの、デモンストレーションが示すように、双方向コミュニケーションの促進を目的としたアプリケーションの基本的な枠組みを示すことができた。

今後は、デモンストレーションで示された問題点の解決に取り組むとともに、ライブ配信型遠隔授業の実態分析を進めながら、意思表示の強調方法の改善や、頷きと首振り以外の意思表示の検討も行っていきたい。

謝辞 実験にご協力していただいた立命館大学映像研究科の方々に感謝の意を示します。本研究の一部は JSPS 科研費 21K12004 の助成を受けたものです。

参考文献

- 1) 神奈川大学 教育支援センター: 「遠隔授業の有効性と課題」に関する調査アンケート: 学生向け 集計結果 (学外公開用), (https://www.kanagawa-u.ac.jp/att/20645_48473_010.pdf) (2020/9/4) (2021/07/18 取得)
- 2) 立命館大学 教育学部 教務課: 学びと成長 レポート 特別号, (<http://www.ritsumei.ac.jp/file.jsp?id=472823>) (2021/07/18 取得)
- 3) 神奈川大学 教育支援センター: 「遠隔授業の有効性と課題」に関する調査アンケート: 教員向け 集計結果 (学外公開用), (https://www.kanagawa-u.ac.jp/att/20645_48474_010.pdf) (2020/9/4) (2021/07/18 取得)
- 4) Apple 社: iPhone や iPad Pro でミー文字を使う (<https://support.apple.com/ja-jp/HT208986>) (2021/07/25 取得)
- 5) LINE Corporation: アバタの作り方・遊び方 (<https://guide.line.me/ja/account-and-settings/account-and-profile/avatar.html>) (2021/07/25 取得)
- 6) 亀岡 嵩幸, 佐瀬 一弥, 吉元 俊輔, 蜂須 拓, 嵯峨 智, 黒田 嘉宏, 矢内 智大, 石丸 翔也, 川口 碧, 清川 清, ヤエム ヴィボル,

- 白井 暁彦: 学会オンライン化・VR 開催の幕開け, 日本バーチャルリアリティ学会誌, Vol. 25, No. 2, p. 35-43 (2020/06/30)
- 7) 山本健太, 尾上聡, 田中一晶, 中西英之: 「動き」と「外見」どちらの伝達が重要か: アバタがビデオを代替する可能性, 研究報告 ヒューマンコンピュータインタラクション, vol.2011-HCI-144, No. 11, p.1-7 (2011/07/21)
 - 8) 瀬島吉裕, 渡辺富夫, 山本倫也: うなずき反応モデルを重畳した VirtualActor を介する身体的コミュニケーションの合成的解析, 日本機械学会論文集 C 編, Vol.75, No.758 (2009)
 - 9) 平林 信隆: コロナ禍における大学のオンライン授業に対する新入生の認識についての探索的研究, 共栄大学研究論集, No.19, pp.55-66 (2021-03)
 - 10) Dlib FaceLandmark Detector: (<https://enoxsoftware.com/dlibfacelandmarkdetector/documentation/q-a/>) (2021/07/18 取得)