



ELHR 開閉眼抽出法を用いた自己アバターの鏡像認知実験

Mirror Image Recognition Experiment of Self-Avatars Using ELHR Blink-Triggered Extraction Method

戸田壮駿¹⁾, 田中久弥²⁾

Masatoshi TODA, and Hisaya TANAKA

- 1) 工学院大学大学院 工学研究科情報学専攻 (〒192-0015 東京都八王子市中野町 2665-1, em23030@ns.kogakuin.ac.jp)
2) 工学院大学 情報学部情報デザイン学科 (〒192-0015 東京都八王子市中野町 2665-1, hisaya@cc.kogakuin.ac.jp)

概要: 開閉眼の変化はメタバースにおけるアバターコミュニケーションの促進に有効である。また、システムによる自動的な変化よりもユーザの生体情報に基づく変化の方が、より効果が期待される。本研究では独自の開閉眼抽出指標である ELHR を Web カメラから取得し、アバタに反映させるシステムを構築した。そのシステムを用いて鏡像認知実験を行った結果、自動的な瞬目付与よりも認知スコアが向上したことが確認された。

キーワード: アバターコミュニケーション, MediaPipe, 開閉眼, メタバース

1. 序論

新型コロナの世界的流行後、メタバースというインターネット上の 3D 空間におけるアバタを用いたコミュニケーションが一般的になった。東京都では子供たちの新たな学び場として“バーチャル・ラーニング・プラットフォーム”というメタバースを運用を行うことを目指している[1]。また、大日本印刷は情報発信や販促活動の場所として“バーチャル秋葉原”というメタバースシステムを公開した[2]。メタバースを用いることで川北らはカウンセリングの相談意欲が変化したということを示している[3]。また、船木らはアバターをフェイストラッキングで動かすことでアバターへの親近感を持ちやすいことを示している[4]。そのため、アバタをフェイストラッキングを用いてリアルタイムで動かすことは、メタバース上でのコミュニケーションにおいて重要であると考えられる。しかし、フェイストラッキングのリアルタイム処理は重いため、幅広くメタバースにおいて使用されていない。

今日のメタバースシステムの傾向には 2 種類あるためそれぞれを説明する。1 つ目は、カスタマイズ性に重きを置いたシステムである。具体的には、VRchat や NeosVR, Virtual cast が挙げられる。このジャンルに属するメタバースシステムでは、ユーザ側で幅広いことが作成出来ることに加えて Vive Pro Eye のような特定の器具を用いることで顔の非言語情報をアバタに付与することができるが特定のハードウェアを必要とする。もう 1 つは、動作環境の軽量化に重きを置いたシステムである。具体的には cluster や door, Vket cloud などが挙げられる。このシス

テムに属するメタバースシステムでは、スマートフォン等で動作する程の軽量動作を実現している。しかし、顔の非言語を用いて動作することは対応していない。

そこで筆者らは ELHR という新たな開閉眼の抽出を提案し、MediaPipe を用いて実装し、提案手法が他の手法と比較してリアルタイムでの精度の優位性を示した[5]。本研究では ELHR によって得られた開閉眼をアバタに付与し、その動作がユーザにとって意図通りの投影がされているかについてアバタを自分の顔の動作を鏡のように反映させる実験である鏡像認知実験を行うことで確認することを目的とする。

2. 鏡像アバターの作成方法

2.1 頭部動揺についての抽出方法

頭部動揺は 3 軸あり、x 軸を首の横振り、y 軸を頭の傾げとする、z 軸を首の縦振り。x 軸は、左右の目頭と目尻の幅を抽出して、左右の比率を見ることで首の横振りの回転の度合いを推論している。x 軸の角度を算出する式を式 1 に示す。y 軸は眉間と顎の中央の幅の距離を算出することで頭を傾げたときの角度を推論している。y 軸の角度を算出する式を式 2 に示す。z 軸は鼻尖と鼻柱の距離の比率を見ることで首の縦振りの角度を推論している。z 軸の角度を算出する式を式 3 に示す。なおこの角度は MediaPipe で顔を捉えられていることが必須であり、角度の誤差としては 5%未満であり、アバタへの表出を行う際に誤った動作がなく自然な動作であることを確認した。

$$\theta_x = \frac{\text{Left eye wide}}{\text{Right eye wide}} \quad (1)$$

$$\theta_y = \frac{\text{glabella wide distance}}{\text{thin center wide distance}} \quad (2)$$

$$\theta_z = \frac{\text{tip of the nose} - \text{bridge of nose}}{\text{nasal bridge} - \text{center of the middle fossa}} \quad (3)$$

2.2 開閉眼特徴量 ELHR の抽出方法

開閉眼抽出は筆者らの先行研究で構築した ELHR という計算式を用いた[5]。ELHR とは、瞼の外側の距離(EL_out)と瞼の内側の距離(EL_in)の比率である。ELHR の計算方法のフローを図 1 に示す。この時の顔が傾いた角度を“ θ_z ”とし、角度の基準は正面を見たときとする。その角度を用いて距離を計算する。外側の計算方法を式 4、内側の計算方法を式 5、ELHR の計算方法を式 6 にそれぞれ示す。開閉眼の取得には、実際の開眼時と閉眼時の二つの状態のデータを用いて Min-Max 法で正規化する必要がある。正規化を行うことで、0 が閉眼状態、1 が開眼状態となる。この時の正規化の式を式 7 に示す。

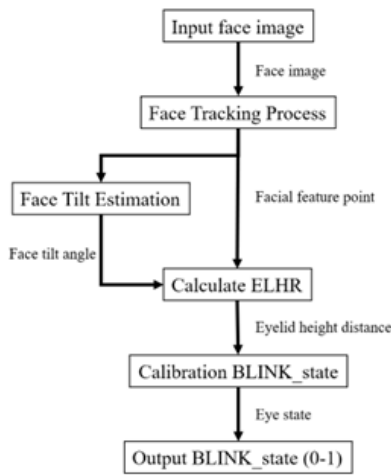


図 1 ELHR の算出フロー図[5]

$$EL_{out} = \frac{EL_{out_upper} - EL_{out_lower}}{\theta_z} \quad (4)$$

$$EL_{in} = \frac{EL_{in_upper} - EL_{in_lower}}{\theta_z} \quad (5)$$

$$ELHR = \frac{Eyelid_in_distance}{Eyelid_out_distance} \quad (6)$$

$$Blink_state = \frac{ELHR - ELHR_Min}{ELHR_Max - ELHR_Min} \quad (7)$$

2.3 アバタ表情の制御方法

本研究において、被験者の顔画像から頭部動揺と ELHR を用いて開閉眼をリアルタイムで取得して、アバタへの付与をリアルタイムで行った。本研究で用いた 3D モデルは汎用的な 3D モデル規格である VRoid に基づいたモデルを使用した。まず、頭部動揺についてだが、3D アバタの頭部

を制御するボーンである Head ボーンに 3 軸の回転軸を代入することで実現した。次に、開閉眼については、アバタのオブジェクトにポリゴンの移動として記録されている Blendshapes の瞬きを制御するパラメータに得られた開閉眼を代入した。なお、この時の開閉眼は自然な開閉眼に見せるために、左右それぞれの開閉眼の平均を取ったデータを使用した。この開閉眼の抽出を行いアバタに付与させた例を図 2 に示す。

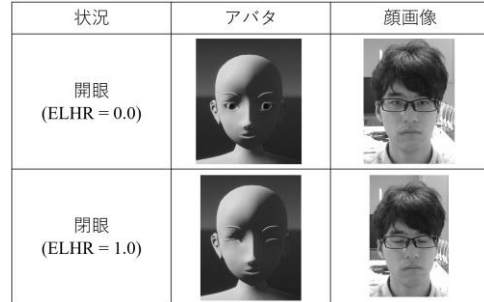


図 2 ELHR を付与させたアバタと顔画像

3. 実験

3.1 実験協力者

被験者は、モニターの映像がはっきりと見える視力を有した 19 歳～22 歳の工学院大学の学生 13 名とした。そのうち、眼鏡の着用者は 4 名であった。本実験で用いる開閉眼抽出のシステムは筆者らの先行研究によって、眼鏡の着用の有無によって大きく開閉眼の抽出精度は 6%以内である[5]。そのため眼鏡の着用の有無によってデータを分けなかった。被験者には実験の前に実験概要を説明し、不明点がないことを確認したことを書面にて署名もらった。なお本実験は工学院大学の倫理審査“新しいインターフェース開発のための心理生体計測 2021-A-29”に準拠して行った。

3.2 実験環境

実験システム概要は図 3 に示す。被験者の顔画像の入力には Web カメラ (BUFFALO 社の BSWBK10H) を用いた。その時の解像度は VGA 画質 (640px×480px) とし、30fps での記録を行った。開閉眼抽出には python 3.9.13 と MeidaPipe 0.10.10 を用いて作成し、exe ファイルにビルドした。

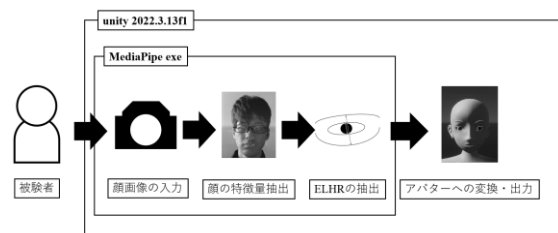


図 3 実験システム図

exeファイルにした理由としては、処理速度の向上によって、安定した処理を実現するためである。アバタ提示部分にはゲームエンジンであるunity(2022.3.13.f1)を用いて作成し、開閉眼抽出を行ったexeファイルからの出力を受け取る形でリアルタイムに動作させた。本実験の提示にはOSにwindows 10 home, CPUにintel製のcore i7 12700K, GPUにNvidia製のRTX A4000を搭載したデスクトップPC, 24インチのFHDモニター(BenQ社のGL2450-T)を用いた。

3.3 実験条件

アバタ提示方法は3種類ある。それぞれ、アバタ静止条件、瞬きランダム条件、瞬き+頭部動揺条件である。この3種類の提示を1セットとして1セットを実験として実施した。また、順序効果の影響を無くすために提示する条件はランダムで行った。また、使用したアバタは被験者の好みの影響を軽減するために真っ白な物を使用した。瞬きランダム条件における瞬きはプログラムで制御し、開閉間隔は、一般的な瞬き間隔である4秒とした[6]。瞬き+頭部動揺条件においてアバタには、左右の開閉眼、頷きと顔の横振りを表す二軸の頭部動揺を反映させた。例えば、開閉眼を表すblink stateが閉眼を表す1.00だった場合、アバタ静止条件では0.00であり、瞬きランダム条件では1.00、瞬き+頭部動揺条件では1.00であった。また、blink stateが閉眼を表す0.26だった場合、アバタ静止条件では0.00であり、瞬きランダム条件では1.00、瞬き+頭部動揺条件では0.26であった。この時の瞬きランダム条件では実際の開閉眼と自動で付与された瞬きが異なることを示している。この2つの条件を実際にアバタに付与させたときの違いを図4に示す。









入力顔画像	アバタ静止条件	瞬きランダム条件	瞬き+頭部動揺条件
 Blink state = 1.00	 Blink state = 0.00	 Blink state = 1.00	 Blink state = 1.00
 Blink state = 0.26	 Blink state = 0.00	 Blink state = 1.00	 Blink state = 0.26

図4 実験条件ごとのアバタ反映の違い

3.4 実験タスク

鏡像認知実験の比較実験として、部分的にアバタのフィードバックを付与した3つの条件を設定し共通のタスクを行った。タスクとしては、提示したタイミングでの眼の開閉をそれぞれ3回ずつ行った。このタスクは、開閉眼抽出の際に用いるキャリブレーションの処理であるが、全ての条件において実施した。処理が成功したことを確認した後に、被験者の前にアバタを表示した。次に、被験者の任意のタイミングで、瞬きかウインクをさせて、アバタへのリアルタイムの表出を確認させた。また、モニター上のアバタが確認できる範囲で、左右を見渡すように顔を振り

頷きをそれぞれ二回ずつさせた。そして、タスクが終わった段階で被験者に“アバタが自分の意図したとおりに動作しているか”と、“アバタがまるで鏡のように感じた”という2つの質問に対して、1は“強くそう思わない”, 2は“ややそう思わない”, 3は“どちらかといえばそう思わない”, 4は“中立”, 5は“どちらかといえばそう思う”, 6は“やや思う”, 7は“強くそう思う”という7件法の主観評価をした。鏡像実験の様子を図5に示す。



図5 鏡像実験の様子

4. 結果

結果として採用したのは2セット目と3セット目のデータだけを使用した。1セット目のデータを利用しなかった理由としては、アバタ静止条件において、被験者の半数が挙動をエラーと誤解し、実験が一時的に中断したためである。得られたデータは7段階で得られて2回の点数について、合計値を求めた。

まずは、“アバタが自分の意図したとおりに動作しているか”についてのアンケート結果を示す。合計値の値が小さいほど、被験者がアバタの動作を意図的に動作したと感じたと評価した結果である。その結果はアバタ静止条件では合計値の平均が0.08点、瞬きランダム条件での合計値の平均が1.38点、瞬き+頭部動揺条件での合計値の平均は8.69点となった。このことから、瞬き+頭部動揺条件が他の条件と比較して合計の平均が大きく低下し、被験者の意図した通りの挙動をしたと評価されたと明らかになった。これらのデータに対してフリードマン検定とボンフェローニの多重比較を用いて検定したところ、被験者の瞬き+頭部動揺条件と比較して1%の有意差があることが認められた。次に、“アバタがまるで鏡のように感じた”についてのアンケート結果を示す。合計値の値が小さいほど、被験者がアバタの動作を見て鏡であるように感じたとして評価した結果である。その結果はアバタ静止条件では合計の平均が0.15点、瞬きランダム条件では合計の平均値が0.85点、瞬き+頭部動揺条件では合計の平均値が7.31点となった。このことから、瞬き+頭部動揺条件が他の条件と比較して平均値が大きく低下したことが明らかになった。このデータに対してフリードマン検定

とボンフェローニの多重比較を用いて検定したところ、瞬き+頭部動揺条件と瞬きランダム条件には1%の有意差があることが認められた。この結果を図6と図7にまとめる。

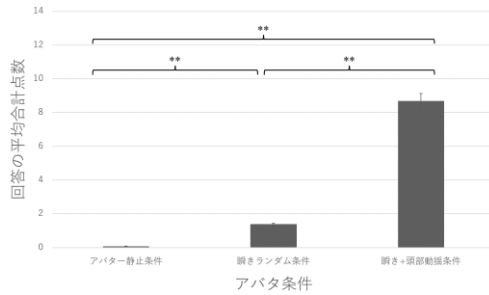


図6 “アバタが自分の意図したとおりに動作しているか”の合計評価得点

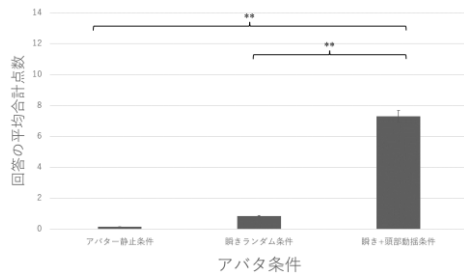


図7 “アバタがまるで鏡のように感じた”の合計評価得点

5. 考察

被験者の瞬き+頭部動揺条件が他の条件と比較して最も被験者の意図通りの動作をしていると評価された。これは、鏡像認知実験として、アバタに最も多い非言語情報を付与したことためであると考えられる。また採用した開閉眼抽出のシステムが筆者らの先行研究に示す通りに精度良くリアルタイムでの解析を行うことを実現したためだと考えられる。このことから本研究で開発したシステムはユーザの意図した通りにアバタを制御することが考えられる。また、鏡のように感じたかという評価については、他の2つの条件と比較して有意差があることが示された一方で、最も分散が大きく、点数差にすると7点差が生じている。これは、被験者の顔とは全く異なる姿のアバタと背景を使用したため、そのアバタへの自己投影性が低下したことが原因と考えられる。加えて、アバタに付与する際に MediaPipe で解析されたデータを

そのまま適用したため、微細なノイズが、アバタ提示の際の細かな揺れに繋がってしまい、その結果鏡のように感じなかったという被験者が低評価をつけている傾向にあった。より高評価にするために、移動平均などの動作を補完関数を用いてノイズを除去する必要である。また、口の動きや、顔色の変化などその他の顔の非言語情報を付与させてアバタがより鏡像に近づける必要があると考えられる。

6. 結論

本研究では、ELHR 開閉眼抽出のシステムを用いて、鏡像認知実験を行いユーザの意図した通りに動作するのかを評価させた。その結果、全ての被験者において瞬き+頭部動揺条件がユーザの意図した通りの動作をしたと高い評価された。今後としては、実際のメタバースシステムに組み込むことを行い、実際にメタバースでの実験を行いその有用性について示すことが必要である。

参考文献

- [1] 東京都教育委員会: 仮想空間を活用した新たな児童・生徒支援を開始します 「バーチャル・ラーニング・プラットフォーム」について; https://www.kyoiku.metro.tokyo.lg.jp/press/press_release/2022/release20221202_01.html, (2024年7月5日参照)
- [2] 大日本印刷: VIRTUAL AKIHABARA; <https://virtual-akihabara-web.com/>, (2024年7月5日参照)
- [3] 川北 輝, 大西 徹, 石原 茂和, 橋本 健汰, 金井 秀明: カウンセラーアバターの外観とユーザーの相談状況の違いが相談意欲に及ぼす影響の検討, 日本感性工学会論文誌, 2022年21巻3号 p. 267-274, <https://doi.org/10.5057/jjske.TJSKE-D-22-00001>
- [4] 舟木 烈, 物部 寛太郎: フェイストラッキングを用いたアバターの親近感を高める手法に関する研究, 第27回日本バーチャルリアリティ学会大会論文集 (2022年9月)
- [5] Masatoshi TODA, Hisaya TANAKA: Construction of an open/closed eye extraction system for live capture, International Journal of Affective Engineering, 2023, <https://doi.org/10.5057/ijae.IJAE-D-23-00030>
- [6] L.G.CARNEY, R.M.HILL: THE NATURE OF BLINKING PATTERNS, Acta Ophthalmologica, 1982, <https://doi.org/10.1111/j.1755-3768.1982.tb03034.x>