

IMU 搭載ヘッドホンを用いた注視位置の簡易推定

Estimating gaze position using IMU-equipped headphone

森 友吾[†]
Yugo Mori[†]

佐藤 光起^{††}
Koki Sato^{††}

松下 光範[†]
Mitsunori Matsushita[†]

1 はじめに

スマート照明やインテリジェント空調などの Internet of Things (IoT) 機器の普及に伴い、我々の周辺に計算機で制御可能なデバイスが増加している。これらのデバイスはインターネットを介して接続されたスマートフォンやタブレットによって操作することができ、細やかな設定が可能になっている。さらに、これらのデバイスの操作を直観的に行うため、音声による操作手法も提案されている [5]。

しかし、こうしたデバイスが同一空間に多数存在する環境下では、意図したデバイスを選択する場合にそのデバイスを選択することが必ずしも容易ではないという問題がある。例えばスマート照明が複数設置されている環境の場合、個々のスマート照明には各々識別子が割り当てられているものの、ユーザは、意図したスマート照明の識別子が何であるかを把握していなければ操作することができない。

そこで本研究では、「操作対象を見る」という行為に着目し、デバイスの選択を行う方法を検討する。ユーザが操作したい対象を見て確認する際に生じる頭の動きに基づいて、注目対象を推定できることが知られている [8]。この知見に照らすと、対象を注視する際に伴って生じる頭部姿勢の動きを検出し、頭部姿勢から操作対象となるデバイスを推定することで、ユーザが操作対象を指定するプロセスを簡略化できると期待される [9]。頭部姿勢を用いた対象の選択としては Kytöらの手法 [4] などが提案されている。これらの手法は、高い精度での識別が可能であるが、カメラや頭部装着型のデバイスなど、専用機器を用意する必要があるため使用できる状況が限定される。この課題を解決するため、広く普及しつつある AirPods (©Apple Inc.) などの慣性計測センサー (Inertial Measurement Unit; 以下, IMU) 搭載型ヘッドホンを用いて操作対象を直感的に指定できる手法の実現をめざす。

提案手法では、IMU 搭載ヘッドホンから取得した回転角に基づいてユーザの頭部姿勢が示す先の座標を簡易に推定する。本稿では、ユーザが実際に注視した位

置と、提案手法によりシステムが推定した座標とを比較する実験を通じてその有効性を検証するとともに、提案手法の活用方法について議論する。

2 関連研究

これまでに、頭部姿勢を用いた様々なインタラクショ手法が提案されている。ここでは、頭部姿勢の代表的な検出手法であるカメラを用いた手法、IMUを用いた検出手法、および身体動作を用いた対象推定方法とその活用法について述べる。

2.1 カメラによる頭部姿勢の推定と活用

ユーザの様子をカメラで撮像することで頭部位置を検出する手法がある。

Stiefelhagen らは、全方位カメラを用いて人の頭部を撮像することで頭部姿勢を推定し、会議中の人の注目位置を検出する手法を提案している [8]。Hueber らは、スマートフォンのカメラを用いて頭部姿勢の位置と向きを検出し、ジェスチャとして対応づけることで、コピー・ペーストなどのショートカット機能を実行するシステムを提案している [2]。

これらの手法はカメラの視野内で行われたジェスチャ動作を認識対象とするため、ユーザはカメラの視野範囲内に居る必要があるという制約がある。さらにカメラを用いることは、プライバシー上の懸念も伴う。

2.2 IMU を用いた頭部姿勢の推定と活用

IMU を備えたデバイスをユーザの頭部に装着することで頭部姿勢を検出する手法もある。

Kim らは、IMU を搭載したイヤホンの加速度情報を取得し、機械学習を用いることで、学習者がモニタと机のいずれに注目しているのかを推定する手法を提案している [3]。Panda らは、ヘッドホンに IMU, LiDAR, タッチセンサなどのセンサを追加することで、ジェスチャや身体の動き、視線、環境文脈などを取得し、ユーザインタラクションに活用することを提案している [6]。この手法において IMU は、頭部の回転を検出することで、デバイスを切り替えたり、自動操作を行ったりするために用いられる。

これらの手法ではユーザに装着したセンサを用いるため、カメラのように認識できる範囲が限定されない利点がある。また、用いる情報は IMU から取得される角度や加速度であるため、カメラによる撮影画像を用

[†] 関西大学大学総合情報学部, Faculty of Informatics, Kansai University

^{††} 関西大学大学院総合情報学研究科, Graduate School of Informatics, Kansai University

〒 569-1095 大阪府高槻市霊仙寺町 2 丁目 1-1

いることに比べてプライバシーの問題が軽減される。しかし、これらの手法では事前に対応づけられた動きのパタンに基づいて対象を選択するため、選択できる操作対象の数が事前に対応付けられたジェスチャパタンの数に制限される。

2.3 身体動作を用いた対象指定

手の動きや手に持ったデバイスを対象に向けることで、操作対象の選択を可能としている手法も提案されている。Bolt らの研究を始めとし、手の動作で機器を指定し、音声によって指定した対象を操作する手法が提案されている [1]。Sato らの研究では、スマートフォンに搭載されているセンサからデバイスの回転角を用いることで、スマートフォンのポインティング位置を推定し、IoT の機器の指定を可能としている [7]。しかしこの手法では、機器指定を行った後、スマートフォンと操作したい対象を交互に見ながら操作する必要があり、操作が煩雑である。

3 デザイン指針

ここまで述べてきたように、頭部姿勢を用いたインタラクションにおいて、「ユーザが意図する対象に顔を向ける」動作を検出することにより注視している座標を推定し、それに基づいて操作対象を指定できれば、ユーザの操作プロセスを簡略化できると期待される。一方で頭部姿勢の推定を行う際にカメラなどの専用機器を用いることは、頭部姿勢の推定できる状況が制限される。特にカメラやマイクを用いることは、画像や音声を取得することがユーザにとって心理的な負担を生じさせる恐れがあり、日常利用に適さない。本研究ではこれらの背景を踏まえて

1. 頭部の動きを捉えることができること。
2. ユーザの注視位置が簡易的に推定できること。
3. 汎用されるデバイスを用いて実現できること。
4. カメラやマイクを用いることなく実現できること。

の4項目をシステム要件として定義した。

4 実装

本研究では、3章で定義したシステム要件を満たすために、IMU を搭載した汎用のヘッドホンを用いて注視位置の推定を行う。IMU 搭載ヘッドホンは、ユーザの頭部の回転角情報を取得する。取得した回転角情報に基づいて、Sato らの手法 [7] を用いて位置推定を行う。Sato らの手法では、座標が既知の基準点に対してユーザがキャリブレーションを行うことで、ユーザが示している座標を推定する。本研究では、IMU 搭載ヘッドホンによって得られた頭部の回転角情報に基

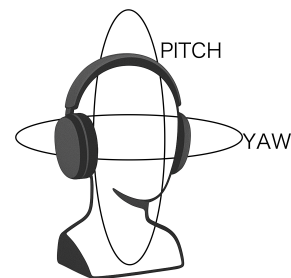


図 1: ヘッドホンで取得できる回転角

づきキャリブレーションを行うことで、ユーザが注視している座標を推定する。

4.1 システム構成

提案システムは、ユーザが装着する IMU 搭載のヘッドホン (AirPods Max) と、計算処理を行う PC (MacBook Air (M2, 2022)) とにより構成される。これらのデバイスは Bluetooth を介して接続され、相互に通信を行う。本システムは、オーバーヘッド型ヘッドホン (ヘッドホン) に限らずインイヤー型ヘッドホン (イヤホン) による実装も可能であるが、本稿ではユーザ実験を想定し、着脱が容易なオーバーヘッド型ヘッドホンを用いて実装を行った。

提案システムにおいてヘッドホンの回転角は Swift の CMHeadphoneMotionManager¹ を使用して取得する。これにより、ヘッドホンの姿勢情報 (roll, pitch, yaw) をリアルタイムに取得できる。得られたヘッドホンの姿勢情報は PC アプリケーションに送信され、推定処理が行われる。推定処理を行う PC アプリケーションは Swift を用いて作成した。

4.2 推定手法

ヘッドホンを装着したユーザはキャリブレーションとして平面左上の第 1 の基準点を注視してマウスをクリックすることで、基準点を注視した際の頭部姿勢の回転角を登録する。次にユーザは、平面右下の第 2 の基準点を注視して同様に登録操作を行う。

推定処理を行う PC アプリケーションは、キャリブレーション時に登録された頭部姿勢の回転角と、現在の回転角の情報とを用いて、注視している座標を推定する。

$$x = w_2 + \frac{w_1 - w_2}{\alpha_2 - \alpha_1}(\alpha - \alpha_1) \quad (1)$$

$$y = h_1 + \frac{h_2 - h_1}{\beta_2 - \beta_1}(\beta - \beta_1) \quad (2)$$

注視座標の推定処理は、式 (1) によって線形補間を行うことで実行する。なお座標系は、左上を原点と

¹ <https://developer.apple.com/documentation/coremotion/cmheadphonemotionmanager> (2025/5/26 確認)。

し、原点から右方向に x 軸を、下方向に y 軸を取る。式 (1) において x, y は注視していると推定される x, y 座標を、 α, β は現在の **pitch, yaw** 値を、 w_1, h_1 は平面左上の第 1 基準点における x, y 座標を、 w_2, h_2 は平面右下の第 2 基準点における x, y 座標を、 α_1, α_2 は第 1 基準点、第 2 基準点において登録された **yaw** 値を、 β_1, β_2 は第 1 基準点、第 2 基準点において登録された **pitch** 値を表す。

5 実験

提案手法によってユーザの頭部姿勢に基づいて推定される座標と、実際に注視している座標との間には誤差があることが想定される。提案手法は頭部姿勢に基づいており、注視において用いられる眼球の動作は考慮していないためである。そこで本稿では、実験を通じて提案手法の精度や特徴を明らかにするとともに、提案手法の利用可能性について議論する。

5.1 実験手法

実験では、現実空間の平面上に設置したターゲットの座標と、実験協力者がターゲットに対してターゲティング（目標を注視する行為）を行った際に推定した注視位置の座標との差を比較をすることで検証を行う。実験参加者は提案手法におけるキャリブレーションを行った後、壁に配置された 9 つの点に対してターゲティングを行う。実験システムは、実際のターゲットを注視した時の座標を推定する。

実験では幅 240cm、高さ 240cm の、壁に設置した平面を対象に検証を行った。平面上にはターゲットとして半径 10cm の円を 9 つ、図 2 に示す位置に配置した。実験参加者はこの平面に正対して直立し、各ターゲットに対して指示された順序でターゲティングを行う。実験参加者の直立する位置は、平面からそれぞれ 100cm、200cm、300cm 離れた 3 地点とし、立つ位置の順序は実験対象者ごとにランダムに行った。

実験は、実験参加者が IMU 搭載ヘッドホン装着し、トラックパッドを把持した状態で行った。実験参加者は、所定の位置に直立し、キャリブレーションとして平面上に設置した左上の第 1 基準点を注視した後、手元のトラックパッドをクリックする。右下の第 2 基準点についても同様にキャリブレーションを行う。キャリブレーションが完了したことを確認した後、実験参加者に注視するターゲットをランダムな順序で指示する。実験参加者は指示された注視対象を注視した後、トラックパッドをクリックする。システムはトラックパッドがクリックされた時の **pitch, yaw** の値、及び座標の推定値を記録する。それぞれの直立位置において、各ターゲットが 2 回ずつ選択されるよう指示

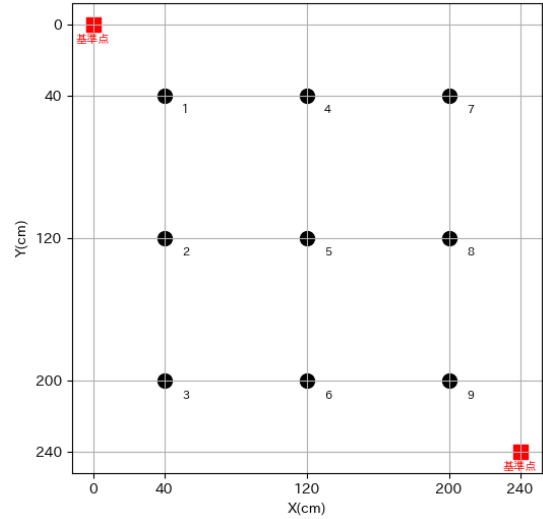


図 2: 実験環境におけるターゲットの配置

を行う。すべての地点において同様の手順で計測を行い、実験協力者毎に計 54 回 (9 ターゲット $\times$ 2 回 $\times$ 3 地点) のターゲティング結果を測定した。実験中の様子はカメラにより録画するとともに、ユーザの身長を口頭での質問により記録した。

5.2 実験結果

5.2.1 概要

実験は情報系の学部・大学院に所属する学生 25 名を対象とした。実験では、実際のターゲットの位置と、それを実験対象者が注視した際にシステムが計算した推定位置および回転角である **pitch, yaw** の値を取得した。計測したデータにおいて、左上の基準点の設定を行った際の **pitch, yaw** 値が、右下の基準点の **pitch, yaw** 値より小さい値を取る実験参加者が 1 名存在した。通常、本実験条件では基準点の位置関係から、左上を注視した際の **pitch, yaw** 値は右下を注視した際の値よりも大きくなるため、その値が正常に記録されていないこのデータを外れ値として除外した。

5.2.2 実験結果の分析

ターゲット $T_i (i \in [1, 9])$ の座標を (X_i, Y_i) 、IMU ヘッドホンにより計測したその推定座標 E_{T_i} を $(\hat{x}_{T_i}, \hat{y}_{T_i})$ とするとき、推定座標のターゲットからのずれを変位 $(\delta X_i, \delta Y_i)$ 、推定座標のターゲットからの誤差を座標軸毎の絶対誤差 $(\Delta X_i, \Delta Y_i)$ 、距離 R_{T_i} とし、算出には以下の式を用いた。

$$(\delta X_i, \delta Y_i) = (\hat{x}_{T_i} - X_i, \hat{y}_{T_i} - Y_i) \quad (3)$$

$$(\Delta X_i, \Delta Y_i) = (|\delta X_i|, |\delta Y_i|) \quad (4)$$

$$R_{T_i} = \sqrt{(\delta X_i)^2 + (\delta Y_i)^2} \quad (5)$$

α_1, α_2 は左上第 1 基準点、第 2 基準点において登録された **yaw** 値を、 β_1, β_2 は右下第 1 基準点、第 2 基準

表 1: ターゲティング距離ごとの誤差と回転差の平均

距離	誤差平均 [cm]			回転差平均 [°]	
	δX	δY	R_T	P	Y
100	19.8	32.5	41.8	34.6	61.5
200	28.8	37.9	53.0	23.8	36.3
300	27.0	48.8	60.4	19.1	27.8

表 2: ターゲット位置ごとの距離 R_T の平均

		左	中央	右
上	id	①	④	⑦
	R_T	53.3	49.6	54.4
中央	id	②	⑤	⑧
	R_T	49.1	47.8	49.2
下	id	③	⑥	⑨
	R_T	55.7	53.4	53.2

点において登録された **pitch** 値を表し、左上の第 1 基準点の **pitch**, **yaw** 値と、右下の第 2 基準点の **pitch**, **yaw** 値との差を回転差 ΔP , ΔY とし、算出には以下の式を用いた。

$$\Delta P = \beta_2 - \beta_1 \quad (6)$$

$$\Delta Y = \alpha_2 - \alpha_1 \quad (7)$$

表 1 はターゲティング平面からの距離ごとに、 ΔX_i , ΔY_i , R_{T_i} , ΔP , ΔY それぞれの平均である δX , δY , R_T , P , Y を算出した結果を示している。図 3, 図 4 はそれぞれ 100cm, 300cm 地点における実験結果をプロットしたものであり、表 3, 表 4 は 100cm, 300cm におけるターゲットの位置ごとの変位 δX_i , δY_i の平均である δX , δY 求めた結果を示す。また、表 2 にはターゲットの位置ごとの R_{T_i} の平均 R_T を求めた結果を示す。

6 議論

6.1 ターゲティング距離に関する考察

結果から、ユーザのターゲティング距離が遠くなるに従い、誤差の値も増加することが確認された。この原因として、距離が遠いほどキャリブレーション時に必要となる頭部の回転差が小さくなり、頭部のわずかな回転でも推定される座標が大きく変化するためと考えられる。

6.2 縦方向と横方向の誤差に関する考察

さらに、 x 座標と y 座標の各々の絶対誤差については、いずれの距離においても x 座標の絶対誤差が y 座標の絶対誤差に比べて小さく、これは **pitch** (上下) 方向よりも **yaw** (左右) 方向のターゲティングの精度が高いことを示唆している。

回転差の平均は、いずれの距離においても **pitch** よりも **yaw** の方が大きい値になっているが、実験環境は 240cm×240cm の平面であり、上下左右の大きさ

表 3: 距離 100cm の変位 δX , δY の平均

		左	中央	右
上	id	①	④	⑦
	δX	-6.7	-2.8	6.9
	δY	-3.7	-14.8	-11.3
中央	id	②	⑤	⑧
	δX	-7.7	5.0	13.3
	δY	14.4	25.7	12.5
下	id	③	⑥	⑨
	δX	-4.9	4.6	11.2
	δY	26.1	38.1	27.6

表 4: 距離 300cm の変位 δX , δY の平均

		左	中央	右
上	id	①	④	⑦
	δX	12.8	-1.0	-6.4
	δY	33.5	9.9	21.3
中央	id	②	⑤	⑧
	δX	3.0	2.2	-18.8
	δY	26.0	24.5	27.9
下	id	③	⑥	⑨
	δX	9.1	2.3	-8.8
	δY	13.6	34.3	28.7

は等しく設定していた。これは、人は **pitch** 方向に比べ **yaw** 方向により大きな角度で頭部姿勢を動かす傾向があることを示唆している。

ターゲットの位置ごとの絶対誤差では、中心が最も誤差が小さく、上下左右の端で絶対誤差が大きくなった。ここから、ターゲットの位置が中心から離れるにつれて絶対誤差が大きくなることが推察される。

6.3 眼球運動の影響に関する考察

変位について、図 3, 4 及び表 3, 4 に示すように、100cm の距離では推定された注視位置が実際のターゲットの位置に比べ中心側に分布し、300cm の距離では実際のターゲットの位置に比べ外側に分布する傾向が見られた。この要因として、ユーザのターゲティング動作の特性が影響を与えていることが考えられる。

人は頭部の動きと眼球の動きを組み合わせることでターゲットの注視を行っている。正面から大きく外れたターゲットの注視では、頭部への負担や可動域の問題から、眼球の動きによる寄与が大きくなると考えられる。提案手法ではこのような眼球の動きを検出していないため誤差が生じていると推測される。

実験環境では基準点が最も中心から離れた場所に位置しており、実験参加者はキャリブレーションを行う際に最も回転動作を行う必要がある。一方、ターゲットは全て基準点よりも内側に存在しており、注視するために必要な回転量が小さい。このことから、実験参加者は、基準点を見る際に最も眼球を動かして注視を行っており、ターゲットを見る際には比較的眼球を動かさず、頭部の回転で注視を行っていると考えられる。ターゲティングにおいて、キャリブレーション時に取

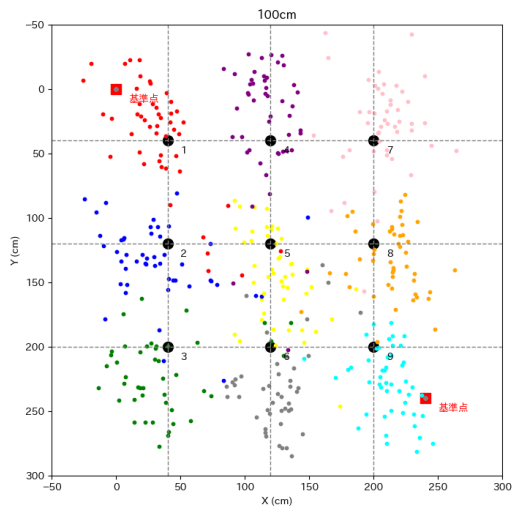


図3: 距離 100cm からのターゲティングの分布図

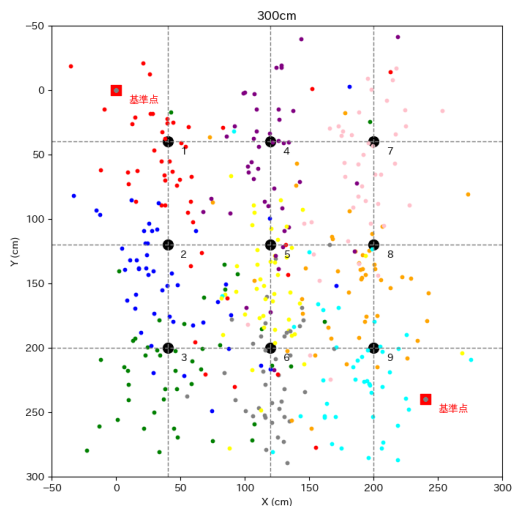


図4: 距離 300cm からのターゲティングの分布図

得される回転角は理論値の回転角よりも小さな値を取り、ターゲットの注視の際の回転角ではキャリブレーション時と比べ回転角が理論値に比較的近い値を取るため、注視位置がターゲットの位置よりも外側に分布することとなる。

図5には100cmの距離における、yaw方向の基準点間の理論上の角度と、実際に計測された回転差の平均を示す。100cmの距離では注視する平面との距離が近く、基準点を見る際に、頭部の動きに加え、眼球の動きによる影響が大きいと考えられる。

一方で基準点より内側にあるターゲットを見る際には、頭部を十分に動かせるため、基準点を見る際に比べ眼球の動きが小さい。すなわち、ターゲットを見る際は基準点を見る際に比べ相対的に頭部姿勢の変化による影響が大きい。結果、基準点を注視した際に眼球を大きく動かしてキャリブレーションを行った際の値に基づいて線形補間を行う提案手法は、眼球を比較的小さく動かしてターゲットを注視する際の推定におい

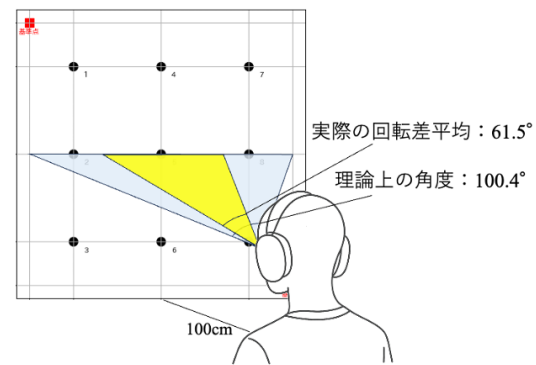


図5: 100cmにおける理論上の角度と実際の平均

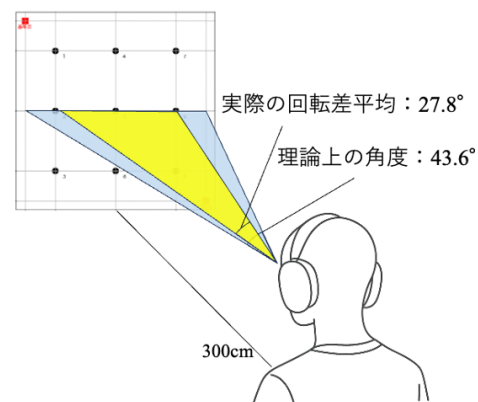


図6: 300cmにおける理論上の角度と実際の平均

て、推定結果がユーザが実際に注視するターゲットより外側になったと考えられる。

図6には、300cmの距離におけるyaw方向の基準点間の理論上の角度と、実際に計測された回転差の平均を示す。300cmの距離では実験参加者と注視する平面との距離が比較的遠いことから、基準点は比較的实验参加者の中央に近い場所に位置することとなる。そのため、300cmの距離では眼球の動きによる影響が小さく、100cmの距離で見られた推定注視位置が外側へ分布する傾向が見られなかったと考えられる。

このように、注視する対象がユーザから見てどの角度に位置するのかによってターゲット位置と推定された注視位置との誤差に傾向がみられる。こうした傾向に基づいて、ユーザが注視する角度に応じて推定位置を補正することで、精度の向上が期待される。

6.4 本システムの応用について

本システムは、ユーザが注視する位置を簡易に推定することが可能であり、推定された注視位置は、ユーザの選択操作のためのUIや、ユーザの興味位置を取得するUIとしての活用が考えられる。以下では、想

定される本システムの応用方法を例示する。

IoT 機器の操作に応用することで、注視によって対象機器を選択し操作することが可能となり、迅速かつ容易な選択操作を実現できる。また、共創的な作業においては、複数人がそれぞれ注視している対象を可視化することができ、各ユーザの関心対象を把握しながら、円滑で相互的な協調操作が可能となる。PC 操作に活用することで、注視位置に応じてマウスカーソルを移動させたり、注視中のウィンドウに対する入力操作を可能としたりすることで、特に複数ディスプレイ環境における直観的かつシームレスな GUI 操作が実現する。このように、本手法は多様な状況において操作性・協調性の向上に寄与するインタフェースとしての活用が可能であると考えられる。

7 おわりに

本稿では、IMU 搭載ヘッドホンによる注視位置の簡易推定手法について提案した。実験の結果から、ユーザの注視対象との距離が遠くなるほど推定される注視位置の座標と注視対象との座標の誤差が大きく、ターゲットごとの絶対誤差に着目すると、中心における誤差が上下左右における誤差に比べて小さいことが確認された。変位については特に yaw 方向について、ターゲット距離が近づくにつれ、推定される注視位置が実際のターゲットに比べて外側に分布することが確認された。提案システムは、IoT 操作、共創システム、カーソル操作などの、注視することを対象としたインタラクションに活用することが想定される。

一方で、本稿では 240cm×240cm の平面に対し 100cm, 200cm, 300cm のターゲット距離で実験を行ったが、例えば PC を利用する際のようなユーザとターゲットの距離がより近い条件や、映画館、ライブ会場といった、より広い空間を想定した条件については検証を行っていない。

また、提案システムは、キャリブレーション時に登録された回転角に基づいて推定を行うため、キャリブレーション後に移動すると推定される座標に誤差が生じるが、その具体的な影響については検証していない。

今後は、これらについても検証を行うことで、さらなる活用方法について検討する。

参考文献

[1] Bolt, R. A.: “Put-that-there” Voice and gesture at the graphics interface, *Proc. 7th annual conference on Computer graphics and interactive techniques*, pp. 262–270 (1980).

[2] Hueber, S., Cherek, C., Wacker, P., Borchers, J. and Voelker, S.: Headbang: Using head

gestures to trigger discrete actions on mobile devices, *Proc. 22nd International Conference on Human-Computer Interaction with Mobile Devices and Services*, pp. 1–10 (2020).

- [3] Kim, D., Min, C. and Kang, S.: Towards automatic recognition of perceived level of understanding on online lectures using earables, *Adjunct Proc. 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proc. 2021 ACM International Symposium on Wearable Computers*, pp. 158–164 (2021).
- [4] Kytö, M., Ens, B., Piumsomboon, T., Lee, G. A. and Billingham, M.: Pinpointing: Precise head-and eye-based target selection for augmented reality, *Proc. 2018 CHI conference on human factors in computing systems*, pp. 1–14 (2018).
- [5] Noda, K.: Google Home: smart speaker as environmental control unit, *Disability and rehabilitation: assistive technology*, Vol. 13, No. 7, pp. 674–675 (2018).
- [6] Panda, P., Nicholas, M. J., Nguyen, D., Ofek, E., Pahud, M., Rintel, S., Gonzalez-Franco, M., Hinckley, K. and Lanier, J.: Beyond audio: Towards a design space of headphones as a site for interaction and sensing, *Proc. 2023 ACM Designing Interactive Systems Conference*, pp. 904–916 (2023).
- [7] Sato, K. and Matsushita, M.: Object Manipulation by Absolute Pointing with a Smartphone Gyro Sensor, *Proc. ACM Symposium on Spatial User Interaction*, No. 31 (2019).
- [8] Stiefelhagen, R.: Tracking focus of attention in meetings, *Proc. 4th IEEE International Conference on Multimodal Interfaces*, pp. 273–280 (2002).
- [9] Wang, Y., Ding, J., Chatterjee, I., Parizi, F. S., Zhuang, Y., Yan, Y., Patel, S. and Shi, Y.: Faceori: Tracking head position and orientation using ultrasonic ranging on earphones, *Proc. 2022 CHI Conference on Human Factors in Computing Systems*, pp. 1–12 (2022).