SEEING IS NOT PLANNING: WHEN AND HOW EGOCENTRIC SIGNALS HELP HUMAN MOTION FORECASTING

Anonymous authors
Paper under double-blind review

ABSTRACT

Egocentric signals such as eye gaze and scene geometry are increasingly fed to human motion forecasting models, yet the reported gains are small and inconsistent across benchmarks. We ask *when and why* these signals help. On a controlled testbed spanning human–object (MoGaze), human–scene (GIMO) and human–human (EgoBody) interaction with two state-of-the-art forecasters, we train each network with and without egocentric input and study the gain. We find three answers. **When:** the gain peaks at input and output windows that cover the lag between gaze and body motion, which differs across datasets; a poorly chosen window eliminates the gain or makes it negative. **Where:** the gain concentrates in specific motion states and joints and is diluted or cancelled by aggregate metrics; on a pick-and-place benchmark the gain at the acting wrist is nearly twice the reported aggregate. **What:** current datasets under-supply the information egocentric signals need; oracle probes that provide the missing information substantially increase the gain; on EgoBody, a simple wearer status oracle raises the Global MPJPE gain from 4% to 14%. But what is missing is highly dataset-specific. Together, the three findings reframe a near-zero aggregate gain not as evidence that egocentric signals are useless, but as a consequence of poorly matched windows, indiscriminate aggregation and incomplete data.

1 INTRODUCTION

People look before they move: they fixate an object well before the hand reaches it (Land et al., 1999; Johansson et al., 2001; Hayhoe & Ballard, 2005), and in everyday object interaction their eye movements lead full-body motion (Hu et al., 2024c). Head-mounted devices now make this anticipatory signal observable, recording eye gaze, head pose, egocentric video and scene geometry alongside the wearer's body (Zhang et al., 2022; Pan et al., 2023; Grauman et al., 2024; Ma et al., 2024). A growing line of work therefore feeds such *egocentric signals* into 3D human motion forecasting (Figure 1, Part I), using gaze as an intent cue, a scene index, a leading signal or a pointer to affordances (Zheng et al., 2022; Lou et al., 2024; Hu et al., 2024a; Yu et al., 2025), and exploiting egocentric object boxes and headset proprioception in the same spirit (Escobar et al., 2025).

Yet the evidence that these signals help is thin and inconsistent. Most of these works introduce a new architecture together with the new input, so their reported improvements conflate the two. On MoGaze, roughly half of the improvement of GazeMotion and HOIMotion over the strongest prior method is already achieved by their pose-only variants, leaving 3.7–3.9% MPJPE attributable to the egocentric input (gaze, or object boxes with head orientation); on ADT, the two papers even disagree on whether eye gaze outperforms head orientation (Hu et al., 2024a;b). On GIMO, adding gaze to four backbones reduces the trajectory endpoint error by 5.8–19.6% but the endpoint MPJPE by only 0.5–3.2%, and adding scene and gaze increases one backbone's 2-second MPJPE in one scene by about 20% (Lou et al., 2024).It remains unclear whether this reflects a weak signal or the way the signal is measured, timed and supplied. Settling this question is not only of academic interest. In embodied AI, robots that collaborate with, assist or navigate around people must anticipate their motion from onboard sensors (Cao et al., 2020; Wang et al., 2021), and head-mounted devices are increasingly used to teleoperate robots and to collect demonstration data for imitation learning (Zhang et al., 2018; Cheng et al., 2024). If the egocentric stream captured during data

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

collection is poorly timed, under-informative or misaligned with the downstream prediction horizon, the resulting policy inherits the same blind spots. A clear diagnosis of *when*, *where* and *what* egocentric signals contribute therefore doubles as practical guidance for sensor configuration and dataset design in human–robot collaboration, teleoperation and assistive robotics.

Following Lu et al. (2026), who resolved conflicting reports on proprioception in robot policies by asking *when* a modality fails rather than *whether* it helps, we ask: *when and why do egocentric signals help human motion forecasting?* We break this into three sub-questions (Figure 1, Part IV), each ruling out a shallower explanation before turning to a deeper one.

- **When does the gain appear?** The interval between gaze first landing on a target and the body reaching it differs substantially across datasets, and predicted future gaze helps more than past gaze (Hu et al., 2024a).
- **Where does the gain appear?** If egocentric cues mainly reveal where a person will go or what a hand will reach, metrics that pool all joints and samples may average away a gain confined to a few. Indeed, gaze reduces root-trajectory error far more than articulation error (Lou et al., 2024), and mainly affects the final poses of a prediction (Zheng et al., 2022).
- **What information is missing?** A gaze point says where a person looks, not what they will do there (Yu et al., 2025); in social interaction, gaze dwells on the partner's upper body, leaving the lower body out of view (Zhang et al., 2022). No model can exploit information the data do not contain.

We answer these questions with a controlled testbed (Figure 1, Parts II–III; Section 3) that pairs three datasets, covering human–object, human–scene and human–human interaction, with two state-of-the-art forecasters, and trains each network with and without egocentric input. We find that the gain is not absent but diluted, mistimed and under-supplied. Our contributions are:

- A controlled analysis of egocentric signals in 3D human motion forecasting across three interaction types that isolates the contribution of the modality from that of the architecture.
- **When** (Section 4): the gain peaks when the input and output windows cover the gaze-to-motion lag, which differs across datasets.
- **Where** (Section 5): the gain concentrates on specific joints and motion states and is diluted by aggregate metrics.
- **What** (Section 6): an information-sufficiency probe shows current datasets under-supply the information egocentric signals need, and the gain grows substantially once supplied.

2 Related Work

3D human motion forecasting. Deterministic forecasting of future poses has progressed from recurrent models (Fragkiadaki et al., 2015; Jain et al., 2016; Martinez et al., 2017) to graph convolutional networks (Mao et al., 2019; Li et al., 2020; 2022; Ma et al., 2022), Transformers (Mao et al., 2020; Aksan et al., 2021; Xu et al., 2023) and MLPs (Guo et al., 2023). Stochastic methods instead sample multiple futures with variational or diffusion models (Yuan & Kitani, 2020; Chen et al., 2023; Barquero et al., 2023) and score the best sample; we study the deterministic setting, where a single forecast is compared directly with the ground truth. To reduce long-horizon ambiguity, later work adds external context such as scenes (Cao et al., 2020; Mao et al., 2022; Xing et al., 2024) or other people (Wang et al., 2021; Guo et al., 2022). Evaluation almost always averages MPJPE over joints and samples.

Egocentric signals for motion forecasting. Building on datasets such as MoGaze (Kratzer et al., 2020) and GIMO (Zheng et al., 2022), most egocentric forecasters feed the raw signal to the network and leave its use to learned fusion: gaze enters through cross-modal attention (Zheng et al., 2022; Lou et al., 2024) or diffusion conditioning (Yan et al., 2024), and headset poses or egocentric video drive estimation and forecasting of the wearer's own body (Li et al., 2023; Yi et al., 2025; Escobar et al., 2025; Patel et al., 2025). A few methods first convert the signal into more explicit information, such as the fixated object (Razali & Demiris, 2022), objects near the viewport center (Hu et al.,

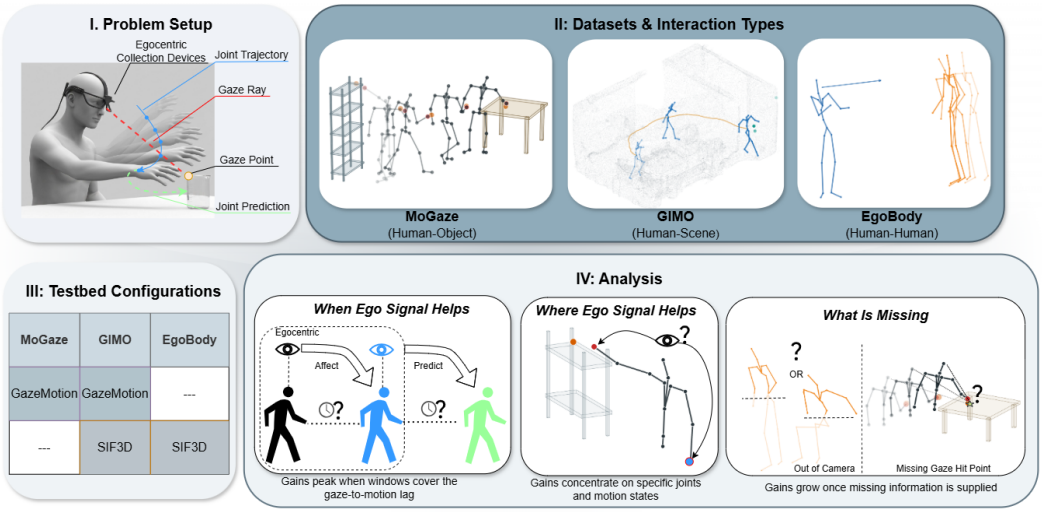

Figure 1: **Overview of our study.** (I) Egocentric motion forecasting predicts future body motion from past poses together with signals captured by a head-mounted device, such as eye gaze. (II) We cover three interaction types: human–object (MoGaze (Kratzer et al., 2020)), human–scene (GIMO (Zheng et al., 2022)) and human–human (EgoBody (Zhang et al., 2022)). (III) Four testbed configurations pair each dataset with a state-of-the-art forecaster: GazeMotion (Hu et al., 2024a) for short-horizon forecasting with gaze, and SIF3D (Lou et al., 2024) for long-horizon forecasting with gaze and scene; GIMO is run with both for cross-model validation. Each configuration keeps the architecture fixed and toggles only the egocentric input. (IV) We ask *when*, *where* and *what*: the gain from egocentric signals is not absent, but mistimed, diluted and under-supplied.

2024b), predicted future gaze (Hu et al., 2024a) or the affordances of the gazed object (Yu et al., 2025). Our probe in Section 6 belongs to this second group, but uses it as a diagnostic of the data rather than as a model component. Within-network ablations in this literature are reported in aggregate, yet they already suggest condition dependence: on GIMO, gaze reduces MultimodalNet's translation error but worsens its path MPJPE on both test splits (Zheng et al., 2022).

Understanding modality contributions. Adding a modality can fail to help: a multimodal network may underperform its best unimodal counterpart (Wang et al., 2020), and a dominant modality can suppress the learning of the others (Peng et al., 2022; Huang et al., 2022; Du et al., 2023). These works locate the failure in optimization and treat a modality's contribution as uniform, static and informative. Three lines relax these assumptions, one for each of our questions. Aggregate metrics can hide systematic subgroup behavior (Oakden-Rayner et al., 2020) (Where). In robot policies that combine vision and proprioception (Wang et al., 2024; Octo Model Team et al., 2024), the value of each modality changes across task phases, and joint training can let proprioception suppress vision in exactly the phases that need it (Lu et al., 2026) (When). Finally, measured gains may not come from the added modality at all (Hessel & Lee, 2020), depending on how much unique task information it carries (Liang et al., 2023) (What).

3 Datasets and Controlled Testbed

3.1 Datasets and Testbed

For our findings to hold beyond a single setting, we need data in which the egocentric signal and the full-body motion are recorded simultaneously and accurately, across the interaction types in which egocentric sensing is expected to matter. We select MoGaze (Kratzer et al., 2020), GIMO (Zheng et al., 2022) and EgoBody (Zhang et al., 2022), which cover human–object, human–scene and human–human interaction respectively; each provides synchronized full-body ground truth, head-mounted eye gaze and 3D scene geometry.

Table 1: Testbed configurations.

	Dataset	Forecaster	Input → output	Rate	Egocentric input
C1	MoGaze	GazeMotion	0.33 s → 1 s	30 Hz	Gaze direction
C2	GIMO	GazeMotion	0.33 s → 1 s	30 Hz	Gaze direction
C3	GIMO	SIF3D	3 s → 5 s	2 Hz	Gaze point + scene
C4	EgoBody	SIF3D	4 s → 1.5 s	10 Hz	Gaze point + scene

Body motion. MoGaze tracks 50 reflective markers with an optical motion-capture system and provides the body as a kinematic skeleton, from which joint positions are obtained by forward kinematics. GIMO records motion with an IMU suit, fits SMPL-X (Pavlakos et al., 2019) to the resulting joints, and uses the headset pose to fix the head. EgoBody fits SMPL-X to both interacting people from a multi-view RGB-D rig, using 2D keypoints, depth, and scene-contact terms. For GIMO and EgoBody, joint positions are decoded from the SMPL-X parameters (global translation, orientation and body pose) through the body model.

Egocentric signals. All three datasets record eye gaze with a head-mounted eye tracker whose pose is known in the world frame. MoGaze tracks the eye-tracking headset with motion-capture markers and calibrates gaze directly in world coordinates, yielding a gaze direction per frame. GIMO and EgoBody use the HoloLens 2 eye tracker; the 3D gaze point is defined as the intersection of the gaze ray with the scanned scene. The scene itself is available as object meshes in MoGaze and as scanned meshes in GIMO and EgoBody.

Movements. The three datasets differ in what drives the motion. In MoGaze, participants set and clear a table and move plates, cups, a jug and a bowl between shelves, producing 1,627 pick-and-place actions in a fixed workspace. In GIMO, participants start away from a target and walk to it to perform an intent-driven activity, such as sitting down, reaching for an object, or opening or picking up something, yielding 217 sequences from 11 subjects across 19 scenes. In EgoBody, two people interact following scripted scenarios, such as games, teaching, arguing or casual conversation, in 125 sequences from 36 subjects across 15 scenes; only the camera wearer carries the headset, and the interaction partner wears no device.

Testbed. On these datasets we build four configurations (Table 1). The two forecasters represent the two protocols common in this literature: GazeMotion (Hu et al., 2024a) forecasts one second ahead at a high frame rate from poses and gaze, while SIF3D (Lou et al., 2024) forecasts several seconds ahead at a low frame rate from poses, gaze and a scanned scene. Each forecaster covers two datasets, and GIMO is run with both so that findings on it can be cross-checked across models. We do not pair SIF3D with MoGaze or GazeMotion with EgoBody because MoGaze contains no scene point cloud while EgoBody's sampling rate is too low for the DCT representation of GazeMotion. In EgoBody, the forecast target is the wearer. Unless stated otherwise, data processing, segmentation and splits follow the official GazeMotion and SIF3D implementations. The only departure is C4, for which we set the windows to 4 s → 1.5 s at 10 Hz to fit EgoBody, since SIF3D was not originally evaluated on it. On MoGaze, we follow the leave-one-subject-out protocol over the six subjects with gaze and average the six folds; every configuration is trained with five random seeds.

3.2 Formulation and Metrics

Notation. We call the person wearing the headset the *wearer* and, in EgoBody, the other person the *partner*. A *pose* $x_\tau \in \mathbb{R}^{J\times 3}$ holds the 3D positions of J joints at frame τ in the world frame, and the *root* is the pelvis, indexed by r. The *local pose* expresses each joint relative to the root. *Egocentric signals* e_τ are quantities captured from the wearer's head-mounted device: a gaze direction (C1) or a gaze point (C2, C3, C4), together with the scene point cloud. Given observed poses $X = x_{t-H+1:t}$ and egocentric signals $E = e_{t-H+1:t}$ over an input window of H frames, a forecaster f predicts the future poses $Y = x_{t+1:t+F}$ over an output window of F frames, i.e. $\hat{Y} = f(X, E)$.

Egocentric gain. For each configuration, we train two instances of the same forecaster with identical architecture, hyperparameters and data: f_{ego}, which receives E, and $f_{\text{no-ego}}$, in which the ego-

Table 2: Overall egocentric gain (mm).

	Global MPJPE			Local MPJPE			RDE		
	w/o ego	w/ ego	Δ (G)	w/o ego	w/ ego	Δ (G)	w/o ego	w/ ego	Δ (G)
C1	80.53	77.83	2.70 (3.35%)	52.12	50.59	1.53 (2.94%)	140.66	135.17	5.49 (3.90%)
C2	86.39	86.27	0.12 (0.13%)	47.79	47.17	0.62 (1.29%)	153.73	150.59	3.14 (2.04%)
C3	809.38	617.24	192.15 (23.74%)	163.42	162.47	0.95 (0.58%)	1158.88	741.76	417.12 (35.99%)
C4	92.30	86.27	6.03 (6.53%)	51.95	51.38	0.57 (1.10%)	119.69	111.97	7.72 (6.45%)

centric input branch is removed. For an error metric $\mathcal{E}$, the egocentric gain is

$$G = \frac{\mathcal{E}_{\text{no-ego}} - \mathcal{E}_{\text{ego}}}{\mathcal{E}_{\text{no-ego}}}, \tag{1}$$

so that $G > 0$ means the egocentric input reduces error. Because both models share the architecture, G measures the contribution of the input alone. Sections 4–6 apply Equation equation 1 to restricted errors: G_s over test samples in motion state s, G_j over joint j, and $G(H, F)$ with both models retrained at windows H and F. Since window length changes task difficulty, windows are compared through G rather than raw error.

Metrics. We report three metrics throughout, each averaged over test samples:

$$\text{Global MPJPE} = \frac{1}{FJ} \sum_{\tau=t+1}^{t+F} \sum_{j=1}^{J} \left\| \hat{x}_{\tau,j} - x_{\tau,j} \right\|_2, \tag{2}$$

$$\text{Local MPJPE} = \frac{1}{FJ} \sum_{\tau=t+1}^{t+F} \sum_{j=1}^{J} \left\| (\hat{x}_{\tau,j} - \hat{x}_{\tau,r}) - (x_{\tau,j} - x_{\tau,r}) \right\|_2, \tag{3}$$

$$\text{RDE} = \left\| \hat{x}_{t+F,r} - x_{t+F,r} \right\|_2. \tag{4}$$

Global MPJPE measures how well the motion is predicted in the world frame, combining where the body goes and how it is posed. Local MPJPE removes the root trajectory and measures how well the predicted posture matches the ground truth. The root destination error (RDE) measures only where the body ends up at the last predicted frame, and therefore reflects whether the forecaster captures the wearer's movement intention.

3.3 Headline Observation

Table 2 reports the overall gain. Egocentric input barely improves the body pose: the gain in Local MPJPE stays below 3% in every configuration and at most 1.3% in three of them. The gain is also inconsistent. It differs widely across configurations, even between C2 and C3 on the same dataset, and across seeds, where it turns negative for some runs; we therefore average over five seeds throughout. Where a larger gain does appear, it comes from the root trajectory rather than the body pose, a pattern that Section 4 relates to the window length.

> **Finding 0.** Egocentric input yields a small and unreliable gain, and barely improves the body pose in any configuration.

Whether this reflects a weak signal or conditions that aggregate numbers hide is the question of Sections 4–6.

4 When Does the Gain Appear?

Egocentric signals do not act instantaneously. There are two temporal gaps between the signal and the motion it informs, and the model's windows must cover both. The first is the *gaze-to-motion lag*: gaze lands on a target before the body begins to respond, so the input window H must reach far enough into the past to capture the informative gaze. The second is the *prediction-effectiveness*

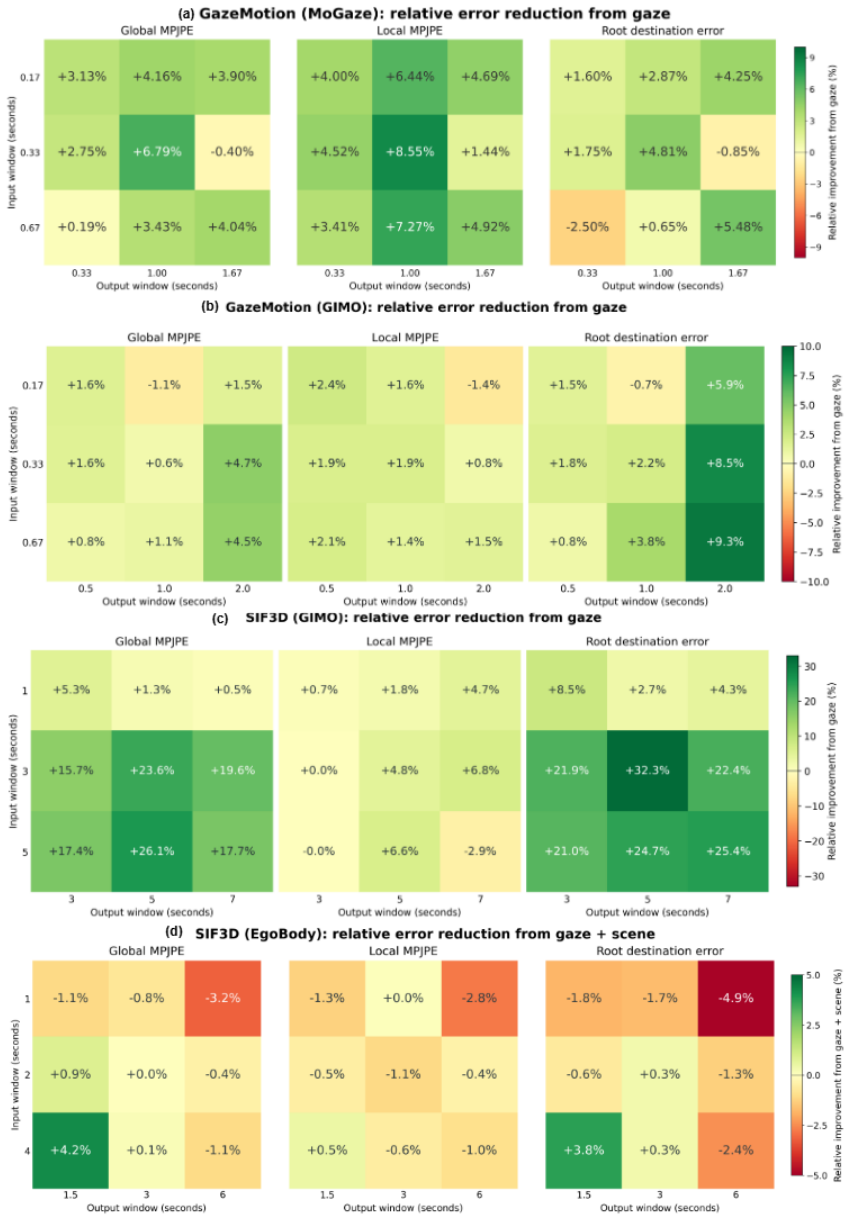

Figure 2: **When the gain appears.** Gain G as a function of input and output window (in seconds) for all four configurations. Each cell is a separately trained model pair.

lag: that past gaze is most useful for predicting a specific future horizon, so the output window F must reach far enough ahead to cover the motion the gaze anticipated. If either window is too short, the cue falls outside the model's scope; if either is too long, the relevant signal is diluted by uninformative frames. We retrain each configuration at a grid of (H, F) values and compute G at each cell (Figure 2).

The gain peaks at a particular (H, F) and falls or turns negative at others: on MoGaze (a) the Global gain peaks at $(0.33\,\text{s}, 1\,\text{s})$ and drops to near zero at shorter output windows; on EgoBody (d) it is positive only at the shortest output and reverses at longer ones.

The peak location tracks the gaze-to-motion lag of each dataset. On MoGaze, the gain peaks near $F \approx 1\,\text{s}$. On GIMO, the optimal output window is correspondingly longer. GIMO serves as a cross-verification here: tested on the same data with different models, GazeMotion (b) with its 1-second

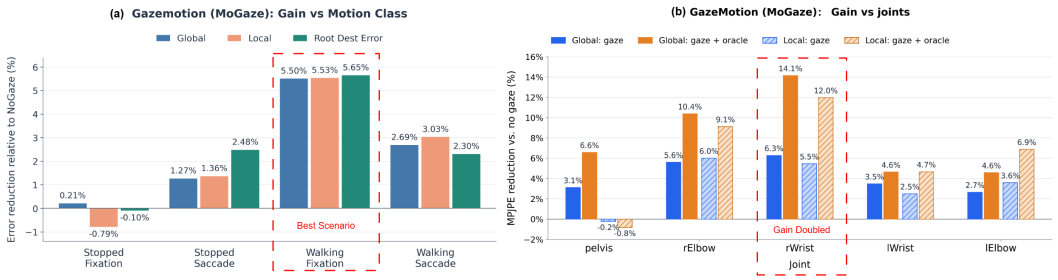

Figure 3: **Where the gain appears** (MoGaze, C1). (a) Gain by motion state: walking or stopped, with fixation or saccade. (b) Gain by joint in Global and Local MPJPE, with gaze alone and with gaze plus a pseudo stopspot (Section 6).

output window falls well short of this lag and captures almost no gain, while SIF3D (c) with its 5-second window covers it and yields gains above 20% in Global MPJPE and above 30% in RDE. The near-zero gain of C2 in Table 2 is therefore not a property of the data or the signal, but of the window. On EgoBody (d), where social interaction involves no distant reaching target, gaze appears to inform only the immediate response, and extending the output window introduces frames the signal says nothing about, turning the gain negative.

Throughout, Local MPJPE remains insensitive to the window, reinforcing that egocentric input improves where the body goes, not how it is posed.

> **Finding 1.** The gain peaks at input and output windows that cover the gaze-to-motion lag and the prediction-effectiveness lag respectively, both of which differ across datasets. A poorly chosen window eliminates the gain or makes it negative.

We now ask which motion states and joints benefit from egocentric input.

5 Where Does the Gain Appear?

A small aggregate gain can mean a signal that is weak everywhere, or one that is informative in some situations and useless or even harmful in others. Aggregate metrics cannot tell these apart, yet only the second calls for changing where the signal is used. We distinguish them on MoGaze (C1) by stratifying the gain by motion state and by joint (Figure 3).

Motion states. A frame is labeled walking when the root speed exceeds 0.30 m/s and stopped when it drops below 0.15 m/s, and fixation when the eye-tracking confidence exceeds 0.7 and saccade otherwise; each sample takes the state of its last observed frame. The gain peaks at 5.5% for walking with fixation, about 1.6 times the aggregate, and vanishes for stopped fixation, where the local gain even turns negative (-0.79%). While walking, the wearer fixates the place or object ahead before reaching it (Kratzer et al., 2020); while standing, they fixate the object already at hand, which the pose history already implies. Since the strata differ in sign, the aggregate not only dilutes the gain but partly cancels it.

The same pattern holds on EgoBody (C4; Appendix D): the gain rises with translation speed (Stationary $+1.97\%$ $\rightarrow$ Moving $+7.81\%$ Global) but turns negative for Low posture (-2.00% Global, -9.16% RDE), where gaze says little about the relevant body configuration. As on MoGaze, strata of opposite sign partly cancel in the aggregate.

Joints. The gain concentrates on the right arm, which performs the pick-and-place: 6.3% in Global and 5.5% in Local MPJPE at the right wrist, and 5.6% and 6.0% at the right elbow, against 3.35% and 2.94% over all joints (details in Appendix B). The local pose of the acting arm thus improves even though the aggregate Local MPJPE barely does.

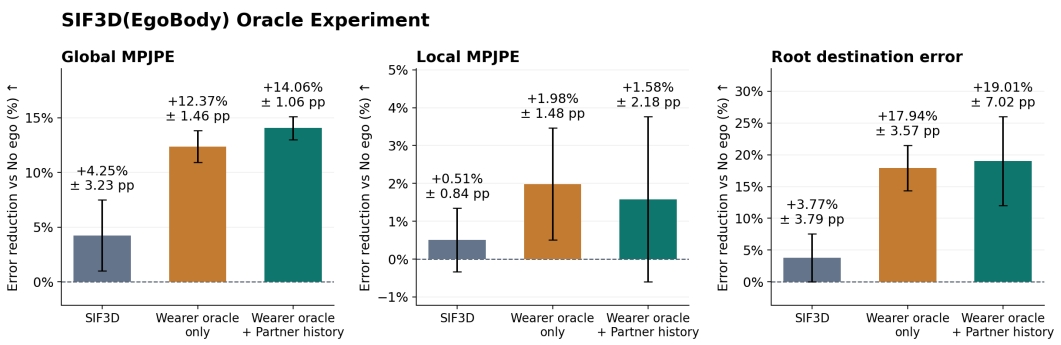

Figure 4: **Information-sufficiency probe on EgoBody (C4).** Gain over the no-ego baseline. Error bars show ± 1 std over five seeds.

> **Finding 2.** The gain from egocentric input concentrates where gaze is informative, when walking with fixation and at the acting arm, and vanishes or reverses elsewhere, so aggregate metrics dilute and partly cancel it.

The oracle bars in Figure 3(b) already hint that the data leave information on the table. The next section examines what is missing.

6 What Is Missing?

Even under the best conditions identified by Findings 1 and 2, the gain remains modest. The bottleneck may lie not in the model but in the data: the egocentric signal may lack information that the motion requires. We test this with an *information-sufficiency probe*: we inject oracle information that a richer egocentric signal could in principle provide, and measure whether the gain grows. If it does, the current data under-supply the needed information. We design a probe for MoGaze and EgoBody; GIMO is discussed at the end.

MoGaze: pseudo stopspot. A gaze direction tells where the wearer looks but not how far away the target is, because MoGaze does not record where the ray lands. We construct a *pseudo stopspot* $s \in \mathbb{R}^3$ from the full clip: near the terminal slow-down of a reaching action, we project the acting wrist onto the recorded gaze ray and take the projected point as s, fixed for the clip. Since s uses the future stopping location, it is oracle information. It is injected by constructing, for each frame t and joint j, a relation vector $r_j(t) = \text{joint}_j(t) - s$, adding 21 relation nodes (63 dimensions) to the input graph. With the pseudo stopspot, the gain at the right wrist rises from 6.3% to 14.1% in Global and from 5.5% to 12.0% in Local MPJPE, while the gain at the left wrist rises by only 1.1 and 2.2 points (Figure 3b): the missing depth information benefits exactly the acting hand.

EgoBody: future wearer state. In human–human interaction, the wearer's future motion depends on social dynamics that the current gaze and scene only partially capture. We probe this by injecting the wearer's future state as categorical labels covering translation, turning, posture and arm activity, computed per frame from the fitted body model (details in Appendix A). In the *wearer oracle only* condition, only the wearer's future 15-frame labels are added. In *wearer oracle + partner history*, the partner's past 40-frame labels are also provided, with the last observed state held constant into the prediction interval; the partner receives no future oracle.

Figure 4 shows the results. The baseline SIF3D gain is 4.25% in Global MPJPE and 3.77% in RDE[1], both with standard deviations that cross zero. The wearer oracle alone raises the gain to 12.37% and 17.94%, and adding the partner's history brings it to 14.06% and 19.01%. Local MPJPE remains low throughout. A per-state breakdown (Appendix D) confirms that the oracle probes specifically

[1]The oracle experiments use the same C4 protocol as Table 2 but were run in a separate batch; the small difference in the baseline gain (4.25% vs. 6.53% in Table 2) reflects seed-level variation, consistent with the instability noted in Section 3.3.

repair the states where the baseline signal hurts, converting negative gains into positive ones.The jump from baseline to oracle shows that the current egocentric signals substantially under-supply information about future wearer intent. The further improvement from the partner's history shows that the wearer's motion is partly shaped by the partner's behavior, information that is observable in principle from the headset but not yet exploited.

GIMO. On GIMO, we have not found an oracle that reliably improves the gain. The scenes contain large obstacles such as sofas and beds that the wearer navigates around using peripheral vision rather than forward gaze. The path around these obstacles has little correlation with the gaze fixation point, so probes based on gaze depth do not help. We also tested alternative encoding strategies and injection points within the network, but the official SIF3D configuration already performs best, suggesting that the bottleneck lies on the data side rather than in the architecture. The bottleneck on GIMO appears to be scene-navigation rather than gaze depth, a qualitatively different gap.

> **Finding 3.** Current egocentric data under-supply the information that motion forecasting needs. When the missing information is provided through oracle probes, the gain grows substantially. What is missing, however, is highly dataset-specific, shaped by both the interaction type and the data collection setup, and researchers should account for these gaps when choosing an existing dataset or designing a new one.

7 Conclusion

We asked when and why egocentric signals help human motion forecasting. Across three interaction types, two forecasters and four controlled configurations, we found that a near-zero aggregate gain is not evidence that egocentric signals are useless, but a consequence of poorly matched windows, indiscriminate aggregation and incomplete data. The findings suggest four practical guidelines:

1) *Match the window to the lag.* The input and output windows should cover the gaze-to-motion and prediction-effectiveness lags of the target dataset, which differ across interaction types.

2) *Direct the signal to where it helps.* Egocentric input should be routed to the body parts and motion states it informs, such as the acting arm during reaching, rather than applied uniformly.

3) *Report gains per state and joint.* Aggregate metrics hide conditional effects; stratified reporting reveals where a method actually improves and where it only adds noise.

4) *Audit the data before adding a modality.* Before designing a new fusion module, check whether the dataset supplies the information the modality is supposed to carry; if it does not, the module cannot help regardless of its architecture.

Limitations. Our analysis uses two forecasters and three datasets. Although these span three interaction types and two temporal regimes, the findings may not transfer to architectures with fundamentally different fusion strategies, such as diffusion-based models, or to datasets with richer egocentric streams, such as egocentric video. The oracle probes on MoGaze and EgoBody use information that is not available at test time, so they establish an upper bound on how much the gain *could* grow, not a deployable improvement; translating them into test-time features, for example by training a gaze-depth estimator or a wearer-state predictor, is an open problem.

Future work. Guided by these findings, we are collecting an indoor locomotion dataset with head-mounted depth video and lower-body motion capture. Preliminary results show that depth video improves lower-limb prediction, and that preprocessing the point cloud to distinguish geometric primitives such as edges further increases the gain. We will release the dataset and use it to validate the diagnostic framework developed here on a richer egocentric stream.

REFERENCES

Emre Aksan, Manuel Kaufmann, Peng Cao, and Otmar Hilliges. A spatio-temporal transformer for 3D human motion prediction. In *International Conference on 3D Vision (3DV)*, 2021.

German Barquero, Sergio Escalera, and Cristina Palmero. BeLFusion: Latent diffusion for behavior-driven human motion prediction. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023.

Zhe Cao, Hang Gao, Karttikeya Mangalam, Qi-Zhi Cai, Minh Vo, and Jitendra Malik. Long-term human motion prediction with scene context. In *European Conference on Computer Vision (ECCV)*, 2020.

Ling-Hao Chen, Jiawei Zhang, Yewen Li, Yiren Pang, Xiaobo Xia, and Tongliang Liu. HumanMAC: Masked motion completion for human motion prediction. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023.

Xuxin Cheng, Jialong Li, Shiqi Yang, Ge Yang, and Xiaolong Wang. Open-TeleVision: Teleoperation with immersive active visual feedback. In *Conference on Robot Learning (CoRL)*, 2024.

Chenzhuang Du, Jiaye Teng, Tingle Li, Yichen Liu, Tianyuan Yuan, Yue Wang, Yang Yuan, and Hang Zhao. On uni-modal feature learning in supervised multi-modal learning. In *International Conference on Machine Learning (ICML)*, 2023.

Maria Escobar, Saturnino Maldonado-Bascón, Ali Furkan Biten, and Francesc Moreno-Noguer. EgoCast: Forecasting egocentric human pose in the wild. In *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2025.

Katerina Fragkiadaki, Sergey Levine, Li Fei-Fei, and Jitendra Malik. Recurrent network models for human dynamics. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2015.

Kristen Grauman, Andrew Westbury, Lorenzo Torresani, Kris Kitani, Jitendra Malik, Triantafyllos Afouras, Kumar Ashutosh, et al. Ego-Exo4D: Understanding skilled human activity from first- and third-person perspectives. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.

Wen Guo, Xiaoyu Du, Shuang Wang, Francisco Massa, Gérard Medioni, and Mayu Otani. Multi-person extreme motion prediction. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

Wen Guo, Yuming Du, Xi Shen, Vincent Lepetit, Juan-Manuel Perez-Rua, and Gérard Medioni. Back to MLP: A simple baseline for human motion prediction. In *IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2023.

Mary Hayhoe and Dana Ballard. Eye movements in natural behaviour. *Trends in Cognitive Sciences*, 9(4):188–194, 2005.

Jack Hessel and Lillian Lee. Does my multimodal model learn cross-modal interactions? It's harder to tell than you might think! In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2020.

Zhiming Hu, Syn Xu, and Andreas Bulling. GazeMotion: Gaze-guided human motion forecasting. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2024a.

Zhiming Hu, Syn Xu, and Andreas Bulling. HOIMotion: Forecasting human motion during human–object interactions using egocentric 3D object bounding boxes. *IEEE Transactions on Visualization and Computer Graphics*, 2024b.

Zhiming Hu, Syn Xu, and Andreas Bulling. Pose2Gaze: Eye-body coordination during daily activities for gaze prediction from full-body poses. *IEEE Transactions on Visualization and Computer Graphics*, 2024c.

Yu Huang, Junyang Lin, Chang Zhou, Hongxia Yang, and Longbo Huang. Modality competition: What makes joint training of multi-modal network fail in deep learning? (Provably). In *International Conference on Machine Learning (ICML)*, 2022.

Ashesh Jain, Amir R. Zamir, Silvio Savarese, and Ashutosh Saxena. Structural-RNN: Deep learning on spatio-temporal graphs. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

Roland S. Johansson, Göran Westling, Anders Bäckström, and J. Randall Flanagan. Eye–hand coordination in object manipulation. *The Journal of Neuroscience*, 21(17):6917–6932, 2001.

Philipp Kratzer, Simon Bihlmaier, Niteesh Balachandra Midlagajni, Rohit Prakash, Marc Toussaint, and Jim Mainprice. MoGaze: A dataset of full-body motions that includes workspace geometry and eye-gaze. *IEEE Robotics and Automation Letters*, 6(2):367–373, 2020.

Michael Land, Neil Mennie, and Jennifer Rusted. The roles of vision and eye movements in the control of activities of daily living. *Perception*, 28(11):1311–1328, 1999.

Jiaman Li, C. Karen Liu, and Jiajun Wu. Ego-Body Pose Estimation via Ego-Head Pose Estimation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023.

Maosen Li, Siheng Chen, Yangheng Zhao, Ya Zhang, Yanfeng Wang, and Qi Tian. Dynamic multiscale graph neural networks for 3D skeleton based human motion prediction. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

Maosen Li, Siheng Chen, Ya Zhang, Yanfeng Wang, and Qi Tian. Skeleton-parted graph scattering networks for 3D human motion prediction. In *European Conference on Computer Vision (ECCV)*, 2022.

Paul Pu Liang, Amir Zadeh, and Louis-Philippe Morency. Quantifying & modeling multimodal interactions: An information decomposition framework. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

Zhenyu Lou, Qianyu Qi, Yiming Liu, Tongda Li, Shuai Ding, and Yan Wang. Multimodal sense-informed forecasting of 3D human motions. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024.

Yichao Lu et al. When would vision-proprioception policies fail in robotic manipulation? *arXiv preprint arXiv:2602.12032*, 2026.

Lingni Ma et al. Nymeria: A massive collection of multimodal egocentric daily motion in the wild. In *European Conference on Computer Vision (ECCV)*, 2024.

Tiezheng Ma, Yongwei Nie, Chengjiang Long, Qing Zhang, and Guiqing Li. Progressively generating better initial guesses towards next stages for high-quality human motion prediction. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

Wei Mao, Miaomiao Liu, Mathieu Salzmann, and Hongdong Li. Learning trajectory dependencies for human motion prediction. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019.

Wei Mao, Miaomiao Liu, and Mathieu Salzmann. History repeats itself: Human motion prediction via motion attention. In *European Conference on Computer Vision (ECCV)*, 2020.

Wei Mao, Miaomiao Liu, and Mathieu Salzmann. Contact-aware human motion forecasting. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

Julieta Martinez, Michael J. Black, and Javier Romero. On human motion prediction using recurrent neural networks. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

Luke Oakden-Rayner, Jared Dunnmon, Gustavo Carneiro, and Christopher Ré. Hidden stratification causes clinically meaningful failures in machine learning for medical imaging. In *ACM Conference on Health, Inference, and Learning (CHIL)*, 2020.

Octo Model Team, Dibya Ghosh, Homer Walke, Karl Pertsch, Kevin Black, Oier Mees, Sudeep Dasari, Joey Hejna, Tobias Kreiman, Charles Xu, Jianlan You, et al. Octo: An open-source generalist robot policy. In *Robotics: Science and Systems (RSS)*, 2024.

Xiaqing Pan, Nicholas Charron, Yongqian Yang, Scott Peters, Thomas Whelan, Chen Kong, et al. Aria Digital Twin: A new benchmark dataset for egocentric 3D machine perception. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023.

Dhruvil Patel et al. UniEgoMotion: A unified model for egocentric motion reconstruction, forecasting, and generation. *arXiv preprint*, 2025.

Georgios Pavlakos, Vasileios Choutas, Nima Ghorbani, Timo Bolkart, Ahmed A. A. Osman, Dimitrios Tzionas, and Michael J. Black. Expressive body capture: 3D hands, face, and body from a single image. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

Xiaokang Peng, Yake Wei, Andong Deng, Dong Wang, and Di Hu. Balanced multimodal learning via on-the-fly gradient modulation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

Haziq Razali and Yiannis Demiris. Using eye gaze to forecast human pose in everyday pick and place actions. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2022.

Jiashun Wang, Huazhe Xu, Medhini Narasimhan, and Xiaolong Wang. Multi-person 3D motion prediction with multi-range transformers. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2021.

Lirui Wang, Xinlei Chen, and Kaiming He. Scaling proprioceptive-visual learning with heterogeneous pre-trained transformers. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024.

Weiyao Wang, Du Tran, and Matt Feiszli. What makes training multi-modal classification networks hard? In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020.

Chaoyue Xing, Wei Mao, and Miaomiao Liu. Scene-aware human motion forecasting via mutual distance prediction. In *European Conference on Computer Vision (ECCV)*, 2024.

Chenxin Xu, Robby T. Tan, Yuehua Tan, Siheng Chen, Yanfeng Wang, Xinchao Wang, and Yilun Yu. Auxiliary tasks benefit 3D skeleton-based human motion prediction. In *IEEE/CVF International Conference on Computer Vision (ICCV)*, 2023.

Haodong Yan, Zhiming Hu, and Andreas Bulling. GazeMoDiff: Gaze-guided diffusion model for stochastic human motion prediction. In *Pacific Graphics*, 2024.

Brent Yi, Vickie Engel, Kaiyu Lee, et al. Estimating body and hand motion in an ego-sensed world. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025.

Jiange Yu et al. Vision-guided action: Enhancing 3D human motion prediction with gaze-informed affordance in 3D scenes. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2025.

Ye Yuan and Kris Kitani. DLow: Diversifying latent flows for diverse human motion prediction. In *European Conference on Computer Vision (ECCV)*, 2020.

Siwei Zhang, Qianli Ma, Yan Zhang, Zhiyin Qian, Taein Kwon, Marc Pollefeys, Federica Bogo, and Siyu Tang. EgoBody: Human body shape and motion of interacting people from head-mounted devices. In *European Conference on Computer Vision (ECCV)*, 2022.

Tianhao Zhang, Zoe McCarthy, Owen Jow, Dennis Lee, Xi Chen, Ken Goldberg, and Pieter Abbeel. Deep imitation learning for complex manipulation tasks from virtual reality teleoperation. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2018.

Yang Zheng, Yanchao Yang, Kaichun Mo, Jiaman Li, Tao Yu, Yebin Liu, C. Karen Liu, and Leonidas J. Guibas. GIMO: Gaze-informed human motion prediction in context. In *European Conference on Computer Vision (ECCV)*, 2022.

A Body-State Label Definitions

In the EgoBody experiments (Section 6.3 of the main paper), we derive per-frame discrete body-state labels from the fitted 3D body meshes. Specifically, for every frame we compute kinematic quantities from the SMPL-X root translation, root orientation, and joint angles, and map them to the categorical channels listed in Table 3. These labels serve as the "ground-truth body state" used to construct the self-state prior injected into SIF3D.

Table 3: Body-state channels and their categories derived from EgoBody's 3D body fits. Each channel is computed per-frame using kinematic thresholds on the SMPL-X parameters.

State Channel	**Categories**
Translation	Stationary, Slow-moving, Moving
Turning	Not turning, Turning
Body posture & elevation	Upright, Low posture, Mid posture, Descending, Ascending
Left arm	Quiet, Active, Extending, Retracting
Right arm	Quiet, Active, Extending, Retracting

The *Translation* channel is determined by thresholding the frame-to-frame root displacement: displacements below 5 mm/frame are labelled *Stationary*, between 5–20 mm/frame *Slow-moving*, and above 20 mm/frame *Moving*. The *Turning* channel is binarised from the yaw angular velocity of the pelvis. *Body posture & elevation* combines the hip–shoulder angle with the vertical velocity of the root joint to distinguish static postures (Upright / Low / Mid) from transitions (Descending / Ascending). The two *arm* channels are computed from the shoulder–elbow –wrist kinematic chain: *Quiet* denotes negligible joint-angle change, *Active* denotes movement without large reach change, *Extending* and *Retracting* capture monotonic increases or decreases in hand-to-torso distance over a sliding window.

B Per-Joint Results on MoGaze

Table 4 reports the complete per-joint prediction errors on MoGaze under three gaze configurations: *no gaze* (motion-only baseline), *normal gaze* (the standard GazeMotion setting), and *gaze + stop-spot (oracle)*, where an oracle stop-spot prior is additionally provided. Each cell shows the Global / Local MPJPE in millimetres.

Several patterns stand out. First, the benefit of gaze is *joint-selective*: the largest absolute reductions appear in the right wrist ($136.72 \rightarrow 128.14 \rightarrow 117.38$ mm, Global) and right elbow ($107.21 \rightarrow 101.20 \rightarrow 96.07$ mm), i.e., the joints most closely linked to the reaching action that gaze anticipates. Second, the oracle stop-spot prior yields a further uniform reduction across all joints, confirming that the existing gaze signal under-supplies the intent information that the model could exploit (cf. Section 6.2). Third, the *base* joint (root) errors decrease from 63.49 to 59.38 mm, indicating that gaze also improves global trajectory prediction, not only local articulation.

C Action-Category Breakdown on GIMO (SIF3D)

Table 5 breaks down the egocentric-signal gain G on GIMO (using SIF3D as the backbone) by action category. We segment each test sequence into four mutually exclusive phases: (1) *stay*—the subject remains largely stationary near a piece of furniture; (2) *approach*—the subject walks toward a target; (2b) *arrive*—a subset of *approach* capturing the final deceleration phase just before the subject reaches the target; and (3) *free walk*—unconstrained locomotion with no clear goal object.

The results reveal pronounced heterogeneity. The *arrive* phase exhibits the highest gains across all metrics (e.g., +43.12% Global MPJPE, +45.35% Root endpoint), consistent with the intuition that egocentric scene features become most informative when the subject is about to interact with a nearby object and must adjust trajectory and posture accordingly. In contrast, the *stay* category shows the smallest gains (+22.76% Global MPJPE), as the body is largely stationary and egocentric cues add little beyond the motion-only baseline. Notably, the Local MPJPE improvements are

Table 4: Per-joint prediction errors on **MoGaze**. Each cell reports **Global / Local** MPJPE (mm). Best result per joint in **bold**.

Joint	No Gaze	Normal Gaze	Gaze + Stop-spot (Oracle)
base	63.494 / 0.000	61.592 / 0.000	**59.381 / 0.000**
pelvis	63.036 / 5.066	61.069 / 5.077	**58.888** / 5.108
torso	63.713 / 21.156	61.981 / 20.944	**60.165 / 20.497**
neck	66.468 / 31.961	64.518 / 31.668	**61.981 / 30.171**
head	71.902 / 41.613	69.444 / 41.494	**65.934 / 39.424**
linnerShoulder	65.646 / 30.244	63.766 / 29.734	**61.472 / 28.267**
lShoulder	78.922 / 54.031	76.483 / 52.103	**73.853 / 49.020**
lElbow	88.318 / 65.763	85.976 / 63.398	**84.270 / 61.236**
lWrist	100.679 / 76.266	97.173 / 74.358	**96.004 / 72.709**
rinnerShoulder	66.112 / 30.197	64.125 / 29.890	**61.641 / 28.698**
rShoulder	81.302 / 53.768	78.598 / 52.650	**75.177 / 50.543**
rElbow	107.209 / 81.023	101.199 / 76.153	**96.073 / 73.636**
rWrist	136.720 / 110.202	128.138 / 104.180	**117.378 / 97.007**
lHip	66.966 / 19.573	64.914 / 18.905	**62.959 / 17.744**
lKnee	79.920 / 50.120	77.251 / 48.725	**73.937 / 45.710**
lAnkle	81.793 / 70.366	80.046 / 68.657	**77.576 / 66.195**
lToe	89.832 / 77.708	86.597 / 75.371	**83.160 / 72.800**
rHip	66.700 / 19.570	64.665 / 18.907	**61.884 / 17.754**
rKnee	79.597 / 51.431	77.685 / 50.407	**73.501 / 47.990**
rAnkle	83.238 / 73.150	81.842 / 71.281	**78.097 / 68.706**
rToe	89.580 / 79.188	87.442 / 77.837	**81.771 / 73.956**

modest across all categories ($\leq$3.63%), indicating that egocentric signals primarily benefit *global trajectory* prediction rather than local joint articulation on this dataset.

Table 5: Egocentric-signal gain G (%) on **GIMO** (SIF3D), broken down by action category. Higher is better. Best result per metric in **bold**.

Category	# Win.	Global MPJPE	Local MPJPE	Root Traj.	Root Endpt.
1 stay	582	+22.76%	+0.90%	+23.55%	+33.52%
2 approach	1,010	+27.30%	+3.58%	+28.58%	+36.42%
2b arrive	275	**+43.12%**	+1.40%	**+42.63%**	**+45.35%**
3 free walk	194	+32.63%	**+3.63%**	+33.78%	+39.91%
Overall	**2,061**	+29.29%	+2.57%	+30.18%	+37.40%

These category-level results corroborate **Finding 1** (Section 4): the overall gain is a mixture of highly heterogeneous per-category gains. When reported only as an aggregate, the strong signal in the *arrive* phase is diluted by the near-zero improvement in the *stay* phase, producing the modest headline numbers observed in Section 3.3.

D Motion-State Breakdown on EgoBody (SIF3D)

Table 6 breaks down the egocentric-signal gain on EgoBody (C4) by the body-state channels defined in Appendix A. Each test window is assigned the state of its last predicted frame. Three input conditions are compared: the baseline SIF3D with gaze and scene (*SIF3D*), the model augmented with oracle future wearer state only (*Wearer oracle*), and the model augmented with both the wearer oracle and the partner's past state history (*Oracle + Partner*). Each cell reports the gain G in Global / Local / Root-destination-error (RDE) format.

Several patterns emerge. First, the baseline gain is strongly state-dependent: it rises with translation speed (Stationary $+1.97\% \rightarrow$ Moving $+7.81\%$ Global) and is higher during turning ($+6.48\%$) than

during straight motion (+3.92%). Second, the gain turns negative in specific states—Low posture (−2.00% Global, −9.16% RDE) and right-arm Retracting (−1.96% Global)—showing that egocentric input actively hurts in conditions it does not inform, paralleling the stopped-fixation finding on MoGaze (Section 5). Third, the oracle probes lift gains across nearly all states. In particular, they convert the Low-posture deficit from −2.00% to +5.54% (wearer oracle) and +4.33% (with partner), and the right-arm Retracting deficit from −1.96% to +5.90%, confirming that the negative gains stem from missing information rather than an inherently harmful signal.

Table 6: Per-state egocentric gain $G(\%)$ on **EgoBody** (C4, SIF3D). Each cell shows **Global / Local / RDE**. Negative values are highlighted in **bold**.

Channel	Endpoint state	# Win.	SIF3D	Wearer oracle	Oracle + Partner
Translation	Stationary	995	+1.97 / **–2.37** / +0.14	+11.46 / **–1.18** / +20.73	+13.16 / **–2.61** / +25.39
	Slow-moving	390	+5.68 / +4.20 / +2.77	+12.19 / +6.43 / +13.79	+13.26 / +7.03 / +10.70
	Moving	126	+7.81 / +5.29 / +10.45	+15.24 / +6.17 / +20.17	+18.08 / +8.34 / +22.20
Turning	Not turning	1,403	+3.92 / **–0.15** / +3.23	+12.44 / +1.26 / +18.42	+13.63 / +0.51 / +19.61
	Turning	108	+6.48 / +6.14 / +6.80	+11.80 / +8.14 / +15.09	+16.92 / +10.77 / +15.49
Body posture & elevation	Upright	1,077	+4.27 / +1.76 / +3.48	+12.46 / +1.61 / +16.84	+14.71 / +2.12 / +18.90
	Low posture	247	**–2.00** / **–6.79** / **–9.16**	+5.54 / **–2.21** / +5.25	+4.33 / **–5.79** / +7.27
	Mid posture	167	+8.88 / +4.59 / +8.97	+15.03 / +4.73 / +21.12	+16.63 / +3.94 / +18.23
	Descending	9	+8.12 / +2.51 / +3.34	+21.47 / +21.50 / +25.05	+29.13 / +28.26 / +43.19
	Ascending	11	+2.51 / **–1.87** / +2.63	+26.84 / +23.47 / +46.07	+25.71 / +21.43 / +43.18
Left arm	Quiet	1,362	+4.20 / +0.08 / +3.95	+12.68 / +1.52 / +18.04	+15.00 / +0.93 / +20.12
	Active	121	+5.01 / +3.08 / +2.58	+11.12 / +4.92 / +16.74	+8.31 / +5.69 / +9.60
	Extending	17	+3.20 / +2.40 / +2.35	+7.07 / +3.29 / +20.80	+11.22 / +5.13 / +29.59
	Retracting	11	+0.18 / +2.61 / +0.86	+6.08 / +3.04 / +15.74	**–0.95** / +0.76 / **–2.27**
Right arm	Quiet	1,327	+4.25 / +0.20 / +3.78	+13.13 / +1.74 / +18.78	+15.55 / +1.10 / +21.29
	Active	150	+4.43 / +2.34 / +3.65	+7.60 / +3.15 / +9.77	+5.94 / +4.74 / +2.20
	Extending	18	+7.61 / +1.15 / +10.23	+16.38 / +4.26 / +36.18	+12.81 / +1.32 / +30.19
	Retracting	16	**–1.96** / **–0.87** / **–4.61**	+5.90 / +1.03 / +10.51	+0.31 / **–2.36** / +4.12