

Hands on the Bars: A Multimodal Gesture Elicitation Study for Motorcycle Handlebars Comparing Visual, Tactile, and Conceptual Consensus

Marvin Reuter
Technische Hochschule Köln
Cologne, Germany
marvin.reuter@th-koeln.de

David Petersen
Technische Hochschule Köln
Cologne, Germany
david.petersen@th-koeln.de

Kevin Krüger
Brehmer GmbH & Co. KG
Wiehl, Germany
k.krueger@brehmer.gmbh

Moritz Schmidt
Brehmer GmbH & Co. KG
Wiehl, Germany
m.schmidt@brehmer.gmbh

Matthias Böhmer
Technische Hochschule Köln
Cologne, Germany
matthias.boehmer@th-koeln.de

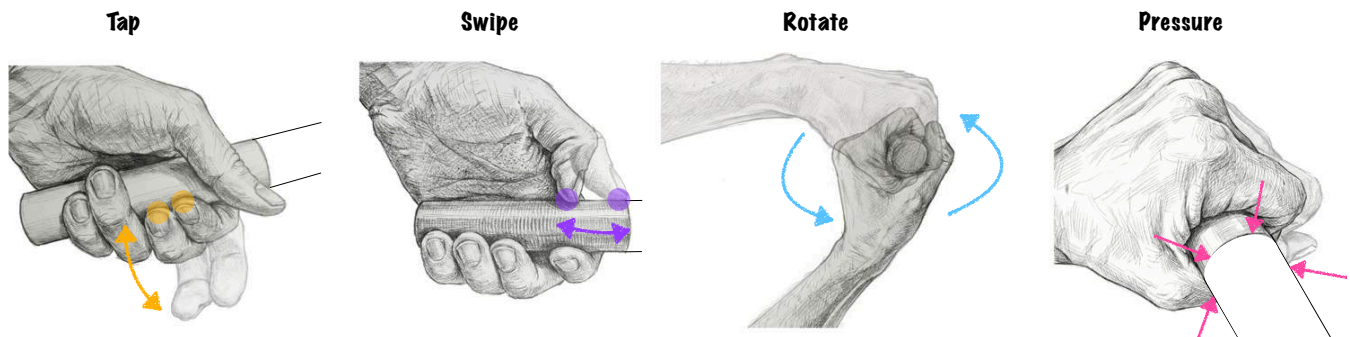


Figure 1: We propose a multimodal approach to evaluate user-defined gestures for motorcycle handlebars, such as tapping, swiping, rotating, or applying pressure. By analyzing the consensus of these gestures, we derived a gesture set optimized for interacting with motorcycle handlebars.

Abstract

The increasing digitalization of motorcycle cockpits presents a safety-critical challenge: when interacting with apps and services for communication, entertainment or navigation, one has to take away the hands from the steering handlebar. Hence, such secondary tasks compete with the primary task of riding the motorcycle. We address this conflict by systematically eliciting, structuring, and evaluating a user-defined and safe gesture set for touch interactions on the motorcycle handlebar. A Gesture Elicitation Study was conducted with 20 participants in an experimental setup. Gesture proposals were captured multimodally using video and capacitive sensors. The analysis of user consensus was performed on three levels: conceptual (based on a developed taxonomy), visual-algorithmic (based on hand trajectories), and tactile-algorithmic (based on capacitive sensor data). The main result reveals a discrepancy between

the analysis methods: only the visual consensus on motor execution correlated significantly with users' subjective ratings. This work provides two main contributions: first, it delivers the methodological insight that concrete motor execution is a crucial factor in deriving user-defined gesture sets. Second, it presents an initial, empirically validated gesture set as a foundation for designing future interaction concepts in the motorcycle context.

CCS Concepts

• **Human-centered computing** → **User studies**; *Interaction devices*; *Gestural input*.

Keywords

Gesture elicitation study, gesture input, motorcycle, handlebars



This work is licensed under a Creative Commons Attribution 4.0 International License. *NordiCHI '26, Vaasa, Finland*

© 2026 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-2373-5/26/10

<https://doi.org/10.1145/3829807.3829847>

ACM Reference Format:

Marvin Reuter, David Petersen, Kevin Krüger, Moritz Schmidt, and Matthias Böhmer. 2026. Hands on the Bars: A Multimodal Gesture Elicitation Study for Motorcycle Handlebars Comparing Visual, Tactile, and Conceptual Consensus. In *Proceedings of the 14th Nordic Conference on Human-Computer Interaction (NordiCHI '26)*, October 03–07, 2026, Vaasa, Finland. ACM, New York, NY, USA, 10 pages. <https://doi.org/10.1145/3829807.3829847>

1 Introduction

The interaction between rider and motorcycle is shifting from purely mechanical control to managing digital cockpits. Modern powered two-wheelers are increasingly equipped with TFT displays, navigation systems, and infotainment features. While these systems offer comfort, they introduce safety risks if they compete with the primary driving task, that is, keeping their hands on the steering wheel.

Unlike in cars, where drivers can temporarily take hands off the steering wheel for mid-air gestures or touch interactions, motorcycling requires a permanent physical coupling for stability and control. Statistics show that the risk of fatal accidents for motorcyclists is drastically higher compared to car occupants, making distraction and physical stability critical factors [11]. Taking a hand off the handlebar to operate a touch screen or small buttons is inherently unsafe, especially during dynamic riding maneuvers. Current alternatives like helmet headsets still require manual input, for example through buttons on the helmet itself or suffer from reliability issues in noisy environments when using voice input. Given that maintaining physical contact with the handlebars is paramount for stability, the handles themselves present a promising surface for integrating additional interaction capabilities.

While Gesture Elicitation Studies have established interaction sets for AutomotiveHCI and CyclingHCI [2, 4, 20, 27], the specific context of powered two-wheelers remains underexplored. Motorcycles differ from bicycles in speed and ergonomics, most notably the throttle on the right handle, which constraints right-hand interactions. Transferring existing gesture sets is therefore not directly feasible. To address this gap, we conducted a multimodal gesture elicitation study with 20 participants using a realistic stationary motorcycle setup. We elicited gestures for safety-critical, comfort, and infotainment functions. Beyond the elicitation metrics proposed by Vatavu and Wobbrock [22] and Vatavu [21], we employed a novel analysis approach by comparing three levels of consensus: conceptual (what users intend to do), visual (how users move), and tactile (what sensors measure).

In our research we address the following research questions:

- Which gestures can be used to interact with a motorcycle without conflicting with the primary task of riding it?
- How can a set of gestures be elicited and validated in a context where physical safety is critical?

With this research, we make two contributions: First, we present a user-elicited gesture set specifically designed for the ergonomics of powered two-wheelers, prioritizing safety. Second, we provide a comparative analysis of conceptual (agreement rate), visual (trajectory-based), and tactile (sensor-based) consensus. We demonstrate that visual consensus is a more reliable indicator of user perception than conceptual agreement rates.

2 Related Work

We build upon established methodologies in gesture elicitation and extend interaction research from the car and cycling domains to powered two-wheelers.

2.1 Gesture Elicitation and Analysis Metrics

The Gesture Elicitation Study methodology, formalized by Wobbrock et al. [25], has become standard for designing user-centered gesture sets. A core component of this method is the calculation of a consensus metric to identify the most guessable gestures [21–23, 26]. As suggested by Wobbrock et al. [25], this is achieved using the Agreement Rate (AR), which relies on researchers manually classifying gestures into semantic categories.

However, Vatavu [21] argues that the agreement rate differs from study to study due to different taxonomies defined by researchers. They introduced the *Dissimilarity Consensus*, a metric that analyzes the actual geometric and kinematic similarity of gesture trajectories using Dynamic Time Warping [19]. This distinction is critical for our study: while AR captures what users intend to do (the mental model), dissimilarity measures capture how they actually perform it. Our work provides an empirical comparison of these two metrics in a safety-critical environment.

Furthermore, elicitation studies often suffer from Legacy Bias, where users transfer interaction patterns from familiar technologies to new contexts [17]. We analyze this bias to distinguish between non-experienced riders transferring touch gestures and experienced riders transferring mechanical operating patterns.

2.2 Gestural Interaction in Transportation

To reduce visual distraction and maintain rider safety, exploring gestural inputs and haptic feedback has become an important area in vehicle research [3, 12, 18]. Building on this premise, performing human-computer interaction tasks while operating a vehicle has critical safety implications.

In the car domain, research has explored mid-air gestures to reduce visual demand and control infotainment systems [7, 16]. Angelini et al. [2] performed an elicitation study for gestures on the steering wheel, identifying a vocabulary of tapping, swiping, and pressing gestures. Research by Lee et al. [14] further confirmed the potential of these interactions. These studies demonstrate the feasibility of using gestures on the steering wheel to interact with infotainment systems without requiring the driver to reach into mid-air. However, while car drivers can temporarily release their grip on the steering wheel or change their hand position without immediately losing stability, motorcyclists face much stricter physical constraints.

In the CyclingHCI, early research highlighted the fundamental dangers of smartphone use while riding, demonstrating that it leads to increased cognitive load and diminished riding performance [6, 10]. In response, in subsequent studies researchers explored the suitability for gestures as a safer input modality while riding a bicycle: Tan et al. [20] proposed *BikeGesture*, a set of micro-gestures for urban cycling through elicitation and performance studies. Their findings confirmed both the technical feasibility and user acceptance of such interactions. Similar user centered approaches were followed by Caon et al. [4], Xiao and He [27], and Geißler et al. [8], who developed gesture sets specifically tailored to cyclists.

Other works expanded this focus to specialized interaction forms: Matviienko et al. [15] investigated the feasibility of complex tasks,

such as text messaging during the ride, while Dancu et al. [5] introduced Head-Up Displays (HUDs) to enhance safety and provide visual feedback for gestural inputs.

While these studies provide valuable insights into gesture-based interactions, their findings cannot be directly applied to all two-wheelers without further investigation. Despite the visual similarity of the input device, the transfer of interaction concepts to a different ergonomic and dynamic context requires the consideration of the specific vehicle characteristics. While bicycles share the handlebar form factor with motorcycles, the control dynamics and ergonomics differ. Cyclists operate at lower speeds and do not manage a throttle with their right hand. The rotating throttle on motorcycles permanently occupies the rider's right hand, creating an asymmetric interaction constraint that previous cycling studies do not address.

Confirming this research gap, a systematic literature review by Villarreal-Narvaez et al. [24] found no reported gesture elicitation studies for powered two-wheelers. Thus, no empirical elicitation study has specifically targeted the powered two-wheeler context, leaving a gap regarding which interactions are perceived as safe and effective while motorcycling.

3 Study Design

We conducted a user study guided by a gesture elicitation protocol by Wobbrock et al. [26] which was extended by the production method proposed by Morris et al. [17]. To ensure external validity while maintaining experimental control, the study was performed on a stationary motorcycle with a simple, single frame simulation setup.

3.1 Study Setup

The study setup consisted of a KTM 790 Adventure motorcycle¹, fixed on a stand to ensure stability while allowing participants to take an authentic riding posture, see Figure 2. A 50-inch display positioned in front of the motorcycle presented a first-person driving situation to provide visual context.

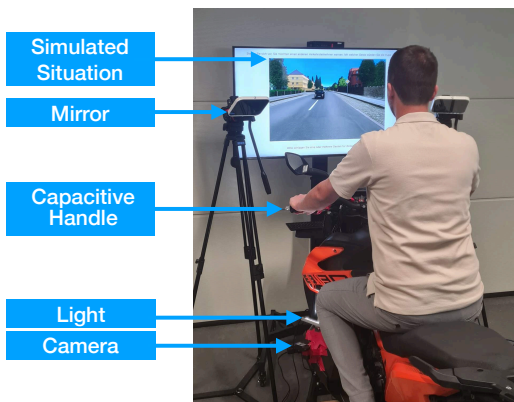


Figure 2: Annotated study setup. The setup is symmetrical, i.e. the mirror, grip, camera and light is present on both the left and right side of the motorcycle.

¹https://en.wikipedia.org/wiki/KTM_790_Adventure

We instrumented the handlebars with two complementary systems, on the right-hand handlebar and left-hand handlebar, to capture interactions multimodally.

3.1.1 Tactile Capture. To investigate gestures on the handlebar without interfering with the primary riding task, we developed a custom capacitive sensing prototype integrated into the handlebars. The tactile capture employs self-capacitance technology [9], chosen for its possible integration into the original handlebar design (e.g., the current heating foil concept). This allowed for a physical integration that preserved the original handlebar diameter, haptics, and foil concept.

The sensor design is based on the established principle of linear capacitive electrodes². To detect changes in a finger's position as a continuous change in the capacitive coupling, this principle arranges multiple capacitive electrodes along an axis. Each handle is equipped with an array of 14 triangular copper electrodes in a 2×7 configuration (see Figure 3). This specific design was chosen to evaluate sliding gestures while minimizing the electrode count to reduce signal processing overhead for potential embedded applications: as a finger moves across the handle, the shifting contact area between the electrodes allows for the detection of incremental inputs.

In preliminary tests, we iteratively validated this layout using copper foil prototypes to determine the optimal balance between hardware complexity and signal resolution for both simple swiping and complex multi-finger patterns. The final sensor, implemented as a flexible printed circuit (FPC) and wrapped around the handle, recorded data at a sampling rate of 20 Hz.

3.1.2 Visual Capture. Two cameras (Logitech c930e) were mounted on the motorcycle to record the usage of the handlebars from a bottom view. A mirror was used to simultaneously capture the top view with the same camera, ensuring full visibility of finger movements for manual analysis.

3.2 Referents

Selecting the actions for which gestures should be proposed is a fundamental component of a gesture elicitation study because it defines the scope of the gesture vocabulary. In the following, we use the term *referent* for these actions, as introduced by Wobbrock et al. [25]. Our goal was to cover a broad spectrum of typical interactions performed while riding, encompassing both vehicle control and infotainment system. We based our selection on the referents defined by Xiao and He [27] and extended them to include interactions specific to the motorcycle context. To elicit a diverse range of gesture inputs, the selected referents represent distinct interaction categories: binary functions (e.g., toggling music playback), stepped control (e.g., adjusting volume), and spatial navigation tasks (e.g., zooming a GPS map). Because interaction methods like using smartphones or infotainment systems require riders to take their hands off the handlebars, which presents a high safety risk, we included a larger proportion of infotainment-related referents. Furthermore, we included an open-ended task allowing participants to define

²<https://www1.microchip.com/downloads/aemDocuments/documents/TXFG/ApplicationNotes/ApplicationNotes/Capacitive-Touch-Sensor-Design-Guide-DS00002934-B.pdf>

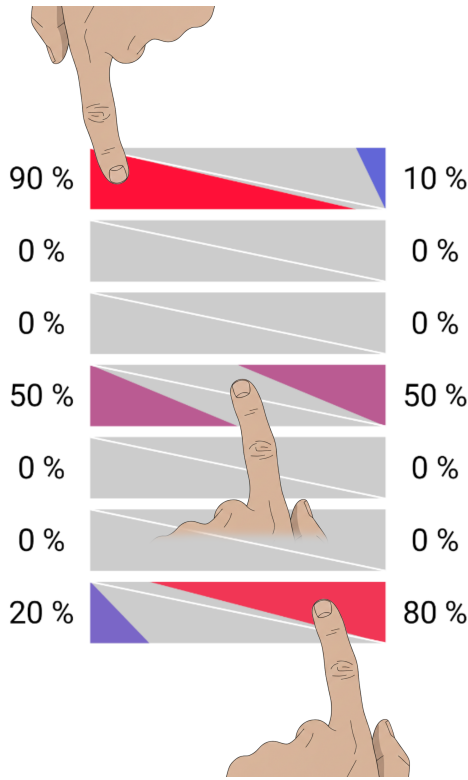


Figure 3: Capacitive sensor design with 7 layers of paired triangles. Conceptual visualization of how touching the sensor on different positions results in sensor activation percentages.

an additional referent of their choice and propose a corresponding gesture. This approach aimed to uncover novel use cases and alternative gesture mappings beyond our initial selection. The study included a total of 12 defined referents (see Table 1). During the elicitation process, these referents were presented both verbally and textually. To provide contextual grounding, each referent was accompanied by a static image depicting a realistic riding scenario. Finally, to mitigate order effects, the presentation sequence of the referents was randomized for each participant.

Infotainment		Safety	Comfort	Misc
Play music	Repeat nav. announ.	Turn signal (left)	Cruise control	User-defined
Accept call	Increase volume	Turn signal (right)	Grip heating	
Reject call	Zoom GPS map	Hazard lights		

Table 1: Overview of gesture referents by category.

3.3 Procedure

The study followed a within-subjects design and took on average 28 ($SD = 7.63$) minutes per participant. To ensure a consistent procedure across all sessions, we developed a desktop application using Compose Multiplatform for executing our study procedure, collecting information from the participants, recording the two time-coded camera videos and logging sensor data simultaneously³.

The procedure was divided into three clearly defined phases: an introductory phase for participant preparation, the gesture elicitation phase for data generation, and a concluding phase.

During the introductory phase, participants provided informed consent regarding data recording and completed a demographic questionnaire. Subsequently, the purpose of the study was explained: participants were explicitly instructed to generate multiple gesture proposals for each referent and think-aloud. Crucially, they were asked to keep their hands on the handlebars while performing the gestures. This constraint was implemented to prioritize safety by eliciting interactions that do not require releasing the grips. Furthermore, technical details regarding the underlying sensor technology were withheld to prevent technical bias, ensuring that participants' proposals were not artificially constrained by assumptions about the system's capabilities. To familiarize participants with the study, a practice run was conducted using a task not included in the study.

The gesture elicitation phase followed an adapted approach based on Morris et al. [17]. In this protocol, participants first generated a divergent set of possible gestures and subsequently identified their preferred proposal. Once selected, participants performed the gesture repeatedly; this procedure aligns with the methodology described by Vataavu [21] to ensure consistency. Finally, participants evaluated their favored gesture on 7-point Likert scales based on four criteria: goodness of fit, complexity, memorability, and safety.

3.4 Participants

We recruited 20 participants (19 male, 1 female) aged 18 to 67 years ($M = 34$, $SD = 14.7$), and 18 of them were right-handed, 1 was left-handed, and 1 was ambidextrous. The recruitment strategy aimed for a mixed-experience sample to investigate the influence of domain knowledge on interaction preferences. Based on self-reported riding frequency, the sample consisted of:

- Riders ($N=12$): Participants with experience in operating powered two-wheelers (motorcycles or scooters).
- Non-Riders ($N=8$): Participants with no prior experience in operating powered two-wheelers, relying primarily on mental models from bicycling or general technology usage.

3.5 Analysis

In our analysis, we adopted a twofold approach: a conceptual analysis to classify each proposed gesture based on a taxonomy and an algorithmic analysis based on data recorded from the camera and capacitance sensors. This methodology allows us to compare participants' intended gestures (conceptual) with their physical execution (algorithmic). Furthermore, we incorporated participant ratings and riding experience to investigate potential discrepancies

³Source code: <https://github.com/moxdlab/HandsOnTheBars>

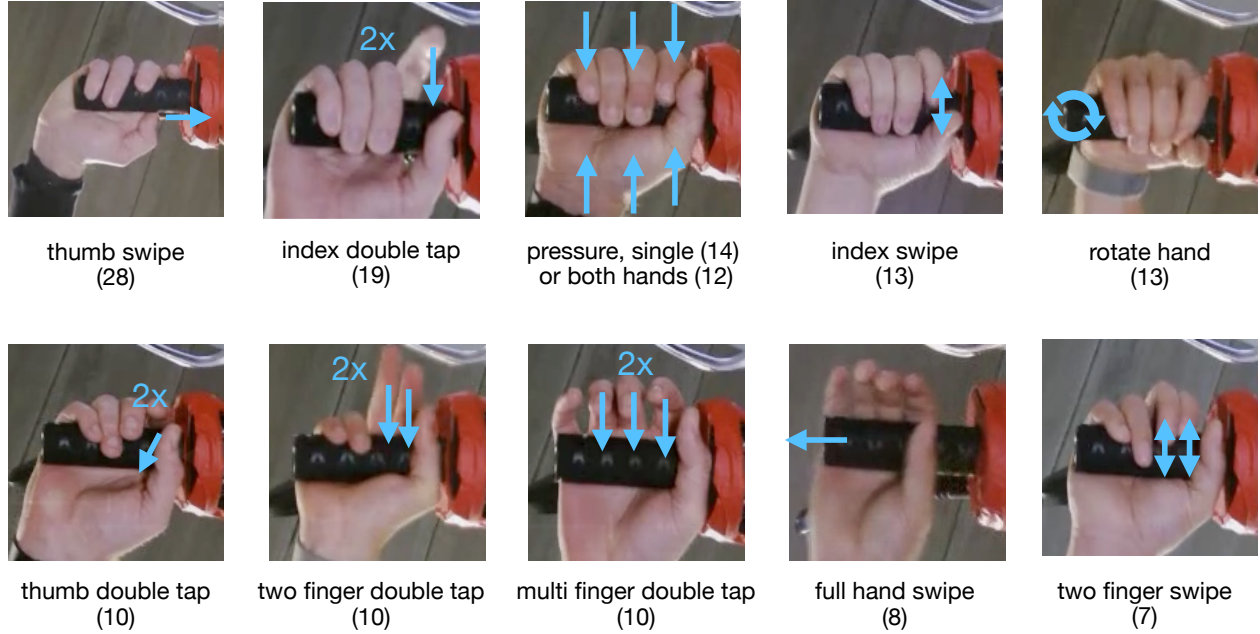


Figure 4: Most frequently proposed specific gestures (gestures that have been proposed at least 6 times). The number in parentheses represents the frequency of the proposed gesture. All gestures in the graphics shown were performed with the left hand.

and to determine whether subjective evaluations correlate with the objective agreement and similarity of the gestures.

Video recordings were manually classified. We developed a taxonomy based on Wobbrock et al.’s [26] taxonomy to classify gestures by form, handedness, and direction. To ensure the reliability of the taxonomy and classification, two raters independently classified a subset (20%) of the gestures. The resulting inter-rater reliability (Cohen’s $\kappa \geq 0.89$) indicates a high reliability in all dimensions, see Table 2. A subsequent discussion clarified the discrepancies and refined the taxonomy.

Based on this classification, we calculated the agreement rate (AR), which quantifies the degree to which participants chose the same semantic category for a referent [22].

To measure consensus based on actual motor execution rather than semantic labels, we also applied the Dissimilarity Consensus (DC) metric [21].

- Visual trajectories: To analyze the kinematics of the elicited gestures, we employed the computer vision framework MediaPipe Hands [28]. This allowed us to extract 21 3D hand

landmarks from the video footage frame-by-frame, resulting in spatial trajectories.

- Tactile profiles: The raw capacitive values from the 14 sensor channels (per handle) formed temporal touch profiles.

For both modalities, we used dynamic time warping to compute the pairwise dissimilarity between all gesture samples for a given referent. Dynamic time warping allows for comparing time-series data of varying speeds, making it ideal for gesture analysis. We then calculated the consensus growth rates (r) to quantify how strongly the group converged on similar motion patterns. We utilized the growth rate r derived from the τC curve because it provides a standardized, scale-independent measure of how quickly consensus forms among participants. This makes the metric highly robust against individual outliers in the dynamic time warping distances [23].

4 Results

Across the 20 participants and 12 referents, 238 trials were performed. During the elicitation phase, participants proposed a total

Dimension	Classes	Cohen’s Kappa (κ)	Rating acc. to Landis and Koch [13]
Form	Swipe, Tap, Rotate, Pressure change, Pinch, Combination	0.89	Almost Perfect
Handedness	Unimanual, Bimanual symmetric, Bimanual asymmetric	1.00	Almost Perfect
Direction	Directional, Undirected, Path-oriented	1.00	Almost Perfect

Table 2: Taxonomy for the categorization of gestures and their rating.

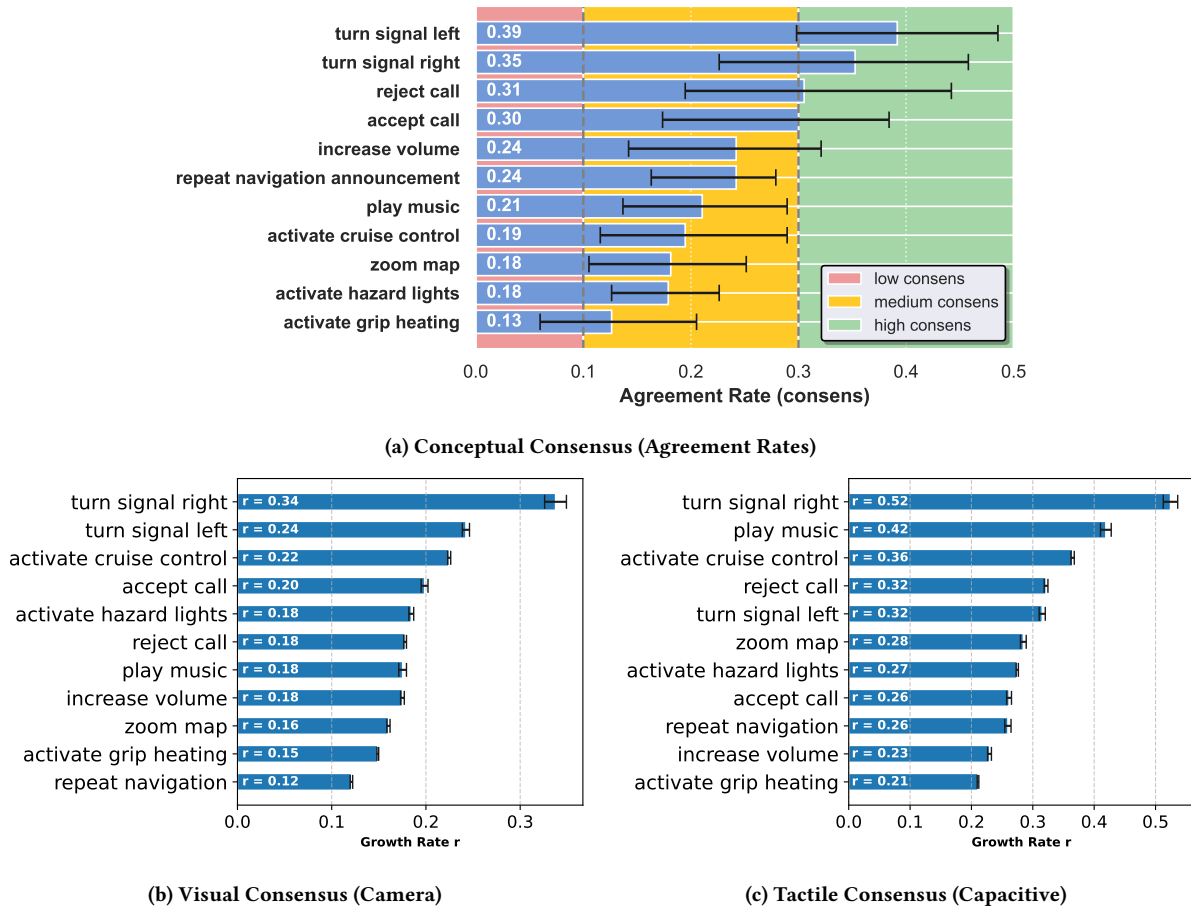


Figure 5: Agreement rate (AR) for conceptual analysis and growth rates (r) for visual and tactile analysis. Error bars indicate 95% CI.

of 400 gestures. From this, 238 preferred gestures were selected by the participants and subsequently included in the final analysis. The recorded data comprises 16.1 hours of video material.

The analysis was conducted in two steps. First, all selected gestures were manually classified based on the taxonomy (see Table 2). Second, we performed a consensus analysis to quantify agreement and investigated correlations with user ratings.

4.1 Gesture Preferences

We aggregated the distinct gesture types to determine the most practical gestures for a gesture set. The most frequently recurrent gestures included interactions with the thumb, index finger, two fingers, multiple fingers, or the whole hand; the most frequently used gesture forms were tapping, swiping, squeezing, and rotating. Figure 4 shows the most popular gestures, ordered by frequency.

4.2 Gesture Characteristics and Influence of Experience

To identify patterns and similarities among the proposed gestures, we analyzed the frequency of specific gesture characteristics. In

terms of form, swiping (33%) and tapping (27%) were the most frequently proposed interactions. Regarding the active body part, participants primarily utilized the whole hand (27%) and the thumb (22%). Analysis of the used hands revealed that the vast majority of gestures were performed unimanual (84%). Specifically, the left hand was the dominant effector (57%), while the right hand accounted for 28%. Bimanual interactions made up 16% of the proposals, all of which were performed symmetrically. Regarding directionality, 55% of the elicited gestures were static (undirected), while 45% involved a specific directional vector (e.g., swiping or rotating).

We investigated the influence of domain knowledge on gesture choice by correlating self-reported riding experience with gesture characteristics (see Figure 6a). Spearman's rank correlation revealed a significant positive correlation between motorcycle experience and the use of pressure-increasing gestures ($\rho = .17, p = .010$). Experience was negatively correlated with interactions such as tapping ($\rho = -.16, p = .011$), pressure-decreasing ($\rho = -.15, p = .025$) and rubbing ($\rho = -.13, p = .05$). This indicates that there are differences in the gestures proposed by motorcyclists and non-riders.

4.3 Consensus Analysis

To quantify the level of consensus among participants, we analyzed the three distinct metrics: conceptual, visual, and tactile consensus.

4.3.1 Conceptual Consensus. We calculated the agreement rate (AR) for each referent using the formula formalized by Vatavu and Wobbrock [22], categorizing the results into low, medium, and high (see Figure 5a). The highest conceptual consensus was observed for the referents *Activate turn signal (left)* ($AR = 0.39$) and *Activate turn signal (right)* ($AR = 0.35$), which were classified as high agreement. The referent *Activate heated grips* reached the lowest score ($AR = 0.13$). The mean agreement rate across all referents was $AR \approx 0.25$, indicating a medium level of agreement.

4.3.2 Visual Consensus. Figure 5b shows the calculated growth rates based on the Dynamic Time Warping algorithm. All calculated growth rates were statistically significant ($p < .01$) indicating the fit between the logistic function to the r-C curve. The *Activate turn signal (right)* referent achieved the highest growth rate ($r = 0.34$), followed by *Activate turn signal (left)* ($r = 0.24$). *Activate heated grips* ($r = 0.15$) and *Repeat GPS navigation announcement* ($r = 0.12$) showed the lowest growth rates.

The growth rates indicate that participants exhibited the highest kinematic similarity for turn signal referents. In contrast, referents such as activate heated grips and repeating navigation announcements resulted in the lowest growth rates, reflecting high divergence in motor execution among participants.

4.3.3 Tactile Consensus. We applied the same dissimilarity analysis to the temporal profiles recorded by the capacitive sensors ($p < .01$ for all growth rates). *Activate turn signal (right)* has the highest growth rate ($r = 0.52$), *Start music* has the second highest growth rate ($r = 0.42$). The referents *Increase volume* ($r = 0.23$) and *Activate heated grips* ($r = 0.21$) reveals the lowest growth rates (see Figure 5c).

The activate turn signal right referent elicited the most consistent tactile profiles, implying a highly uniform grip interaction. In contrast, increasing the volume and activating heated grips showed the lowest growth rates, suggesting that users proposed different gestures for these referents.

4.4 Gesture Ratings

As shown in Table 3, participants rated the selected gestures high across all dimensions (scale 1-7): Safety ($M = 5.75, SD = 1.13$), Complexity ($M = 5.68, SD = 1.15$), Memorability ($M = 5.68, SD = 1.17$), Goodness of Fit ($M = 5.29, SD = 1.38$). The average thinking time (time until proposing the first gesture) was 32.4 seconds ($SD = 20.2$).

To investigate whether the referents have an influence on the rating, we performed a one-way ANOVA between the referent and rating dimensions. The one-way ANOVA revealed a significant effect of the referent on *Goodness of Fit* ($F(10, 207) = 6.08, p < .001$) and *Memorability* ($F(10, 207) = 4.69, p < .001$). Post-hoc tests (Tukey's HSD) showed that the referents *Activate turn signal (right)* and *Activate turn signal (left)* were rated significantly higher than *Play music*, *Repeat navigation announcement*, *Activate cruise control* and *Activate heated grips*. Table 4 summarizes the results of the post-hoc analysis.

The analysis indicates that the gestures used to activate the turn signals were rated significantly higher than the other referents in terms of memorability and goodness of fit.

Dimension	Mean (M)	SD	Median	Cronbach's Alpha (α)
Goodness of Fit	5.29	1.38	6.0	0.87
Safety	5.75	1.13	6.0	0.79
Complexity	5.68	1.15	6.0	0.86
Memorability	5.68	1.17	6.0	0.87
Thinking Time (ms)	32360	20248	27780	–

Table 3: Descriptive statistics and internal consistency (Cronbach's α) for each evaluation dimension (scale 1-7) and thinking time (in ms).

4.5 Correlation of Metrics with User Perception

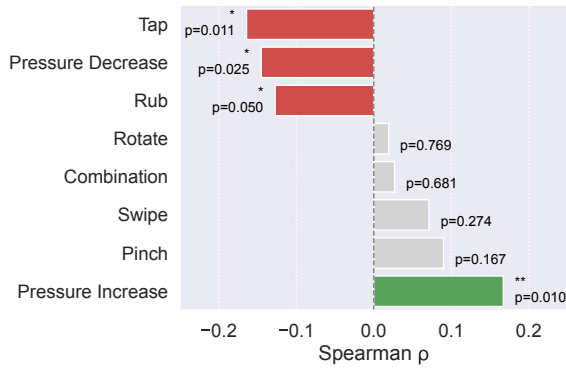
In evaluating how well the distinct consensus metrics predict user satisfaction, we correlated the agreement rate, visual growth rates, and tactile growth rates with the subjective ratings (see Table 5). The visual consensus showed significant positive correlations with user ratings: Higher visual similarity correlated significantly with

Task	Memorability		Goodness of Fit	
	vs. TS Left	vs. TS Right	vs. TS Left	vs. TS Right
Accept Call	n.s.	1.20 [0.09, 2.31]*	1.57 [0.27, 2.87]**	1.65 [0.37, 2.93]**
Grip Heating	n.s.	1.30 [0.19, 2.41]**	1.67 [0.37, 2.97]**	1.75 [0.47, 3.03]***
Play Music	1.42 [0.30, 2.55]**	1.65 [0.54, 2.76]***	1.72 [0.42, 3.02]**	1.80 [0.52, 3.08]***
Reject Call	n.s.	1.35 [0.24, 2.46]**	1.52 [0.22, 2.82]**	1.60 [0.32, 2.88]**
Repeat Nav.	1.37 [0.25, 2.50]**	1.60 [0.49, 2.71]***	1.92 [0.62, 3.22]***	2.00 [0.72, 3.28]***
Cruise Control	n.s.	n.s.	n.s.	1.30 [0.02, 2.58]*

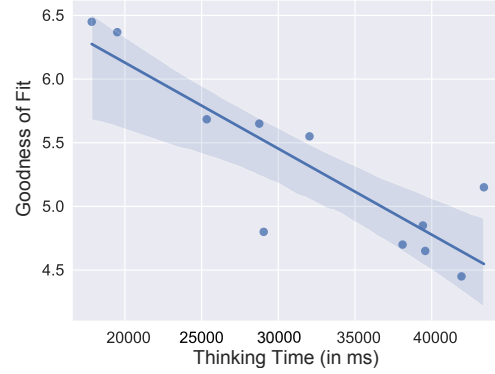
Note: TS = Turn Signal. Values represent ΔM [Lower CI, Upper CI].

* $p < .05$, ** $p < .01$, *** $p < .001$. n.s. = not significant ($p > .05$).

Table 4: Comparison of Tasks vs. Turn Signals: Mean Differences (ΔM) and 95% Confidence Intervals



(a) Correlation between motorcycle driving experience and gesture form.



(b) Correlation between thinking time and goodness of fit.

Figure 6: Correlation analysis of user experience and performance, showing (a) the impact of driving experience on gesture form and (b) the relationship between thinking time and goodness of fit.

Dimension	Conceptual Consensus (AR)	Visual Consensus (DC)	Tactile Consensus (DC)
Memorability	0.47	0.80**	0.40
Goodness of Fit	0.46	0.73**	0.38
Complexity	0.33	0.63*	0.39
Safety	-0.01	0.47	0.28
Thinking Time	-0.55	-0.64*	-0.28

Table 5: Pearson-Correlations (r) for the three consensus metrics averaged per task (N=12). Significance levels: * $p < .05$, ** $p < .01$

Memorability ($r = .80, p < .01$), *Goodness of Fit* ($r = .73, p < .01$), and *Complexity* ($r = .63, p < .05$). Furthermore, visual consensus was negatively correlated with thinking time ($r = -.64, p < .05$). Consequently, the kinematics of the gesture are significantly correlated to the participants' ratings.

5 Discussion

The results of this study provide a comprehensive view of user-defined gestures for motorcycle handlebars. While specific referents showed high agreement across all metrics (conceptual, visual, and tactile), others demonstrated discrepancies between users' semantic intent and their physical execution. Beyond these theoretical insights, the findings of this study served as the foundation for synthesizing the user proposals into a concrete gesture set.

5.1 Methodological Implication

One finding of this work is the discrepancy between conceptual consensus and visual consensus. While standard elicitation metrics (agreement rate) successfully categorize the semantic intent of users [21, 26], they failed to predict the subjective quality of the gesture. In contrast, visual consensus was a predictor for goodness of fit, complexity, and memorability. This suggests that in safety-critical domains, users may agree on a concept, but if the motor execution varies (low visual consensus), the interaction is perceived as less suitable. For future gesture elicitation studies, we argue that researchers should not solely rely on semantic classification but

must analyze the homogeneity of the physical motion to validate a gesture set.

5.2 Design Implications

Based on the proposed gestures, the ratings, and the specific constraints of the motorcycle cockpit (throttle control, clutch), we derive the following design guidelines for future powered two-wheelers interfaces:

- (1) Due to the permanent engagement of the right hand with the throttle, the gesture set must focus primarily on the left hand to maintain vehicle stability.
- (2) We observed that experienced riders prefer physical metaphors (pressure/pushing), while non-riders prefer digital metaphors (tapping). A robust gesture set should combine these: using physical-style gestures for vehicle functions (e.g., swiping for turn signals, mimicking the mechanical switch) and digital-style gestures for infotainment (e.g., tapping for music), utilizing the users' existing mental models from both domains.
- (3) To prevent accidental activation due to vibrations or road bumps, we propose using double-taps for critical discrete functions, requiring a higher level of intentionality than single taps.

Furthermore, the absence of a significant correlation between visual consensus and perceived safety indicates that high consensus does not inherently guarantee a safe interaction. This suggests

that while users may converge on identical gestures, it does not automatically translate to a subjective feeling of safety while riding a motorcycle.

Although we identified certain correlations between participants' domain knowledge and their proposed gestures, we chose to develop one unified gesture set. This choice was motivated by the consistently low correlation coefficients found in all dimensions during analysis, which imply that domain-specific differences are not strong enough to support separate gesture sets.

Regarding visual-spatial tasks such as zooming the GPS map, we note that despite the availability of a gesture, such visually demanding interactions should ideally be restricted to stationary situations to maintain rider safety.

Based on the study results and resulting design guidelines, we propose the set of gestures for interacting with a motorcycle handle, as summarized in Table 6. A visual representation of the core gesture forms (swipe, tap, rotate, and pressure) is illustrated in Figure 1.

Action (Referent)	Recommended Gesture
Turn signal (left/right)	Horizontal thumb swipe (left hand)
Activate hazard lights	Symmetric multi-finger double tap (both hands)
Accept call	Two-finger double tap (left hand)
Reject call	Full hand swipe outward (left hand)
Play music	Thumb double tap (left hand)
Adjust volume (+/-)	Vertical index finger swipe (left hand)
Activate grip heating	Squeeze both grips (bimanual pressure)
Zoom GPS map	Rotate left grip (Throttle-like motion)
Activate cruise control	Squeeze left grip (unimanual pressure)
Repeat navigation announcement	Full Hand double tap (left hand)

Table 6: Proposed gesture set mapping motorcycle interactions to specific gestures. Referents (like increase and decrease volume) were grouped because of their symmetric execution.

5.3 Comparison with AutomotiveHCI and CyclingHCI

Our results both align with and differentiate from existing research in AutomotiveHCI and CyclingHCI. Similar to car steering wheel interactions [2], we found a strong preference for thumb- and pressure-based gestures. However, while car drivers can temporarily release their grip for certain dominant gestures without immediate danger [2], the safety-critical physical coupling on a motorcycle strictly prevents this.

Compared to CyclingHCI, our findings confirm the fundamental preference for micro-gestures that allow users to maintain permanent vehicle control [4, 20]. Yet, the specific motor executions differ significantly: while cyclists frequently use individual fingers or

include brake levers for input [4, 20], our participants predominantly utilized the thumb or the whole hand exclusively on the grip. Most importantly, motorcycle interactions were heavily restricted to the left side. This divergence highlights the unique ergonomic constraints of motorcycles, demonstrating that while the principle of handlebar micro-gestures transfers from cycling, powered two-wheelers require their own empirically derived gesture set.

5.4 Limitations

Our study has limitations that should be addressed in future work:

First, the gender ratio (19 male, 1 female) was imbalanced. While this might be comparable to the current demographics of motorcycle / license ownership [11], future developments - especially eventual gesture recognition systems - must ensure evaluation on a diverse sample population to ensure optimal performance for all potential users.

Second, the study was conducted in a laboratory environment with a fixed, standing motorcycle for safety reasons. Real-world riding introduces factors like vibration, wind, and a higher cognitive load, which could affect gesture execution performance. Furthermore, participants did not wear motorcycle gloves to allow for precise computer vision tracking. Future iterations should evaluate the gesture set in real-world use with protective clothing.

Finally, the lack of correlation between tactile consensus and user ratings might be attributed to the chosen sensor design. A higher-resolution sensor grid could detect more subtle pressure patterns that align better with how users perceive them.

6 Conclusion

With this work, we presented a multi-modal gesture elicitation study specifically targeting the constraints of powered two-wheelers. By conducting the gesture elicitation study, we were able to elicit gestures and propose a specific gesture set for interacting with functions, apps, and services of a motorcycle. We identified a clear preference for unimanual, left-hand interactions that allow riders to maintain a secure grip on the handlebars, respecting the safety-critical nature of riding a motorcycle.

Methodologically, we showed that the agreement rate failed to predict user preference. Instead, the visual consensus based on kinematic trajectory similarity emerged as a superior predictor for perceived complexity, goodness of fit, and memorability. These findings advocate for a paradigm shift in gesture research: moving from purely semantic categorization to the analysis of motor execution. The key reason for this integration is our finding that high conceptual agreement (what users intend to do) can mask high kinematic variance (how they physically perform it). If motor execution is not analyzed, designers risk implementing systems that fail in real-world usage, as the physical variance of a gesture significantly reduces the user's perceived safety and goodness of fit.

Future work must develop a system that can recognize these hands-on-bar gestures and then evaluate them in dynamic riding scenarios with protective gear. As a further step, this work should be extended to the systematic elicitation of haptic gesture sets and feedback patterns (e.g., as discussed by Alvina et al. [1]).

Acknowledgments

This work was funded in part by the German Federal Ministry for Economic Affairs and Energy under grant KK5028804FM4. The photos of the hands shown in the teaser figure were modified using generative AI.

References

- [1] Jessalyn Alvina, Shengdong Zhao, Simon T. Perrault, Maryam Azh, Thijs Roumen, and Morten Fjeld. 2015. OmniVib: Towards Cross-body Spatiotemporal Vibrotactile Notifications for Mobile Phones. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. Association for Computing Machinery, New York, NY, USA, 2487–2496. doi:10.1145/2702123.2702341
- [2] Leonardo Angelini, Francesco Carrino, Stefano Carrino, Maurizio Caon, Omar Abou Khaled, Jürgen Baumgartner, Andreas Sonderegger, Denis Lalanne, and Elena Mugellini. 2014. Gesturing on the Steering Wheel: a User-elicited taxonomy. In *Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '14)*. Association for Computing Machinery, New York, NY, USA, 1–8. doi:10.1145/2667317.2667414
- [3] Niccolò Baldanzini, Giacomo Bencini, and Marco Pierini. 2011. Design and Preliminary Testing of a Haptic Handle for Powered Two Wheelers. *European Transport Research Review* 3, 1 (2011), 1–9. doi:10.1007/s12544-010-0044-z
- [4] Maurizio Caon, Rico Süsse, Benoit Grelier, Omar Abou Khaled, and Elena Mugellini. 2019. Gesturing on the Handlebar: A User-Elicitation Study for On-Bike Gestural Interaction. In *Proceedings of the 20th Congress of the International Ergonomics Association (IEA 2018)*, Sebastiano Bagnara, Riccardo Tartaglia, Sara Albolino, Thomas Alexander, and Yushi Fujita (Eds.). Springer International Publishing, Cham, 429–439. doi:10.1007/978-3-319-96071-5_46
- [5] Alexandru Dancu, Velko Vechev, Adviye Ayça Ünlüer, Simon Nilson, Oscar Nygren, Simon Eliasson, Jean-Elie Barjonet, Joe Marshall, and Morten Fjeld. 2015. Gesture Bike: Examining Projection Surfaces and Turn Signal Systems for Urban Cycling. In *Proceedings of the 2015 International Conference on Interactive Tabletops & Surfaces (ITS '15)*. Association for Computing Machinery, New York, NY, USA, 151–159. doi:10.1145/2817721.2817748
- [6] Dick De Waard, Ben Lewis-Evans, Bart Jelijs, Oliver Tucha, and Karel Brookhuis. 2014. The effects of operating a touch screen smartphone and other common activities performed while bicycling on cycling behaviour. *Transportation Research Part F: Traffic Psychology and Behaviour* 22 (Jan. 2014), 196–206. doi:10.1016/j.trf.2013.12.003
- [7] Tanja Döring, Dagmar Kern, Paul Marshall, Max Pfeiffer, Johannes Schöning, Volker Gruhn, and Albrecht Schmidt. 2011. Gestural interaction on the steering wheel: reducing the visual demand. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. Association for Computing Machinery, New York, NY, USA, 483–492. doi:10.1145/1978942.1979010
- [8] Daniel Geißler, Hymalai Bello, Esther Zahn, Emil Woop, Bo Zhou, Paul Lukowicz, and Jakob Karolus. 2024. Head 'n Shoulder: Gesture-Driven Biking Through Capacitive Sensing Garments to Innovate Hands-Free Interaction. *Proc. ACM Hum.-Comput. Interact.* 8, MCHI (Sept. 2024), 265:1–265:20. doi:10.1145/3676510
- [9] Tobias Grosse-Puppenthal, Christian Holz, Gabe Cohn, Raphael Wimmer, Oskar Bechtold, Steve Hodges, Matthew S. Reynolds, and Joshua R. Smith. 2017. Finding Common Ground: A Survey of Capacitive Sensing in Human-Computer Interaction. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. Association for Computing Machinery, New York, NY, USA, 3293–3315. doi:10.1145/3025453.3025808
- [10] Graeme Horsman and Lynne R. Conniss. 2015. Investigating evidence of mobile phone usage by drivers in road traffic accidents. *Digital Investigation* 12 (March 2015), S30–S37. doi:10.1016/j.diin.2015.01.008
- [11] International Transport Forum. 2024. Road Safety Annual Report 2024. <https://www.itf-oecd.org/road-safety-annual-report-2024>
- [12] Francisco Kiss, Robin Boldt, Bastian Pflöging, and Stefan Schneegass. 2018. Navigation Systems for Motorcyclists: Exploring Wearable Tactile Feedback for Route Guidance in the Real World. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–7. doi:10.1145/3173574.3174191
- [13] J. Richard Landis and Gary G. Koch. 1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics* 33, 1 (1977), 159–174. Publisher: International Biometric Society. doi:10.2307/2529310
- [14] Sang Hun Lee, Se-One Yoon, and Jae Hoon Shin. 2015. On-wheel finger gesture control for in-vehicle systems on central consoles. In *Adjunct Proceedings of the 7th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '15)*. Association for Computing Machinery, New York, NY, USA, 94–99. doi:10.1145/2809730.2809739
- [15] Andrii Matvienko, Jean-Baptiste Durand-Pierre, Jona Cvancar, and Max Mühlhäuser. 2023. Text Me if You Can: Investigating Text Input Methods for Cyclists. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, 1–7. doi:10.1145/3544549.3585734
- [16] Keenan R. May, Thomas M. Gable, and Bruce N. Walker. 2017. Designing an In-Vehicle Air Gesture Set Using Elicitation Methods. In *Proceedings of the 9th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '17)*. Association for Computing Machinery, New York, NY, USA, 74–83. doi:10.1145/3122986.3123015
- [17] Meredith Ringel Morris, Andreea Danielescu, Steven Drucker, Danyel Fisher, Bongshin Lee, m. c. schraefel, and Jacob O. Wobbrock. 2014. Reducing legacy bias in gesture elicitation studies. *Interactions* 21, 3 (May 2014), 40–45. doi:10.1145/2591689
- [18] Manoj Prasad, Paul Taele, Daniel Goldberg, and Tracy A. Hammond. 2014. HaptiMoto: Turn-by-Turn Haptic Route Guidance Interface for Motorcyclists. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. Association for Computing Machinery, New York, NY, USA, 3597–3606. doi:10.1145/2556288.2557404
- [19] H. Sakoe and S. Chiba. 1978. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 26, 1 (Feb. 1978), 43–49. doi:10.1109/TASSP.1978.1163055
- [20] Yanke Tan, Sang Ho Yoon, and Karthik Ramani. 2017. BikeGesture: User Elicitation and Performance of Micro Hand Gesture as Input for Cycling. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, Denver Colorado USA, 2147–2154. doi:10.1145/3027063.3053075
- [21] Radu-Daniel Vatavu. 2019. The Dissimilarity-Consensus Approach to Agreement Analysis in Gesture Elicitation Studies. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3290605.3300454
- [22] Radu-Daniel Vatavu and Jacob O. Wobbrock. 2015. Formalizing Agreement Analysis for Elicitation Studies: New Measures, Significance Test, and Toolkit. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, Seoul Republic of Korea, 1325–1334. doi:10.1145/2702123.2702223
- [23] Radu-Daniel Vatavu and Jacob O. Wobbrock. 2022. Clarifying Agreement Calculations and Analysis for End-User Elicitation Studies. *ACM Trans. Comput.-Hum. Interact.* 29, 1 (Jan. 2022), 5:1–5:70. doi:10.1145/3476101
- [24] Santiago Villarreal-Narvaez, Jean Vanderdonckt, Radu-Daniel Vatavu, and Jacob O. Wobbrock. 2020. A Systematic Review of Gesture Elicitation Studies: What Can We Learn from 216 Studies?. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference (DIS '20)*. Association for Computing Machinery, New York, NY, USA, 855–872. doi:10.1145/3357236.3395511
- [25] Jacob O. Wobbrock, Htet Htet Aung, Brandon Rothrock, and Brad A. Myers. 2005. Maximizing the guessability of symbolic input. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems*. ACM, Portland OR USA, 1869–1872. doi:10.1145/1056808.1057043
- [26] Jacob O. Wobbrock, Meredith Ringel Morris, and Andrew D. Wilson. 2009. User-defined gestures for surface computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. Association for Computing Machinery, New York, NY, USA, 1083–1092. doi:10.1145/1518701.1518866
- [27] Yiqi Xiao and Renke He. 2019. The Handlebar as an Input Field: Evaluating Finger Gestures Designed for Bicycle Riders. In *Advances in Human Aspects of Transportation (Advances in Intelligent Systems and Computing)*, Neville Stanton (Ed.). Springer International Publishing, Cham, 648–659. doi:10.1007/978-3-319-93885-1_59
- [28] Fan Zhang, Valentin Bazarevsky, Andrey Vakunov, Andrei Tkachenka, George Sung, Chuo-Ling Chang, and Matthias Grundmann. 2020. MediaPipe Hands: On-device Real-time Hand Tracking. arXiv:2006.10214 [cs]. doi:10.48550/arXiv.2006.10214