



Designing Intent Communication for Agent-Human Collaboration

Yi Li
TU Wien
Vienna, Austria
yi.5.li@tuwien.ac.at

Francesco Chiossi
Ludwig Maximilian University of
Munich (LMU)
Munich, Germany
francesco.chiossi@ifi.lmu.de

Helena Anna Frijns
Interdisciplinary Transformation
University Austria (IT:U)
Linz, Austria
helena-anna.frijns@it-u.at

Jan Leusmann
Ludwig Maximilian University of
Munich (LMU)
Munich, Germany
jan.leusmann@lmu.de

Julian Rasch
Ludwig Maximilian University of
Munich (LMU)
Munich, Germany
julian.rasch@ifi.lmu.de

Robin Welsch
Aalto University
Helsinki, Finland
robin.welsch@aalto.fi

Philipp Wintersberger
Interdisciplinary Transformation
University Austria (IT:U)
Linz, Austria
philipp.wintersberger@it-u.at

Florian Michahelles
TU Wien
Vienna, Austria
florian.michahelles@tuwien.ac.at

Albrecht Schmidt
Ludwig Maximilian University of
Munich (LMU)
Munich, Germany
albrecht.schmidt@ifi.lmu.de

Abstract

As autonomous agents, from self-driving cars to virtual assistants, become increasingly present in everyday life, safe and effective collaboration depends on human understanding of agents' intentions. Current intent communication approaches are often rigid, agent-specific, and narrowly scoped, limiting their adaptability across tasks, environments, and user preferences. A key gap remains: existing models of what to communicate are rarely linked to systematic choices of how and when to communicate, preventing the development of generalizable, multi-modal strategies. In this paper, we introduce a multidimensional design space for intent communication structured along three dimensions: Transparency (what is communicated), Abstraction (when), and Modality (how). We apply this design space to three distinct human-agent collaboration scenarios: (a) bystander interaction, (b) cooperative tasks, and (c) shared control, demonstrating its capacity to generate adaptable, scalable, and cross-domain communication strategies. By bridging the gap between intent content and communication implementation, our design space provides a foundation for designing safer, more intuitive, and more transferable agent-human interactions.

CCS Concepts

• Human-centered computing → HCI design and evaluation methods.

Keywords

Human-Robot Interaction, Intent Communication, Multimodal Interaction, Design Space

ACM Reference Format:

Yi Li, Francesco Chiossi, Helena Anna Frijns, Jan Leusmann, Julian Rasch, Robin Welsch, Philipp Wintersberger, Florian Michahelles, and Albrecht Schmidt. 2025. Designing Intent Communication for Agent-Human Collaboration. In *24th International Conference on Mobile and Ubiquitous Multimedia (MUM '25)*, December 01–04, 2025, Enna, Italy. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3771882.3771894>

1 Introduction

Advances in artificial intelligence, sensing, and robotics have integrated intelligent agents into everyday life [46]. Examples range from self-driving cars [2], cleaning robots [31], and delivery drones [25] to virtual assistants that coordinate our schedules. As these systems become increasingly autonomous, humans must understand and predict their future actions to enable effective collaboration and seamless integration into individual and societal workflows. Without effective intent communication, even the most sophisticated agents risk being misunderstood or mistrusted, impairing collaboration, increasing cognitive load, and potentially compromising safety in critical applications. Traditional approaches to intent communication in agent-human systems face significant limitations due to their heavy reliance on context-specific designs and narrowly defined intent content. For example, while external Human-Machine Interface (eHMI) have been extensively researched in automated driving contexts [9], it remains unclear whether these findings can be effectively transferred to other forms of automation, such as humanoid or household robots. This context dependency prevents these approaches from scaling across different tasks, environments, and user preferences. The core challenge lies in the absence of adaptable, generalized communication strategies that can convey intent dynamically and intuitively across diverse applications. Without



This work is licensed under a Creative Commons Attribution 4.0 International License. *MUM '25, Enna, Italy*
© 2025 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2015-4/25/12
<https://doi.org/10.1145/3771882.3771894>

such strategies, current systems remain practical only within their specific domains, creating substantial barriers to the large-scale deployment of intelligent agents across varied real-world contexts.

Previous research has explored intent communication across multiple domains, including Human-Robot Interaction (HRI) [37], autonomous vehicles [19], and digital assistants [27]. Researchers have established unified communication models to standardize intent content between agents and humans and have classified communication modalities into visual, auditory, and haptic categories. However, a critical gap remains: the missing connection between standardized communication models, communication timing, and communication modalities. This gap prevents the formation of comprehensive guidelines for designing adaptable, multi-modal systems that achieve both robustness and accessibility.

This paper addresses this challenge by introducing a multidimensional design space for intent communication that enables designers to systematically develop communication strategies based on three key dimensions: **(1) what an agent communicates (transparency)**, **(2) when it communicates (abstraction)**, and **(3) how it communicates (modality)**. We demonstrate the design space's versatility and value by applying it across three distinct collaboration scenarios: bystander interaction, cooperative tasks, and shared control situations, showing how it can guide both the design and evaluation of communication strategies. Unlike existing taxonomies dependent on applications, this design space enables scalable reasoning and cross-domain applications by encouraging novel combinations of information, abstraction, and modality tailored to specific user-agent contexts. It establishes a conceptual foundation for future agent-human communication research by supporting generalization and knowledge transfer across domains. We conclude by discussing key design considerations and observed challenges to guide future improvements of this design space.

2 Related Work

To properly motivate our proposed study, we first define our terminology before reviewing literature on current agent-human intent communication strategies and scenarios.

In this paper, we define agents as autonomous AI systems capable of sensing, reasoning, and acting to achieve goals, mirroring core attributes of human cognition. Prior to the emergence of Large Language Models (LLMs), foundational work by Russell and Norvig [39] categorized AI agents into five archetypes based on their ability to learn and reason: (1) Simple reflex agents (e.g., thermostats executing condition-action rules); (2) Model-based reflex agents (e.g., robot vacuums navigating via dynamic floor plan mapping); (3) Goal-based agents (e.g., AlphaGo's strategic gameplay); (4) Utility-based agents (e.g., autonomous vehicles optimizing safety and efficiency), and (5) Learning agents (e.g., adaptive recommendation systems like Netflix). Throughout this paper, we refer to agents generally as entities of autonomous AI systems without specifying their particular type.

Agent-Human Collaboration refers to interactive systems where autonomous agents and humans engage in coordinated activities within shared environments or toward common objectives [8].

Drawing from Pascher et al.'s [37] systematic framework, we define agent-human collaboration as encompassing three distinct interaction modalities:

- **Bystander:** Agents and humans pursue independent goals but share a physical or digital environment (e.g., pedestrians interacting with autonomous vehicles).
- **Cooperation:** Agents and humans collaborate toward a shared goal through divided subtasks (e.g., cobots assembling products with factory workers).
- **Shared Control:** Agents and humans jointly govern a task in real time (e.g., LLM-assisted playwriting for co-creating screenplays).

2.1 Current Intent Communication Strategies

In human psychology, intent refers to the mental state guiding purposeful actions toward a desired goal [5, 43]. Translating this to agent-human collaboration, intent communication involves conveying an agent's purpose, anticipated actions, and internal states to users. Unlike human-to-human exchanges, where shared context and social cues facilitate understanding, agent-human intent communication hinges on system transparency and adaptability to the agent-human relationship, which shapes how users perceive and trust autonomous behaviour.

In the field of Human-Robot Interaction (HRI), Pascher et al. [37] develop a comprehensive intent communication model that captures the key factors influencing intent communication strategies, derived from a scoping review of robot motion intent communication. This model identifies several critical dimensions that shape communication design: the robot type (robotic arm, humanoid, or mobile robot), the communication location (on-robot, on-world, or on-human), the intent type (motion, attention, state, or instruction), the information characteristics (spatial or temporal), and the human's role (collaborator, observer, coworker, or bystander). Once these factors are established within the communication framework, the actual communication typically occurs through various signal modalities, including auditory, visual, or haptic feedback.

Paralleling developments in HRI, the field of eHMI research has established similar communication frameworks. Dey [9] proposes a taxonomy that encompasses intent content specifications including: (1) vehicle states, such as yielding, driving, initiating movement, or remaining stationary; (2) spatial positioning of information communication, whether on the vehicle itself, integrated into infrastructure, embedded within roadways, or delivered directly to road users through their personal devices; and (3) audience configurations, targeting single receivers (unicast), multiple specific recipients (multicast), or multiple unspecified recipients (broadcast).

Despite the existence of taxonomies in HRI and eHMI for representing the design space of intent communication, previous frameworks focus primarily on communication content (i.e., what to communicate), while significant gaps remain in understanding communication modality (i.e., how to communicate) and timing (i.e., when to communicate). This limitation highlights the need to identify universal principles for designing intent communication that integrate content, modality, and timing in agent-human interaction.

Xu et al. [50] propose the XAIR framework for designing AI explanations in AR. The framework contains the similar problem space dimensions of when, what, and how to explain. However, their framework is AR- and explanation-specific, while ours focuses on intent communication in Human-Agent Collaboration scenarios.

In the long term, we aim to develop a unified framework for intent communication that adapts to context, ensuring intelligent agents are not just functional, but intuitive partners in human environments. This paper takes the first step towards this long-term goal by proposing a design space of intent communication to guide the exploration of such a framework.

3 Design Space for Intent Communication

In this section, we introduce a structured design space for intent communication in agent-human collaboration, building upon existing explicit signaling approaches [6, 37] in the field of automation. This design space serves as a theoretical foundation for analysing, designing, and evaluating implicit communication strategies in future collaborative systems. The design space comprises three intersecting dimensions: (1) Transparency level; (2) Task abstraction level; (3) Communication Modality. Each dimension reflects a critical aspect of intent communication—what to convey, when to convey it, and how to convey it—to foster effective and trustworthy agent-human collaboration. We display the Application, Solution and Design spaces in Figure 1.

3.1 Dimension 1: Transparency Level

The first dimension of this design space is system transparency level, which describes what internal information an agent communicates—distinct from but designed to support human Situational Awareness (SA). We adopt Endsley’s three-level SA framework [13] as a design scaffold: Transparency Level 1 exposes current system states (sensor readings, operational status) to enable human perception of the agent’s condition; Transparency Level 2 reveals reasoning processes (goals, constraints, decision logic) to facilitate comprehension of why the agent acts; and Transparency Level 3 projects future intentions (planned actions, predicted outcomes) to support human anticipation. This structure treats transparency as a controllable design property [4] while acknowledging that actual SA depends on additional factors, including timing, modality, user attention, and task demands [45], among others.

For example, a warning alert notifying users of an agent’s error state represents Transparency Level 1 (exposing current system condition), a verbal message from a driving assistant during bad weather conditions revealing the reasoning behind adaptive assistance represents Transparency Level 2 (explaining decision logic), while the projection of a cleaning robot’s motion trajectory informing users of the agent’s future plan represents Transparency Level 3 (projecting intentions) [28, 36]. While transparency alone does not guarantee SA, it provides the informational foundation necessary for SA development. Effective intent communication will therefore convey information at the appropriate transparency level to avoid both cognitive overload and undertrust of the agent[21].

3.1.1 How to Analyze and Evaluate Transparency Levels. Effective analysis and evaluation of transparency levels in human-agent

systems requires multi-faceted approaches that combine real-time behavioral assessment with structured measurement techniques.

In collaborative manufacturing, SA levels are analyzed through real-time sensing of operator attention using eye and head gaze tracking [11, 35]. Studies demonstrate that gaze features correlate strongly with standard SA measures, such as the Situational Awareness Rating Technique (SART) and Situation Awareness Global Assessment Technique (SAGAT), and can predict handover timing and overall task performance [11, 35]. This enables robots to adjust their actions in synchrony with human attentional states, supporting SA-Level 1 (perception) assessment through observable attention allocation patterns.

From a measurement perspective, Schuster and Jentsch [41] classified SA assessment methods into three categories: implicit behavioral indicators, subjective ratings, and direct probe techniques. They emphasize that no single measure captures SA comprehensively across all three levels, making multi-method approaches essential for accurate evaluation. Direct probe techniques (e.g., SAGAT) can assess comprehension (SA-Level 2) and projection (SA-Level 3) by querying operators’ understanding of system goals and anticipated future states.

Advanced evaluation frameworks employ probabilistic belief modeling [30] that integrates multimodal sensory inputs under uncertainty to maintain structured, temporally framed, and epistemically annotated representations of the robot’s knowledge state. These frameworks formalize *what the system knows* at each transparency level, providing the computational foundation for assessing whether the system’s internal transparency design supports appropriate SA development in human operators.

Collectively, these evaluation approaches support intent communication by making explicit what the agent knows, believes, and anticipates at each transparency level. This allows for a systematic assessment of whether system transparency translates into appropriate human SA and informs adaptive communication strategies.

3.1.2 How to Design for Different Transparency Levels. Effective design for SA-based intent communication requires systems that adapt transparency levels to match users’ current understanding and task demands. Recent research demonstrates that SA requirements in HRI are inherently dynamic, with required SA levels fluctuating based on environmental complexity, robot autonomy, and human roles [42]. Mismatches between required and actual SA can lead to operational inefficiencies, delays, or cognitive overload, highlighting the need for user-adaptive interface design that tailors system transparency to support appropriate human SA development.

The appropriate transparency level depends significantly on users’ familiarity with the task being performed. In highly familiar tasks, users possess strong mental models that enable them to infer agent intent from minimal perceptual cues alone (Transparency Level 1), recognizing what the agent will do based on observable actions. In less familiar tasks where mental models are weak or absent, higher transparency levels must be communicated explicitly, requiring systems to reveal their reasoning processes (Transparency Level 2) or project future states (Transparency Level 3) to support user comprehension and coordination. This familiarity-transparency relationship implies that adaptive systems should assess user expertise and modulate informational depth accordingly.

Task-phase considerations are critical for transparent design. Research indicates that transparency content should align with task phases [37]: planning or high-uncertainty phases benefit from proactive, higher-level intent communication (Transparency Levels 2-3) that supports mental model alignment and anticipatory coordination, while execution phases require more concise, reactive cues (Transparency Level 1), particularly in time-critical contexts where excessive information may impair performance. Signaling uncertainty can improve trust calibration when presented to support decision-making while maintaining operator confidence, provided it frames uncertainty as decision support rather than system unreliability.

Design strategies should leverage established SA detection methods, including gaze-based estimation [11, 35], multi-method SA measurement techniques [41], and dynamic SA frameworks [42], to implement adaptive transparency systems. These systems monitor human SA in real time and adjust transparency dynamically: increasing informational detail when SA drops below task requirements by providing reasoning explanations and outcome projections (Transparency Levels 2-3), while reducing detail when SA is sufficient to prevent cognitive overload. This closed-loop approach ensures that system transparency supports appropriate SA development without overwhelming users.

To summarize, we recommend implementing adaptive transparency levels that:

- Modulate intent information granularity in real-time based on SA indicators and task phase,
- Prioritize essential information during high-tempo situations to prevent cognitive overload,
- Include uncertainty cues when relevant to decision-making and risk assessment, and
- Calibrate transparency depth as operator expertise and mental models mature.

This design approach ensures intent communication remains contextually relevant and cognitively efficient, directly supporting SA maintenance and effective agent-human collaboration.

3.2 Dimension 2: Task Abstraction Level

The second dimension of this design space is the task abstraction level, which concerns the temporal and conceptual abstraction of the task at hand. This dimension includes operational, tactical, and strategic levels, as commonly found in both cognitive models and robotics planning literature [4, 49]. The operational level corresponds to immediate, concrete actions; the tactical level represents short-term planning and sequencing; and the strategic level reflects broader goals and adaptive reasoning. These three levels capture different temporal aspects of intent. By combining them with the three SA levels, we can categorize context-specific intent content based on what information is needed and when.

3.2.1 How to Analyze and Evaluate Task Abstraction Levels. Systematic analysis of task abstraction levels requires examining the temporal and conceptual complexity inherent in different HRI domains. Research demonstrates that HRI tasks can be categorized into three primary domains with distinct abstraction requirements: navigation [26, 48], object manipulation [1, 32], and collaborative interactions [22]. Navigation tasks involve robots communicating

movement intentions within shared environments, ranging from street crossings [48] to guided assistance [26] and industrial space navigation [3]. These tasks primarily operate at operational and tactical levels, focusing on immediate spatial decisions and short-term path planning. Object manipulation tasks center on communicating handling intentions, for example, grasping [32], handover coordination [32], and item retrieval [29, 33], which typically requires operational-level precision with tactical-level sequencing. Collaborative interaction tasks integrate multiple communication modes to achieve complex shared goals, such as assembly operations in cooperative or shared control tasks, demanding communication across all abstraction levels to convey immediate actions, subtask coordination, and strategic objectives.

3.2.2 How to Design for Different Task Abstraction Levels. Effective design requires systematic alignment of intent communication with the temporal and conceptual abstraction demands of specific tasks. This alignment ensures information is conveyed with appropriate granularity, supporting human understanding and coordination while avoiding cognitive overload.

Researchers should begin by systematically analyzing their target task to identify required abstraction levels and decompose communication requirements into distinct components:

- Operational communication addresses immediate, concrete actions the agent will perform (e.g., initiating movement, grasping objects);
- Tactical communication conveys short-term plans or action sequences linking operational steps toward near-term goals (e.g., planned path segments, ordered assembly steps); and
- Strategic communication encompasses long-term goals, overarching objectives, and adaptive reasoning guiding overall task direction.

After defining the sub-components, researchers should ensure internal consistency within each abstraction level by avoiding the blending of details from different levels within a single communication cue. This approach maintains clear conceptual boundaries and enables smooth transitions between abstraction levels as tasks progress.

3.3 Dimension 3: Communication Modality

The third dimension of the proposed design space is communication modality, which addresses the sensory channel through which intent is conveyed: visual, auditory, haptic, and potentially more speculative modalities such as smell and temperature. Common modalities differ significantly in accessibility, expressiveness, and environmental suitability. Visual modalities (lights, displays, gestures) offer spatial precision but can be ineffective in cluttered spaces [17], while auditory signals (speech, alarms) can cut through visual noise but may be drowned out in loud settings or cause irritation [34]. Haptic feedback (vibrations, force feedback) provides discrete, localized prompts but is often less expressive [7]. Multi-modal communication, combining channels either simultaneously or sequentially, can improve robustness and reduce cognitive effort through redundancy, yet may also lead to sensory overload or conflicting signals [16, 40]. The most effective intent communication results from matching modality choices to the context, the user's state, and the task demands.

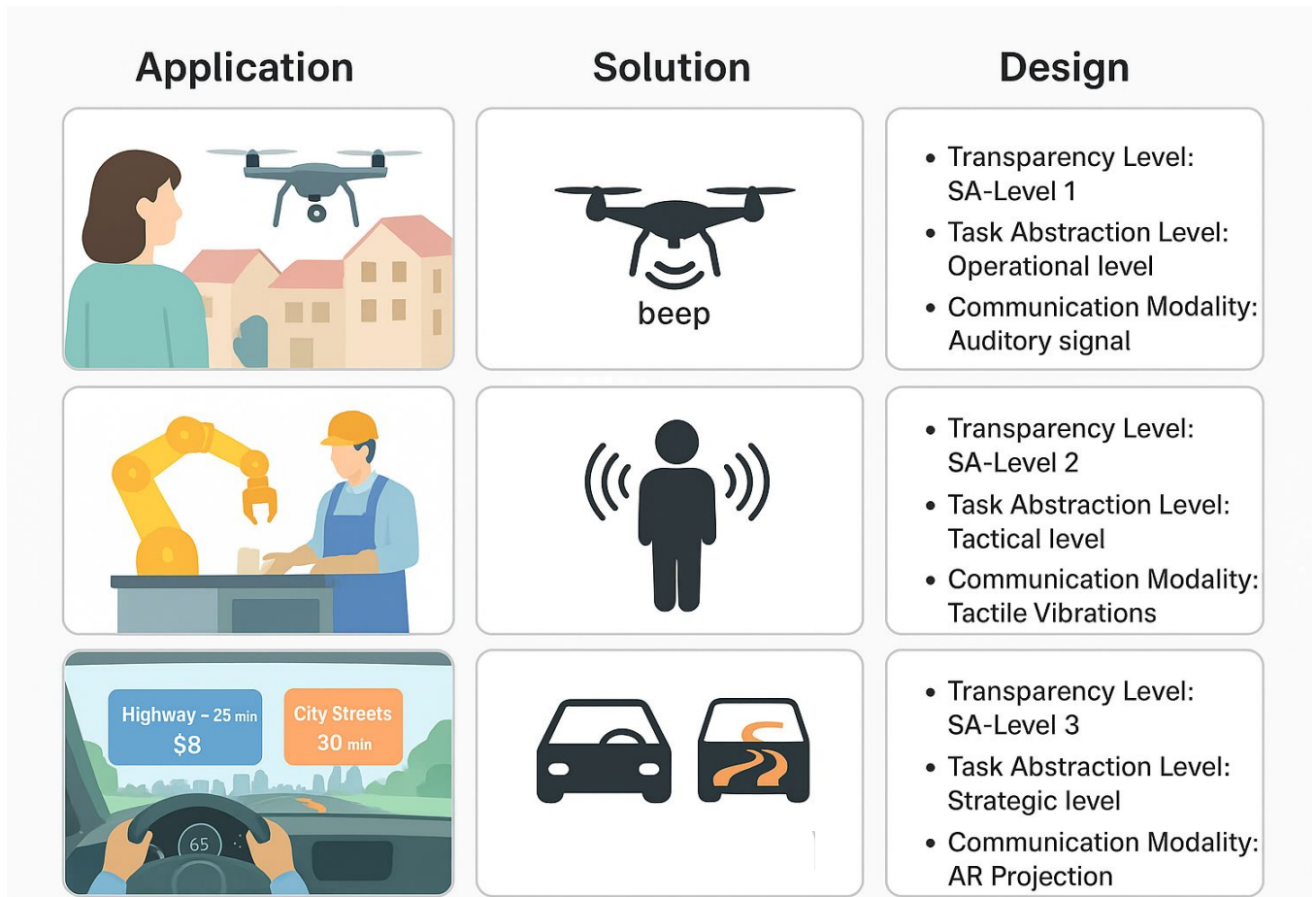


Figure 1: Illustration of three distinct application scenarios. From top to bottom: (1) Bystander Interaction, (2) Cooperative Task, and (3) Shared Control. Image generated using GPT-5.

3.3.1 How to Analyze and Evaluate Communication Modality. Systematic analysis of communication modality requires both systematic classification frameworks and domain-specific investigations. Comprehensive modality analysis employs established classification frameworks that capture the breadth of communicative channels. Feine et al. [14] provide a validated taxonomy for conversational agents, categorizing verbal, visual, auditory, and invisible cues into ten subcategories grounded in interpersonal communication theory and validated through empirical methods, offering a systematic framework for evaluating social communicative elements. Applied studies demonstrate domain-specific approaches to modality evaluation. Industrial contexts require mapping human-cobot scenarios to specific communicative modes including proxemics, kinesics, haptics, speech, and visual signals [18], while automated driving research employs comprehensive taxonomies that systematically code visual, auditory, and haptic concepts to characterize vehicle intent signaling effectiveness [10]. These domain-specific frameworks enable researchers to systematically evaluate which modality combinations are most effective for particular interaction contexts and task requirements.

This research demonstrates that communication modalities represent strategic design choices rather than arbitrary selections. Whether formalized through taxonomies or explored in applied settings, studies reveal how different channels, including unintended cues, influence intent reception. Across domains, modality functions as a central factor in human interpretation of agent actions, making systematic evaluation essential for effective intent communication design.

3.3.2 How to Configure Communication Modality. Effective modality configuration requires Situated Modality Coordination: the deliberate combination and calibration of communication channels to match environmental conditions, task demands, and user cognitive states while avoiding sensory overload or conflicting signals. This approach recognizes that both deliberate signals and incidental cues [20] influence user interpretation, requiring coherent orchestration across modalities to maintain intent clarity.

Drawing from established frameworks [10, 15, 23], modality configuration should consider signal origin, deliberateness, and clarity alongside environmental and task constraints. Domain-specific research [18, 44] demonstrates that modality selection must be driven

by communicative purpose within specific collaboration contexts rather than fixed mappings. Effective Situated Modality Coordination involves:

- Mapping modality choices to environmental constraints (noise levels, visual clutter) to ensure signal detectability;
- Leveraging complementary channel strengths, for example, haptics for immediate localized alerts, visuals for persistent spatial information;
- Integrating deliberate and incidental cues so unintended signals (motor sounds, idle motions) reinforce rather than contradict intended messages;
- Sequencing modalities to manage attention by introducing high-salience cues first followed by detailed information through slower channels;
- Empirically testing multimodal configurations to detect conflicts or cognitive overload before deployment; and
- Adjusting modality emphasis based on task phases or user states while maintaining predictable patterns for user orientation.

Properly configured modalities should form a unified communicative act that matches interaction context, fosters shared awareness, minimizes ambiguity, and facilitates effective agent-human collaboration through strategically coordinated sensory channels.

4 Design Space Application

To demonstrate the practical value of the proposed three-axis design space for intent communication in human-agent collaboration, we applied it to three distinct human-agent collaboration scenarios: bystander interaction, cooperative tasks, and shared control. The design space is conceived as a cube whose axes represent, first, the depth of information disclosed according to the transparency hierarchy; second, the abstraction level of the task in question—operational, tactical, or strategic; and third, the communication modality, namely visual, auditory, or haptic. Treating each intent message as a point inside this design space enables systematic reasoning about what the agent reveals, when it does so, and through which sensory channel.

The first application involves a bystander scenario where a delivery drone operates in a residential neighborhood for daytime delivery mission. As the drone descends toward a designated yard location to drop off a package, it emits a brief but distinctive auditory beep at regular intervals to alert nearby pedestrians and residents of its immediate presence and ongoing delivery activity. Here the cue reveals the drone’s basic perception of the drop-off location (SA-Level 1) at an operational level via auditory signals. This auditory communication is particularly effective in public residential spaces where visual attention may be divided, allowing pedestrians to quickly recognize the drone’s presence, thereby supporting safe coexistence in shared airspace.

The second application concerns a cooperation scenario where a collaborative robot assists a human in an industrial drilling task. As the robot approaches completion of the alignment phase and prepares for synchronous motion with the human operator, it delivers tactile vibrations through wearable devices worn by the worker, signaling readiness for coordinated action. Here the cue exposes the

robot’s comprehension of the joint task state and timing requirements (SA-Level 2) at a tactical level with haptic feedback to convey operational readiness. This haptic communication allows the human operator to understand the current progress of the alignment task and anticipate the transition to the next phase of coordinated drilling, providing a communication method that is both private to the individual operator and robust against the ambient noise typical of industrial workshops.

The third application features a shared control scenario where an autonomous vehicle suggests a highway route to the airport under time pressure, but the driver prefers taking city streets. Before the trip begins, the vehicle projects its proposed route onto the roadway using augmented reality display technology, showing side-by-side route options with clear time and cost trade-offs for each alternative. Here the cue projects the vehicle’s future planning decisions and strategic reasoning for route selection (SA-Level 3) at a strategic level through visual projection signals. This visual communication enables the driver to evaluate route options, allowing them to accept the vehicle’s recommendation or override the autonomous system’s suggestion, maintaining human authority while benefiting from the vehicle’s advanced sensing and planning capabilities.

As these examples occupy distinct cells of the design space, they illustrate how designers can consciously distribute communicative effort across space, time, and modality rather than defaulting to a single channel or information tier. Moreover, the exercise exposes sparsely populated regions—strategic-level haptics, for instance—that may hold untapped potential in supervisory contexts or environments with constrained sensory channels.

4.1 Analytical Benefits and Design Implications

Mapping a product’s intent messages often reveals overreliance on particular modalities, such as vision, or neglect of deeper SA levels. For example, automated vehicles frequently use visual eHMI like displays or lights as the primary intent communication channel, despite evidence that such reliance can limit accessibility and robustness, especially in visually cluttered or low-visibility environments [38, 47]. Moreover, research in multimodal interaction consistently finds that systems relying on a single modality, especially vision, risk exclusion of users in diverse real-world contexts and miss opportunities for more robust, redundant, and user-centered communication strategies [12, 24].

The geometric view, therefore, supports more principled trade-offs between informativeness and cognitive economy. Because intent “chunks” are labeled solely by their coordinates in the design space, designers can also transplant successful combinations across domains: an SA-Level 1-auditory cue that excels in a warehouse frequently transfers to healthcare with minimal adaptation, while SA-Level 3-visual projections suit strategic planning tasks in comparatively static settings. Empty regions of the design space flag research opportunities where novel combinations could address unmet user needs.

Moreover, the model encourages generalization: by abstracting communication strategies from specific use cases, it becomes possible to derive cross-domain principles. For instance, SA-Level

Table 1: Critical considerations when implementing the design space

Consideration	Impact for Designers
Trust vs. Cognitive Load	Increase transparency only to the point it enhances calibrated trust; beyond that threshold, extra detail becomes distracting and counter-productive.
Temporal Coordination	Align the timing of SA-Level cues with users' decision windows: deliver SA-Level 1 in real time, defer SA-Level 3 until cognitive lulls.
Ergonomics and Accessibility	Choose modalities that suit the sensory context and combine channels to offset visual, auditory, or haptic limitations.
Generalizability	Tag intent messages solely by SA-Level depth, task horizon, and modality to create reusable chunks that transfer across domains.
Explicit vs. Implicit Signalling	Balance overt notifications (e.g., speech) with subtler motion or light patterns to maintain task flow while preserving interpretability.
Cultural & Ethical Adaptability	Tailor colour semantics, privacy levels, and autonomy cues to local norms; use the design space as a scaffold for responsible adaptation.

1–auditory combinations may consistently excel in dynamic environments, while SA-Level 3–visual cues may be more suited to strategic planning in static settings.

5 Design Considerations and Observed Challenges

Applying the design space across the three scenarios and several pilot deployments unearthed six recurring considerations that merit attention during implementation, which we summarize in Table 1.

First, the tension between transparency and cognitive load remains acute. Rich explanations undoubtedly foster calibrated trust, yet higher-level SA-Level disclosures can overwhelm attention if delivered too often or at inopportune moments. Judicious pacing is therefore essential.

Second, task abstraction is as critical as transparency. SA-Level 1 cues typically require delivery within milliseconds of an impending action, whereas SA-Level 3 messages can be deferred until natural cognitive lulls, thereby avoiding premature or stale updates.

Third, the modality axis foregrounds ergonomic and accessibility concerns. Visual messages excel when sightlines are clear, auditory signals dominate public but visually cluttered spaces, and haptics prove invaluable when both sight and hearing are compromised. Multimodal blends can further mitigate the limitations of any single channel.

Fourth, the design space's abstract structure promotes generalisability. When tagging intent messages only with SA-Level depth, task horizon, and modality, engineers can implement open libraries that can be used from factory floors to autonomous vehicles without extensive rewrites.

Fifth, the design space accommodates both explicit and implicit signaling. Designers may layer overt speech with subtler kinematic or light cues, balancing interpretability against continuity of task flow.

Finally, cultural and ethical variation cannot be ignored. Colour semantics, privacy norms, and expectations of autonomy differ across regions and sectors; the design space does not dictate solutions but offers a structured basis for reasoning about adaptation and compliance.

Overall, the design space enables systematic reasoning about intent communication strategies in terms of informational depth, temporal dynamics, modality selection, and contextual fit. By making explicit the dimensions along which communication design can be evaluated and optimized, the design space advances toward a more principled, adaptable, and user-centered approach to agent-human collaboration.

6 Conclusion

This paper introduces a multidimensional design space that systematically connects the three fundamental dimensions of intent communication: what agents communicate (transparency), when they communicate (abstraction), and how they communicate (modality). For each dimension, we provide guidelines to analyze, evaluate, and design the current factor to better practice intent communication. We conceptualize intent as modular components that occupy distinct positions within this design space, providing a practical foundation for developing and evaluating transparent agent behaviors across domains. Applied across three distinct collaboration scenarios, this design space enables a more systematic approach to designing intent communication strategies across agent-human systems. We consider this design space as a generative foundation for ongoing research, evaluation, and tool development. Our long-term goal is to develop a unified framework that supports transparent, trustworthy, and adaptive collaboration between humans and intelligent agents across the expanding landscape of agent-human interaction.

Acknowledgments

This research was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under project number 529719707, "Multimodale Kommunikation von Intention in autonomen Systemen", and by the Austrian Science Fund (FWF) under project number I6682-N as part of the joint DFG–FWF collaboration on multimodal intention communication in autonomous systems.

References

- [1] Rasmus Andersen, Ole Madsen, Thomas Moeslund, and Heni Ben Amor. 2016. Projecting Robot Intentions into Human Environments. In *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*.

- doi:10.1109/ROMAN.2016.7745145
- [2] Claudine Badue, Rănik Guidolini, Raphael Vivacqua Carneiro, Pedro Azevedo, Vinicius B Cardoso, Avelino Forechi, Luan Jesus, Rodrigo Berriel, Thiago M Paixao, Filipe Mutz, et al. 2021. Self-driving cars: A survey. *Expert systems with applications* 165 (2021), 113816.
 - [3] Ravi Teja Chadalavada, Henrik Andreasson, Robert Krug, and Achim J. Lilienthal. 2015. That's on my mind! robot to human intention communication through on-board projection on shared floor space. In *2015 European Conference on Mobile Robots (ECMR)*. 1–6. doi:10.1109/ECMR.2015.7403771
 - [4] Jessie YC Chen and Michael J Barnes. 2014. Human-agent teaming for multirobot control: A review of human factors issues. *IEEE Transactions on Human-Machine Systems* 44, 1 (2014), 13–29. doi:10.1109/THMS.2013.2293535
 - [5] Francesco Chiossi, Luke Haliburton, Changkun Ou, Andreas Martin Butz, and Albrecht Schmidt. 2023. Short-Form Videos Degrade Our Capacity to Retain Intentions: Effect of Context Switching On Prospective Memory. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) (CHI '23). Association for Computing Machinery, New York, NY, USA, Article 30, 15 pages. doi:10.1145/3544548.3580778
 - [6] Francesco Chiossi, Julian Rasch, Robin Welsch, Albrecht Schmidt, and Florian Michahelles. 2025. Designing Intent: A Multimodal Framework for Human-Robot Cooperation in Industrial Workspaces. *arXiv preprint arXiv:2506.15293* (2025).
 - [7] Francesco Chiossi, Steeven Villa, Melanie Hauser, Robin Welsch, and Lewis Chuang. 2022. Design of on-body tactile displays to enhance situation awareness in automated vehicles. In *2022 IEEE 9th International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)*. IEEE, 1–6. doi:10.1109/CIVEMSA53371.2022.9853690
 - [8] Nazli Cila. 2022. Designing human-agent collaborations: Commitment, responsiveness, and support. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–18. doi:10.1145/3491102.3517500
 - [9] Debargha Dey. 2020. Taming the eHMI jungle: A classification taxonomy to guide, compare, and assess the design principles of automated vehicles' external human-machine interfaces. 7, 100174 (2020), 24.
 - [10] Debargha Dey, Azra Habibovic, Andreas Löcken, Philipp Wintersberger, Bastian Pflöging, Andreas Riemer, Marieke Martens, and Jacques Terken. 2020. Taming the eHMI Jungle: A Classification Taxonomy to Guide, Compare, and Assess the Design Principles of Automated Vehicles' External Human-Machine Interfaces. *Transportation Research Interdisciplinary Perspectives* 7 (2020), 100174. doi:10.1016/j.trip.2020.100174
 - [11] Amir Dini, Cornelia Murko, Saeed Yahyanejad, Ursula Augsdörfer, Michael Hofbauer, and Lucas Paletta. 2017. Measurement and Prediction of Situation Awareness in Human-Robot Interaction based on a Framework of Probabilistic Attention. In *Proceedings of the 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. doi:10.1109/IROS.2017.8206301
 - [12] Elias Dritsas, Maria Trigka, Christos Troussas, and Phivos Mylonas. 2025. Multimodal interaction, interfaces, and communication: a survey. *Multimodal Technologies and Interaction* 9, 1 (2025), 6. doi:10.3390/mti9010006
 - [13] Mica R Endsley. 2017. From here to autonomy: lessons learned from human-automation research. *Human factors* 59, 1 (2017), 5–27. doi:10.1177/0018720816681350
 - [14] Jasper Feine, Ulrich Gnewuch, Stefan Morana, and Alexander Maedche. 2019. A Taxonomy of Social Cues for Conversational Agents. *International Journal of Human-Computer Studies* 132 (2019), 138–161. doi:10.1016/j.ijhcs.2019.07.009
 - [15] Jasper Feine, Ulrich Gnewuch, Stefan Morana, and Alexander Maedche. 2019. A taxonomy of social cues for conversational agents. *International Journal of human-computer studies* 132 (2019), 138–161.
 - [16] Jesse Fox and Andrew Gambino. 2021. Relationship development with humanoid social robots: Applying interpersonal theories to human-robot interaction. *Cyberpsychology, Behavior, and Social Networking* 24, 5 (2021), 294–299. doi:10.1089/cyber.2020.0181
 - [17] Ramsundar Kalpagam Ganesan, Yash K Rathore, Heather M Ross, and Heni Ben Amor. 2018. Better teaming through visual cues: how projecting imagery in a workspace can improve human-robot collaboration. *IEEE Robotics & Automation Magazine* 25, 2 (2018), 59–71. doi:10.1109/MRA.2018.2815655
 - [18] Stephanie Gross and Brigitte Krenn. 2024. A Communicative Perspective on Human–Robot Collaboration in Industry: Mapping Communicative Modes on Collaborative Scenarios. *International Journal of Social Robotics* 16, 7 (2024), 1315–1332. doi:10.1007/s12369-023-00991-5
 - [19] Azra Habibovic, Victor Malmsten Lundgren, Jonas Andersson, Maria Klingegård, Tobias Lagström, Anna Sirkka, Johan Fagerlönn, Claes Edgren, Rikard Fredriksson, Stas Krupenia, et al. 2018. Communicating intent of automated vehicles to pedestrians. *Frontiers in psychology* 9 (2018), 1336. doi:10.3389/fpsyg.2018.01336
 - [20] Frank Hegel, Sebastian Gieselmann, Annika Peters, Patrick Holthaus, and Britta Wrede. 2011. Towards a Typology of Meaningful Signals and Cues in Social Robotics. In *20th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 72–78. doi:10.1109/ROMAN.2011.6005288
 - [21] Kevin Anthony Hoff and Masooda Bashir. 2015. Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors* 57, 3 (2015), 407–434. doi:10.1177/0018720814547570
 - [22] Guy Hoffman and Cynthia Breazeal. 2004. Collaboration in Human-Robot Teams. *Collection of Technical Papers - AIAA 1st Intelligent Systems Technical Conference 2* (09 2004). doi:10.2514/6.2004-6434
 - [23] Patrick Holthaus, Trenton Schulz, Gabriella Lakatos, and Rebekka Soma. 2023. Communicative Robot Signals: Presenting a New Typology for Human–Robot Interaction. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction (HRI '23)*. ACM, 132–141. doi:10.1145/3568162.3578631
 - [24] Yongquan 'Owen' Hu, Jingyu Tang, Xinya Gong, Zhongyi Zhou, Shuning Zhang, Don Samitha Elvitigala, Florian 'Floyd' Mueller, Wen Hu, and Aaron J. Quigley. 2025. Vision-Based Multimodal Interfaces: A Survey and Taxonomy for Enhanced Context-Aware System Design. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 758, 31 pages. doi:10.1145/3706598.3714161
 - [25] Robin Kellermann, Tobias Biehle, and Liliann Fischer. 2020. Drones for parcel and passenger transportation: A literature review. *Transportation Research Interdisciplinary Perspectives* 4 (2020), 100088. doi:10.1016/j.trip.2019.100088
 - [26] K. L. Koay, G. Lakatos, D. S. Syrdal, M. Gácsi, B. Bereczky, K. Dautenhahn, A. Miklósi, and M. L. Walters. 2013. Hey! There is someone at your door. A hearing robot using visual communication signals of hearing dogs to communicate intent. In *2013 IEEE Symposium on Artificial Life (ALife)*. 90–97. doi:10.1109/ALIFE.2013.6602436
 - [27] Nicolas Lair, Clement Delgrange, David Mugisha, Jean-Michel Dussoux, Pierre-Yves Oudeyer, and Peter Ford Dominey. 2020. User-in-the-loop adaptive intent detection for instructable digital assistant. In *Proceedings of the 25th International Conference on Intelligent User Interfaces*. 116–127. doi:10.1145/3377325.3377490
 - [28] John D Lee and Katrina A See. 2004. Trust in automation: Designing for appropriate reliance. *Human factors* 46, 1 (2004), 50–80.
 - [29] Jan Leusmann, Ludwig Felder, Chao Wang, and Sven Mayer. 2025. Understanding Preferred Robot Reaction Times for Human-Robot Handovers Supported by a Deep Learning System. In *2025 20th ACM/IEEE International Conference on Human-Robot Interaction (2025-03-06) (HRI)*. IEEE/ACM. <https://sven-mayer.com/wp-content/uploads/2025/01/leusmann2025understanding.pdf>
 - [30] Pierre Lison, Carsten Ehrler, and Geert-Jan M. Kruijff. 2010. *Belief Modelling for Situation Awareness in Human-Robot Interaction*. Technical Report. German Research Centre for Artificial Intelligence (DFKI).
 - [31] Rajesh Kannan Megalingam, Shree Rajesh Raagul Vadivel, Sai Smaran Kotaprolu, Bagathi Nithul, Devisetty Vijay Kumar, and Gaurav Rudravaram. 2025. Cleaning robots: A review of sensor technologies and intelligent control strategies for cleaning. *Journal of Field Robotics* (2025). doi:10.1002/rob.22515
 - [32] Rhys Newbury, Akansel Cosgun, Tysha Crowley-Davis, Wesley P. Chan, Tom Drummond, and Elizabeth Croft. 2022. Visualizing Robot Intent for Object Handovers with Augmented Reality. *arXiv:2103.04055 [cs.RO]* <https://arxiv.org/abs/2103.04055>
 - [33] Monica Nicolescu and Maja Matari. 2005. Task learning through imitation and human-robot interaction. *Imitation and Social Learning in Robots, Humans and Animals: Behavioural, Social and Communicative Dimensions* (01 2005). doi:10.1017/CBO9780511489808.027
 - [34] Bastian Orthmann, Iolanda Leite, Roberto Bresin, and Ilaria Torre. 2023. Sounding robots: Design and evaluation of auditory displays for unintentional human-robot interaction. *ACM Transactions on Human-Robot Interaction* 12, 4 (2023), 1–26. doi:10.1145/3611655
 - [35] Lucas Paletta, Amir Dini, Cornelia Murko, Saeed Yahyanejad, and Ursula Augsdörfer. 2019. Estimation of Situation Awareness Score and Performance Using Eye and Head Gaze for Human-Robot Collaboration. In *Proceedings of the 2019 Symposium on Eye Tracking Research and Applications (ETRA '19)*. doi:10.1145/3314111.3322504
 - [36] Raja Parasuraman and Victor Riley. 1997. Humans and automation: Use, misuse, disuse, abuse. *Human Factors* 39, 2 (1997), 230–253. doi:10.1518/00187209778543886
 - [37] Max Pascher, Uwe Gruenefeld, Stefan Schneegass, and Jens Gerken. 2023. How to Communicate Robot Motion Intent: A Scoping Review. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg Germany). ACM, 1–17. doi:10.1145/3544548.3580857
 - [38] Fabienne Roche, Anna Somieski, and Stefan Brandenburg. 2019. Behavioral changes to repeated takeovers in highly automated driving: effects of the takeover-request design and the nondriving-related task modality. *Human factors* 61, 5 (2019), 839–849. doi:10.1177/0018720818814963
 - [39] Stuart J Russell and Peter Norvig. 2016. *Artificial intelligence: a modern approach*. Pearson.
 - [40] Nadine B Sarter. 2006. Multimodal information presentation: Design guidance and research challenges. *International journal of industrial ergonomics* 36, 5 (2006), 439–445.
 - [41] David Schuster and Florian Jentsch. 2011. Measurement of Situation Awareness in Human-Robot Teams. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 55. 1496–1500. doi:10.1177/1071181311551311
 - [42] Hashini Senaratne, Leimin Tian, Pavan Sikka, Jason Williams, David Howard, Dana Kulić, and Cécile Paris. 2025. A Framework for Dynamic Situational

- Awareness in Human-Robot Teams: An Interview Study. *ACM Transactions on Human-Robot Interaction* (2025). doi:10.1145/3743693
- [43] Paschal Sheeran. 2002. Intention—behavior relations: a conceptual and empirical review. *European review of social psychology* 12, 1 (2002), 1–36.
 - [44] Yoshiaki Shiokawa, Winnie Chen, Aditya Shekhar Nittala, Jason Alexander, and Adwait Sharma. 2025. Beyond Vacuuming: How Can We Exploit Domestic Robots' Idle Time?. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '25)*. ACM, 1–17. doi:10.1145/3706598.3714266
 - [45] Rebecca Stewart, Neville A Stanton, Don Harris, Chris Baber, Paul Salmon, Mel Mock, Kerry Tatlock, Linda Wells, and Alison Kay. 2008. Distributed situation awareness in an airborne warning and control system: application of novel ergonomics methodology. *Cognition, Technology & Work* 10, 3 (2008), 221–229. doi:10.1007/s10111-007-0094-8
 - [46] Sirko Straube and Frank Kirchner. 2024. Robotics and autonomy: From space robots to applications in daily life. *Research Outreach* 140 (2024), 24–27.
 - [47] Sule Tekkesinoglu, Azra Habibovic, and Lars Kunze. 2025. Advancing Explainable Autonomous Vehicle Systems: A Comprehensive Review and Research Roadmap. *J. Hum.-Robot Interact.* 14, 3, Article 39 (March 2025), 46 pages. doi:10.1145/3714478
 - [48] Atsushi Watanabe, Tetsushi Ikeda, Yoichi Morales, Kazuhiko Shinozawa, Takahiro Miyashita, and Norihiro Hagita. 2015. Communicating robotic navigational intentions. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 5763–5769. doi:10.1109/IROS.2015.7354195
 - [49] Philipp Wintersberger. 2020. Automated driving: Towards trustworthy and safe human-machine cooperation. (2020). <https://epub.jku.at/obvulihs/content/titleinfo/5108376> Doctoral Thesis.
 - [50] Xuhai Xu, Anna Yu, Tanya R. Jonker, Kashyap Todi, Feiyu Lu, Xun Qian, João Marcelo Evangelista Belo, Tianyi Wang, Michelle Li, Aran Mun, Te-Yen Wu, Junxiao Shen, Ting Zhang, Narine Kokhlikyan, Fulton Wang, Paul Sorenson, Sophie Kim, and Hrvoje Benko. 2023. XAIR: A Framework of Explainable AI in Augmented Reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg Germany). ACM, 1–30. doi:10.1145/3544548.3581500