How People Envision Robot Interruption Behaviors

Jan Leusmann [1], Maximiliane Windl [1], Michael Gienger [2], Chao Wang [2], Sven Mayer [3,4]

***Abstract*— In everyday human–robot interaction, robots may need to interrupt people while they are engaged in ongoing activities, for example, to ask questions, provide reminders, or warn about potential issues. Designing appropriate robot interruptions is challenging, as their effectiveness and social acceptability depend on a range of contextual factors. We conducted an in-person interview study in which 18 participants embodied a robot and demonstrated how it should interrupt them in six scenarios that varied in task type (communicative vs. physical) and urgency (low, medium, high). Our results revealed 15 interruption behavior clusters and 4 themes describing participants' expectations. We found that participants' expectations differed regarding whether they wanted robots to interrupt them at all in low-urgency situations, often depending on whether the robot was perceived as a social companion or a task-oriented assistant. As urgency increased, participants generally preferred more immediate and pronounced interruptions that combined modalities such as visual cues, gestures, speech, and, occasionally, physical touch.**

I. Introduction

A robot that collaborates with a person will sometimes need to interrupt them. It may need to ask for information, provide a reminder, suggest a correction, or warn about a problem that requires action. Yet interruptions are not neutral information deliveries; they are socially and temporally delicate interaction acts. Their appropriateness depends not only on what the robot wants to communicate, but also on when and how. Humans navigate such situations using shared social conventions, including turn-taking, gaze, timing, and prosody [1], [2]. However, for robots, it remains unclear how to design interruption behaviors that are both effective and socially acceptable across different tasks and urgency levels.

Prior work has addressed robot interruptions primarily through two complementary approaches. First, Human-Computer Interaction (HCI) research on interruptibility prediction models *when* and *whether* humans can be interrupted across smartphones [3], [4], driving [5], [6], and immersive contexts [7], [8]. They typically treat the form of the interruption as fixed, with urgency remaining implicit. Second, human–robot Interaction (HRI) examines *how* robots should interrupt through predefined behaviors like minimal nonverbal cues [9], [10], multi-party strategies [11], and timing-sensitive turn-taking [12], [13], [14], pointing out that poorly-timed interruptions harm performance [15], [16] and frustrate users [17]. However, the designers' strategies leave critical gaps in understanding how users envision interruption behaviors across different tasks and urgencies.

To address this gap, we focus on the user perspective and investigate how people envision robot interruptions. We conducted an in-person interview study in which 18 participants embodied a robot and demonstrated how they would want the robot to interrupt them across six scenarios varying in task type (communicative vs. physical) and urgency (low, medium, high). Using open and axial coding we derived 15 interruption behavior clusters and found that participants implicitly structured interruptions into three stages: attention-grabbing, waiting, and reasoning. We further analyzed the interviews using reflexive thematic analysis [18] and identified 9 scenario considerations and 9 modality affordances that shape how interruption behaviors are envisioned.

Taken together, our work contributes (1) a set of user-elicited robot interruption behaviors, (2) a set of scenario considerations and modality affordances for designing interruptions, and (3) a view of interruptions as context-sensitive, multi-stage interaction processes. These findings provide a foundation for designing robot interruptions that are both effective and socially appropriate.

II. Related Work

Next, we discuss interruptions in HCI, HRI, and nonverbal robot communication through user-centered design.

A. Interruptions in Ubiquitous Computing

Work on interruptions in HCI and ubiquitous computing has shown that interruptions can impose cognitive, affective, and performance costs that vary across context [19], [20]. Users are often reachable by a device but not necessarily receptive, motivating models of activity, workload, and task structure [4], [19], [21], [22]. Prediction approaches span smartphones that model location and activity for interruption intensity [3], [4], cars that use sensor-based workload detection [5], [6], and immersive/HMD contexts that leverage motion data [7], [8]. While prior work focuses primarily on *when and whether* to interrupt with fixed modalities and implicit urgency, we investigate *how* robots should behave across systematically varied contexts and urgency levels.

B. Robot-Initiated Interruptions and Turn-Taking

By evaluating a small set of predefined behaviors, prior work has found that minimal nonverbal cues [9], [10] and multi-party strategies like hand-raising and waiting [11] are perceived as more polite than immediate speech. Timing proves critical across contexts: temporal coordination shapes

[1]LMU Munich, Munich, Germany {jan.leusmann, maximiliane.windl}@ifi.lmu.de
[2]Honda Research Institute EU, Offenbach am Main, Germany chao.wang@honda-ri.de
[3]TU Dortmund University, Dortmund, Germany
[4]Research Center Trustworthy Data Science and Security, Dortmund, Germany info@sven-mayer.com

perception [12], mistimed signals during physical manipulation increase human errors [13], and proactive smart speakers gain acceptance through context-awareness [23]. In contrast, poor interruptions harm performance [15], [16] and frustrate users [17], while conversational adaptation requires intent classification [24]. These designer-specified approaches leave a gap that our user-centered approach addresses.

C. User-Centered Design of Robot Behaviors

Nonverbal cues play a central role in managing interruptions and turn-taking in both human-human and human–robot interaction. To bring this seamless human-human interaction to HRI, prior work has proposed frameworks to translate human codes into robot signals for attention/engagement [25], with gaze sustaining attention [26] and turn-taking requiring multimodal coordination [14], [27]. Recent work has also begun exploring how people envision and co-design robot behaviors through user-centered methods. The expression elicitation approach focuses on creating understandable expression through embodied elicitation [28], as robot morphology shapes pre-interaction social expectations [29]. Users show mixed preferences between human-mimicry and transparent cues [30], highlighting the need to elicit mental models for delicate actions like interruptions. While prior work evaluates designer-specified strategies, we systematically elicit user expectations across task types and urgency through embodied demonstrations and qualitative analysis.

III. User Study

We investigated how people envision robot interruptions during two collaborative tasks and six urgency levels. We focus on understanding participants' expectations and mental models of robot interruptions. Thus, we conducted in-person, semi-structured interviews in which they embodied a robot.

A. Study Design

We positioned participants in front of a robot whose movements they controlled via immersive extended reality. We presented six collaborative human-robot scenarios in which the robot intended to interrupt the user. For each scenario, participants were asked to imagine themselves in the situation and demonstrate how they would want the robot to interrupt them. We used a within-subject design and counterbalanced the order of scenarios across participants using a balanced Latin square to mitigate ordering effects.

B. Scenarios

We designed collaborative HRI scenarios along two dimensions: Task Type and Urgency. All interruption reasons can be classified as competitive [1], as they change the topic of the current interaction. Task Type distinguishes the nature of the collaborative activity. The interaction can either be *physical collaboration* or *verbal communication* [31]. Urgency captures how time-critical the robot's interruption is. We defined urgency as the "time until action is required." *High-urgency* scenarios require immediate action to prevent negative consequences, *medium-urgency* scenarios require action in the near future, and *low-urgency* scenarios allow for flexible timing. Each scenario described a concrete, everyday situation and specified why the robot wanted to interrupt.

1) Online Pre-Study: We conducted an online survey to confirm the reliability of the urgency levels, where we presented participants with scenarios and asked them to rate their urgency on a 101-point Visual Analog Scale using the following statement: "It is very urgent that the robot interrupts me right now" (0 = Strongly disagree, 100 = Strongly agree). We used a within-subject design with randomized order. We recruited Prolific participants residing in European countries (to mitigate cultural differences). The survey took an average of 4 minutes, which we compensated with 9£ per hour. Each participant completed a consent and demographic form before rating the six scenarios and one attention check.

In total, we collected data from 219 participants in two batches ($N_1 = 115$, $N_2 = 104$). We removed entries from participants who failed the attention check ($N = 13$) and removed low-effort responses using an outlier filter ($M \mp 2 \times SD$), resulting in $N_1 = 81$ and $N_2 = 80$. Based on the first batch results, we found that both low- and high-urgency scenarios were stable, whereas the communicative medium scenario was perceived as too urgent ($M = 70.36$, $SD = 26.76$). Thus, we slightly reworded it to reduce urgency. In the second batch, the results were stable, indicating a good spread of perceived urgency levels for the three different scenarios for each Task Type and similar perceived urgency for each Urgency level. The resulting perceived urgency values for the communicate scenario are: low ($M = 8.79$, $SD = 11$), medium ($M = 55.72$, $SD = 25.86$), high ($M = 97.1$, $SD = 5.12$) and for the physical scenario: low ($M = 16.45$, $SD = 17.94$), medium ($M = 64.55$, $SD = 23.32$), high ($M = 96.78$, $SD = 5.67$). The final scenarios used in the main study are shown in Table I.

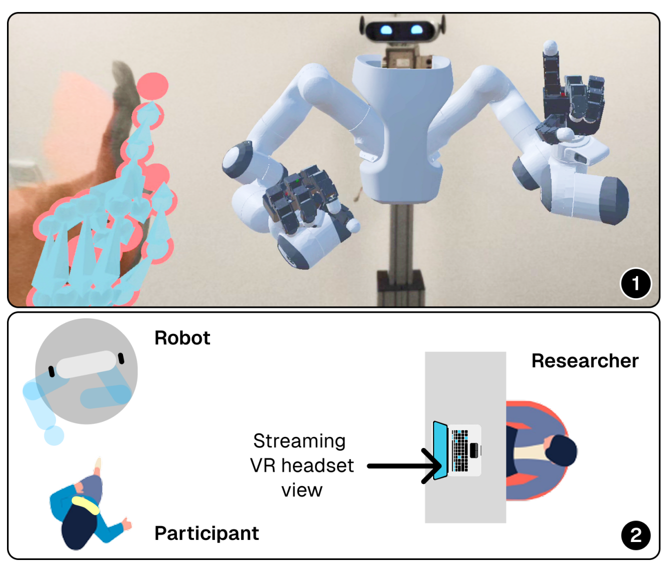

Fig. 1: 1. The hybrid robot setup, with a real head and augmented arms. 2. Study setup.

TABLE I: Scenarios used in the pre-study and interview study, varying task type and urgency.

	Preamble	**ID**	**Urgency**	**Description**
Communicative	You are at home and have just cooked dinner. You are now sitting at your table, eating. During that, you are having a casual conversation with your robot, explaining your cooking recipe as it listens. While you are talking, the robot ...	S1	Low	... remembers that it wanted to ask you whether you plan to celebrate your birthday next year.
		S2	Medium	... remembers that you mentioned earlier this week that you ordered a package that needs to be returned. The robot realizes that the return deadline is in the coming days.
		S3	High	... notices that the wrong stove is currently turned on and that a plastic spatula is resting on it. If nothing changes, the spatula will start to melt.
Physical	You are at home, assembling your newly bought wooden chair with your robot. Both you and the robot are currently attaching a leg to the chair without talking. While you are attaching the leg, the robot ...	S3	Low	... wonders whether the chair's wood is the same type as your table.
		S5	Medium	... notices there is a slightly easier order in which you can assemble the chair. If you continue assembling the chair in this order, the next steps will require more effort, but still lead to a correctly built chair.
		S6	High	... notices that you are about to attach it in the wrong position. If you screw it in at this point, it will leave a visible hole in the chair, and the chair will break as soon as someone sits on it.

C. *Apparatus*

To enable participants to envision the robot's interruption behavior, we built a hybrid physical-virtual robotic platform in which participants wore an Oculus Quest Pro and embodied a robot avatar consisting of a humanoid upper body with two arms, a head, and a face; see Figure 1. Participants' head, eye, and upper-body movements were mapped in real time onto the robot. The head and face screen are real, while the hands and arms are AR overlays of Franka Research 3 arms with Allegro hands V5. We used augmented arms for embodiment due to safety concerns about re-targeting arm movement, while still allowing participants to observe the robot's arm movement.

D. *Procedure*

After welcoming the participants, we explained the study goals and procedure. After that, we obtained written informed consent and collected demographics via a survey. We then explained the study procedure in detail, asked them to attach a microphone, and introduced them to the apparatus. When participants reported feeling comfortable in augmented reality, we asked them to make the robot perform a waving gesture, a thumbs-up gesture, and a nodding gesture to familiarize themselves with the robot's control. We then started video and audio recordings and began the main study.

For each of the six scenarios, we first read it aloud to participants and asked them to immerse themselves in the described situation. We then asked participants to demonstrate and explain how they would want the robot to express its intent to interrupt them. After this, we conducted a semi-structured interview with the topics: Appropriateness of the interruption, scenario factors informing the interruption behavior, reasons for the choice of modalities, unfitting modalities, additional interruption possibilities outside the robots' capabilities, comparison to how humans would interrupt here, and a reflection.

After all scenarios, we conducted a final semi-structured interview in which we asked about the general rules for designing interruptions across scenarios, whether participants think their designed behaviors can be translated to different robot morphologies, their reasons, and comparison to borrowing behaviors from human interruptions, and whether there are situations in which only humans or robots are allowed to interrupt. The study took on average 45 minutes.

E. *Participants*

We recruited 18 participants (4 female, 14 male) from 7 nationalities. We recruited 9 participants with substantial prior experience or academic backgrounds in HRI/robotics (E) and 9 novices (N) to ensure a wide spread of visions. Participants rated their own expertise in robotics ($M_E = 4.33 \pm 0.5$, $M_N = 2.33 \pm 0.87$), HRI ($M_E = 4.22 \pm 0.83$, $M_N = 2.0 \pm 1.0$), HCI ($M_E = 3.67 \pm 0.71$, $M_N = 2.33 \pm 1.0$), and UX-Design ($M_E = 2.67 \pm 0.1.23$, $M_N = 2.0 \pm 0.87$).

F. *Data Analysis*

We transcribed the interviews using atrain [32] and manually corrected them. We analyzed the data using thematic analysis [18] and atlas.ti. Two researchers independently coded 4 interviews and met to discuss and create a codebook. Afterward, each independently coded 50% of the remaining interviews. They met multiple times to merge all codes and identify themes, resulting in 175 codes and 4 core themes.

Further, we extracted all envisioned interruption behaviors and used open and axial coding to cluster them [28]. The 18 participants suggested 126 interruption behaviors. Two authors independently conducted open coding on 20% of the behaviors and then met to discuss them and establish a codebook. One author then coded the remaining behaviors. We then used axial coding to cluster semantically similar behaviors, yielding 15 unique behaviors. See OSF for further information.

IV. Results

For each scenario, we asked participants if an interruption would be appropriate. We found that 31% found it appropriate to interrupt in S1, S2 (47%), S3 (100%), S4 (50%), S5 (100%), and S6 (100%). In the following, we present the results from our interview analysis and behavior coding.

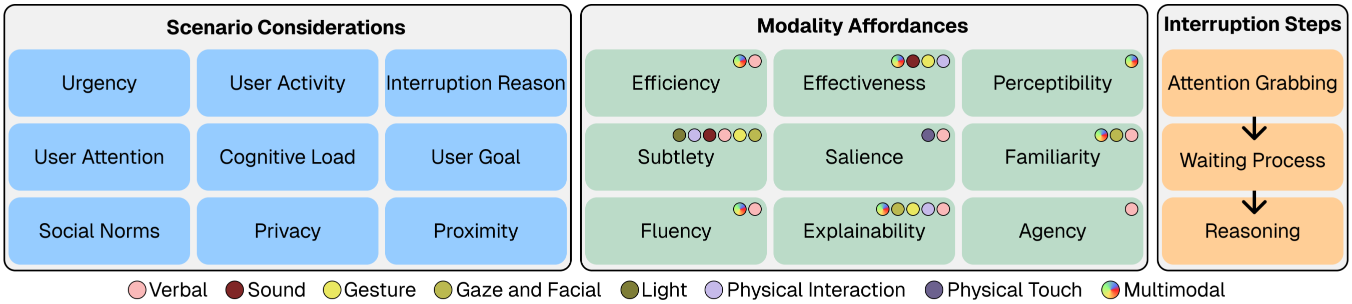

Fig. 2: Framework for designing robot interruptions, scenario considerations, modality affordances, and interaction steps. The colored modalities are not exclusive but based on probes from our results and should be just seen as potential suggestions.

A. *Scenario Factors & Considerations*

We identified nine scenario considerations that inform interruption behavior. Moreover, participants suggested and discussed eight different modalities: verbal, sound, gestures, gaze, facial expressions, light and visual notifications, physical interaction, physical touch, and multimodal combinations.

1) Urgency: The urgency of a situation is an important factor to consider. Participants agreed that urgency is highest in physically dangerous situations, in situations that could result in physical damage, and in time-critical situations. Participants agreed that behaviors should always respect the urgency level (E3, E5, E6, N1, N6, N8, N9): interruptions in urgent situations should happen immediately (E2, E5, N7, N9), be highly intrusive (E7, E9, N1, N3, N6, N8), capture the users attention quickly (E1, E3, E7), may override ongoing interactions (E1, N4), should be clear and umambiguous (E1), and should be more robot-like (E2). In contrast, interruptions for non-urgent situations should be subtle (E2, E3, E6, N9), and wait for appropriate timing (E1, E2, E3, E5, N2, N3).

2) Interruption Reason: The interruption reason, namely its importance and content, is another important factor. Participants stated that the interruption reason must be clearly communicated (E3), that non-task related questions are less important and should have low-intensity (E1), that optimization-based interruptions should only occur when there is a large gain (E3, N3), that non-important questions should be subtle or avoided (N2, N4), and that robots should avoid too many non-immediately beneficial questions (N5), as those could be annoying (E3, E8).

3) User Attention: Interruptions should first capture the user's attention before conveying information (N1). Thus, designers should select modalities based on their effectiveness in capturing attention within the given context (E6), particularly using auditory modalities when mutual attention has not yet been established (E1, E3) or when other channels are occupied (E1, E6, N7, N9). When the user is not attending to the robot, designers can employ auditory signals such as verbal cues or sounds (E1, E3, N9). In situations where attention must be captured rapidly, designers may instead use stronger modalities, such as physical touch (E3). The robot should use non-verbal cues to show deviating attention (E3), can use gestures to visually grab (N3, E7) and steer attention (E2, E7, N9), and should use non-occupied modalities in physical tasks (e.g., verbal) to get attention (E1, N3, N8, N9).

4) Cognitive Load: Interruptions should match the user's cognitive load (E3, E8, N7, N9): If the user's current load is high, interruptions should be subtle (N7); if the cognitive load is low, the robot can immediately verbally interrupt (N9).

5) User Activity: Participants identified the user's current activity as an important factor for designing interruptions. They distinguished whether the user is currently speaking, emphasizing that ongoing speech should generally not be interrupted and that users prefer to finish their utterances before the robot intervenes (E7, N3, N6). Activities can further be differentiated into conversational, physical, and idle contexts, each requiring different interruption strategies (E1, N7, N9). Participants also highlighted that interruptions should respect the current conversation context, as unrelated interruptions were perceived as disruptive and unnatural (E2, N2, N4, N5, N8). Moreover, interruption design depends on whether the activity is carried out collaboratively or in parallel, with more direct interruptions being acceptable in shared tasks (E2, E9). Finally, participants emphasized that interruptions are more acceptable when activities can be easily paused or are not time-critical (N6, E6).

6) User Goal: In general, interruptions should respect the user's primary goal and avoid negative impact on task performance (E2, N4, N7). Participants stated that task-oriented interruptions are more acceptable than non-task-oriented ones (E6), and that they are even more acceptable when the robot is necessary for task completion (E3). When the robot is also working on the task, it should use modalities that are not occupied (E1, N3, N8, N9). Non-task-related interruptions should be low intensity (E1).

7) Social Norms: Participants stated that robots, just like humans, should follow social norms. Politeness was named as one of the important factors robots should always follow (E2, E3, E8, N1, N2, N3). In urgent situations, the robot should begin verbal interruptions with an apology (N6). Participants highlighted that robots also require consent before physical contact (E8, N5, N1, E9) and should not invade personal space without prior signaling (N5). Lastly, participants noted that the robot's behavior also strongly depends on the relationship between the user and the robot (N3, N5, E5, E6).

8) Privacy: Participants also highlighted privacy sensitivity as an important factor influencing the design of interruptions. Storing information beyond the current context might lead to discomfort and reduce trust, suggesting that robots should store and access only data relevant to the current context (N5).

9) Proximity: Interruption design should also respect the physical proximity between the robot and user (E2, E6). This affects whether the participant can see or hear the robot; thus, the interruption modalities need to be selected accordingly. The robot should also respect users' personal space (N5).

B. Modality Affordances

We found 9 modality affordances, which reflect their properties and should be used to tackle scenario considerations.

1) Efficiency: To convey information as quickly and precisely as possible, participants preferred the verbal modality (N9, E2, N4, E7), as it allows them to receive answers to questions immediately (E3, N1, N4, N9). Participants also emphasized multimodal interruptions for efficiency, as they amplify attention capture and the interruption is perceived more quickly (E2–E6, E9, N1, N8).

2) Effectiveness: Participants emphasized multiple modalities for effective interruption. Sound is highly effective in capturing attention (E3, E4, N1, N7), gestures are effective to get visual attention (N3, E7) and signal intent (E3), and physical interactions are effective as they are both visible and make sounds through the moving actuators (E3, N2). Again, participants suggested multimodal interruptions to increase effectiveness across channels (E6, N2).

3) Perceptibility: Participants suggested multimodal behaviors (combining the auditory and visual channels) to provide redundancy (E6, E7, N3) and accessibility (N2) in case a user can only use some of their sensory channels.

4) Subtlety: Certain modalities allow for subtle and minimally intrusive interruptions. Gestures were frequently mentioned as enabling less intrusive attention signaling (E1, N6) and as a way to soften the harshness of verbal interruptions (E1). Similarly, gaze was described as minimally distracting (E7), and sound can be designed to remain subtle while still noticeable (N7). Verbal interaction was also perceived as low in physical intrusiveness, although its disruptiveness depends on the delivery (E1–E3, E6, N2, N5–N7, N9). Additionally, physical interaction with objects can be minimally intrusive (E2), indirect interruption strategies, such as inviting the user to initiate interaction, can reduce disruption (E1), and combining modalities such as light with verbal communication can soften the perceived harshness of interruptions (N7).

5) Salience: Participants stated that verbal interruptions are very noticeable, but highlighted physical touch as highly interruptive, requiring immediate reactions (E1, E7, N4, N7).

6) Familiarity: Participants highlighted that certain modalities are perceived as familiar and therefore easier to understand. Verbal interaction was most frequently described as a familiar form of communication (E1–E3, E5, E8, E9, N2, N3, N7, N9), while gaze and facial expressions were also seen as naturally embedded in social interaction (E3, N3). In urgent situations, loud verbal communication was perceived as familiar and appropriate (N6). Additionally, participants noted that combined modalities can support more natural and familiar interactions (E2, E4, E8, N2, N3, N6).

7) Fluency: Participants emphasized that certain modalities facilitate smooth, continuous interaction. Verbal communication enables natural conversational integration and aligns with turn-taking behavior (E1–E3, E5, E6, E8, E9, N2, N3, N7, N9). In addition, combining modalities like head movement and gazing to indicate active listening can help maintain conversational continuity and prevent disruptions (E9).

8) Explainability: Certain modalities facilitate the understanding of interruptions. Verbal inherently combines attention-grabbing with reasoning and information transfer (E3, N1, N4, N9). Gestures can convey contextual meaning through deictic cues (E2, E7, N9) and direct user attention through pointing (E2, E7, N9). Gaze can further guide attention and signal upcoming interaction (E2, E3). Physical demonstration alone can also be sufficient to convey explanations (N4). Additionally, combining verbal and non-verbal modalities enables high explainability by allowing simultaneous demonstration and explanation (E6, N6).

9) Agency: Participants emphasized that interruption modalities can differ in the level of autonomy they grant the user. Verbal interaction allows users to decide independently how and whether to respond to an interruption (N9). In contrast, direct physical interaction was often perceived as limiting user control and overly intrusive (N1).

C. Human vs. Robot

Participants drew heavily on human-to-human interactions when designing their interruption behaviors, as they indicated that 78% of all behaviors borrowed strongly from humans. We report these findings along the three code groups.

1) Acceptance: Participants' acceptance of interruptions differed between humans and robots. In general, humans were granted greater social permission to interrupt, especially in non-critical situations, due to existing social norms and relationships (E1, E3, E5, E7, N2, N5). For example, participants noted that they would be more accepting of humans asking spontaneous or unrelated questions (N4) and of being physically touched by humans (N1). In contrast, participants required stronger justification from robots for interrupting, particularly in non-urgent scenarios. However, in urgent or high-risk situations, participants agreed that both humans and robots are equally permitted to interrupt (E3, N3, N7). Some participants further emphasized that interruptability depends more on the relationship than on whether the interrupter is a human or a robot (E6, N3, N5), whereas others considered interruptability similar, with differences arising mainly from their capabilities (N9).

2) Capability Differences: While some participants suggested that robot behavior should be designed based on how humans interrupt (N9), they also emphasized key differences. Because humans are prone to forgetting, whereas robots typically are not, immediate interruptions are more socially acceptable in human–human interaction (E2, E3, E5–E8, N4, N6). At the same time, robots were often expected to outperform humans in certain aspects. Participants noted that robots could interrupt more intelligently (E2), remain calmer in urgent situations (N3), and detect dangers faster than humans (E3). Robots were also seen as having additional expressive capabilities, including visual signals (e.g., color changes) and artificial sounds (E3, E4). However, humans were perceived as superior at understanding when to interrupt (N7, N9), and participants emphasized that granting autonomy to robots depends on trust in their capabilities (E6). In some cases, participants also trusted robot-generated optimization suggestions more (E2, N2).

3) Behavioral Differences: Participants described differences in the behavioral style of interruptions. Human interruptions were perceived as more variable and dependent on individual traits such as politeness (N6), being more direct or impulsive (N6), or unnecessary (N2, N3). Humans were also described as more likely to use simple or abrupt interruptions, such as directly saying “stop” (N2), or using physical touch (E4, N2, N6). Compared to humans, robots were perceived as more likely to interrupt rather than resolve issues independently (E1, E2, E5, N1, N6–N8). Participants highlighted different uses of modalities between humans and robots. Robots were described as more likely to use additional or non-human-like signals, such as explicit hand gestures to capture attention (N3) or artificial sounds (N7), whereas humans tend to rely on more implicit or socially learned behaviors.

D. Envisioned Robot Interruptions

We found that each interruption consists of three steps: (1) attention-grabbing signal, (2) waiting process, and (3) interruption-reasoning process. We coded all behaviors into these categories, yielding 9 distinct attention-grabbing signals, 4 waiting processes, and 4 interruption-reasoning behaviors, for a total of 34 unique behavior combinations. We then applied axial coding to merge semantically similar combinations, resulting in 15 final interruption behavior clusters, see Table II.

For attention-grabbing, the majority suggested no attention-grabbing behaviors (N=39), followed by auditory (N=25), multimodal (visual + auditory) (N=19), facial expressions and gaze (N=14), gestures (N=12), multiple visual channels (N=8), physical interaction (N=4), and physical touch (N=4).

Participants suggested four different waiting behaviors. The majority of participants suggested not waiting for a user response but directly continuing with the reasoning (N=70). Others suggested waiting until the user finishes their turn or topic (N=26), until they show attention to the robot (N=10), or give the user explicit permission to interrupt (N=10).

Most participants realized the reasoning phase through verbal utterances (N=105). Some suggested using physical demonstrations or interaction (N=3), or combining the utterance with deictic gestures or physical demonstration (N=3).

There are immediate interruptions (I_{01}–I_{09}), where the robot proceeds without waiting for a user response, and waiting-based interruptions (I_{10}–I_{15}), where the robot first signals intent and defers to the user before explaining. In 7 cases, participants suggested no interruptions.

I_{01}–I_{03} use no or auditory attention signals followed by an immediate verbal explanation; I_{01} uses no signal at all, I_{02} precedes speech with a brief sound (e.g., a chime), and I_{03} additionally uses physical demonstration in the reasoning phase. I_{04}–I_{06} use a single non-verbal cue to direct attention before explaining verbally: a gesture such as a raised arm or deictic point (I_{04}, I_{05}), or a gaze shift toward the user or a relevant object (I_{06}). I_{05} additionally combines speech and physical demonstration in the reasoning phase. I_{07} was the most common behavior, appearing across all scenarios. The robot combines facial expressions, color or light changes, and auditory signals to capture attention across multiple channels. Participants highlighted the calibration potential of this behavior: intensity and valence can be adjusted to urgency. I_{08}–I_{09} use a physical signal: I_{08} uses users’ physical touch to stop ongoing motion, while I_{09} acts on the shared task environment rather than the user directly, for instance, by obstructing the assembly. For both, the physical event itself conveys the problem before any verbal explanation.

For waiting-based behaviors, I_{10}–I_{11} emit an auditory signal to indicate pending communication and wait for acknowledgment before proceeding with verbal reasoning or physical demonstration. I_{12}–I_{14} use visual or gestural signals to indicate pending intent, ranging from a combined facial expression and gesture (I_{12}) to a raised hand to signal intent to speak (I_{13}) to a subtle eye color change that relies on the user noticing and inquiring without any active bid for attention (I_{14}). I_{15} produces no attention signal and waits for a natural break before intervening verbally. This turn-taking behavior was among the most commonly envisioned.

TABLE II: The 15 resulting elicited interruption behaviors ordered from most to least intrusive. N is the amount of times one interruption behavior was suggested.

Waiting	Attention	Reasoning	Scenarios	ID	N
None	None	Verbal	S1 - S5	I_{01}	9
	Auditory	Verbal	S1 - S6	I_{02}	18
		Multimodal	S5	I_{03}	2
	Gesture	Verbal	S1, S3, S6	I_{04}	8
		Multimodal	S5	I_{05}	1
	Facial Expr.	Verbal	S2, S4, S6	I_{06}	5
	Multimodal	Verbal	S1 - S6	I_{07}	23
	Touch	Verbal	S5, S6	I_{08}	4
	Interaction	Verbal	S5, S6	I_{09}	4
Wait	Auditory	Verbal	S4, S5	I_{10}	3
		Physical	S5	I_{11}	2
	Multimodal	Verbal	S3, S4	I_{12}	4
	Gesture	Verbal	S1, S2	I_{13}	3
	Facial Expr.	Verbal	S1 - S4	I_{14}	9
	None	Verbal	S1, S2, S4	I_{15}	24

V. Discussion

Next, we discuss the implications of our findings.

A. Context-Sensitive Interruption Design

Our results show that robot interruptions cannot be reduced to a single "one-size-fits-all" behavior, but must be designed relative to the interaction context. Across the 15 elicited behaviors, the same modality (e.g., verbal or auditory signals) was used differently depending on factors such as urgency, user activity, and attention, indicating that interruption design is not about selecting a modality, but aligning behaviors with situational constraints. Figure 2 captures this relationship by showing scenario considerations and modality affordances. Rather than prescribing fixed strategies, our findings suggest reasoning in terms of trade-offs: urgent situations require salience and efficiency, while low-urgency situations favor subtlety and turn-taking. Taken together, this frames interruption design as an orchestration problem, where designers compose context-sensitive strategies by configuring modalities based on their affordances, rather than selecting isolated behaviors.

B. Human-Inspired vs. Robot-Appropriate Interruptions

Participants frequently drew on human-human interruption behavior when envisioning robot interruptions, suggesting that human interaction serves as a familiar reference point, likely shaped by both the humanoid embodiment used in our study and pre-existing social expectations [29]. However, participants did not expect robots to simply copy humans. Compared to humans, robots were expected to interrupt more politely, less intrusively, and with stronger justification, especially in non-urgent situations, aligning with prior work on politeness in interruptions [33]. In particular, many participants preferred that robots wait for an opportune moment rather than interrupt immediately. While humans rely on subtle social cues to identify such moments, this remains technically challenging for robots, which may instead need to signal their intent to interrupt through subtle indications that invite user attention or inquiry. More generally, behaviors acceptable in human-human interaction, such as interrupting mid-sentence, were often rejected for robots, which participants expected to act more deliberately and choose better timing, reflecting preferences for negotiated over immediate interruptions except in urgent cases [34]. At the same time, robots were attributed capabilities beyond humans, such as additional signaling channels and more controlled behavior. Together, this suggests that robot interruptions should be human-inspired but not purely human-like, requiring designs that balance familiarity with robot-specific capabilities and expectations [30].

C. Interruptions as Multi-Stage Behaviors

Building on our findings, we propose to treat robot interruptions as a structured, multi-stage design problem. Rather than defining interruption behaviors as single actions, designers can construct them by configuring three stages with the scenario considerations in mind: attention-grabbing, waiting, and reasoning (Figure 2). This provides a practical design workflow. First, designers select how to establish attention, choosing modalities that match the required level of salience or subtlety. Second, they determine whether to interrupt immediately or defer the interaction, thereby controlling intrusiveness and preserving user autonomy. Third, they define how the interruption reason is conveyed, balancing efficiency, clarity, and explainability. Importantly, each stage can be designed independently but must be coordinated through modality affordances. For example, highly salient attention signals can be paired with brief verbal reasoning in urgent situations, while subtle visual cues can be combined with delayed explanations in low-urgency contexts. The 15 elicited interruption behaviors illustrate how these stages can be combined into coherent interaction strategies; rather than fixed solutions, they serve as design exemplars that capture different trade-offs. Together with the identified modality affordances as a vocabulary for selecting and combining modalities, this provides designers with concrete building blocks that can be adapted and recombined to fit the requirements of a given situation.

D. Limitations & Future Work

While we elicited 15 interruption behavior clusters, we did not evaluate whether these are now perceived as good and understandable interruptions. Future work should validate the understandability. However, we now provide designers with a set of considerations and affordances with which they can create informed interruption behaviors to test. Secondly, we recruited 18 participants for our study. While this might seem like a low number, prior studies had a similar participant amounts [28], [35], and we recruited both experts and novices for generalizability. Lastly, we used a humanoid robot for our study. This for sure had a legacy and familiarity bias on what kind of interruptions participants suggested. However, in our analysis, we abstracted away from merely humanoid interruptions, so future designers can also use our design space as a basis for non-humanoid robots.

VI. Conclusion

We investigated how people envision robot interruption behaviors across six scenarios varying in task type and urgency. We identified 15 interruption behavior clusters, 9 scenario considerations, and 9 modality affordances that shape how users expect robots to interrupt them. Our findings show that interruption design is highly context-sensitive: appropriate behaviors depend on factors such as urgency, user attention, current activity, social norms, and user goals. Beyond this, we found that participants often drew on human-human interruption behavior when envisioning robots, while simultaneously expecting robots to behave differently. Based on these insights, we argue that robot interruptions should be designed not as fixed behaviors, but as context-sensitive, multi-stage interaction processes that combine attention-grabbing, waiting, and reasoning strategies. We suggest that our findings provide a foundation for designing more appropriate, effective, and user-centered robot interruptions.

References

[1] J. A. Goldberg, "Interrupting the discourse on interruptions," *Journal of Pragmatics*," 1990. https://doi.org/10.1016/0378-2166(90)90045-F

[2] D. C. McFarlane and K. A. Latorella, "The scope and importance of human interruption in human-computer interaction design," *Human–Computer Interaction*," 2002. https://doi.org/10.1207/S15327051HCI1701_1

[3] F. Yuan, X. Gao, and J. Lindqvist, "How Busy Are You?: Predicting the Interruptibility Intensity of Mobile Users," in *Proc. of the HI Conference on Human Factors in Computing Systems*. ACM, 2017. https://doi.org/10.1145/3025453.3025946

[4] M. Pielot, B. Cardoso, K. Katevas, J. Serrà, A. Matic, and N. Oliver, "Beyond interruptibility: Predicting opportune moments to engage mobile phone users," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*," 2017. https://doi.org/10.1145/3130956

[5] S. Kim, J. Chun, and A. K. Dey, "Sensors Know When to Interrupt You in the Car: Detecting Driver Interruptibility Through Monitoring of Peripheral Interactions," in *Proc. of the ACM Conference on Human Factors in Computing Systems*. ACM, 2015. https://doi.org/10.1145/2702123.2702409

[6] A. Kim, W. Choi, J. Park, K. Kim, and U. Lee, "Interrupting drivers for interactions: Predicting opportune moments for in-vehicle proactive auditory-verbal tasks," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*," 2018. https://doi.org/10.1145/3287053

[7] K.-W. Chen, Y.-J. Chang, and L. Chan, "Predicting opportune moments to deliver notifications in virtual reality," in *Proc. of the CHI Conference on Human Factors in Computing Systems*. ACM, 2022. https://doi.org/10.1145/3491102.3517529

[8] C. George, P. Janssen, D. Heuss, and F. Alt, "Should I Interrupt or Not?: Understanding Interruptions in Head-Mounted Display Settings," in *Proc. of the 2019 on Designing Interactive Systems Conference*. ACM, 2019. https://doi.org/10.1145/3322276.3322363

[9] S. G. Paul Saulnier, Ehud Sharlin, "Exploring interruption in hri using wizard of oz," in *International Conference on Human-Robot Interaction*. IEEE, 2010. https://doi.org/10.1109/HRI.2010.5453232

[10] P. Saulnier, E. Sharlin, and S. Greenberg, "Exploring minimal nonverbal interruption in hri," in *2011 RO-MAN*, 2011. https://doi.org/10.1109/ROMAN.2011.6005257

[11] O. Palinko, K. Ogawa, Y. Yoshikawa, and H. Ishiguro, "How should a robot interrupt a conversation between multiple humans," in *Social Robotics*. Springer International Publishing, 2018. https://doi.org/10.1007/978-3-030-05204-1_15

[12] G. Hoffman, M. Cakmak, and C. Chao, "Timing in human-robot interaction," in *Proc. of the International Conference on Human-robot Interaction*. ACM, 2014. https://doi.org/10.1145/2559636.2560029

[13] F. Cini, T. Banfi, G. Ciuti, L. Craighero, and M. Controzzi, "The relevance of signal timing in human-robot collaborative manipulation," *Science Robotics*," 2021. https://doi.org/10.1126/scirobotics.abg1308

[14] G. Skantze and B. Irfan. Applying General Turn-taking Models to Conversational Human-Robot Interaction. https://arxiv.org/abs/2501.08946

[15] E. Ganal, S. C. Steinhaeusser, F. Niebling, and B. Lugrin, "Message for You: Observing the Effects of a Social Robot's Interruptions During an Office Task," in *Social Robotics + AI*. Springer Nature Singapore, 2026. https://doi.org/10.1007/978-981-95-2379-5_12

[16] S. Mannem, W. Macke, P. Stone, and R. Mirsky, "Exploring the Cost of Interruptions in Human-Robot Teaming," in *International Conference on Humanoid Robots (Humanoids)*. IEEE, 2023. https://doi.org/10.1109/Humanoids57100.2023.10375236

[17] B. Irfan, S. Kuoppamäki, A. Hosseini, and G. Skantze, "Between reality and delusion: Challenges of applying large language models to companion robots for open-domain dialogues with older adults," *Autonomous Robots*," 2025. https://doi.org/10.1007/s10514-025-10190-y

[18] J. Lazar, J. H. Feng, and H. Hochheiser, "Chapter 11 - analyzing qualitative data," in *Research Methods in Human Computer Interaction (Second Edition)*, second edition ed. Morgan Kaufmann, 2017. https://doi.org/10.1016/B978-0-12-805390-4.00011-X

[19] L. D. Turner, S. M. Allen, and R. M. Whitaker, "Interruptibility prediction for ubiquitous systems: Conventions and new directions from a growing field," in *Proc. of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 2015. https://doi.org/10.1145/2750858.2807514

[20] S. Ohly and L. Bastin, "Effects of task interruptions caused by notifications from communication applications on strain and performance," *Journal of Occupational Health*," 2023. https://doi.org/10.1002/1348-9585.12408

[21] L. D. Turner, S. M. Allen, and R. M. Whitaker, "Reachable but not receptive: Enhancing smartphone interruptibility prediction by modelling the extent of user engagement with notifications," *Pervasive and Mobile Computing*," 2017. https://doi.org/10.1016/j.pmcj.2017.01.011

[22] J. E. Fischer, C. Greenhalgh, and S. Benford, "Investigating episodes of mobile phone activity as indicators of opportune moments to deliver notifications," in *Proc. of the International Conference on Human Computer Interaction with Mobile Devices and Services*. ACM, 2011. https://doi.org/10.1145/2037373.2037402

[23] N. Cha, A. Kim, C. Y. Park, S. Kang, M. Park, J.-G. Lee, S. Lee, and U. Lee, "Hello There! Is Now a Good Time to Talk?: Opportune Moments for Proactive Interactions with Smart Speakers," *Proc. of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*," 2020. https://doi.org/10.1145/3411810

[24] S. Cao, J. Moon, A. Mahmood, V. N. Antony, Z. Xiao, A. Liu, and C.-M. Huang. Interruption Handling for Conversational Robots. https://arxiv.org/abs/2501.01568

[25] J. Urakami and K. Seaborn, "Nonverbal Cues in Human–Robot Interaction: A Communication Studies Perspective," *ACM Transactions on Human-Robot Interaction*," 2023. https://doi.org/10.1145/3570169

[26] M. Li, F. Guo, X. Wang, J. Chen, and J. Ham, "Effects of robot gaze and voice human-likeness on users' subjective perception, visual attention, and cerebral activity in voice conversations," *Computers in Human Behavior*," 2023. https://doi.org/10.1016/j.chb.2022.107645

[27] G. Skantze, "Turn-taking in Conversational Systems and Human-Robot Interaction: A Review," *Computer Speech & Language*," 2021. https://doi.org/10.1016/j.csl.2020.101178

[28] J. Leusmann, S. Villa, T. Liang, C. Wang, A. Schmidt, and S. Mayer, "An approach to elicit human-understandable robot expressions to support human-robot interaction," in *Proc. of the CHI Conference on Human Factors in Computing Systems*. ACM, 2025. https://doi.org/10.1145/3706598.3713085

[29] N. Dennler, C. Ruan, J. Hadiwijoyo, B. Chen, S. Nikolaidis, and M. Matarić, "Design Metaphors for Understanding User Expectations of Socially Interactive Robot Embodiments," *ACM Transactions on Human-Robot Interaction*," 2023. https://doi.org/10.1145/3550489

[30] L. Spillner, J. Rockstroh, P. R. Peña, N. Wenig, N. Zargham, and B. R. Cowan, "Beyond the Illusion: LLMs and the Case for Pragmatic Cues in Conversation," in *Proc. of the ACM Conference on Conversational User Interfaces*. ACM, 2025. https://doi.org/10.1145/3719160.3737614

[31] D. Lee, C. Ott, and Y. Nakamura, "Mimetic communication model with compliant physical contact in human—humanoid interaction," *The International Journal of Robotics Research*," 2010. https://doi.org/10.1177/0278364910364164

[32] A. Haberl, J. Fleiß, D. Kowald, and S. Thalmann, "Take the atrain. introducing an interface for the accessible transcription of interviews," *Journal of Behavioral and Experimental Finance*," 2024. https://doi.org/10.1016/j.jbef.2024.100891

[33] V. Srinivasan and L. Takayama, "Help Me Please: Robot Politeness Strategies for Soliciting Help From Humans," in *Proc. of the CHI Conference on Human Factors in Computing Systems*. ACM, 2016. https://doi.org/10.1145/2858036.2858217

[34] D. C. McFarlane, "Comparison of four primary methods for coordinating the interruption of people in human-computer interaction," *Human–Computer Interaction*," 2002. https://doi.org/10.1207/S15327051HCI1701_2

[35] A. B. V. D. Nguyen, J. Leusmann, S. Mayer, and A. V. Moere, "Eliciting understandable architectonic gestures for robotic furniture through co-design improvisation," in *International Conference on Human-Robot Interaction*. IEEE, 2025. https://doi.org/10.1109/HRI61500.2025.10973858