Perception of Service Robots with Different Genders, Designs, and Interaction Modes in a Simulated Hospital Environment

Sıla Keskin, Nurşah Satılmış, Arda Deniz Ayyıldız

1. Introduction

This study investigates how people perceive hospital service robots that differ in gender presentation, appearance, and interaction style. As service robots become increasingly integrated into healthcare environments, understanding how these features influence user trust, empathy, and acceptance is essential. The research was conducted in a simulated hospital setting and aimed to explore user preferences toward three distinct robot designs through video-based observation and a follow-up survey.

Hospital environments are uniquely sensitive and emotionally demanding. In such settings, service robots must do more than complete functional tasks—they must also be perceived as supportive, trustworthy, and respectful. Existing literature in human-robot interaction suggests that people form strong impressions based on a robot's appearance, gendered voice, and method of communication. These impressions, in turn, affect how users emotionally engage with robots and whether they feel comfortable interacting with them. Therefore, designing robots with social and psychological factors in mind is critical for successful deployment in healthcare.

Three robots were introduced in this study. The first, Hospy, is a non-humanoid robot that relies on speech recognition to understand patient pain levels. The second, ClaraDose, is a feminine humanoid robot that uses a touchscreen interface, allowing patients to rate their pain manually. The third, CareX Screen, is a masculine humanoid robot with similar functionality to ClaraDose, but with a different visual and vocal presentation. Each robot was shown in a short video performing its tasks, and participants were asked to evaluate their preferences and perceptions based on what they observed.

The study is guided by several key research questions: Does the robot's gendered voice influence perceived trust or empathy? Does a humanoid appearance lead to higher acceptance and emotional comfort compared to a non-humanoid design? Do users prefer voice-based interaction or touchscreen input in a hospital context? Are feminine robots seen as more empathetic, while masculine ones are viewed as more competent or professional? Based on previous research and class discussions, we hypothesize that gender cues and interaction modalities significantly influence users' emotional responses, and that humanoid robots may be more favorably received in healthcare settings than non-humanoid alternatives.

2. Background & Related Work

As service robots increasingly become part of healthcare environments, researchers have focused on how social factors such as appearance, gender, and communication style affect user acceptance. Human-Robot Interaction (HRI) studies have shown that emotional cues—both visual and auditory—play a critical role in shaping user perceptions and trust in robots.

One foundational concept in HRI is social presence theory, which suggests that users tend to attribute social characteristics to robots, especially when the robot exhibits humanlike features such as facial expressions, gestures,

or a human voice. This effect, sometimes referred to as anthropomorphism, can lead users to treat robots as social agents rather than mechanical tools [1]. As such, robots designed for hospital environments may benefit from carefully chosen visual traits and communication strategies.

A robot's perceived gender also impacts user responses. Studies have found that feminine-appearing or soft-voiced robots are often viewed as more empathetic and nurturing, while masculine-appearing robots are associated with competence and authority [2]. However, these associations can vary based on cultural expectations and the context of interaction. In healthcare, where empathy is valued, feminine features may improve user satisfaction—yet professional authority may also be attributed to masculine designs.

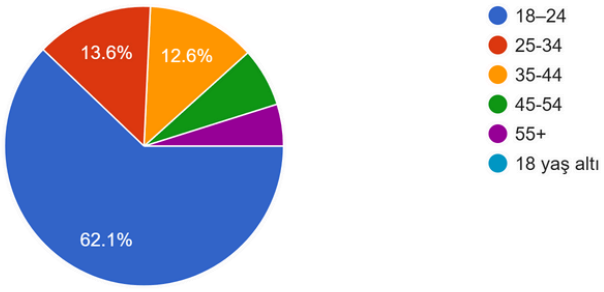
Voice-based interaction versus touchscreen interaction presents another major design choice. While speaking to a robot may feel more natural to some users, it can also create discomfort if the voice or timing feels unnatural [3]. In contrast, touchscreens offer privacy and clarity but may reduce the sense of personal connection. Research suggests that the ideal modality depends on user expectations and the task being performed.

Finally, trust and empathy have emerged as key metrics in HRI evaluation. Trust influences whether users follow a robot's instructions, while empathy affects whether they feel emotionally supported [4]. In healthcare settings, where patients may be vulnerable or in pain, these dimensions are especially relevant.

This study builds on such research by examining how three differently designed hospital robots are perceived in terms of trust, empathy, and interaction style. It also considers how the robot's gendered voice and physical form influence these perceptions, extending prior work on gendered design and social cognition in HRI.

3. Methodology

3.1 Participants



A total of 105 participants took part in the study. The majority of respondents (61.9%) were between the ages of 18–24, followed by 13.3% aged 25–34, 12.4% aged 35–44, and smaller portions in the 45–54 and 55+ age brackets.

In terms of gender identity, 52.4% identified as women, 43.8% as men, while a small number identified as non-binary or third gender or preferred not to disclose their gender.

Participants came from a variety of academic backgrounds. The largest groups were from medicine (35.2%) and law (37.1%), followed by smaller representations from engineering, education, psychology, computer engineering, logistics, and other fields. This diversity provided a broad perspective for evaluating human-robot interaction across disciplines.

To investigate how hospital service robots are perceived based on their appearance, voice gender, and interaction modality, this study adopted a dual-method approach: the development of a simulation-based interaction environment, and the design of a structured user perception survey. This hybrid methodology enabled both system-level prototyping and human-centered evaluation.

3.2 Simulation and System Implementation

The simulation was developed using ROS1 Noetic and Gazebo 11, and was run on Ubuntu 20.04 via VirtualBox on a Windows host. The limited system resources (4GB RAM) significantly influenced the design process, preventing the use of prebuilt large simulation environments such as the AWS Robomaker Hospital World. Consequently, a custom hospital room was manually designed using SDF files, with elements like patient beds, walls, and navigation spaces.

Two distinct service robots were created:

Early development used a scaled-up TurtleBot3 model to increase visual presence. However, due to limitations in realism and humanoid expressiveness, the simulation later transitioned to more advanced humanoid models such as TIAGo, which allowed for better representation of body posture and screen-based interaction.

Hospy, one of the three robots showcased, is a compact, non-humanoid unit designed to interact with patients through speech recognition. It uses voice input to detect pain levels and appears in different versions with both male and female voices to explore how vocal gender may influence user perception.

To simulate realistic interaction, three scenes were recorded:

1. The robot approaching a patient bed.
2. Delivering medication while speaking.
3. Prompting the patient to rate their pain from 1 to 10.

Image I. The video scenes were generated using the simulated hospital environment developed in ROS and Gazebo.

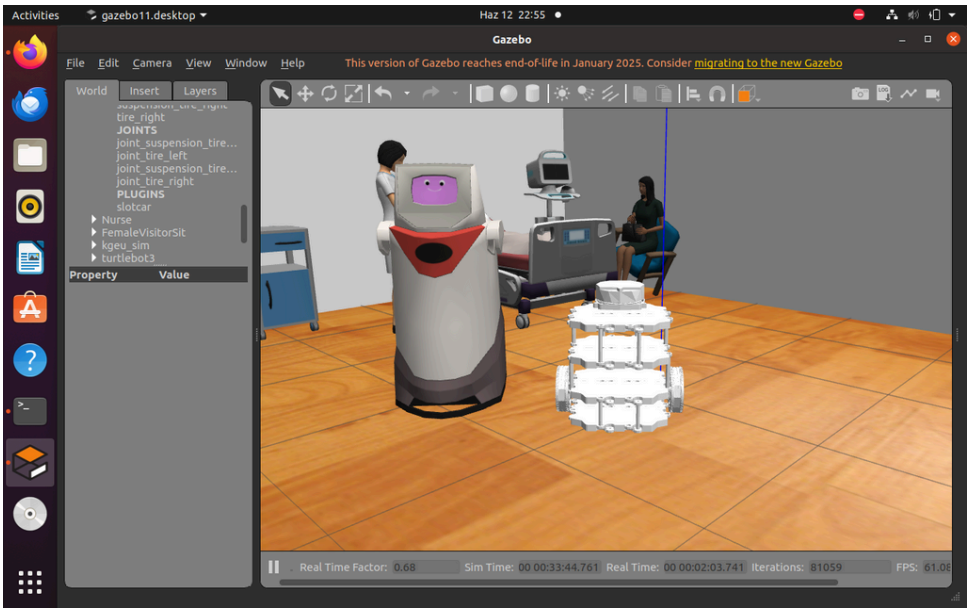


Image II. In this image, the TurtleBot3 Burger robot is shown at four times its original size. The default version appeared smaller than a household pet, which made interaction scenarios less visible and less realistic.

Due to rendering limitations inside VirtualBox, OBS Studio was used instead of Gazebo's internal recording tools, which produced black frames. Additionally, select still images from the simulation were enhanced using AI-based video generation tools (e.g., Sora) and external text-to-speech (TTS) services to simulate realistic voices and movements.

3.3 Survey Design and Procedure

The video demonstration presents three uniquely designed hospital service robots operating in a simulated care environment. The simulation showcases their approach toward a patient, their method of delivering medication, and how they prompt patients to report their pain levels. Each robot features distinct visual and interaction characteristics based on both appearance and input modality. Link for the video: Smart Hospital Robot Showdown | Vote for the Best! - Human Robot Interactions

Hospy is a compact, non-humanoid service robot designed to interact with patients using voice-based input. It relies on speech recognition to understand patient responses and is portrayed using both male and female voices in different video scenarios. Its appearance is minimalistic, emphasizing functionality over form.

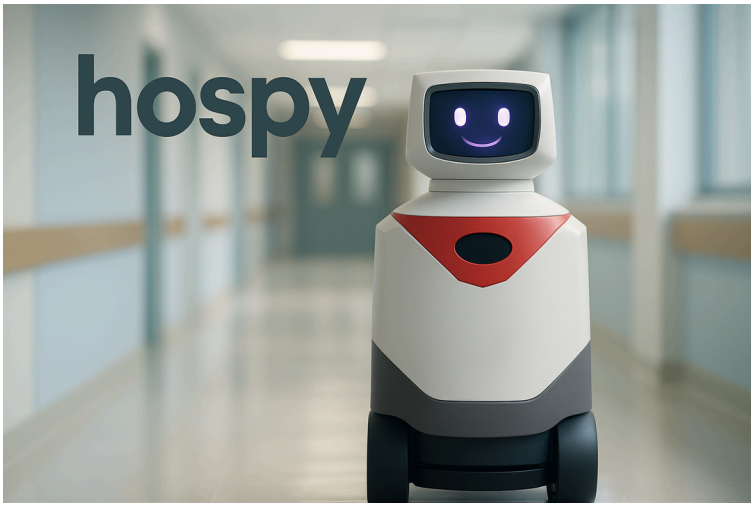


Image III. Hospy

ClaraDose represents a feminine humanoid robot. It features a screen interface on its chest where patients can rate their pain levels by touching virtual buttons. ClaraDose is designed to project warmth and empathy, using a female voice to enhance the perception of emotional support during interaction.

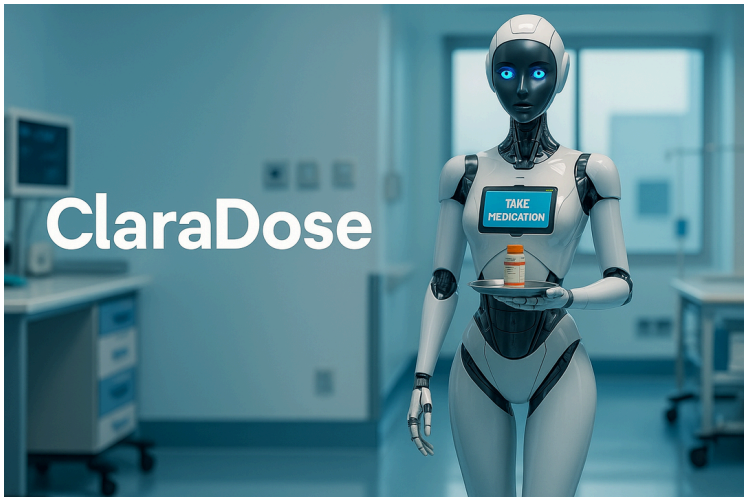


Image IV. ClaraDose

CareX Screen, on the other hand, is designed as a masculine humanoid robot with similar screen-based interaction. The robot is shown in two versions: one with a male voice and another with a female voice, allowing viewers to reflect on how vocal gender influences perception. Its appearance conveys professionalism and stability.

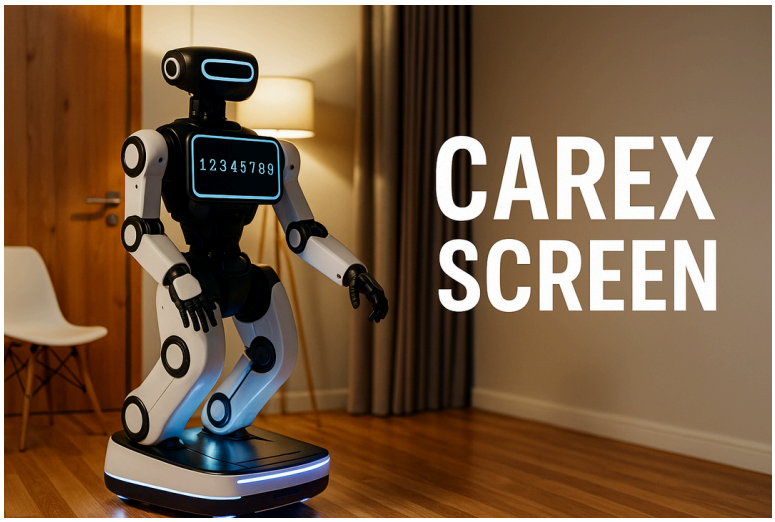


Image V. CareX Screen

Following the simulation, a Google Forms survey was prepared to assess user reactions to the robots. Participants first watched the video demonstration and then responded to a series of 34 structured questions, combining Likert-scale items, multiple-choice questions, and open-ended prompts. The form was distributed online and took approximately 6–8 minutes to complete.

Key focus areas in the survey included:

Demographic data (age, gender identity, academic field), Preference for each robot's voice (male/female), Perceived trustworthiness, empathy, and safety of each robot, Comparison of voice-based vs. screen-based pain input, Overall robot likability and willingness to interact in real hospitals

Participants' responses provided insights into how gendered design cues and modality affected their emotional and functional evaluations of the robots.

3.4 Technical and Experimental Challenges

Numerous challenges shaped the simulation and survey process:

Memory constraints in VirtualBox caused frequent crashes when loading larger simulation environments. Model scaling for TurtleBot3 disrupted physical dynamics and control plugins. Rendering limitations in Gazebo required third-party screen recording via OBS Studio. Manual adjustment of robot models (URDF/Xacro) was time-intensive due to relaunch delays (15–30 minutes per iteration).

Despite these limitations, a fully functional simulation and data collection pipeline was successfully completed, allowing for controlled analysis of human-robot interaction in a hospital setting.

4. Results

4.1 Quantitative Analysis

This section presents the results of a user study conducted to evaluate three hospital service robots—Hospy, ClaraDose, and CareX Screen—with respect to interaction effectiveness, perceived empathy, safety, appearance, and overall user preference. Each robot had a unique design and interaction modality, allowing for comparative assessment based on survey responses from 105 participants.

4.1.1 Comparative Overview

Evaluation Metric	Hospy (Voice-only, Non-humanoid)	ClaraDose (Feminine, Screen-based)	CareX Screen (Masculine, Screen-based)
Appearance (Rated 5/5)	21.0%	45.7%	19.0%
Usefulness of Interaction (5/5)	23.8%	40.0%	19.0%
Empathy/Politeness (5/5)	18.1%	33.3%	30.5%
Safety During Interaction (5/5)	18.1%	29.5%	28.6%
Preferred Voice: Female (%)	40.0%	74.3%	69.5%
Willingness to Re-interact (Yes)	35.2%	48.6%	40.0%

ClaraDose consistently outperformed the other two robots across all measured categories, particularly in appearance, empathy, and interaction usefulness. CareX Screen followed closely, especially in safety and professional perception, while Hospy received moderate scores, being viewed as functional but emotionally limited.

4.1.2 Interaction Type and Voice Preferences

Across all robots, female voices were preferred overall, with 69.5% of participants expressing a preference for female voices regardless of the robot's physical appearance. Interestingly, even for the masculine robot CareX Screen, users preferred a female voice(%40).

In terms of interaction modality, 55.2% of participants found speaking to be the most human-like method, compared to 28.06% who preferred using a touchscreen. However, screen-based robots (ClaraDose and CareX Screen) still received higher scores for perceived usefulness—suggesting that screen interactions may feel more precise or structured, even if speaking is perceived as more human-like.

4.1.3 Motivation and Preference

When asked which robot would motivate them most if used daily:

- ClaraDose: 40.6%
- Hospy: 34%
- CareX Screen: 25.4%

When asked which robot they preferred overall:

- ClaraDose: 44.8%
- Hospy: 30.2%
- CareX Screen: 24.5%

These findings suggest that ClaraDose not only leads in immediate preference but also in sustained engagement potential. Her combination of a feminine voice, human-like form, and touchscreen interaction appears to foster both emotional comfort and perceived reliability—making her a favorable long-term companion in a hospital context. Hospy, while simpler and less human-like, appeals to users who prefer clarity and minimalism, reflecting a preference for robots that "feel like robots" rather than humanoids. Despite its professional design, CareX Screen lags slightly in both motivation and preference, possibly due to a less emotionally engaging presence, even if it was rated highly in safety and structure. Overall, the data reflects a delicate balance between professionalism, emotional resonance, and trustworthiness when designing service robots for real-world healthcare environments

4.2 Qualitative Feedback

4.2.1 Key Thematic Findings

Participant comments revealed several recurring themes across the three robots:

Theme	**ClaraDose**	**CareX Screen**	**Hospy**
Empathy & Comfort	**“Sympathetic”, “like a nurse”**	**“Professional”, “confident”**	**“Plain”, “robotic but okay”**
Trust	**“Voice gave confidence”**	**“Looked like a real doctor”**	**“Did the job, but distant”**
Interaction Preference	**Appreciated screen + feminine tone**	**Clear instructions, but less warmth**	**Simplicity appreciated by some**
Unexpected Insights	**High empathy despite low preference**	**Preferred female voice despite male look**	**Appealed more to engineering students**

4.2.2 Sample Participant Quotes

“ClaraDose reminded me of a nurse—her tone made me feel safer.”

“CareX looked serious and reliable. I trusted him more.”

“Hospy was simple, but not engaging. Good for basic tasks.”,

“Hospy: I did not like the face and height, but I don't want to choose a robot with human like body type, I think it should be genderless(face-bofy-voice wise)”

4.2.3 Observational Trends

Participants from healthcare and education fields tended to favor ClaraDose.

Engineering students appreciated Hospy's simplicity and non-humanlike design.

CareX Screen was preferred by participants who valued professionalism and clarity over warmth.

5. Discussion

The study set out to explore how robot design variables—specifically gendered voice, physical appearance, and interaction modality—influence user perception in a simulated hospital environment. The data collected from 105 participants largely supports the hypotheses introduced in the beginning of the report.

5.1 Gendered Voice and Emotional Response

Consistent with prior research, the female voice was overwhelmingly preferred across all robots, including the masculine-appearing CareX Screen. This supports the hypothesis that users associate feminine voices with greater empathy, soothing presence, and approachability, even when the visual cues do not align. While this trend may be partially influenced by sociocultural expectations around caregiving roles, it highlights the importance of voice design in creating emotionally supportive robotic systems.

5.2 Humanoid vs. Non-Humanoid Design

The study also confirmed that humanoid robots, especially those with more expressive designs like ClaraDose, tend to be rated more favorably in terms of empathy, appearance, and interaction usefulness. ClaraDose significantly outperformed both Hospy and CareX Screen in perceived warmth and re-interaction willingness. Interestingly, Hospy, the only non-humanoid robot, still garnered substantial preference from engineering students and users uncomfortable with overly human-like robots—highlighting a divide in user expectations. Some participants even described humanoid robots as “uncanny” or “threatening,” aligning with the Uncanny Valley Theory discussed in HRI literature.

5.3 Interaction Modality and Trust

Although voice-based interaction was rated as the most “human-like” by the majority of participants, screen-based input methods (used by ClaraDose and CareX Screen) were still rated higher in terms of perceived usefulness. This suggests a trade-off: natural communication via voice enhances emotional engagement, but structured input via touchscreen may be more effective for task accuracy and clarity—especially in medical contexts. This supports previous HRI findings that recommend multi-modal interaction to accommodate different user needs and expectations.

5.4 Alignment with HRI Theories

The results align with concepts from Social Presence Theory and the Technology Acceptance Model (TAM). Participants showed greater trust in robots that matched their expectations of helpfulness, professionalism, or

empathy. For instance, ClaraDose's alignment of visual, auditory, and interaction design led to the highest overall acceptance, suggesting a high perceived usefulness and ease of interaction (core TAM constructs).

Additionally, the theme of trust emerged repeatedly in qualitative responses. Trust was not just about functionality —it was tied to appearance, tone of voice, and how "serious" or "approachable" the robot seemed. This underscores the multidimensional nature of user trust in robots, especially in high-stakes environments like hospitals.

6. Limitations

While this study offers valuable insights into the perception of hospital service robots, several limitations should be acknowledged:

6.1 Technical Limitations

The simulation environment was developed using ROS1 Noetic and Gazebo 11 on a VirtualBox-based Ubuntu system with only 4GB of allocated RAM. This limited hardware capacity constrained several aspects of the simulation:

- Model complexity had to be minimized, restricting the physical realism and range of motion of the robots.
- Environmental richness was compromised; larger, more detailed hospital environments (e.g., AWS Robomaker Hospital World) could not be used due to memory crashes.
- Simulation fluidity and control realism were negatively affected, especially when upscaling robot models (e.g., TurtleBot3).

In addition, audio realism was limited. The text-to-speech (TTS) voices used to simulate the robots lacked emotional nuance and natural speech rhythm, which may have influenced participants' perceptions of empathy and trust.

6.2 Experimental Constraints

Due to time constraints, the experiment included only a small number of pre-scripted scenarios (e.g., medication delivery, pain level inquiry), which may not capture the full spectrum of human-robot interactions in real hospital settings.

Moreover, participants did not interact with the robots in real-time. Their evaluations were based solely on short, edited video clips, which may not fully reflect how users would feel during actual, prolonged or repeated interactions. This format limits the ecological validity of the study and may lead to idealized or overly simplified judgments.

Finally, while the sample size was reasonably broad (n=105), the majority of participants were university students in specific academic fields (e.g., engineering, medicine, law). This may introduce cultural and demographic bias, potentially limiting the generalizability of the findings to broader patient populations in diverse hospital settings.

7.Conclusion

This study highlights the profound impact of design choices on user perception of hospital service robots. Among the three robots tested, ClaraDose emerged as the most preferred, appreciated for its empathetic tone, intuitive interface, and human-like appearance. Hospy, while more basic in design, was favored by users who valued

simplicity or were wary of humanoid robots. CareX Screen was seen as professional and reliable, though sometimes less emotionally engaging.

Key findings include:

- Female voices significantly increase perceived empathy and trust.
- Humanoid robots are often preferred, but some users find them unsettling.
- Voice interaction feels more human, but touchscreens are seen as more reliable in medical contexts.
- User background (e.g., engineering vs. medicine) influences robot preference.

These results have practical implications for the design of socially acceptable robots in healthcare. Designers must carefully balance emotional cues, technical interaction styles, and appearance to meet the diverse needs of users. Furthermore, customizability (e.g., voice gender options, interaction mode settings) may improve robot acceptance across demographics.

References

[1] Lee, Kwan Min, et al. “Can Robots Manifest Personality?: An Empirical Test of Personality Recognition, Social Responses, and Social Presence in Human–Robot Interaction.” *Journal of Communication*, vol. 56, no. 4, 2006, pp. 754–772.

[2] Eyssel, Friederike, and Frank Hegel. “(S)he’s Got the Look: Gender Stereotyping of Robots.” *Journal of Applied Social Psychology*, vol. 42, no. 9, 2012, pp. 2213–2230.

[3] Fischer, Kerstin, et al. “Levels of Embodiment: Linguistic Analyses of Factors Influencing HRI.” *Cognitive Systems Research*, vol. 6, no. 4, 2005, pp. 282–295.

[4] Lemaignan, Séverin, et al. “Learning by Teaching a Robot: The Case of Handwriting.” *Science Robotics*, vol. 3, no. 24, 2017.