Where are we with J.A.R.V.I.S?

An analysis of diegetic Human-AI interaction and design implications for real-world systems

Arina Shahbazi
Linköping University
Norrköping, Sweden

Lonni Besançon
Linköping University
Norrköping, Sweden

Mario Romero
Linköping University
Norrköping, Sweden

Figure 1: Tony Stark and J.A.R.V.I.S. collaboratively investigating an architectural model through an immersive visualization in *Iron Man 2* (2010), 01:22:13 [3].

Abstract

Science fiction invariably depicts human-AI interactions that exceed current technological capabilities. In this paper, we explore the value of such depictions not for their capacity to predict specific technologies, but for their power to methodically inspire interaction patterns that may influence future system design. We test an early-stage method that we call Diegetic Design Extraction (DDE) on a scene from Iron Man 2 (2010) [3], in which Tony Stark collaborates with J.A.R.V.I.S., his artificial research partner, to immersively analyze data. From this scene, we extract design patterns that point to open research challenges today. We specifically focus on turn-taking and the role of silence in conversation, as these are central observations made during our ongoing research on conversational flow in multimodal human-AI interaction. From this framing, we identify five related aspects of J.A.R.V.I.S.'s participation: initiative, turn-taking, functional silence, non-verbal activity, and contextual awareness. The analysis highlights how verbal silence can coexist with continued participation through the visualization, pointing to open design challenges for conversational AI in immersive sensemaking.

Keywords

Human-AI interaction, immersive visualization, science fiction, turn-taking, functional silence, multimodal interaction, Diegetic Design Extraction (DDE)

1 Introduction

Recent advances in large language models (LLMs), mixed reality, and immersive analytics have brought human-AI interaction closer to science fiction than ever before. Systems can now support conversational interaction, visual analysis, and AI-assisted sensemaking [6]. In our own empirical deployment studies, children using immersive AI visualizations reported frustration with interruptions during moments of reflection and with overly verbose system responses [2]. These observations connect to our ongoing research on conversational flow in multimodal human-AI interaction, where we are particularly interested in turn-taking and functional silence. We use turn-taking to refer to how the conversational floor is managed between the human and AI, and functional silence to periods of verbal silence that serve an interactional purpose, such as giving the human space to think or continue acting. In this paper, we explore these aspects of conversational flow through a science-fiction example, asking what the interaction between Stark and J.A.R.V.I.S. can reveal as potential design lessons for future immersive AI systems.

Using our early-stage method for Diegetic Design Extraction (DDE), we selected and analyzed a scene from Iron Man 2 (see Figure 1) [3]. We focus on the interaction between J.A.R.V.I.S. (Just A Rather Very Intelligent System) and Tony Stark during a process of scientific discovery. J.A.R.V.I.S. is an artificial intelligence that interacts through speech and immersive data visualizations. Throughout the scene, it combines voice interaction, immersive visualization,

gesture recognition, selective intervention, and extended periods of silence while supporting Tony's analytical process. Our goal is to uncover the useful characteristics that J.A.R.V.I.S. displays that remain challenging for current systems.

2 Methodology

We apply Diegetic Design Extraction (DDE), a method we are developing to systematically analyze fictional human-AI interactions and extract design lessons that may inform real-world systems. The current steps of DDE are: 1) select a design space; 2) identify relevant science-fiction scenes; 3) define the dimensions of analysis; 4) code the scene; and 5) extract design lessons that can inform the engineering of real-world systems.

The design space in this paper focuses on the workshop topic: human-centered AI support for immersive data sensemaking. Therefore, we were interested in fictional AI systems that not only execute commands, but also support a human in exploring, interpreting, or reasoning about information through an immersive interface. In this work, we consider AI guides or companions that help humans learn, solve problems, or investigate while showing some degree of agency, for example, by asking questions or providing relevant information.

To generate a list of sci-fi scenes for analysis, the authors independently identified media that fit the design space defined above. We combined these lists, discussed candidates to add or remove, and also used an LLM as an additional source of potential scenes. This process resulted in more than 60 candidates. We discussed these candidates based on their relevance to human-centered AI support for immersive data sensemaking, particularly whether they involved multimodal human-AI interaction, an immersive interface, and an AI showing some degree of agency. We chose the sequence from Iron Man 2 (2010) between 01:21:20 and 01:23:40 because it brings these characteristics together within one scene (see Figure 1).

In the scene, Stark investigates a physical architectural model in conversation with J.A.R.V.I.S. through an immersive visualization. Although J.A.R.V.I.S. primarily interacts through voice, it is integrated into the surrounding laboratory and visualization systems, allowing interaction through speech, visualization, and gesture. The goal in the scene is to uncover an element discovered and hidden in the model by Stark's father. Stark leads the overall investigation, while J.A.R.V.I.S. generates and manipulates the visualizations, interprets verbal statements and gestural actions, provides relevant information when necessary, and occasionally takes initiative. Thus, the scene provides a compact example of multimodal human-centered collaboration with an AI in an immersive environment.

Dimensions of analysis are the analytical perspectives used to examine an interaction, defining which features, behaviors, and relationships are observed, coded, and interpreted during the analysis. For these dimensions, we first recorded observable actions in the scene, including verbal statements, periods of verbal silence, gestures, stillness, and changes to the visualization.

We then interpreted these observations in relation to five aspects: initiative, turn-taking, functional silence, non-verbal activity through view transformations, and contextual awareness. *Initiative* captures when J.A.R.V.I.S. contributes without an immediate explicit command. *Turn-taking* captures when J.A.R.V.I.S. takes, holds, or yields the conversational floor in relation to Stark. *Functional silence* captures when J.A.R.V.I.S. remains verbally silent for an interactional purpose, such as giving Stark space to think or act. *Non-verbal activity* captures how J.A.R.V.I.S. contributes through view transformations rather than speech. And *Contextual awareness* captures J.A.R.V.I.S.'s apparent sensitivity to the broader interaction context, such as distinguishing between speech directed toward it and Stark thinking aloud or speaking to himself.

The authors' analyses were merged into a single table after discussion. The rows represent short interaction intervals, while the columns describe Stark's actions, J.A.R.V.I.S.'s actions, the conversational state, and the analytical aspect(s) represented by each interaction. Table 1 shows a 20-second segment from the analyzed scene to provide an example of the analytical process.

From the dimensions of analysis, we synthesize design lessons that present open challenges to current human-AI systems.

3 Results

We highlight evidence of five related aspects: 1) initiative; 2) turn-taking; 3) functional silence; 4) non-verbal activity; and 5) contextual awareness.

Initiative. J.A.R.V.I.S. regulates its participation not only by responding to Stark's commands, but also by taking initiative when appropriate. Most notably, during a sequence of filtering operations, J.A.R.V.I.S. asks, "What is it you are trying to achieve, sir?" (01:22:25). This shifts the interaction from carrying out individual requests to clarifying Stark's overall objective, prompting him to explain that he is trying to rediscover a new element. This creates the possibility for J.A.R.V.I.S. to support the reasoning process with greater autonomy.

Turn-taking. Throughout the scene, J.A.R.V.I.S. manages its participation by carefully choosing when to take the conversational floor. Stark holds the floor for much of the interaction while exploring the visualization, thinking aloud, and giving instructions. J.A.R.V.I.S. generally avoids verbally interrupting these sequences, but briefly takes the floor when it asks about Stark's broader goal before allowing Stark to continue. This allows Stark to maintain control of the reasoning process while still receiving support from J.A.R.V.I.S., reflecting the balance between user control and system initiative described in mixed-initiative interaction [4].

Functional silence. J.A.R.V.I.S. remains verbally silent for extended periods while Stark explores, thinks aloud, performs repetitive filtering operations, and interacts with the visualization. This silence gives Stark space to think, observe, and continue acting without verbal interruption. Importantly, silence does not imply inactivity. While remaining verbally silent, J.A.R.V.I.S. continues supporting the interaction through changes to the visualization.

Non-verbal activity. The interaction relies heavily on non-verbal activity. J.A.R.V.I.S. appears capable of integrating spoken language, gestural interaction, body movement, task context, and the evolving state of the visualization to respond to Stark's flow of actions and explicit commands. It continuously adapts the visual environment as Stark speaks and gestures, allowing the interaction to progress without requiring a verbal response to every action. A surprising

Table 1: Merged analysis of a 20-second segment (01:22:19–01:22:39) from the selected *Iron Man 2* scene (01:21:20–01:23:40).

Timestamp	Stark's actions	J.A.R.V.I.S.'s actions	Conversational state	Analytical aspect(s)
01:22:19–01:22:25	Stark studies the visualization, pauses, and then quietly asks J.A.R.V.I.S. to remove the footpaths.	Without speaking, J.A.R.V.I.S. removes the footpaths from the visualization.	Stark keeps the floor while J.A.R.V.I.S. supports him through a visual response.	Functional silence, Turn-taking, Non-verbal activity
01:22:25–01:22:33	Stark continues examining the visualization and explains that he is trying to rediscover a new element. His voice becomes quieter as he continues thinking aloud.	While updating the visualization, J.A.R.V.I.S. asks, "What is it that you are trying to achieve, sir?"	At the start of the interval, J.A.R.V.I.S. takes the floor to ask about Stark's broader goal, then yields it back to him.	Initiative, Turn-taking, Contextual awareness
01:22:33–01:22:39	Stark continues giving instructions while gesturing toward the visualization.	J.A.R.V.I.S. responds to Stark's speech and gestures by updating the visualization without speaking.	Stark retains the floor while J.A.R.V.I.S. follows Stark's ongoing multimodal input through view transformations.	Functional silence, Turn-taking, Non-verbal activity

consequence of this behavior is that not once in the scene does Stark verify that J.A.R.V.I.S. is listening, understanding, or processing a request. One possible explanation is that J.A.R.V.I.S. continuously shows its engagement through visible action. Highlighting, filtering, animation, and transformation of visual objects provide ongoing feedback that it is responding to Stark without requiring explicit verbal confirmation. The main takeaway is that the bulk of the conversation occurs through the visualization itself.

Contextual awareness. The interaction also suggests contextual awareness that extends beyond individual commands to how Stark is engaging with J.A.R.V.I.S. and the visualization. J.A.R.V.I.S. appears capable of distinguishing between reflective thought and interaction directed toward it. For instance, at the end of the interaction, Stark says, "Dead for almost 20 years, and still taking me to school [...] Thanks Dad" (01:23:16), and J.A.R.V.I.S. holds back from acting on or responding to it, suggesting that the statement is treated differently from speech directed toward the system.

4 Discussion

An unexpected conclusion emerges from our analysis. The most futuristic aspect of J.A.R.V.I.S. is not its embodied immersive visualization. Current systems support spatial interfaces to explore 3D data, gesture-based interaction, and AI-assisted analysis [1] [6]. For example, SpatialTouch shows how users can fluidly interact with spatial data visualizations across 2D and 3D environments [7]. Related work has combined mixed reality with generative AI to support collaborative analytical tasks [5]. What remains open is the seamless integration of these capabilities into a smooth human-AI collaboration.

The scene suggests that the main gap between current systems and J.A.R.V.I.S. is not a single missing technology, but how different forms of participation are coordinated during the interaction. In particular, J.A.R.V.I.S.'s verbal silence does not mean a pause in its participation. It can refrain from speaking while continuing to support Stark through view transformations. The visualization therefore becomes part of the conversational flow rather than a separate output channel. For real-world systems, the challenge is not only deciding what to say or visualize, but also when speech, silence, or visual action is the appropriate form of participation. This combination may offer useful design directions for future human-centered AI systems for immersive sensemaking of data.

5 Conclusion

Using Diegetic Design Extraction, we analyzed a human-AI interaction scene from Iron Man 2 and discussed its value in the context of recent work in conversational immersive analytics. The analysis suggests that contemporary systems are approaching J.A.R.V.I.S. in several technical dimensions, including multimodal interaction, immersive visualization, and AI-assisted workflows. However, the central challenge remains conversational agency.

J.A.R.V.I.S. stands out because it appears to coordinate when to speak, when to remain silent, and when to act through the visualization. This points to conversational agency as an important challenge for future human-centered immersive AI. Rather than relying on continuous verbal exchange, such systems may need to coordinate speech, silence, and visual action while allowing the human to maintain control of the reasoning process.

Acknowledgments

This work was partially supported by the Marcus and Amalia Wallenberg Foundation (2023.0130).

References

[1] Lonni Besançon, Anders Ynnerman, Daniel F. Keefe, Lingyun Yu, and Tobias Isenberg. 2021. The State of the Art of Spatial Interfaces for 3D Visualization. *Computer Graphics Forum* 40, 1 (2021), 293–326. doi:10.1111/cgf.14189

[2] Emelie Fälton, Isabelle Strömstedt, Mathis Brossier, Andreas Göransson, Konrad J. Schönborn, Amy Loutfi, Erik Sundén, Mujtaba Fadhil Jawad, Yadgar Suleiman, Johanna Björklund, Mario Romero, Anders Ynnerman, and Lonni Besançon. 2026. Children's Expectations, Engagement, and Evaluation of an LLM-enabled Spherical Visualization Platform in the Classroom. In *EuroVis 2026 - Education Papers*, Jillian Aurisano, Carolina Nobre, and Xinhuan Shu (Eds.). The Eurographics Association. doi:10.2312/eved.20261003

[3] Jon Favreau. 2010. Iron Man 2. Film. Distributed by Paramount Pictures.

[4] Eric Horvitz. 1999. Principles of Mixed-Initiative User Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 159–166. doi:10.1145/302979.303030

[5] Vahid Pooryousef, Maxime Cordeil, Lonni Besançon, Richard Bassed, and Tim Dwyer. 2024. Collaborative Forensic Autopsy Documentation and Supervised Report Generation Using a Hybrid Mixed-Reality Environment and Generative AI. *IEEE Transactions on Visualization and Computer Graphics* 30, 11 (Nov. 2024), 7452–7462. doi:10.1109/TVCG.2024.3456212

[6] Chaoming Wang, Valeria Garro, Veronica Sundstedt, Yan Hu, and Prashant Goswami. 2026. Immersive Analytics Meets Artificial Intelligence: A Systematic Review. *Computational Visual Media* 12, 1 (Feb. 2026), 1–34. doi:10.26599/CVM.2025.9450478

[7] Lixiang Zhao, Tobias Isenberg, Fuqi Xie, Hai-Ning Liang, and Lingyun Yu. 2025. SpatialTouch: Exploring Spatial Data Visualizations in Cross-Reality. *IEEE Transactions on Visualization and Computer Graphics* 31, 1 (Jan. 2025), 897–907. doi:10.1109/TVCG.2024.3456368