Proceedings of the Design Society, Volume 6: DESIGN 2026
https://doi.org/10.1017/pds.2026.10589

Engineering design through dialogue: a method for analysing speech-based human-AI conversation

Kieran Gunn✉, Ross Brisco, Alexander "Freddie" Holliman, Erfu Yang and Rebecca Macfie

University of Strathclyde, United Kingdom

✉ kieran.gunn.2019@uni.strath.ac.uk

ABSTRACT: Speech-capable AI systems introduce new possibilities for communication and collaboration in design, yet methods for analysing human-AI interactions through speech remain limited. This paper proposes and applies a method for analysing conversational interactions in speech-based human-AI design activity. Grounded in conversation analysis, this method reveals how conversational structure and designer roles emerge through spoken interaction, offering an analytical framework for examining communication, cognition, and collaboration in design.

KEYWORDS: artificial intelligence (AI), communication, design analysis, utterance, generative

1. Introduction

The rapid development of artificial intelligence (AI) has introduced new opportunities for communication and collaboration within engineering design practice. While recent research in generative AI has typically focused on the development and use of text-based and image-based AI systems (Liu, 2023; Tembine et al., 2025), recent advances in speech-capable interfaces have expanded the modalities through which designers can interact with AI (Rayan et al., 2024). Despite extensive work on the technical development and application of AI systems in engineering design (Edgecomb et al., 2025), the role of speech-based AI interaction remains largely underexplored compared to that of text and image. This paper presents exploratory research into how designers engage in spoken interaction with a speech-capable generative AI system during design tasks. By comparing human-human (H-H) and human-AI (H-AI) spoken interactions, this study analyses differences in conversational structure, utterance types, and interactional roles. This represents an early-stage methodological contribution, proposing an approach for analysing spoken interaction as the medium through which design activity is articulated and structured. This research situates this analysis of spoken interaction within a broader discussion of how AI systems may support collaborative design.

2. Research background

Research attention of AI capabilities has increased since the commercial release of generative and multimodal systems such as ChatGPT, Gemini, and Copilot (Bubeck et al., 2023). Within engineering design, studies examining the use and assessment of AI in practice have revealed both affordances and limitations in creative and human-centred contexts (Dall'Asta, 2025; Srinivasan et al., 2024). As AI systems have developed, communication has expanded beyond just text and images to include speech as a modality for interaction. This highlights the need to examine how spoken interaction structures and organises design activity, and how it shapes the roles assumed by designers when engaging with speech-capable AI. This section provides background and justification for a proposed method of analysing spoken interaction between designers and AI systems.

https://doi.org/10.1017/pds.2026.10589 Published online by Cambridge University Press

2.1. Artificial intelligence, speech and design

The development of large language models and multimodal interfaces has made spoken dialogue a mode of interaction alongside text and images. (Zhang et al., 2023). In this context, speech-based interaction refers to the use of speech as the primary means of communication between humans and AI. The established literature on speech-capable AI systems highlights speech as an emerging yet marginal modality in design research.

For example, Fernandes et al. (2024) presented a GPT-powered assistant for Building Information modelling (BIM), which demonstrated functionality but focused on single-function and command-based queries rather than sustained dialogue. Chandrasegaran et al. (2023) analysed spoken human-human design conversations using linguistic and machine-learning methods to observe designers' tentativeness and explanation. However, they did not analyse AI as an interactive design partner. This means that their findings do not consider how an AI partner may influence interactions in design conversations.

Research also examines the integration of AI technologies into design education to enhance student learning engagement and creativity. For example, Chui et al. (2024) introduces a framework to enhance engineering students' practical skills in generative AI. However, speech was used only to transcribe students' discussions, rather than to enable any form of interaction. This means that the framework does not account for how differences between textual and spoken interactions with AI systems may alter the flow and structure of collaboration. Also, Al Omoush et al. (2023) introduced a speech-based pedagogical agent for STEM education that supported spoken interactions. While the agent demonstrated speech functionality, the context remained in educational support rather than in engineering design practice. This means that the study could not explore how a speech-based agent could influence designers' roles, design reasoning, and the progression of design activities.

Across these studies, speech is treated as input, output, or data, rather than as a sustained interactional medium. As a result, core conversational features, such as turn-taking and behavioural roles, are missing from research on speech-based AI interactions within engineering design. This suggests that, while speech-capable AI systems are advancing technologically, there is limited exploration of how these systems can effectively converse with designers in engineering design contexts.

2.2. Toward a speech-based Interaction analysis method

The limitations identified in existing research emphasise the need for a method capable of capturing and analysing the conversational dynamics of speech-based human-AI interaction in design. The approach proposed in this paper draws on established work in human-AI interaction and conversational analysis (CA) to examine how spoken interaction structures design activity when an AI is introduced.

Previous research on human-AI interaction has largely focused on evaluating performance, usability, and task outcomes rather than interaction. For example, Shi et al. (2023) reviewed 93 papers on human-AI collaboration, identifying roles such as AI assisting designers, designers assisting AI, and designer-AI collaboration. However, this review does not consider how these roles are produced through spoken interaction. Similarly, Kadenhe et al. (2025) reviewed co-creation and co-design AI frameworks, but did not reference speech as an interactional modality. This suggests a gap between research on the technical developments of AI systems and the exploration of AI as an interactive partner.

The term "human-AI collaboration" is not consistently treated as equivalent to human collaboration. It is important to note that while many studies refer to collaboration between humans and AI, others question whether collaboration is actually taking place, given the lack of stake, responsibility, or agency on the side of the AI system. Shneiderman (2021) and Cabitza et al. (2021) cautions against framing AI systems as "teammates, partners, or collaborators", arguing that AI risks eroding agency, responsibility, and accountability. From this perspective, interaction with AI may assume conversational or collaborative forms without constituting collaboration. In Computer-Supported Cooperative Work (CSCW), collaboration is commonly understood to involve communication, coordination, and cooperation (Ellis et al., 1991; Schmidt & Simone, 1996). Spoken interaction plays an important role in this process as designers articulate their intentions, negotiate decisions, and align their actions through dialogue (Matthews & Heinemann, 2012; McDonnell, 2012). Analysing spoken interactions, therefore, provides insight into how design activity is organised, without assuming shared agency or collaboration.

Conversational Analysis (CA) provides a foundation for observing turn-taking, interactional structure, and organisation (Sacks et al., 1974). Within design research, Matthews & Heinemann (2012) demonstrate that CA can respecify design methods as social actions, carried out through people's

https://doi.org/10.1017/pds.2026.10589 Published online by Cambridge University Press

interactions. Ekström & Stevanovic (2023) builds on this idea and shows that social action consists of small, sequential moments that produce larger structures and relationships. This suggests that turns and utterances shape the structure of an activity, and positions conversations as a way to examine design activity. Although CA was developed for human-human social interaction, this research adopts CA principles for examining spoken design activity, without assuming social interaction.

This research does not look to establish whether collaboration occurs between human and AI participants; rather, it examines how spoken interaction is structured when AI is introduced into design activity. The aim of this research is to develop and apply a method for analysing spoken interaction in speech-based human-AI, enabling the analysis of conversational structure and interactional roles within spoken, turn-based interaction through a comparative examination of human-human and human-AI conversations in design activity

3. Methodology

This section outlines the development and testing of the proposed method within a controlled environment. This approach involves capturing and analysing spoken conversation, the structure of design activity, and a coding framework to identify changes in conversational structure, behaviours, and the roles of humans and AI partners.

This research adopts an exploratory approach grounded in conversational analysis and design research. This approach looks to observe, interpret, and analyse how speech-based AI systems engage in dialogue with human designers during design activity. A study was designed to analyse quantitative data on participants' conversations during design tasks. Quantitative measures were used to characterise conversational features, including the total number of turns, frequency of utterances per turn, and transitions between design stages.

An empirical study was designed to apply and test the proposed analytical method within a controlled design setting. It outlines the rationale for the study design, the structure of the sessions, and the framework principles used to ensure comparability between human-human and human-AI interactions.

To evaluate the proposed method, a comparative experimental design was created to compare the effects of AI interaction in a design environment. Two conditions were defined:

Human-Human interaction (H-H): Two human participants interact verbally to complete a design task. This serves as the baseline condition to capture natural design conversation.

Human-AI interaction (H-AI): One human participant and one speech-capable generative AI system (ChatGPT, Advanced Voice Mode) interact verbally to complete a design task. This condition enables the analysis and comparison of AI participation in design activity.

ChatGPT, in Advanced Voice Mode, was selected for this study because it supports sustained spoken dialogue for turn-based interaction. A comparison of AI systems regarded ChatGPT as the reference point for high-quality voice interaction against other systems (Stefanelli, 2025). This selection does not assume equivalence between the conversational abilities of the AI system and those of human designers. Future work will examine questions concerning the comparability of human design activity and speech-capable AI interactions are areas of future work.

To observe and analyse each condition, two design tasks were developed to enable natural design reasoning and interaction solely through speech. As such, each task had its own design brief, which was a short, open-ended design problem for participants to work through.

Design brief 1: "Domestic food waste is a serious problem due to global food shortages and socio-economic imbalances. Generate concepts for novel and feasible products that may reduce unnecessary food wastage in the home"

Design brief 2: "Chores such as cooking and cleaning may be difficult for wheelchair users due to space and height limitations. Generate concepts for novel and feasible products that may facilitate domestic chores for wheelchair users"

Both design briefs were adopted from (Campbell et al., 2024), an fMRI study of professional product design engineers. Both briefs were selected for their real-world, relatable scenarios, which enable participants to generate a variety of solutions that have been employed in design cognition research.

The method used in this research, which instructs participants to use only speech during the design tasks, is the 'Sit On Your Hands' method developed by Rebecca Macfie at the University of Strathclyde, originally designed for the study of design cognition and mental imagery. This method involves teams developing a design solution through discussion, without writing, sketching, or physical design artefacts.

https://doi.org/10.1017/pds.2026.10589 Published online by Cambridge University Press

Established research demonstrates that sketching and gestures play a crucial role in design cognition by externalising thought and relieving working memory (Tversky, 2002, 2010). While this method limits the use of external representations known to support design reasoning, ideation, and problem-solving, it was intentionally adopted to focus on spoken interaction as the primary medium of design activity. This was considered appropriate for the aim of this exploratory study, analysing conversational structure and interactional behaviour in spoken design interactions.

3.1. Design session structure

Each session followed a consistent structure to ensure comparability of the two conditions. Table 1 shows the sequence of stages within the sessions. This sequence was chosen to allow for a direct comparison between conditions, where completing the H-H interaction task first provided a baseline of natural conversation in a design task before participants interacted with the AI system.

Table 1. Overview of session stages and durations for H-H and H-AI tasks

Stage	Duration	Description
Welcome and Introduction	5 minutes	Participants were welcomed to the room, briefed on the purpose and procedure of the tasks, and asked to provide consent for participation.
H-H Interaction Task	25 minutes	Participants were asked to complete a set of activities.
H-AI Interaction Task	25 minutes	Participants were asked to complete a set of activities.

To maintain consistency across both tasks and conditions, the lead researcher read the task instructions aloud before each design task:
"You will work together to complete a design task. You are to develop a design solution to the design brief solely through discussion. No form of physical design is allowed, including writing, sketching or bodily gestures. Read everything aloud as you think or do, including the design brief and activities."
Table 2 presents a sequence of eight design activities that each participant was required to complete throughout the task. Activities 1-4 were given a maximum of 20 minutes to complete. Activities 5 & 7 were included to ensure that both participants had the chance to verbalise their thoughts and discuss their process for completing the following activities. Activities 6 & 8 were included to deliver the participants' final solutions.

Table 2. Design activities and durations for H-H and H-AI tasks

Activity Number	Activity	Duration
1	Discuss the brief	20 minutes
2	Generate solutions	
3	Select solution	
4	Develop solution	
5	Prepare to present	1 minute
6	Present solution	1 minute
7	Prepare to justify	1 minute
8	Justify solution	1 minute

Participants were recruited from the Department of Design, Manufacturing & Engineering Management (DMEM) at the University of Strathclyde. Eligible participants were drawn from the final year and recent graduates across DMEM courses: Product Design Engineering (PDE), Product Design and Innovation (PDI), Sports Design Engineering (SDE), and Manufacturing Engineering with Management (MEM). Final-year participants were invited from the cohort of bachelor's and master's students enrolled in formal education in design methods and processes. Recent graduates were invited only if they had completed a DMEM undergraduate course within the past 2 years, to ensure retention of design

https://doi.org/10.1017/pds.2026.10589 Published online by Cambridge University Press

knowledge. There were 10 participants in total, including four PhD students, four master's students, and two bachelor's students.
All participants were provided with information about the session, its conditions, and activities prior to their participation. All participants were provided with participant information sheets and a consent form, in line with the University of Strathclyde's policy.
Each session was held across two rooms. The H-H sessions were held in one room for both participants. The H-AI sessions were held in separate rooms, allowing human participants to converse with their AI partner without interference. For each participant, the sessions' equipment included a laptop or tablet running ChatGPT Advanced Voice Mode, as well as a recording device to capture audio.

3.2. Analytical framework

The analytical framework integrates quantitative conversational analysis of design conversations to evaluate the differences in conversational structure between H-H and H-AI sessions. This framework supports the research on how participants used speech to complete design tasks, both with a human partner and a speech-capable AI system.
To analyse the H-H and H-AI tasks, three data sources were collected for quantitative analysis of the conversational interactions. Complete recordings of verbal dialogue for each participant, across both conditions, were collected for transcription and coding. The logs of participants' conversations with ChatGPT were collected, which included the participants' and AI's verbatim prompts and responses. These logs were not intended for generating transcripts of the tasks, but rather to verify the accuracy of audio-to-text transcription.
All session recordings were transcribed verbatim. The audio-video tool 'Riverside.fm' (October 2025) was used to generate complete transcriptions. All audio recordings were stored on a secure server once a complete anonymised transcription of dialogue could be generated and verified.
All transcripts were segmented into turns and utterances by the lead researcher prior to coding. Transcripts were converted from.txt files through Excel, formatted by the following columns:
[Condition (H-H/H-AI)], [Speaker], [Timestamp], [Utterance Text], [Utterance type], [Design Stage]

3.2.1. Utterance-type coding framework

Within this research, an utterance is not just a linguistic expression but an active response to an ongoing exchange (Haye & Larrain, 2011). This definition treats an utterance as a single, continuous piece of speech. The coding framework was developed through an inductive process, informed by the structure of the data. The coding scheme for utterance types contains eleven codes, each representing a distinct conversational function or purpose. Table 3 details the set of codes and definitions used to code and analyse each participant's utterances.

Table 3. Utterance-type coding scheme

Code	Definition (Cambridge Dictionary, 04/11/2025)
Agree	To have the same opinion
Clarify	To make something clear or easier to understand
Direct	To control or be in charge of an activity, organisation, etc.
Disagree	To not have the same opinion, idea, etc.
Evaluate	To judge or calculate the quality, importance, amount, or value of something
Exclaim	To say or shout something suddenly because of surprise, fear, pleasure, etc
Explain	To make something clear or easy to understand by describing or giving information
Question	To ask a question
Reflect	To show or to express or be a sign of something
Inform	To give information or to teach someone about something
Suggest	To mention an idea, a possible plan, or an action for other people to consider

https://doi.org/10.1017/pds.2026.10589 Published online by Cambridge University Press

3.2.2. Design-stage coding framework

In addition to utterance-type coding, each transcript utterance was coded by design stage to mark the progression of the participants through the design tasks. Table 4 details the set of codes and descriptions used to code and analyse the design-stage progression throughout the task. From the instruction to participants to use only speech throughout the tasks, the procedure for applying codes to turns and utterances was based on participants' verbal cues.

Table 4. Design-stage coding scheme

Code	Description
Brief	Framing or discussing the design problem, activities or tasks
Ideate	Generating ideas, concepts, or features
Evaluate	Comparing or judging previous ideas, concepts, or features
Develop	Elaborating or refining the selected idea, concept, or feature
Deliver	Summarising or concluding task outcomes

3.2.3. Analysis procedure

The analysis procedure involves multiple steps to capture quantitative data. This approach enables ease of coding and analysis, based on both the conversational structure and content of the transcripts. This procedure comprises three analytical steps:

Step 1 - Utterance-Type Coding:

Each transcript was coded using the utterance-type coding schemes (Table 3). Each participant's turn was separated into individual utterances, and each utterance was assigned an utterance type. More than one code may be applied to utterances with multiple functions or purposes.

Step 2 - Design-Stage Coding:

Each transcript was coded using the Design-stage coding scheme (Table 4). Each utterance was assigned a design stage based on its content and context in relation to the set list of activities for both tasks. This approach allows the researcher to map utterance types to stages in the design process.

Step 3 - Quantitative Analysis of Transcripts:

Coded data was combined into one document to identify differences and frequency distributions of utterance types and design stages. Comparative analysis between the H-H and H-AI sessions was conducted to reveal differences in conversational structure and interactions.

Given the exploratory nature of this study, all transcripts were segmented and coded by the lead researcher. Inter-rater reliability was not assessed at this stage as the purpose of this study was to explore and propose an analytical approach rather than to validate a finalised coding scheme. The coding framework was developed inductively and refined as additional transcripts were analysed, allowing new utterance functions and purposes to inform the coding scheme's development.

4. Results

The results outline how the structure of conversation differed when participants interacted with a speech-capable AI system. Quantitative measures of conversational turns, utterances, and the distribution of utterances by design stage were collected to observe these differences across conditions.

4.1. Turns and utterances by design stage

Table 5 summarises the conversational metrics for each design stage in both the H-H and H-AI sessions. Each row for each design stage shows the total number of turns and utterances made by all ten participants, along with the calculated average number of utterances per turn. The totals row at the bottom of the table contains the combined number of turns, utterances, and average utterances per turn across the sessions.

Across all design stages, the total turns made in H-AI sessions were fewer than those in H-H sessions, resulting in a 46% decrease. There is also a notable difference in the total number of utterances in H-AI sessions compared with H-H sessions, with a 50% decrease overall. However, taking the total number of utterances from both sessions in relation to the total number of turns from each, the average number of

https://doi.org/10.1017/pds.2026.10589 Published online by Cambridge University Press

utterances per turn from the H-AI sessions was larger than that of the H-H sessions. The average number of utterances per turn from the H-AI sessions increased by 21% compared to the H-H sessions. These results show that participants took fewer turns to complete the H-AI design tasks, but within those turns, there were more utterances.

Table 5. Conversational metrics across H-H and H-AI sessions

Design Stage	H-H Turns	H-H Utterances	H-H: Avg. Utterances Per Turn	H-AI Turns	H-AI Utterances	H-AI: Avg. Utterances Per Turn
Brief	34	61	1.79	20	38	1.90
Ideate	199	342	1.72	62	125	2.02
Evaluate	74	140	1.89	38	77	2.03
Develop	204	346	1.70	53	116	2.19
Deliver	55	90	1.64	57	138	2.42
Total	566	979	1.75	230	494	2.11

4.2. Distribution of utterance types

Table 6 details the average percentage differences (Δ%) of each utterance type across the five design stages and the overall design process. These averages were calculated by subtracting the percentage share of each utterance type in the H-H session from its share in the H-AI session. Positive values indicate a higher proportion of the utterance type during H-AI interaction, where negative values indicate a reduction in its use. The right-most column of the table shows this overall average percentage difference for each utterance type.

Table 6. Average percentage difference in utterance type per design stage

Utterance Type	Brief (Δ%)	Ideate (Δ%)	Evaluate (Δ%)	Develop (Δ%)	Deliver (Δ%)	Overall (Δ%)
Agree	−6.17	−15.41	−10.00	−10.47	−7.34	−9.88
Clarify	−1.64	0.14	−2.14	0.85	−0.39	−0.64
Direct	−1.29	0.86	5.06	2.86	5.75	2.65
Disagree	5.26	−2.42	−2.53	0.27	−0.43	0.03
Evaluate	2.63	0.57	1.75	2.58	2.51	2.01
Exclaim	−5.57	−1.03	7.08	1.14	5.41	1.41
Explain	−0.65	5.06	0.65	5.42	−6.86	0.72
Inform	2.11	−5.56	−16.49	−18.34	−18.02	−11.26
Question	6.60	12.81	14.42	12.27	13.24	11.87
Reflect	0.00	0.51	0.00	−0.58	0.72	0.13
Suggest	−1.29	4.47	2.21	4.00	5.41	2.96

Between the H-H and H-AI sessions, notable differences were observed through the distribution of utterance types throughout the five stages of the design process. The overall average differences show an increase in ‘Question’ (11.87%), ‘Suggest’ (2.96%), and ‘Direct’ (2.65%) utterances during the H-AI tasks, suggesting a more exploratory and directive role by participants. The overall differences also show a decrease in ‘Agree’ (-9.88%) and ‘Inform’ (-11.26%) during H-AI tasks, suggesting a reduction in affirmation and fewer factual exchanges from participants. Other utterance types had minor differences, suggesting that these utterances were relatively stable across both tasks.

https://doi.org/10.1017/pds.2026.10589 Published online by Cambridge University Press

5. Discussion

The reduction in total turns and utterances during H-AI tasks suggests a shift in how participants used their conversation when interacting with a speech-capable AI system. Despite fewer overall interactions, participants produced a higher average number of utterances per turn, suggesting that each exchange with the AI system was longer and more elaborate. This has been observed in established research where participants adapted their communication to make points more straightforward or to ensure that AI understood their meaning or intent (Luger & Sellen, 2016). The longer turns with the AI indicate that the interaction was less exploratory and more one-sided, with participants guiding the conversation and the progression of the task. This suggests an increased cognitive effort and workload placed on participants to maintain a coherent and effective exchange with the AI (Schmidhuber et al., 2021). Within these turns, the distribution of utterance types indicates that the behaviour of human participants also shifted during interaction with the AI. The increase in 'Question', 'Suggest', and 'Direct' in H-AI interaction indicates that participants assumed a more directing and commanding role, guiding the AI rather than engaging in a partnership. In contrast, the decrease in 'Agree' and 'Inform' utterances in H-AI interaction indicates that participants gave less affirmation to the AI and shared fewer factual details or pieces of information. This aligns with established research showing that humans tend to assume conversational control when a partner provides limited feedback or fails to sustain natural interaction (Clark & Brennan, 1991; Galland et al., 2022). These results suggest that participants adjusted their communication to accommodate the conversational limitations of the AI, resulting in fewer turns but more structured and intentional exchanges.

These interactional differences appeared to influence how participants assumed roles in the design process. Rather than engaging as design partners, participants treated their interaction with the AI as a reactive tool, assuming responsibility for guiding the conversation and maintaining task progression. This shift in human roles aligns with the idea that communication adapts to the constraints of the medium (Clark & Brennan, 1991). When grounding cues are limited, participants take on more communicative effort to manage the task and maintain a shared understanding. These observations of human-AI interactions highlight how differences in grounding effort influence the structure of the interaction and how human designers assume different roles between tasks.

These findings demonstrate that speech-based AI interactions within design affect the conversational structure and roles in human-AI interaction. This highlights that integration of speech-capable AI systems into design practice requires not only continued technical development of such systems, but also exploration into conversational structure and cognitive effort within human-AI interaction.

While established research has identified roles for AI systems in engineering design, such as generators, evaluators, or decision-makers (Alqahtani et al., 2023), examining how these roles manifest in spoken interaction remains an area for future work. The findings from this study reflect how interactional roles emerged within this specific turn-based, task environment. As such, they should be interpreted in relation to this configuration rather than representing all possible roles of AI participation in design.

Furthermore, these findings contribute to discussions on collaborative design activity by clarifying the role of communication within design work. Changes in conversational structure and role distribution suggest that participants adapted their communication to manage the AI's limitations, rather than engaging in shared coordination or responsibility. These findings do not demonstrate collaboration between humans and AI. However, analysing these spoken interactions provides valuable insights into how design activity is organised when paired with an AI system, and how speech-capable AI systems may support or facilitate collaborative design processes.

6. Limitations

This study represents an early-stage methodological exploration of speech-based human-AI interaction in engineering design, and the findings should be interpreted appropriately.

All transcript segmentation and coding were conducted by the lead researcher using an inductively developed coding framework. Inter-rater reliability was not assessed at this stage, as the purpose of the study was to explore and propose an analytical approach rather than to validate a final coding scheme. As the framework was derived from a single study, some utterance categories may be broad or context-dependent (e.g. Reflect), and coding decisions may reflect task-specific interpretations.

The participant sample was small and drawn from a single university department. Although appropriate for piloting the proposed method, the results should be interpreted as exploratory.

https://doi.org/10.1017/pds.2026.10589 Published online by Cambridge University Press

The experimental design was not fully counterbalanced, with participants completing the H-H task before the H-AI task, and different design briefs being used across conditions. Although this structure enabled initial comparison of conversational behaviour across design problems, it may have influenced the observed interaction.

Finally, the study intentionally excluded writing, sketching, and gesture to isolate spoken interaction as the primary medium of design activity. As a result, participants' design activity was observed under these constraints; accordingly, the observed behaviours should be interpreted with this in mind.

7. Conclusion

This paper proposes an exploratory method for analysing spoken interactions through speech-based human-AI design activity. The comparison of human and human-AI sessions suggests that AI participation in collaborative tasks altered the conversational structure and assumed roles in design tasks. Speech interactions between human designers and the AI system enabled natural exchanges but also highlighted challenges in maintaining a shared understanding throughout the design tasks. As part of a broader research focus on collaborative design, the proposed method provides a foundation for further exploration into how speech-based AI systems may support or shape collaborative design activity through dialogue. Future work will extend and validate this approach at a larger scale across broader participant pools from design domains and multimodal design settings.

References

Al Omoush, M. H., Salih, S. E., Kishore, S., & Mehigan, T. (2023). Interactive Multimodal Learning: Towards Using Pedagogical Agents for Inclusive Education. 2023 IEEE International Humanitarian Technology Conference, IHTC 2023, IEEE Communications Society; IEEE Engineering Projects in Community Service (EPICS); IEEE Power and Energy Society-. http://dx.doi.org/10.1109/IHTC58960.2023.10508848

Alqahtani, T., Badreldin, H. A., Alrashed, M., Alshaya, A. I., Alghamdi, S. S., bin Saleh, K., Alowais, S. A., Alshaya, O. A., Rahman, I., Al Yami, M. S., & Albekairy, A. M. (2023). The emergent role of artificial intelligence, natural learning processing, and large language models in higher education and research. *Research in Social and Administrative Pharmacy*, 19(8), 1236–1242. https://doi.org/10.1016/j.sapharm.2023.05.016

Bubeck, S., Chandrasekaran, V., Eldan, R., Gehrke, J., Horvitz, E., Kamar, E., Lee, P., Lee, Y., Li, Y.-F., Lundberg, S. M., Nori, H., Palangi, H., Ribeiro, M. T., & Zhang, Y. (2023). Sparks of Artificial General Intelligence: Early experiments with GPT-4. ArXiv. https://www.semanticscholar.org/paper/Sparks-of-Artificial-General-Intelligence%3A-Early-Bubeck-Chandrasekaran/8dbd57469bb32e6d57f23f5e765bf1c9ac8e080c

Cabitza, F., Campagner, A., & Simone, C. (2021). The need to move away from agential-AI: Empirical investigations, useful concepts and open issues. *International Journal of Human-Computer Studies*, 155, 102696. https://doi.org/10.1016/j.ijhcs.2021.102696

Campbell, G., Hay, L., Gilbert, S., McTeague, C., Coyle, D., & Grealy, M. (2024). Functional activity and connectivity during ideation in professional product design engineers. *Design Studies*, 91–92, 101247. https://doi.org/10.1016/j.destud.2024.101247

Chandrasegaran, S., Salah, A. A., & Lloyd, P. (2023). Constructing design activity in words: Exploring linguistic methods to analyse the design process. *Design Studies*, 86, 101182. https://doi.org/10.1016/j.destud.2023.101182

Chui, C. K., Yang, L., & Kao, B. (2024). Empowering Students in Emerging Technology: A Framework for Developing Hands-on Competency in Generative AI with Ethical Considerations. ASEE Annual Conference and Exposition, Conference Proceedings. https://doi.org/10.18260/1-2–47250

Clark, H. H., & Brennan, S. E. (1991). *Grounding in communication*. (L. B. Resnick, J. M. Levine, & S. D. Teasley, Eds; pp. 127–149). American Psychological Association. https://doi.org/10.1037/10096-006

Dall'Asta, J. (2025). Integrating Artificial Intelligence into the Human-Centered Design Process: Enhancing Creativity and User-Centricity in Architectural Education. https://doi.org/10.54941/ahfe1005908

Edgecomb, I. M., Brisco, R., Gunn, K., & Holliman, A. "Freddie". (2025). Artificial Intelligence in engineering design: An industry perspective. *Proceedings of the Design Society*, 5, 641–650. https://doi.org/10.1017/pds.2025.10078

Ekström, M., & Stevanovic, M. (2023). Conversation analysis and power: Examining the descendants and antecedents of social action. *Frontiers in Sociology*, 8. https://doi.org/10.3389/fsoc.2023.1196672

Ellis, C. A., Gibbs, S. J., & Rein, G. (1991). Groupware: Some issues and experiences. *Communications of the ACM*, 34(1), 39–58. https://doi.org/10.1145/99977.99987

https://doi.org/10.1017/pds.2026.10589 Published online by Cambridge University Press

Fernandes, D., Garg, S., Nikkel, M., & Guven, G. (2024). A GPT-Powered Assistant for Real-Time Interaction with Building Information Models. In *BUILDINGS* (Vol. 14, Issue 8). MDPI. https://doi.org/10.3390/buildings14082499

Galland, L., Pelachaud, C., & Pecune, F. (2022). Adapting conversational strategies in information-giving human-agent interaction. *Frontiers in Artificial Intelligence*, 5. https://doi.org/10.3389/frai.2022.1029340

Haye, A., & Larrain, A. (2011). What is an utterance? In *Dialogicality in Focus: Challenges to Theory, Method and Application* (pp. 33–52). Nova Science Publishers.

Kadenhe, N., Al Musleh, M., & Lompot, A. (2025). Human-AI Co-Design and Co-Creation: A Review of Emerging Approaches, Challenges, and Future Directions. *Proceedings of the AAAI Symposium Series*, 6(1), 265–270. https://doi.org/10.1609/aaaiss.v6i1.36061

Liu, V. (2023). Beyond Text-to-Image: Multimodal Prompts to Explore Generative AI. Conference on Human Factors in Computing Systems - Proceedings, ACM SIGCHI; Apple; Bloomberg; Google; NSF; SIEMENS-. http://dx.doi.org/10.1145/3544549.3577043

Luger, E., & Sellen, A. (2016). 'Like Having a Really Bad PA': The Gulf between User Expectation and Experience of Conversational Agents. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, 5286–5297. https://doi.org/10.1145/2858036.2858288

Matthews, B., & Heinemann, T. (2012). Analysing conversation: Studying design as social action. *Design Studies, Studying Design in Practice*, 33(6), 649–672. https://doi.org/10.1016/j.destud.2012.06.008

McDonnell, J. (2012). Accommodating disagreement: A study of effective design collaboration. *Design Studies*, 33(1), 44–63. https://doi.org/10.1016/j.destud.2011.05.003

Pande, A., Patil, R., Mukkemwar, R., Panchal, R., & Bhoite, S. (2024). Comprehensive Study of Google Gemini and Text Generating Models: Understanding Capabilities and Performance.

Rayan, J., Kanetkar, D., Gong, Y., Yang, Y., Palani, S., Xia, H., & Dow, S. P. (2024). Exploring the Potential for Generative AI-based Conversational Cues for Real-Time Collaborative Ideation. *Proceedings of the 16th Conference on Creativity & Cognition*, C&C '24, 117–131. https://doi.org/10.1145/3635636.3656184

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A Simplest Systematics for the Organization of Turn-Taking for Conversation. *Language*, 50(4), 696–735. JSTOR. https://doi.org/10.2307/412243

Schmidhuber, J., Schlögl, S., & Ploder, C. (2021). Cognitive Load and Productivity Implications in Human-Chatbot Interaction (p. 6). https://doi.org/10.1109/ICHMS53169.2021.9582445

Schmidt, K., & Simone, C. (1996). Coordination mechanisms: Towards a conceptual foundation of CSCW systems design. *Computer Supported Cooperative Work (CSCW)*, 5(2), 155–200. https://doi.org/10.1007/BF00133655

Shi, Y., Gao, T., Jiao, X., & Cao, N. (2023). Understanding Design Collaboration Between Designers and Artificial Intelligence: A Systematic Literature Review. *Proceedings of the ACM on Human-Computer Interaction*, 7(CSCW2), 1–35. https://doi.org/10.1145/3610217

Shneiderman, B. (2021). Human-centered AI. *Issues in Science and Technology*, 37(2), 56–61.

Srinivasan, K., Puliyanda, A., Thosar, D., Bhakte, A., Singh, K., Addo, P., Srinivasan, R., & Prasad, V. (2024). Artificial intelligence and machine learning at various stages and scales of process systems engineering. *The Canadian Journal of Chemical Engineering*, 103, 1004–1035. https://doi.org/10.1002/cjce.25525

Stefanelli, G. (2025, June 16). Voice Function in AI Chatbots: Comparison of ChatGPT, Meta AI, Gemini, Microsoft Copilot, and Claude for Natural Voice Interactions. Data Studios Exafin. https://www.datastudios.org/post/voice-function-in-ai-chatbots-comparison-of-chatgpt-meta-ai-gemini-microsoft-copilot-and-claude

Tembine, H., Bamia, I., NDong, M., Coulibaly, B., Traore, O. I., Traore, M., Sanogo, Moussa, Sangare, M. E., Kante, S., Yongueng, D. N., Ali, H. T., Tiomoko, M., Laleye, F., Djehiche, B., Dipama, W. E., Saje, I. B., Ibrahim, H. M., Sanogo, Moumini, Nininahazwe, M. C., … Coulibaly, B. L. A. (2025). Breaking the Barriers of Text-Hungry and Audio-Deficient AI. https://doi.org/10.48550/ARXIV.2506.02443

Tversky, B. (2002). What do Sketches Say about Thinking. https://www.semanticscholar.org/paper/What-do-Sketches-Say-about-Thinking-Tversky/058b706b39debfef498fd9a51c89b1df47a63804

Tversky, B. (2010). Visualizing Thought. *Topics in Cognitive Science*, 3, 499–535. https://doi.org/10.1111/j.1756-8765.2010.01113.x

Zhang, D., Li, S., Zhang, X., Zhan, J., Wang, P., Zhou, Y., & Qiu, X. (2023). SpeechGPT: Empowering Large Language Models with Intrinsic Cross-Modal Conversational Abilities. In H. Bouamor, J. Pino, & K. Bali (Eds), *Findings of the Association for Computational Linguistics: EMNLP 2023* (pp. 15757–15773). Association for Computational Linguistics. https://doi.org/10.18653/v1/2023.findings-emnlp.1055

https://doi.org/10.1017/pds.2026.10589 Published online by Cambridge University Press