

Empathy as interactional accomplishment in clinical interactions with a conversational agent

Spencer Hazel¹, Adam Brandt¹, Yajie Vera He²,
Ernest Lim^{2,3}, Jared Joselowitz², Zachary Ellis²

¹Newcastle University

²Ufonia Limited

³University of York

¹{spencer.hazel, adam.brandt}@newcastle.ac.uk

²{vh, el, jj, ze}@ufonia.com

Abstract

As healthcare services deploy AI to automate patient-facing communication, concerns persist about the interactional work through which empathy is made relevant. We examine empathy not as an internal state but as an interactional accomplishment, asking how patients display orientations to an LLM-powered voice assistant's turns as (non-)empathic in real clinical telephone calls. Using Conversation Analysis (CA) to analyse post-cataract surgery follow-up calls conducted by AI-powered voice assistant Dora (Ufonia), we compare patient responses across earlier and later system versions.

Earlier calls show minimal, delayed, prosodically closed responses to wellbeing enquiries, consistent with treating Dora as a transactional information-gathering device. Later calls more often feature socially rich formats, for example colloquial upgrades, gratitude tokens, occasional return enquiries, and increased turn-final rising intonation, suggesting patients hear Dora's talk as socially implicative and thus opening space for affiliative/empathetic uptake. We discuss implications for CA-informed conversation design and for evaluating "empathy" via participant orientations in situ rather than post-hoc self-report.

1 Introduction

At a time when healthcare systems around the world are under pressure to increase productivity and deliver more efficient services, many organisations are turning to AI-driven tools to support everyday clinical work. While AI is widely used in areas such as diagnostic support, it is increasingly being introduced to assist with communication tasks, including interactions that traditionally require substantial staff time. Tools such as chatbots or voice-based conversational agents can offer immediate relief to clinical teams who are struggling to meet the demands of patient-centred care. However, automating patient-facing communication also intro-

duces risks. An AI system may be able to gather information or complete transactional tasks, yet removing human clinicians from these interactions can affect the overall quality of care that patients experience and expect.

While these transactional capabilities meet technical goals, they often fail to replicate the relational work of human clinicians, particularly the demonstration of empathy, here *interactionally displayed and oriented-to* rather than *felt*, in telephone-based encounters. The challenge of automating empathy is compounded by the fact that interactional practices for affiliation are not universal but are socialised within specific linguacultural groups (Brandt and Hazel, 2025). One area where this is particularly noticeable concerns empathy in telephone-based clinical encounters. Human clinicians routinely demonstrate forms of understanding and affiliation that support patient trust and engagement, whereas a machine is incapable of feeling or responding to patients' disclosures in the same way.

The challenge of automating such relational work is rooted in the complex "interaction-implicative norms" that govern human affiliation (Brandt and Hazel, 2025). We have previously argued that naturalness in voice user interfaces is achieved by basing design on empirical observations of human-human interactional patterns (Hazel and Brandt, 2023). However, as conversational agents move beyond performing basic transactional tasks and toward emulating empathetic clinical interaction, we must account for what we call the "capability" dynamic, namely the system's ability to meet the interactional demands of the encounter and the user's ability to adapt to the limitations of the system. Mismatches on either side constrain the interaction until the system's functionality or the user's interactional competence develops sufficiently to overcome these limits (Brandt and Hazel, 2025).

This paper examines this issue through our work on Dora, an LLM-powered voice-based conversational assistant designed to carry out routine telephone calls to patients on behalf of hospital teams in the United Kingdom and North America. The project brings together Conversation Analysts from the field of Applied Linguistics and conversation designers at the developer to explore how empathy is displayed and perceived in human clinical interactions, and how these practices might inform the development of an automated agent for comparable tasks.

Our analysis focuses on how users respond to Dora's turns during particular sequences of talk. These responses allow us to identify the kinds of conversational moves that are treated by users as socially engaged, even though they are produced by an automated system. Elsewhere in the related field of human-robot interaction, Pelikan et al. (2020, p.461) argue that from a Conversation Analytic perspective, "robotic emotions exist as soon as participants treat robots as having them". Following the same principle, seeking to understand social interaction through the participants own internal logic (CA's *emic* approach), our analysis seeks to identify where users, in the formatting of their turns-at-talk, display an orientation to the system as user/patient-centred rather than simply transactional. In doing so, we address how interactional practices shape the experience of empathy in human-machine communication and what this means for designing conversational AI in healthcare settings.

2 The turn to LLMs in conversational agents for institutional interaction

The widespread adoption of Large Language Model (LLM) technology to power conversational agents continues apace. LLMs are increasingly performing the role previously reserved for conversation designers, namely that of determining the speech output of the automated system. The rationale behind this move is that an LLM used well can be more responsive to users and lend the interaction a richer conversational user experience. How this compares with previous deterministic NLU-powered models is still relatively unexplored, including how the greater versatility in system output impacts the talk produced by the users.

Institutional talk comes in many guises, and participants to a social interaction enact the institutionality of the encounter into being through how they

modify their contributions (Heritage and Greatbatch, 1991). This includes a respecification of baseline conversational practices, orienting to normative choices in e.g., speakership rights, lexical choices, turn-formatting and turn-taking models. It is the participants' formatting choices in their interactions with one another that make a news interview recognisable as a news interview, a clinical consultation recognisable as that and not something else, a service encounter in a coffee bar recognisably different from a help-desk encounter at a local government office.

Some institutional interactions allow participants some freedom to introduce off-task elements into the conversation without this threatening to disrupt the institutionality of the exchange. Service encounters in bars or hair salons, for example, can include quite extensive forays into often quite intimate sharing of personal details, ones that have little relevance to the service transaction. Elsewhere, institutional interaction may reserve phases where 'small talk' is oriented to as appropriate, for example in clinical interactions where general enquiries into wellbeing for example may precede a move into the clinical phase of the encounter.

Institutional activities may have the goal of eliciting more precise information from the client, and here greater versatility in user responses may present problems, both to a human or to an automated system, with relevant information more difficult to glean from the client's talk. In such activities, ensuring the client provides information in clear, simple formats requires the provider to keep the client on task. However, greater system responsiveness may enhance the sense of conversational flow and therefore improve the user experience.

Considering automated conversational systems, early empirical research suggests that LLM-powered voice assistants offer great potential and additional conversational affordances (Ramaul et al., 2024), and "advanced conversational skills and versatility" (Kim et al., 2024). However, findings from such preliminary studies are not based upon users' conduct in their real-life engagements with conversational agents. This paper addresses this gap by examining whether users themselves *treat* a system as conversational, by examining their interactional behaviour in situ. This makes LLM-based agents a particularly consequential site for studying how 'empathy' is interactionally inferred, because increased responsiveness can alter what patients treat as expected next actions.

In this paper, we report on initial findings from a Conversation Analytic study of an LLM-powered voice-based conversational agent, which has been developed by a digital health start-up based in the UK. This automated clinical assistant performs routine calls on behalf of clinical teams in healthcare settings and is already in use in hospitals across the UK. Research has provided evidence for the safety (Meinart et al., 2024), patient acceptability (Khavandi et al., 2020) and feasibility (Hatamnejad et al., 2024) of Dora. Further development of the LLM-powered version is presently in progress in the UK, and clinical trials are underway in North America. Here, we present preliminary findings of analysis of those trial calls, building on previous research on the model (Brandt et al., 2024).

To assess whether the increased responsiveness from LLM adoption is consequential *for participants*, we use Conversation Analysis to examine how patients' next turns display their understanding of Dora's prior turns.

3 Conversation Analysis for Conversational Agents

The majority of Human-Computer Interaction (HCI) research relies on experimental settings and/or self-report data to analyse interactions between human users and AI-powered conversational agents. That is, users are either research participants placed into contrived scenarios and/or are post-hoc subjected to interviews or surveys, to gain insight of their perceptions of their experience.

To complement this work, and among other discourse and linguistic analysis approaches, Conversation Analysis (CA) has emerged as a means of analysing human use of conversational technologies in real-world, non-experimental, settings (Mlynář et al., 2025; Stokoe et al., 2024). Much of this research to date has focused on social robots, chatbots and conversational agents that are based on a deterministic model of pre-determined speech outputs scripted by a conversation designer.

The findings of such studies suggest that users do not typically treat such systems as 'humanlike' or 'natural conversation'; rather, there is evidence that users adapt their talk to the perceived abilities of the system. For example, in their analysis of use of Google Home devices, Due and Lüchow (2024) describe user utterances delivered to the devices as 'VUI-speak' (referring to the acronym for voice user interfaces). Similarly, Avgustis et al.

(2021) suggest that users' utterances to a telephone-based voice agent for local governmental services in Russia are the spoken equivalent of search engine entries.

At present, we are aware of few studies which have employed CA to examine human-LLM voice interactions (e.g., Albert et al., 2025; Ivarsson and Lindwall, 2023). In both cases, data analysed were publicly available recordings of appointment booking telephone calls made by Google's now defunct personal assistant, Duplex (Leviathan and Matias, 2018). In hearing these calls, there is no evidence that the human interlocutor is even aware that they are speaking to a conversational agent, let alone adapting their talk accordingly. However, there have been doubts about the authenticity of these demo Duplex calls (Natale, 2021).

4 The study

In our current study, we examine an LLM-powered voice-based conversational agent. The context of our human-agent telephone calls is the healthcare sector, and is non-experimental (that is, the human-agent interactions have not been set up for the purpose of this study). For this study, we have analysed clinical trial and user testing telephone calls between patients/testers and Dora.

The context of the clinical trial calls are post-cataract surgery follow-ups. The patients have recently had cataract surgery, and the purpose of a follow-up call is to check the patient's recovery after the operation. Since these are clinical trial calls, no diagnosis is made as a result of the call, and the patient will subsequently have their usual clinician-led appointment.

5 Methodology

The analysis follows the trajectory of our previous work, which established the CADENCE (Conversation Analytic Design for Enhanced Natural Conversation Experience) model as a framework for using human-human interactional patterns to inform the design of voice-based agents (Hazel and Brandt, 2023). Following multiple repeated listens to the calls, and following the CA principle of 'unmotivated looking', an area of interest emerged in the complexity of patients' responses to Dora's questions in later versions of the model, as we outline in the analysis below.

Those segments of the calls were transcribed according to the Jeffersonian transcription conven-

tions (Jefferson, 2004, - see table at the end of this section for overview)) and analysed using CA methodology (Sacks et al., 1974). This involves a turn-by-turn sequential analysis, focusing on how each utterance is designed, and how each turn-at-talk displays an understanding of the preceding turn, while generating grounds for a relevant next turn. The extracts presented here have been selected to be analytically representative rather than statistically representative, in line with established Conversation Analytic practice. This eschews quantification of conversational patterns, rather adopting an emic approach which foregrounds how participants themselves treat one another’s interactional contributions as normative within the unfolding interaction.

Symbol	Meaning
↑	Marked pitch rise
?	Rising (questioning) intonation
.	Falling (final) intonation
,	Continuing (level) intonation
(0.5)	Pause length in seconds
(.)	Micro-pause (< 0.2 seconds)
::	Sound stretch
hh	Audible outbreath
()	Uncertain/unclear hearing

6 Analysis

To avoid treating LLM ‘empathy’ as a model property inferred from surveys, we examine participant orientations in situ, using sequential evidence from real clinical trial calls. Previously, Brandt et al. (2024) have discussed how user-patients can be found to provide much more truncated responses to Dora questions than when compared with clinician-patient interactions. Moreover, the responses are more delayed. Taken together, they argued that users modified their contributions according to the level of trust they had in the ability of the system to uphold standard conversational practices.

This is illustrated in the following examples, taken from the openings of calls between a previous model of the Dora system (noted as SYS) and patient-users (noted as PAT).

EXCERPT 1

29 SYS: †how are you feeling today.
30 (1.4)
31 PAT: good, hh

EXCERPT 2

32 SYS: †how are you feeling today.

33 (1.8)

34 PAT: good,

EXCERPT 3

51 SYS: †how are you feeling today.

52 (1.6)

53 PAT: i’m okay,

EXCERPT 4

15 SYS: †how are you feeling today.

16 (0.2) overall.

17 (1.8)

18 PAT: good.

EXCERPT 5

19 SYS: †how are you feeling today.

20 (1.0)

21 PAT: good.

EXCERPT 6

29 SYS: †how are you doing today.

30 (0.3)

31 SYS: mister pinter.

32 (0.5)

33 PAT: i’m okay thank you.

EXCERPT 7

14 SYS: †how are you feeling today.

15 (1.6)

16 PAT: feeling better(h).

EXCERPT 8

28 SYS: †how are you doing to[day.]

29 PAT: [yes]

30 (1.1)

31 PAT: (yeah) great,

EXCERPT 9

29 SYS: †how are you feeling today.

30 (1.0)

31 PAT: (oh) very good,

EXCERPT 10

25 SYS: †how are you feeling today.

26 (1.2)

27 PAT: i’m feeling good.

Across all these examples, we note how user responses to Dora’s question are heavily delayed and display minimal response formatting. This is not at all considered problematic in a transactional sense, though it does evidence a lack of conversational flow and limited input from the user.

We note how there is some variation in the question formatting across cases. Where Dora uses the term ‘feeling’, overwhelmingly we find the client responding with ‘good’ or some variant (“I’m feeling good”; “very good”) and one occasion here a “feeling better”. Where Dora uses ‘doing’ in the formatting, we find the clients responding with “I’m okay thank you” and “yeah great”. Turn final

intonation for all responses was either falling or continuing (i.e. flat intonation).

As the Dora user interface has continued to develop, user engagement with the system has evidenced some change. In these later versions, the system talk is produced using a different speech synthesis voice. The voice and tone are hearable as more natural-sounding in their timing and prosody, there is greater variation in intonation, and issues around latency have improved. In the following, we provide a number of excerpts to illustrate how users provide more socially rich responses.

EXCERPT 11

01 SYS: ↑so (.) how are you doing
02 today.
03 mister baker,
04 (1.7)
05 PAT: okay. apart from dry eyes?

EXCERPT 12

01 SYS: okay? (.) ↑how are you doing
02 today.
03 (1.2)
04 PAT: ↑not too bad at all. thank
05 you?

EXCERPT 13

01 SYS: ↑great (0.2) so. how are you
02 doing today missus fisher?
03 (1.7)
04 PAT: great thank you?

What should be immediately apparent is the greater complexity of the client responses. Here, rather than limit the answer to its simplest form, a minimal response to the question, we find speakers produce post-expansions and more complex formatting features in their turns-at-talk. Excerpt 11 sees the patient add a qualifier to the answer, leaving it to the system to infer from the two pieces of information what to log. Excerpt 12 sees the patient providing a colloquially formatted response, suggesting the speaker trusts the system to be able to recognise the intent embedded in the colloquialism.

The last two excerpts also embed a social implicative pragmatic feature of gratitude tokens, suggesting speakers are treating Dora's question as a general enquiry into their wellbeing, rather than an information seeking question relevant to the clinical context of the call. We see this also elsewhere, for example in the following response:

EXCERPT 14

03 PAT: ↑oh I'm really good actually
04 today.
05 thank you for asking?

This evidences an orientation to the social nature of the interaction, and that normative expectations around how to respond to enquiries like this are relevant to the formatting of the talk. Present in the data are also instances of clients expanding on the response with a return enquiry: asking Dora how it is doing.

EXCERPT 15

03 PAT: ↑I'm absolutely fine thank
04 you?
05 how are you.

A further formatting feature found across these data is the turn final intonation contour adopted across these cases. Where previously we found falling or continuing intonation marking the end of the client's response, here we find overwhelmingly a rising intonation.

EXCERPT 16

03 PAT: ↑very well thank ↑you?

EXCERPT 17

03 PAT: ↑I'm all right. thank you::?

In previous cases, the falling or continuing turn-final intonation was hearable as marking the end of information delivery, effectively treating the system as an information logging device. However, the shift in pattern of intonation that we see in the later Dora model suggests that users are increasingly modelling their speech on patterns they would use in human-human interaction. In these contexts, a rising intonation contour can index an orientation to the social interactional dimension of the exchange, withholding sequence closure and inviting further talk.

If this is also the case here, then a range of social interaction implicative norms are arguably being brought into play, including those which index social affiliation. As a consequence, a sense that the users are interacting with the system in ways that allow them to experience the engagement as mutually social affiliative, also unlocks the possibility for them to feel they are participating in an *empathetic* encounter.

7 Conclusion

In this study, we treat empathy not as an internal state but as an interactional achievement, and we have shown how users' next turns provide evidence of whether Dora's enquiries are being heard: as merely information-seeking, or as socially implicative moves that make affiliation possible. Across earlier trial versions of Dora, responses

to “how are you feeling/doing today” were predominantly delayed, minimal, and produced with falling/continuing intonation, consistent with patients orienting to Dora as a constrained, transactional device. In later versions of the system, patients more frequently produced expanded and socially rich responses, including colloquial formulations, gratitude tokens (“thank you”, “thank you for asking”), occasional return enquiries (“how are you”), and a marked shift toward the use of turn-final rising intonation. These features are hearable as calibrating the exchange toward human–human norms of reciprocity and social engagement. Taken together, these patterns suggest that as Dora becomes interactionally more nuanced as a conversational system, the interaction opens up a broader space in which empathy can be ascribed to the system and oriented to by patients.

This study does not attempt to isolate the causal contribution of any single system component. Changes between system versions were introduced as bundled updates (including synthesis voice, latency, underlying LLM, and prompt design), reflecting the realities of iterative development in deployed technologies rather than a controlled experimental design. As a result, the analysis cannot attribute observed behavioural differences to any one technical modification.

Our contribution is therefore descriptive rather than causal. We document observable shifts in patient conduct that co-occurred with successive system updates, without claiming that these shifts were produced by specific components. Future research could address this limitation through A/B testing or factorial experimental designs that isolate individual variables. However, implementing such designs was beyond the scope of the present clinical trial context.

8 Ethics statement

The data for this study comprise real-world clinical follow-up calls between trial patients, human testers, and an automated conversational agent. All patients enrolled in the clinical trial provided written informed consent that explicitly covered the use of anonymized call recordings for research purposes, including academic publication. In addition, during the calls, patients provided documented verbal consent and confirmed that they were happy for their anonymized data to be used for research and training purposes.

To protect participant privacy, all recordings were anonymized in line with Information Commissioner’s Office (ICO) guidance. Specific geographic and institutional identifiers were removed and replaced with generic placeholders. The study was conducted within a clinical trial in which the AI agent did not make primary diagnoses; all patients subsequently received a clinician-led appointment to ensure safety and quality of care. The researchers operated under strict data-sharing agreements between the system developer and the participating healthcare providers.

9 Limitations

This study presents preliminary findings from a specific medical context: post-cataract surgery follow-ups. We note that the patterns of “interactional accomplishment” observed here may not generalize to more acute clinical settings or different linguistic groups where interactional norms for affiliation differ. Furthermore, while we observe a shift in patient orientation based on system updates, we are not claiming empathy is being produced by the system, only that the participants orient to the system in more socially affiliative ways in the interaction.

The clinical context examined in this study, namely post-cataract follow-up calls, represents a relatively routine, low-stakes setting. Patients are typically in a recovery phase rather than experiencing acute distress, and the calls are primarily informational rather than diagnostic. As such, the interactional demands placed on patients and the system may differ substantially from those found in higher-stakes clinical environments. In settings such as oncology consultations, mental health triage, or emergency services, interactions are likely to carry greater emotional valence and place different demands on expressions and interpretations of empathy. The patterns of patient conduct documented here may therefore not transfer directly to contexts where uncertainty, urgency, or distress are more prominent features of the encounter.

Future research should examine whether similar interactional patterns emerge in (a) urgent or acute clinical calls, (b) clinical calls where greater emotional distress is in evidence, and (c) longitudinal designs that track patient conduct across repeated interactions with the same system over time. At the same time, the relative stability of the post-operative context may be considered a

methodological advantage for the present study, as it provides a comparatively controlled baseline in which variables such as emotional distress and diagnostic uncertainty are minimized, allowing shifts in patient conduct to be observed with fewer competing influences.

Acknowledgments

This project is supported by the UK Research and Innovation, Innovate UK Biomedical Catalyst grant (project 10102111).

References

- Saul Albert, William Housley, Rein Ove Sikveland, and Elizabeth Stokoe. 2025. The conversational action test: Detecting the artificial sociality of artificial intelligence. *New Media & Society*, 27(10):5592–5621.
- Juliia Avgustis, Aleksandr Shirokov, and Netta Iivari. 2021. “Please connect me to a specialist”: Scrutinising ‘recipient design’ in interaction with an artificial conversational agent. In *Proceedings of the 18th IFIP TC 13 International Conference on Human-Computer Interaction (INTERACT 2021)*, pages 411–431. Springer.
- Adam Brandt and Spencer Hazel. 2025. Towards interculturally adaptive conversational ai. *Applied Linguistics Review*, 16(2):775–786.
- Adam Brandt, Spencer Hazel, Rory McKinnon, Kleopatra Sideridou, Joe Tindale, and Nikoletta Ventoura. 2024. Educating Dora: Teaching a conversational agent to talk. *Discourse and Communication*, 18(6):905–916.
- Brian L. Due and Louise Lüchow. 2024. VUI-speak: There is nothing conversational about “conversational user interfaces”. In F. Muhle and I. Bock, editors, *Communicative AI in (Inter-)Action: Investigating Human-Machine Encounters Outside the Laboratory*, pages 155–177. Bielefeld University Press.
- Amin Hatamnejad, Aisling Higham, Sohel Somani, Eric S Tam, Ernest Lim, Sarah Khavandi, Nick de Pennington, and Hannah H Chiu. 2024. Feasibility of an artificial intelligence phone call for post-operative care following cataract surgery in a diverse population: two phase prospective study protocol. *BMJ Open Ophthalmology*, 9:e001475.
- Spencer Hazel and Adam Brandt. 2023. Enhancing the natural conversation experience through conversation analysis – a design method. In *HCI International 2023 – Late Breaking Papers. HCII 2023. Lecture Notes in Computer Science*, volume 14054. Springer, Cham.
- John Heritage and David Greatbatch. 1991. On the institutional character of institutional talk: The case of news interviews. In D. Boden and D. H. Zimmerman, editors, *Talk and Social Structure: Studies in Ethnomethodology and Conversation Analysis*, pages 93–137. Polity Press.
- Jonas Ivarsson and Oskar Lindwall. 2023. Suspicious minds: The problem of trust and conversational agents. *Computer Supported Cooperative Work (CSCW)*, 32(3):545–571.
- Gail Jefferson. 2004. Glossary of transcript symbols with an introduction. In G. H. Lerner, editor, *Conversation Analysis: Studies from the First Generation*, pages 13–31. John Benjamins.
- Sarah Khavandi, Ernest Lim, and Guy Mole. 2020. 110 Patient acceptability of telephone follow up after cataract surgery. *BMJ Leader*, 4:A41–A42.
- Callie Y. Kim, Christine P. Lee, and Bilge Mutlu. 2024. Understanding large-language model (LLM)-powered human-robot interaction. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction (HRI '24)*, pages 384–393.
- Yaniv Leviathan and Yossi Matias. 2018. [Google Duplex: An AI system for accomplishing real-world tasks over the phone](#). Google AI Blog.
- Edward Meinart, Madison Milne-Ives, Ernest Lim, Aisling Higham, Selina Boege, Nick de Pennington, Mamta Bajre, Guy Mole, Eduardo Normando, and Kanmin Xue. 2024. Accuracy and safety of an autonomous artificial intelligence clinical assistant conducting telemedicine follow-up assessment for cataract surgery. *eClinicalMedicine*, 73:102692.
- Jakub Mlynář, Lynn De Rijk, Andreas Liesenfeld, Wyke Stommel, and Saul Albert. 2025. AI in situated action: a scoping review of ethnomethodological and conversation analytic studies. *AI & society*, 40(3):1497–1527.
- Simone Natale. 2021. *Deceitful Media: Artificial Intelligence and Social Life after the Turing Test*. Oxford University Press.
- Hannah RM Pelikan, Mathias Broth, and Leelo Keavallik. 2020. “Are you sad, cozmo?” How humans make sense of a home robot’s emotion displays. In *Proceedings of the 2020 ACM/IEEE international conference on human-robot interaction*, pages 461–470.
- Laavanya Ramaul, Paavo Ritala, and Mika Ruokonen. 2024. Creational and conversational AI affordances: How the new breed of chatbots is revolutionizing knowledge industries. *Business Horizons*, 67(5):615–627.
- Elizabeth Stokoe, Saul Albert, Hendrik Buschmeier, and Wyke Stommel. 2024. Conversation analysis and conversational technologies: Finding the common ground between academia and industry. *Discourse and Communication*, 18(6):837–847.