Proceedings Paper:
Jones, A. and Moore, R.K. (2025) HRI-based interview training using the FurHat robot. In: Staffa, M., Cabibihan, J.-J., Siciliano, B., Ge, S.S., Bodenhagen, L., Tapus, A., Rossi, S., Cavallo, F., Fiorini, L., Matarese, M. and He, H., (eds.) Social Robotics + AI. International Conference on Social Robotics + AI 2025, 10-12 Sep 2025, Naples, Italy. Lecture Notes in Computer Science, 16131 (Part 1). Springer Nature Singapore, pp. 584-589. ISBN: 9789819523788. ISSN: 0302-9743. EISSN: 1611-3349.

https://doi.org/10.1007/978-981-95-2379-5_50

HRI-based Interview Training using the FurHat Robot

Adam Jones and Roger K. Moore[0000−0003−0065−3311]

School of Computer Science, University of Sheffield, Sheffield S1 4DP, UK
{ajones29,r.k.moore}@sheffield.ac.uk
https://www.sheffield.ac.uk/cs

Abstract. Navigating the pressures of job interviews can present a significant challenge to an individual as their preparation can critically influence the outcome. While traditional face-to-face (F2F) methods for interview training are still the norm, the evolution of 'artificial intelligence' (AI) presents new opportunities, and a variety of virtual AI-based interview platforms exist to satisfy the need. However, while such platforms offer a wide range of features and broad coverage of interview-style questions, they often lack the physical presence and the capacity for real-time emotional feedback or adaptive response tailoring that is characteristic of real-world F2F scenarios. As a consequence, there is growing interest in the use of physical robots for job interview simulation. This paper presents the development of an advanced AI-based 'Interactive Interview Training System' (IITS) utilising the FurHat robot which goes beyond the feature sets of existing virtual and physical agents to create a more holistic and effective job interview simulation. In particular, FurHat has been configured to allow extensive customisation by the user, such as the ability to define job scenarios from different industry backgrounds and invoke a range of interaction styles. Also, since FurHat's Software Development Kit (SDK) incorporates a virtual animation of the robot, it has been possible to conduct a user-based evaluation which confirmed the merits of physical embodiment in HRI-based interview training.

Keywords: Interview training · FurHat robot · Virtual/physical presence.

1 Introduction

Navigating the pressures of job interviews presents a significant challenge as an individual's preparation can critically influence the outcome. While traditional face-to-face (F2F) methods are still the norm, the evolution of technology – in particular, 'artificial intelligence' (AI) – presents new opportunities for interview training. For example, in a study on using AI to combine exam and interview preparation, participants reported enhanced learning and found the tool's feedback accurate in preparing them for job interviews [4].

A variety of virtual AI-based interview platforms exist (e.g. Yoodli[1] and HireVue[2]). However, while such platforms offer comprehensive features and broad coverage of interview-style questions, they often lack the physical presence and capacity for real-time emotional feedback or adaptive response tailoring that is characteristic of real-world F2F scenarios. Hence, there is interest in the use of physical robots, such as the android ERICA [5], for job interview simulation.

Although systems such as ERICA provide structured and elaborate follow-up questions, they are still relatively rigid and offer limited adaptability or opportunities for fine-tuning. This paper presents the development of an advanced AI-based interviewer based on the FurHat robot [1] that goes beyond the feature sets of virtual platforms like Yoodli and HireVue and the embodied interaction provided by physical agents such as ERICA, to create a more holistic, adaptive and effective job interview simulation. In particular, FurHat has been configured to allow extensive individual customisation by the user, such as the ability to define job scenarios from different industry backgrounds and invoke a range of interaction styles. Also, since FurHat's Software Development Kit (SDK) incorporates a virtual animation of the robot, it has been possible to conduct a short investigation into the relative merits of physical versus virtual embodiment in HRI-based interview training.

2 Motivation

The dynamics of F2F interviews are fundamentally shaped by the physical co-presence of the interviewer and interviewee [3]. This shared physical space allows an interviewer to maintain focus and exert control over the direction and flow of the interview in comparison with mediated methods. A key advantage of this in-person dynamic is the enhanced potential for building rapport; F2F settings enable interviewers to better establish the connection needed to motivate and encourage sustained and detailed engagement from the interviewee.

This personal connection is supported by the frequent use of non-verbal communication inherent in F2F interactions. Important non-verbal cues, such as facial expressions conveying emotion or nods signalling understanding, are often observable F2F but are inevitably missed in many computer-mediated interactions. These cues not only add layers of meaning to the verbal exchange but can actively encourage informants to elaborate on their thoughts.

The immediacy of F2F interaction also allows for responsive and timely probing. Interviewers can ask clarifying or follow-up questions based directly on the unfolding conversation and the interviewee's real-time verbal and non-verbal responses, a process inherently different from the potentially delayed probing in asynchronous communication. Consequently, F2F interviews often capture more spontaneous expressions, with the natural hesitations and conversational fillers (e.g., "*ums*", "*errs*") characteristic of spoken language, reflecting a less reflected stream of thought compared to carefully constructed written replies.

[1] https://yoodli.ai
[2] https://www.hirevue.com/

2.1 AI-based Interview Platforms

To provide some form of assessment of a user's suitability for a role, AI-based interview platforms must evaluate various aspects of their behaviour. Interview practice platforms such as Yoodli provide a variety of feedback insights for the *user*, including measures such as conciseness, key points, filler words, weak words, sentence starters, pacing, eye contact, centring and pauses. In contrast, platforms such as HireVue provide analytics for the *employer*, evaluating the interviewee's hard and soft skills, as well as facial mannerisms to output a final score.

Visual and audio feature extraction is thus important in providing feedback to the user and the employer. Visual extraction can include categories such as emotion, head pose, eye contact and gestures, and audio extraction contains categories such as speaking rate, amplitude and tone [2]. Some researchers have even investigated the use of biometric data to detect subtle cues [6].

2.2 The FurHat-based Interviewer

The FurHat-mediated 'Interactive Interview Training System' (IITS) reported here employed a state-driven approach, utilising the FurHat SDK's flow capabilities, augmented by object-oriented principles in Kotlin. Core intelligence was provided by integrating OpenAI's 'ChatGPT' via a dedicated service layer, enabling dynamic question generation and response handling, thereby overcoming the limitations inherent in static question banks [8]. The main design features were:

Bridging the Physical Presence and Rapport Gap A primary shortcoming of virtual agent-based platforms is their inability to replicate the crucial physical co-presence fundamental to F2F interviews. The work reported here leveraged FurHat's physical embodiment to simulate this sense of shared space. The robot's capacity for realistic human-like features, including nuanced facial expressions and dynamic gaze control, replicates the vital non-verbal cues of F2F settings, aiming to create a more engaging interaction and develop a sense of connection that screen-based agents cannot achieve [7].

Enhancing Customisation Many existing interview practice platforms offer limited scope for tailoring the scenario. FurHat was configured to address this through multi-faceted customisation. First, integration with ChatGPT enabled deep pre-prompting, allowing users to define complex interview contexts, specific roles, industry nuances, and interviewer personalities (e.g., formal, stressed, casual) far beyond typical menu-driven options. Second, FurHat's customisable facial features and voice synthesis permitted visual and auditory alignment with the desired scenario, with a view to increasing relatability and perceived realism.

Enabling Real-Time Adaptation A significant limitation of many interview practice tools is their focus on post-session feedback rather than dynamic interaction. They may analyse gaze or speech patterns afterwards, but often fail to react during the conversation. In contrast, the system reported here is capable of adapting its questioning or non-verbal behaviour in real-time based on the user's input, creating a more dynamic and responsive dialogue akin to a human interviewer.

Incorporating Emotional Recognition and Response Effective human interaction involves recognising and responding to emotional cues, an area often underdeveloped in interview simulation tools. This project uniquely integrated ChatGPT's inferential capabilities directly into FurHat's response generation loop. This enabled the LLM to infer the user's possible emotional state from their language, and for FurHat to respond appropriately through both tailored dialogue and synchronised, emotionally relevant facial expressions and gestures.

3 Experiment

Setup The central aim of the experiment reported here was to investigate the differences in user perception, interaction behaviour, and perceived effectiveness between a physical robot and a virtual agent in the context of job interview training. This was facilitated by running the same IITS on either a virtual FurHat (as displayed in its SDK) or the the physical robot itself – see Fig. 1.

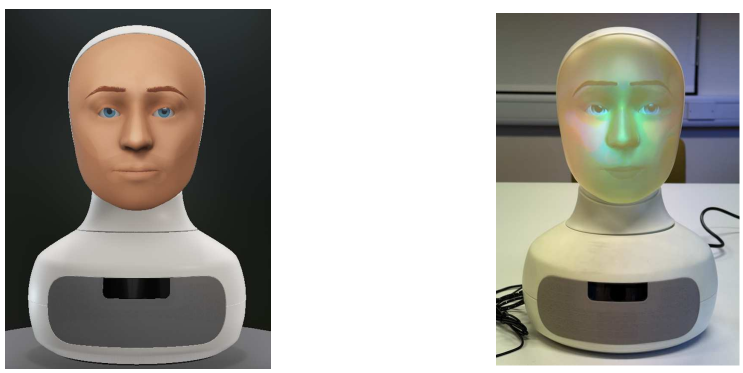

Fig. 1. Virtual FurHat display (on the left) and physical FurHat robot (on the right).

The dependent variables included (i) six *subjective* user experience metrics (measured on a 7-point Likert scale using post-condition questionnaires), (ii) five *objective* interaction measures (automatically logged by the interview system) and (iii) a 5-point 'System Usability Scale' (SUS). In addition, qualitative user feedback was gathered via an open-ended questionnaire.

Participants The participants were aged 18 or over, primarily university students interested in improving their interview skills. The inclusion criteria were (i) an ability to participate in a mock interview conducted in English and (ii) a willingness to interact with both a virtual interface and a physical robot. A total of 10 participants were recruited in the limited time available to conduct the experiment.

3.1 Results

The results ($N = 10$) are presented in Fig. 2 and Table 1. As expected, they indicated a significantly more positive opinion for the physical robot condition compared to the virtual interface with regard to perceived realism, engagement, helpfulness, overall satisfaction, and usability (SUS score). Objective metrics also suggested participants spoke longer and used fewer fillers with the physical robot.

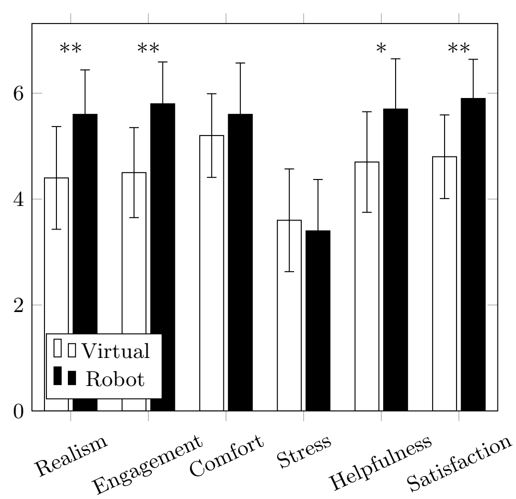

Fig. 2. Participants rated the physical robot significantly higher than the virtual interface on realism, engagement, helpfulness and satisfaction.

Table 1. Participants used significantly fewer fillers and spoke significantly longer utterances when interacting with the Physical Robot compared to the Virtual interface. The System Usability Scale (SUS) score was also significantly higher.

	Virtual	**Robot**
Total Fillers	9.80 (± 2.10)	8.80 (± 2.82) *
Min Duration (s)	5.30 (± 0.68)	5.21 (± 0.70)
Max Duration (s)	95.42 (± 13.48)	102.66 (± 16.25) **
Avg Duration (s)	35.86 (± 4.05)	37.62 (± 3.85) **
Total Duration (s)	295.07 (± 34.87)	312.93 (± 34.41) **
SUS Score	68.50 (± 17.05)	86.25 (± 12.43) ***

Qualitative feedback gathered from participants immediately after the experiment consistently highlighted areas for improvement in the physical robot implementation. In particular, the artificial quality of the robot's voice and occasional unnaturalness in its movements. Although, comfort and stress levels were not found to be significantly different between conditions, comments suggested the physical robot induced a slightly higher, *potentially beneficial*, level of alertness for practice.

4 Conclusion

This paper reports on the successful development of a FurHat-mediated interactive interview training system leveraging modern LLM capabilities to move beyond simpler command-based or scripted dialogues. Notwithstanding the small sample size, empirical evaluation demonstrated that physical embodiment significantly enhances user experience compared to a virtual interface. Participants found the physical robot more realistic, engaging, helpful, and ultimately more satisfying for interview practice. Hence, this work highlights the considerable potential of embodied conversational agents like FurHat as effective tools for professional skills development.

Acknowledgments. This study was conducted by A. Jones (supervised by Prof. R. K. Moore) as part of his BSc. degree in Computer Science at the University of Sheffield.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Al Moubayed, S., Skantze, G., Beskow, J.: The furhat back-projected humanoid head–lip reading, gaze and multi-party interaction. International Journal of Humanoid Robotics **10**(1) (March 2013), https://www.furhatai.com
2. Chou, Y.C., Wongso, F.R., Chao, C.Y., Yu, H.Y.: An ai mock-interview platform for interview performance analysis. In: 10th International Conference on Information and Education Technology. pp. 37–41 (April 2022)
3. Curasi, C.F.: A critical exploration of face-to face interviewing vs. computer-mediated interviewing. International Journal of Market Research **43**(4), 1–13 (2001)
4. Fulk, H.K., Dent, H.L., Kapakos, W.A., White, B.J.: Doing more with less: Using ai-based big interview to combine exam preparation and interview practice. Issues in Information Systems **23**, 204–217 (2022)
5. Inoue, K., Hara, K., Lala, D., Yamamoto, K., Nakamura, S., Takanashi, K., Kawahara, T.: Job interviewer android with elaborate follow-up question generation. In: 22nd ACM International Conference on Multimodal Interaction (ICMI 2020). pp. 1–9 (October 2020)
6. Lee, B.C., Kim, B.Y.: Development of an ai-based interview system for remote hiring. International Journal of Advanced Research in Engineering and Technology (IJARET) **12**, 654–663 (March 2021)
7. Li, J.: The benefit of being physically present: A survey of experimental works comparing copresent robots, telepresent robots and virtual agents. International Journal of Human-Computer Studies **77**, 23–37 (May 2015)
8. Rao S B, P., Agnihotri, M., Babu Jayagopi, D.: Improving asynchronous interview interaction with follow-up question generation. International Journal of Interactive Multimedia and Artificial Intelligence **6**(4), 79 (February 2021)