Enhancing Doctor-Patient Communication in Surgical Explanations: Designing Effective Facial Expressions and Gestures for Animated Physician Characters

Abstract

Paying close attention to facial expressions, gestures, and communication techniques is essential when creating animated physician characters that are realistic and captivating when describing surgical procedures. This paper emphasizes the integration of appropriate emotions, co-speech gestures when medical experts explain the medical procedure, and designing animated characters. We can achieve healthy doctor-patient relationships and improvement of patients' understanding by depicting these components truthfully. We suggest two critical approaches to developing virtual medical experts by incorporating these elements. First, doctors can generate the contents of the surgical procedure with a virtual doctor. Second, patients can listen to the surgical procedure described by the virtual doctor and ask if they have any questions. Our system helps patients by considering their psychology and adding medical professionals' opinions. These improvements ensure the animated virtual agent is comforting, reassuring, and emotionally supportive. Through a user study, we evaluated our hypothesis and gained insight into improvements.

Keywords: healthcare, virtual agents, surgery, computer animation, HCI

1 Introduction

Researchers have actively studied medical chatbot systems with artificial intelligence in various medical fields [1]. We would like to expand the discussion on the medical chatbot system with our suggestions. Medical chatbot systems are playing a positive role in healthcare communication. We utilized these systems to provide patients with complex medical knowledge in a friendly and accessible manner. We integrated a platform for generating content on medical procedures to create a new framework. This approach aims to bridge the gap between the complexity of medical science and the understanding of patients' anxiety. It is essential to ensure that medical information is transmitted in a clear and user-friendly way. The utilization of computer-generated animations in medical education and healthcare communication is experiencing considerable popularity [2]. These animations can effectively convey complex medical information and provide a visually engaging and accessible medium for patients and medical professionals [3]. One critical aspect of creating an animated physician character for surgical explanations is designing appropriate facial expressions and gestures that accurately reflect real-life doctor-patient interactions [4]. We propose a production tool that allows medical professionals to generate surgical processes. Users can generate screenplay format and surgical images or videos for the system rather than simply integrating computer-generated animation into chatbot systems. Subsequently, as the virtual agent automatically speaks lines and performs co-speech gestures, the system concurrently displays images and videos. After the description of the surgical procedure is completed, the patient can request medical information from the virtual agent. In this system, patients can request medical-related questions, and the system provides customized responses to patient's questions. This paper discusses the key factors and

Figure 1: This figure shows the main screen of our system. The loaded script is shown as paragraphs on the right screen. If the patient has any questions during the surgical description, they can input the text in the bottom left input to ask questions.

best practices for generating effective facial expressions and gestures for an animated physician character used in surgical explanations and how to apply the Large Language Model (LLM) to our system.

2 Related works

2.1 *Emotional state and facial expressions*

To establish trust and rapport with patients, an animated physician character should convey empathy, understanding, and calmness through facial expressions [5, 6]. This can be achieved by animating the character with a gentle smile, soft eyes, and raised eyebrows to show attentiveness and concern. Furthermore, maintaining a professional and confident demeanor is essential for establishing trust in the character's expertise [5]. Moreover, the physician's speech should be clear, concise, and easy to understand [7]. The pacing of the explanation should be moderate, allowing the patient enough time to process the information and ask questions [8]. Furthermore, the character should adjust the pacing based on the patient's comprehension, slowing down or elaborating when needed [8].

2.2 *Body language*

The character should maintain an open and approachable posture, with uncrossed arms and a slight forward lean [9]. This posture communicates openness and a willingness to listen to the patient's concerns. In addition, the character's body orientation should be directed toward the patient, reinforcing the physician's attention and interest in the patient's questions or concerns [9]. During the surgical explanation, incorporate gestures that help clarify the information being presented [7]. For example, the animated physician can point to anatomical models, illustrations, or charts to provide visual context for the explanation. Ensure the gestures are fluid and natural, avoiding abrupt or jerky movements [10]. The character's gestures should be appropriate to the information being discussed, emphasizing key points and providing visual aids when necessary [7].

2.3 *Patient interaction*

Include nonverbal cues to show the physician is actively listening and engaged in the conversation [11]. These can include nodding, maintaining eye contact with the patient, and using verbal confirmations like "I see" or "I understand" [11]. These cues can help reinforce the character's attentiveness, building trust and rapport with the patient. In the animation, the physician should address any questions or concerns the patient may have, demonstrating patience and attentiveness [11]. This can be achieved through incorporating responsive facial expressions, gestures and providing clear and understandable answers to the patients' questions [11]. Additionally, the character should encourage patients to ask questions and express their concerns to foster open communication [7].

3 Implentation

3.1 Overall System Architecture

In this study, we developed systems for improving physician-patient communication through virtual agents in surgical settings. Figure 2 shows that the structure is designed to combine generating a surgical procedure content and to provide patients with interaction and comfortable experiences. Key elements of the structure are script modules, voice and gesture-generation elements, and sophisticated virtual agent interaction capabilities, which synergize to create a valuable and reliable virtual environment in real-time. Initially, based on interactions with virtual agents, we envisioned a script module designed to guide content, persona, and emotions. This module paves the way for introducing voice and gesture-generation elements that will bring virtual agents to life in the next phase. These components leverage advanced natural language processing and dynamic, expressive animation to enable agents to engage with patients in ways that closely mimic human interactions. Our virtual agents are a digitized entity designed to deliver surgical information and have meaningful conversations with patients. We applied Retrieval Augmented Generation framework(RAG) with LAMA2 to create virtual physicians designed to enhance patient communication. The RAG framework using LAMA2 allows high-speed performance, contextually relevant responses to patient inquiries, and simulates human-like conversational experiences.

3.2 Script Module

For the virtual character to explain surgical procedures, it must segment the surgical procedure into separate sections. By loading the emotion of the character information in each section, the avatar can explain surgery description like a human being. This process helps interact with patients, and they can easily understand the surgical process. In addition, the script module provides uploading images or videos per paragraph, making it easier for patients to understand information about the surgical procedure.

3.3 Face expression

Facial expressions, including voice generation and real-time conversational actions, help virtual agents express realistic emotions. As shown in Figure 3 and Figure 4, we accurately captured 48 facial expressions for animating virtual characters' faces with FaceWare Studio [12]. With the various facial expressions, the avatar has six basic emotional states for the user interface: anger, disgust, fear, neutrality, happiness, sadness, and surprise. Each emotional state is divided into three levels. This emotional state allows virtual agents to express their emotions in detail. For example, emotional 'happiness' ranges from a simple smile to a broad smile, and each one is distinguished by different degrees of intensity and characteristics.

3.4 Voice Generation

Voice generation is one of the critical components of a system that converts textual content into speech words in scripts. To simulate the natural changes in human speech, an API used by advanced text-to-speech algorithms that can adjust tone, pitch, and speed is required. As an integral part of the system's development, we integrated Naver text-to-speech API [13], allowing the system to convert written text into audible speech, improving interaction and usability. The generated audio synchronizes with the virtual agent's lip movements to create a consistent, lifelike avatar. Using Salsa [14] at Unity Asset, we completed an avatar that can be lip-synced in real-time.

3.5 Gesture Generation

Co-Speech Gesture generation gives the body language an expressiveness and engagement. Earlier

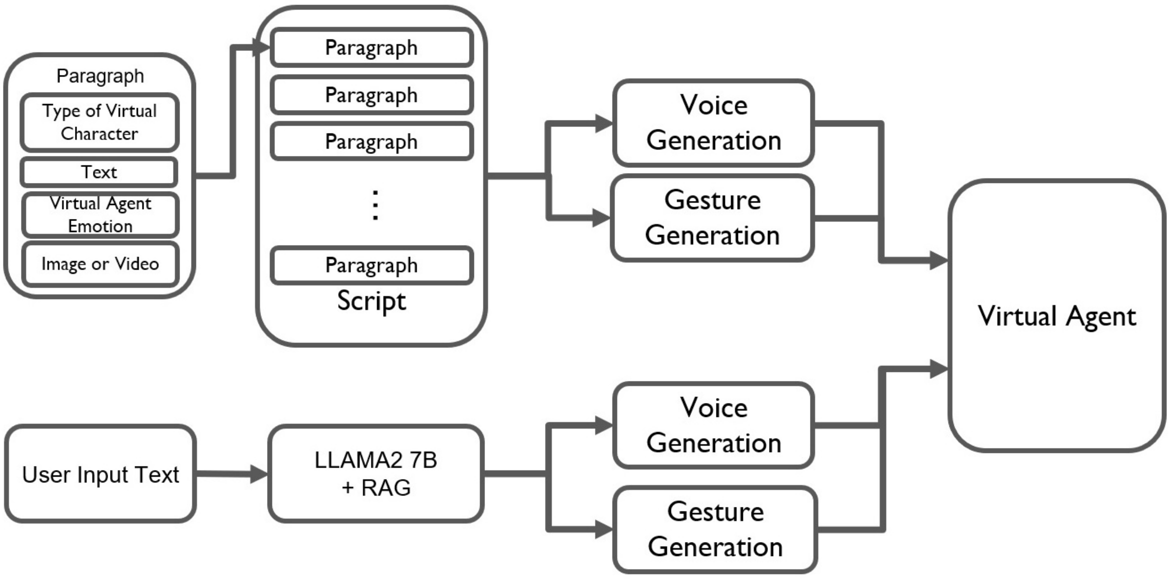

Figure 2: This figure shows the main system. The system has two different types of functions for our virtual agent system. For medical staff using the system, they can generate content about surgical procedures. The virtual agent can answer medical questions about the surgical procedure for patients

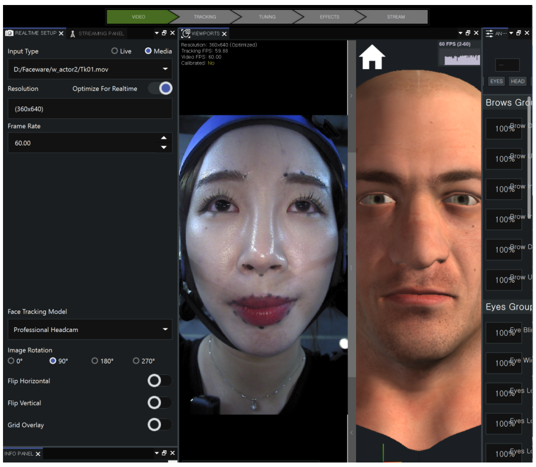

Figure 3: Faceware Studio was used to capture natural human faces. A variety of facial expressions were captured in a short period of time.

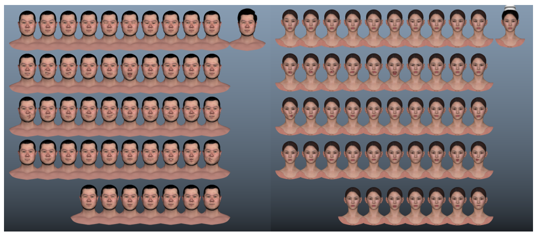

Figure 4: 48 facial expressions for virtual avatars.

studies [15, 16] showed that a text-to-gesture system can be very effective. Our research adopted a multilingual system [17] to convert text into co-speech gestures. This challenge arises from gestures' unpredictable and personal nature, which become even more complex when only text is used as input. Our approach involves collecting pairs of poses and text from videos of people talking. These pairs help us understand how text relates to gestures. We clean up the poses captured from videos and mapped them to 3D gesture units to ensure the gestures are clear and accurate. During operation, we use large language models to process the text, helping the system find the fitting gesture for the given words in real time. This is crucial for interactions that happen without delay. Making the system work across different languages and cultural settings requires videos and gestures from those contexts. We use tools like Unity3D to apply these gestures to any new virtual character, showing the system's flexibility. When dealing with longer texts, we break the text into smaller parts and match each with a gesture unit. These units are then smoothly linked together.

Figure 5: Our system has two avatars : a doctor and a nurse.

3.6 Virtual Agent

Our virtual agents are digital medical experts who interacts with patients through a combination of verbal and nonverbal signals. the virtual Agents are rendered in real time and displayed to the patient to communicate with the patient. We expect that this strategy will have a positive effect on relationship formation. To achieve this, we created realistic a physician and a nurse. Figure 5 shows two virtual characters for surgical process description and Q&A. By allowing the patient to select a virtual avatar and choose a doctor or nurse, we aim to to build an emotional connection between the patient and the avatar.

3.7 User Interaction

Artificial intelligence(AI) has made breakthroughs in diagnosis, treatment planning, patient counseling, and medical research. Recently, developments, especially in natural language processing (NLP), have contributed to healthcare professionals enhancing patient interactions and providing more accurate and rapid healthcare [1]. In this study, we want to combine two leading NLP technologies, Retrieval Augmented Generation (RAG) and Llama2[18], to develop a new system that supports patient interactions and explore its possibilities. As natural language processing advances in computer engineering, the ability to understand and answer user queries greatly influences the weight size of the response-generating model. Rather than requiring a highly developed language model, our research requires a system that delivers textual information accurately. We integrated the Retrieval-Augmented Generation Augmentation (RAG) framework and the Large language model to respond to this need. RAG is a technology that increases the accuracy and reliability of AI models generated by information from external data. This approach not only improves the accuracy of responses but also ensures relevance to the specific context of the queries [19]. This is especially useful in the medical field, requiring complex and professional knowledge. At the beginning of this study, we intended to fine-tune the LLM, specifically Llama2 7B parameter model. We utilized a random sample of 10,000 data from the HealthCareMagic-100k [20] and a dataset of 200 medical terms which is generated with ChatGPT for fine-tuning. We fine-tuned Llama2 7B with 6*A100 GPUs for 10 hours. For the training process, we followed these hyperparameters: learning rate at 2e-4, total batch size of 192, 10 epochs, 1024 tokens for maximum sequence length. We utilized two warmup steps to increase the learning rate gradually. We found that the results did not meet our expectations. The fine-tuned model often delivered inaccurate information or responses that needed more relevance to the questions. We determined that the fine-tuned model is unsuitable for use in healthcare, where accurate information delivery is critical.

Consequently, we used the RAG framework for better results and quality. Yunxiang et al. [21] compared their medical chat model, which is finetuned with the sizeable medical dataset, with ChatGPT using BERTScore. We also adopted BERTScore evaluation method because it closely matches human judgment at the sentence and system level [21]. We compared the fine-tuned Llama2 7B with the RAG framework using the same dataset. Table 2 shows that the answers generated by the RAG framework were significantly more similar to responses in the dataset. The model using the RAG framework scored higher than the fine-tuned model in precision, recall, and F1 scores. This allows the virtual doctor to communicate medical information accurately and effectively.

3.8 User Study

We conducted a user study with 120 participants in Amazon Mechanical Turk [22]. We divided

	RAG with Llama2 7B	Fine-tuned Llama2 7B	T-statistic	P-value
Precision	0.7502 ± 0.1589	0.4306 ± 0.0944	22.00	2.67E-66*
Recall	0.8216 ± 0.1169	0.5798 ± 0.1206	18.32	5.74E-52*
F1 score	0.7791 ± 0.1321	0.4922 ± 0.1007	21.98	3.33E-66*

Table 1: Quantitative comparison was conducted between RAG and Fine-tuned Llama2 7B Using BERTScore. The t-statistics and the p-values presented in the table were calculated using our paired t-test (*: $p < .05$).

	Avatar	Text	T-statistic	P-value
Social presence	4.9821 ± 1.5537	4.3482 ± 1.7016	2.911	0.0039*
trustworthiness	5.4732 ± 1.3217	5.7321 ± 1.0482	-1.624	0.1057
Utility	5.6250 ± 1.2311	5.9196 ± 0.9313	-2.019	0.0445*
Ease of understanding	5.6517 ± 1.2713	6.0357 ± 0.8794	-2.628	0.0091*

Table 2: Quantitative comparison of an avatar surgical description video and text-based a surgical description video (*: $p < 0.05$).

it into two groups of 60 participants to watch two videos, which were about 2 minutes and 30 seconds long. The first group of participants watched a video of a heart surgical procedure with text, pictures, and audio. After that, they watched a video of an avatar explaining the surgical procedure. The second group watched the videos in reverse order. The video, with image, text, and voice, is created using Microsoft PowerPoint slides and the avatar's voice. The video with the avatar is created using Microsoft's PowerPoint slides as images and the 3d doctor avatar in the Unity game engine. We were able to create the survey results with 113 participants, excluding 7 participants who answered insincerely.
Participants evaluated that the surgical description explained by the avatar had a higher social presence than the text-based surgical description ($p = 0.0039$). This result suggests that virtual avatars can generate rapport with people more naturally than the text-based surgical description. We expect that virtual doctors will be able to improve patient participation and satisfaction in the healthcare environment. Participants rated the surgical description explained by the avatar as lower in utility compared to the text-based surgical description ($p = 0.0445$). This suggests that participants may perceive text-based information as more useful than information delivery by the avatar. Lastly, the surgical description of the virtual avatar scored lower for ease of understanding than the text-based surgical description ($p = 0.0091$). This may reflect the perception that information delivery through virtual avatars is more difficult to understand than text-based information delivery. This suggests that improvements in the design and the methodology of the surgical description of virtual avatars are needed.

4 Discussion

Our research is expected to improve communication between patients and physicians in surgical description. An animated doctor character delivers medical information, which might be difficult for patients to understand, in a clear, concise, and understandable way. The system allows patients to ask questions at any time, even though they are not satisfied with the communication with the doctor for some reasons, such as the doctor might not be in fine condition or the patient might not want to share personal stories which are related to his or her disease. This approach is helpful in improving the overall quality of patient life and the quality of medical communication. Our research will elevate the quality of medical services and reinforce a patient-centered approach. We aim to create a virtual doctor avatar that responds in real-time. We focused on research and development to enable creating content for medical procedures using a game engine rather than pre-rendered rendering after writing a script. In particular, various conversational gesture movements are optimized so that even if numerous sentences are entered at once, they can be generated in almost real-time. In addition, we emphasized enhancing patient accessibility to surgical process and post-operative care information by attaching images or videos. By implementing a chatbot system using LLM and RAG, patients can easily ask questions after being briefed on the surgical process. Virtual doctor agents can serve as virtual medical partners who provide continuous patient support and interaction rather than presenting static information. There are several limitations to our research and development. First, continuous updates and verification are required to maximize the accuracy and delivery of information in the process of communicating complex medical procedures and information through virtual medical professionals[23]. Medical experts should update and conduct a thorough verification of new surgical procedures or new medical information. Second, chatbot systems based on LLM have yet to reach a proficiency level that can replace diagnostics or complex medical consultations [21]. Our system cannot completely replace the communication between patients and doctors with limitations. RAG uses pre-stored Vector Score to find document information and generates answers with LLM, so it has the possibility to create various answers depending on the search results [23]. It is difficult to suggest that chatbot systems with RAG are the best approach, considering that patients pose questions while being monitored by doctors. Finally, Patient privacy and data security are critical issues. Using the patient's information to create a rapport with the patient has complex issues both legally and morally [24]. We aim to solve these problems through technological improvement and research to recognize and overcome these limitations.

5 Conclusion

In the medical area, designing facial expressions and gestures for a virtual doctor improves positive communication between doctor and patient. It is an essential factor in building patient understanding and trust in the medical process. Animators can generate animated agents for doctor-patient interaction by following existing research on animation design and communication between doctors and patients. As technology advances and the expansion of computer-generated animation in medical education fields, our research is expected to draw attention and suggest a method for the convergence of medical and computer science.

References

[1] Chetan Bulla, Chinmay Parushetti, Akshata Teli, Samiksha Aski, and Sachin Koppad. A review of ai based medical assistant chatbot. *Research and Applications of Web Development and Design*, 3(2):1–14, 2020.

[2] Jorge G Ruiz, David A Cook, and Anthony J Levinson. Computer animations in medical education: a critical literature review. *Medical education*, 43(9):838–846, 2009.

[3] Richard K Lowe. Animation and learning: Selective processing of information in dynamic graphics. *Learning and instruction*, 13(2):157–176, 2003.

[4] Kyle Johnsen, Robert Dickerson, Andrew Raij, Benjamin Lok, Jonathan Jackson, Min Shin, Jonathan Hernandez, Amy Stevens, and D Scott Lind. Experiences in using immersive virtual characters to educate medical communication skills. In *IEEE Proceedings. VR 2005. Virtual Reality, 2005.*, pages 179–186. IEEE, 2005.

[5] M Robin DiMatteo and David Dante DiNicola. Sources of assessment of physician performance: a study of comparative reliability and patterns of intercorrelation. *Medical care*, pages 829–842, 1981.

[6] Eric B Larson and Xin Yao. Clinical empathy as emotional labor in the patient-physician relationship. *Jama*, 293(9):1100–1106, 2005.

[7] Suzanne Kurtz, Juliet Draper, and Jonathan Silverman. *Teaching and learning communication skills in medicine*. CRC press, 2017.

[8] John Sweller. The redundancy principle in multimedia learning. 2005.

[9] Debra Roter and Judith A Hall. Doctors talking with patients/patients talking with doctors. 2006.

[10] Christian Wallraven, Heinrich H Bülthoff, Steffen Waterkamp, Loes van Dam, and Nina Gaißert. The eyes grasp, the hands see: Metric category knowledge transfers between vision and touch. *Psychonomic bulletin & review*, 21:976–985, 2014.

[11] Wendy Levinson, Cara S Lesser, and Ronald M Epstein. Developing physician communication skills for patient-centered care. *Health affairs*, 29(7):1310–1318, 2010.

[12] Faceware Technologies Inc. Faceware studio. `https://facewaretech.com/faceware-studio/`.

[13] Clova voice. `https://www.ncloud.com/product/aiService/clovaVoice`.

[14] Salsa lip-sync. `https://crazyminnowstudio.com/unity-3d/lip-sync-salsa/`.

[15] Ghazanfar Ali, Myungho. Lee, and Jae-In Hwang. Automatic text-to-gesture rule generation for embodied conversational agents. *Computer Animation and Virtual Worlds*, 31(4-5):e1944, 2020.

[16] Ghazanfar Ali and Jae-In Hwang. Improving co-speech gesture rule-map generation via wild pose matching with gesture units. *SIGGRAPH Asia 2022 Posters*, PartF168147-3:1575–1585, 12 2022.

[17] Ghazanfar Ali, Woojoo Kim, Muhammad Shahid Anwar, Jae-In Hwang, and Ahyoung Choi. Expanding multilingual co-speech interaction: The impact of enhanced gesture units in text-to-gesture synthesis for digital humans. sep 2023. Preprint: https://doi.org/10.21203/rs.3.rs-3350470/v1.

[18] Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, et al. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*, 2023.

[19] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*, 33:9459–9474, 2020.

[20] Healthcaremagic.com. `https://www.healthcaremagic.com`.

[21] Yunxiang Li, Zihan Li, Kai Zhang, Ruilong Dan, Steve Jiang, and You Zhang. Chatdoctor: A medical chat model fine-tuned on a large language model meta-ai (llama) using medical domain knowledge. *Cureus*, 15(6), 2023.

[22] Amazon Mechanical Turk. `https://www.mturk.com/`.

[23] Yunfan Gao, Yun Xiong, Xinyu Gao, Kangxiang Jia, Jinliu Pan, Yuxi Bi, Yi Dai, Jiawei Sun, and Haofen Wang. Retrieval-augmented generation for large language models: A survey. *arXiv preprint arXiv:2312.10997*, 2023.

[24] Thomas C Rindfleisch. Privacy, information technology, and health care. *Communications of the ACM*, 40(8):92–100, 1997.