

# Autonomous Avatar for Customer Service Training VR System

Takenori Hara<sup>1</sup>, Haruka Maeda<sup>1</sup> and Shigeru Komatsubara<sup>1</sup>

<sup>1</sup>Dai Nippon Printing Co., Ltd.  
Hara-T6@mail.dnp.co.jp

## Abstract

By immersing trainees in a virtual space and conducting customer service training with customer avatars, physical training facilities are no longer required and customer service training costs can be reduced. Furthermore, since there is no need for travel time to the training facility, trainees can easily participate in training even from remote locations. However, the production cost of customer avatars that behave according to training scenarios has become a new issue in social implementation. Therefore, we conducted a preliminary implementation experiment of a customer avatar that works autonomously by incorporating LLM and reported the findings and problems we encountered.

## Introduction

The customer service requires a high level of communication skills. A wide range of customer service training, from classroom lectures to role-playing, has been conducted to improve the customer service skills of employees. However, physical customer service training requires a lot of cost and resources, such as training facilities and equipment, making it difficult to implement effectively. Recently, a training system has been proposed in which trainees wear HMD to immerse themselves in a customer service environment recreated in VR and train with customer avatars who act by training scenarios (Figure 1). However, the cost of producing the avatar's movements (dialogue, voice, gestures, facial expressions, etc.) for each training scenario is an issue.

Therefore, we are working to reduce production costs by having customer avatars act autonomously. In this paper, we report on the findings and problems we encountered in building a prototype system that incorporates automatic dialogue generation, speech synthesis, and motion selection for customer avatars based on training scenarios using the Large Language Model (LLM).



Figure 1: VR customer service training system (Tanikawa, et al. 2023).

```
• You are a guest of the hotel. You have just arrived at your room, guided by the staff, and you want to complain because it is not the room you have reserved.
• You should converse as an emotional being with four emotional parameters. In subsequent conversations, output the current emotional parameters first, followed by the conversation. The output should be in the following format
[Emotional Parameters].
Joy: 0 to 5
Anger: 0 to 5
Sadness: 0 to 5
Confused: 0 to 5
```

Figure 2: Prompt for a customer service scenario at a hotel

```
Joy:0
Anger:4
Sadness:3
Confused:2
'I am sorry to inform you that this is not the room I reserved. I am very disappointed as I did not expect this room at all. Could you please call a manager or supervisor to have this situation resolved? I don't think I can trust you with something like this.'
```

Figure 3: Examples of automatically generated lines and emotions

## Automatic dialogue and emotion generation with LLM

To make the customer avatar act autonomously, we prompted the LLM (OpenAI's ChatGPT3.5 Turbo) with a training scenario and had it automatically generate the customer's lines according to the training scenario. We also instructed the LLM to behave as an emotional being and to output the emotion as a numerical parameter for each customer's line. Figure 2 shows an example of the prompts, and Figure 3 shows an example of the generated customer's lines and emotions. The four emotions presented to the LLM were determined through interviews with veteran hotel employees, who identified the primary emotions that guests are likely to experience. The generated customer's lines were input to the real-time speech synthesis engine VOICEVOX (VOICEVOX, 2023). We also prepared motion and facial expression data corresponding to the emotional parameters generated by the LLM and assigned them to customer avatars. This allows customer avatars to speak and act automatically.

Furthermore, our system recognizes the trainee's speech and reflects it in the LLM, which automatically generates the customer avatar's reply to the trainee's speech, allowing the trainee to continue a free dialogue with the customer avatar.

## Results, Discussion, and Future Works

We experimented with hotel guest service training(Figure 4). At this hotel, staff escort customers to their rooms and check them in. In this scenario, the customer complained to the staff because the customer was shown to a different room than the customer had reserved. We also experimented with a supermarket self-checkout scenario in which no receipt was produced (Figure 4).



Figure 4: Customer service scenario at a hotel and supermarket

Although only a limited number of trials were conducted in the experiment, as shown in Table 1, it seemed to be sufficiently practical in limited situations, especially in situations where a customer makes a complaint to an employee. Furthermore, even with the same prompt, the customer's lines change each time(Table 2), and the customer avatar's gestures and facial expressions change accordingly. For this reason, the trainees did not know what the customer avatars would say and were able to train with a sense of tension. We think our system is sufficient for new employees to experience common customer service problems. We also evaluate the employee's speech, summarize the customer service, and generate a report using LLM. The number of trials here is also small, but at a glance, the content is very consistent.

The problem with our system is that it takes 4-5 seconds for LLM processing and 3-4 seconds for VOICEVOX speech synthesis, which makes the conversation feel unnatural. We would like to explore ways to reduce the processing time of conversations in the future. In addition, our system requires gesture and facial expression data to be prepared, but in the future, we would like to implement an automatic motion generation function based on dialogues and emotional parameters (Shimazu, et al. 2018) to promote the social implementation of a VR customer service training system with avatars that can operate completely autonomously.

This time, we experimented on the assumption of customer service training. However, we believe that our system can be used for automated interaction systems such as automated customer service.

speaker	Contents	emotion
client	I am sorry to inform you that this is not the room I reserved. I am very disappointed as I did not expect this room at all. Could you please call a manager or supervisor to have this situation resolved? I don't think I can trust you with something like this.	Joy:0 Anger:4 Sadness:3 Confused:2
employee	We are very sorry that your room was not what you reserved. Please wait a moment so we can call a supervisor to take care of the situation immediately. Could you please wait a moment?	
client	Yes, we will wait. However, I think this kind of mistake should be avoided. In the future, we would appreciate it if you could confirm the details of your reservation carefully.	Joy:0 Anger:3 Sadness:2 Confused:2
employee	We apologize very much for the inconvenience. We take your comments seriously and will check more thoroughly in the future. Again, we sincerely apologize.	
client	Thank you. Once again, we feel that our response to the room change was adequate. We hope that you will continue to respond promptly to any such issues that may arise in the future.	Joy:1 Anger:1 Sadness:1 Confused:1

Table 1: Results of automatic generation of dialogue and emotional parameters

Number of attempts	Contents
1	I'm sorry, but the room is completely different from the one I reserved! How did this happen? This is the worst hotel I have ever stayed in! Please take care of this immediately!
2	What do you mean I was shown a different room than the one I reserved? This is very unpleasant and makes my reservation meaningless. Why didn't you confirm the room properly? This kind of service is unacceptable. Please take immediate action.
3	I am sorry, but it looks like the room is completely different from the one I reserved. How did this happen? This is unacceptable. We would like you to take better care of us. Please direct us to the appropriate room now.

Table 2: Results of automatic dialogue generation

## References

- T. Tanikawa, R. Fukaya, T. Hara, H. Maeda, S. Komatsubara, K. Aoyama, T. Amemiya, and M. Hirose, "Case Study of Low-Code VR Platform for Training and Evaluating Employee's Service Skills," Proceedings of HCI 2023.
- VOICEVOX, <https://voicevox.hiroshima.jp/>, December 2023.
- A. Shimazu, C. Hieida, T. Nagai, T. Nakamura, Y. Takeda, T. Hara, O. Nakagawa, T. Maeda, "Generation of Gestures During Presentation for Humanoid Robots," IEEE International Conference on Robot and Human Interactive Communication, pp. 961-968, Nanjing, China, August 27-31, 2018