Development of an Interactive Live Streaming System for Language Learning

Hong-Yi Pai[1], Ching-Huang Wang[2], Yi-Chen Lai[1]*

[1] *Department of Multimedia Design, National Formosa University, Taiwan*
[2] *Department of Applied Foreign Languages, National Formosa University, Taiwan*
hpai@ms38.hinet.net, chinwang@nfu.edu.tw, thetrue5142@gmail.com

Abstract

In this paper, we propose Virtual Avatar Interactive Live Streaming System (VAILSS), our approach to streaming 3D avatar performances in interactive storytelling applications using multimodal interactive techniques. This study focuses on developing a virtual live interaction system for VTubers in the context of language learning. A live virtual performance from our system consists of three components: 1) an Avatar generation framework, 2) an AI motion capture system, and 3) an interactive storytelling engine. The system integrates artificial intelligence and uses motion capture to identify facial expressions and movements, enabling virtual characters to provide live storytelling services. Additionally, the system allows for bi-directional interaction with players through tablet touch and voice. The system aims to promote multimodal learning channels for the masses. To gain deeper insights into the effectiveness of virtual role-playing in language learning, we organized a three-week workshop to investigate our system's impact on user experience. We extended invitations to 17 night-school freshmen from the NFU Department of Applied Foreign Languages, enrolled in an English class focused on tour guides, to participate in a virtual performance. The experiment results demonstrated that the VAILSS positively affected students' learning outcomes, particularly in enhancing foreign language acquisition.

Keywords: Interactive storytelling, Virtual character, Live streaming, Story teaching

1 Introduction

From the perspective of story teaching, successful storytelling can assist the narrator in effectively conveying the intended message and enhancing emotional communication. When it comes to language learning, performing role imitations can enhance the natural expression of language skills. In conventional foreign language courses, students seldom get the chance to practice language expression through role-playing, and staging performances in the form of drama requires significant resources and production costs, impeding rapid development. Therefore, this paper aims to aid learners in enhancing their language expression abilities and self-confidence by developing a virtual role-playing interactive system.

This paper presents a VAILSS for language learning from the perspective of story teaching [1]. This system differs from usual play learning or serious play development, emphasizing user autonomy. The system accurately represents the behavior and feelings of the storyteller. It provides real-time extraction of information and supporting graphic display, helping learners to scaffold learning activities as they learn new material. Moreover, we are also providing the feasibility of developing virtual endorsements, not only in commercial entertainment. This system draws inspiration from The Little Prince story and incorporates four characters from the narrative as virtual storytellers for immersive performances. These avatars are transformed into tour guides within this immersive system, effectively guiding the audience through personalized live presentations.

2 Literature Review

With the popularity of digitalization, interactive narratives emphasize the immersive experience of the reader/viewer and the user's control of autonomy to join, even beyond the author. The gamification of interactive narratives emphasizes construction, leading Roland Barthes to propose that the author is dead. The authoritarian era of textual authorship is over, and the rise of the reader has made storytelling on different platforms more complex and challenging [2]. Readers have become more reflective and exploratory, making it harder for authors to control the development of their stories and ensure that they conform to the author's original intent. Thus, authorial control and user autonomy are pulled together [3]. Using gamified interactive narratives to create fragments of the story structure, users need to spend more time exploring to understand the original story structure. This system aims to solve the core problem of story structure fragments through the participation of virtual storytelling.

From a comprehensive perspective of language learning, it is essential to facilitate language acquisition in "natural," "authentic," and "complete" language contexts [4]. Ollerenshaw and Lowery argue that storytellers create clear images of the characters in a story and develop the story context through storytelling techniques such as body

*Corresponding Author: Hong-Yi Pai; E-mail: hpai@ms38.hinet.net
DOI: 10.53106/160792642024012501003

movement and voice [5]. This approach significantly enhances children's language-related speaking and listening skills and their ability to express themselves more clearly and develop the habit of listening carefully. Research conducted in domestic settings integrating story teaching into the language education field has revealed increased interest in reading and writing among students. Children are more actively engaged in exploring the worlds depicted in books. Moreover, there has been a notable improvement in verbal communication and listening skills as students become more proficient in expressing their thoughts and develop a habit of attentive listening. Egan argues that "teaching is storytelling" [1], and the multimodal interactive storytelling in this study provides a full context for "story teaching."

Interactive storytelling has created a new literary structure, raised research questions, and ensured essential learning outcomes. Literatures have indicated that the maturation of related devices and application development tools has become increasingly affordable and user-friendly, making learning and entertainment applications feasible and attractive [6]. In addition to traditional storytelling, storytellers and players can participate and influence the storyline, providing new, more engaging, and adaptable types of narratives. However, since the interactive narrative of the game does not focus on learning through oral communication, it presents dynamic storytelling and non-linear event sequences through interactive storytelling, which are determined by the user's actions and choices during the story "reading" process. The gamification does not support children's education in listening to stories and expressing their opinions clearly.

With the popularity of VTuber streaming services, there is growing concern about the role of the storyteller in engaging the audience, i.e., telling stories to others [7]. In traditional face-to-face storytelling, the presence of the speaker and especially the sharing of emotion connect the storytelling to the experience in an unforgettable way. Likewise, the presence and emotion of the speaker are critical to compelling storytelling in the digital realm [8]. Today, digital games facilitate various forms of entertainment and significantly impact our society and culture. Moreover, the role of digital games in education is becoming increasingly important. Game interaction is recommended to enhance students' creativity and engagement and provide them with opportunities to acquire knowledge [9]. To assist regular users in creating virtual reality stories using VAILSS, the storyteller begins by selecting their characters from the virtual character menu. While creating a virtual reality story, facial recognition software reflects the storyteller's facial expressions onto the avatar. By detecting hand gestures through Webcam Motion Capture software, the storyteller can trigger pre-edited emotional effects to make the facial expressions more pronounced.

Additionally, we utilize Unity Render Streaming to enable multiplayer connections and bi-directional communication with minimal hardware and software configuration. Through this study, we demonstrate the importance of emotional embodiment in live virtual reality storytelling to enhance the audience's sense of interaction with the storyteller and increase their enjoyment and engagement. As a result, this technology enhances the immersive storytelling experience, confirming that the global trend in education is shifting towards the utilization of technology and the internet due to advancements in information communication technology (ICT) and evolving learning models [10].

3 Related Work

Research has shown that virtual performances and immersive experiences evoke presence and facilitate narrative appreciation more than traditional paper books [11]. The linear narrative structure of traditional storytelling makes it difficult to achieve an immersive context. In contrast, using interactive games can create a virtual environment that allows players to explore scenes and have ownership of the story. However, to understand the author's concepts, interactive game players need to spend more time interacting, which may be less efficient for learning. Therefore, a role-playing approach where the storyteller acts as the mediator of the interactive narrative can allow the author to take more control of what they want to express while quickly bringing the audience into the storyline. The storyteller's ruling power will be stronger than ordinary non-playable characters (NPCs) in a game. In addition, text and graphics in the scenes can help the player understand what the storyteller is saying. There is a growing call to integrate technology into education to engage students and improve their academic performance. Virtual performance is highly valued for its potential to integrate the real world with virtual objects and engage students in hands-on activities. Experiments have shown that interactive storytelling increases the audience's immersion and presence [12].

According to the literature, controlling virtual characters to make presenters appear in an interactive storytelling system (also known as real-time animation) may help engage the audience and be considered a reproduction of an old storytelling art form and theatrical practice [13]. These technologies allow for the viable development of scripted endorsements in VTuber live presentations and expand storytelling approaches. Additionally, using avatars to express emotions can help children better understand the content of a story. However, we have found less use of them in language education. Previous works have focused on creating animations for characters and depicting their movements. However, the embodiment of emotions in presenters has yet to be fully explored, especially in 3D live storytelling scenarios.

Emotional experience plays a vital role in interpersonal communication [1]. Thomas et al. proposed that successful storytelling requires a balance between the storyteller, the story itself, and the audience [14] to connect the story to the audience through spoken expressions. We aim to use virtual characters' voices and facial expressions to provide an intensely emotional experience for the audience in a virtual live-streaming environment. Although it is still a challenging problem to embody the emotions of the presenter in real-time VR storytelling in an online environment, the system can make capturing facial expressions in real-time through AI more accessible and more realistic than offline manual adjustments. However, it is necessary to enhance training for the hosts' spontaneous emotion transmission, especially when users are not trained to convey their characters' emotions and moods. With such training, it is easier for the system to capture the users' emotional cues [8] automatically.

A proof-of-concept study was conducted in this paper to determine whether character expressions could achieve the intended results. The study involved customizing four virtual avatars (Figure 1) and verifying the effectiveness of expression changes through virtual avatar expressions. For instance, in the story of the pilot, using a sad expression to depict the departure of the little prince can convey the pilot's sense of loss and enable the reader to comprehend the author's emotional intent better.

Hence, proficiency in conveying emotions through facial expressions is crucial for effective interpersonal communication. To ensure the accurate communication of emotional states, it is imperative to extract facial features that play a significant role in the perceptual process [15]. Incorporating facial expressions, along with other modalities like voice and text, can significantly enhance the conveyance of emotional content and captivate the audience in the storytelling experience.

Figure 1. The overall image of four virtual avatars

Virtual characters often have limited facial expressions, meaning other multimodal displays must be used to achieve auxiliary effects. Furthermore, while facial expressions can convey a clear emotional state, they are not always enough to evoke a strong emotional response. Therefore, enriching the expression of emotions with other iconic modalities involving different sensory channels is essential. In addition, these modalities help to specify emotional intensity, which is crucial for compelling storytelling.

For example, when players encounter new items or tasks in games, storytellers can use text and images on the user interface (UI) to provide prompts and explanations to help players better understand and use these contents. These prompts can be displayed in specific locations on the game interface, such as the lower right corner of the screen, with images and text. Additionally, storytellers can engage in dialogue with players to increase the visibility and usability of the prompts. They make the game easier to explore and enhance player engagement and interactivity with the scenes.

In VAILSS, the Webcam Motion Capture software offers a motion assistance function that allows us to blend customized body motions with the original movements. This feature dramatically benefits hosts who struggle with performing naturally, addressing the issue of awkward limb movements. Moreover, these personalized body movements can be added or reduced without pre-programming. VAILSS integrated the Webcam Motion Capture software to enhance control and flexibility in virtual role-playing. As for facial expressions, VSeeFace software handles them, enabling the user to specify the desired expressions using designated buttons.

Regarding interactive scene design, the 3D scenes are organized into chapters, each showcasing distinct content elements. This setup facilitates interaction between storytellers and audiences, enabling bi-direction feedback within the game. The perspective of the virtual scene can be altered by adjusting camera views through rotation and zoom. Additionally, objects within the scene can be manipulated using drag-and-drop functionality. These methods promote exploration, facilitate repeated learning, and establish a framework for learners to imitate. The following image shows the interactive interface of VAILSS in Figure 2.

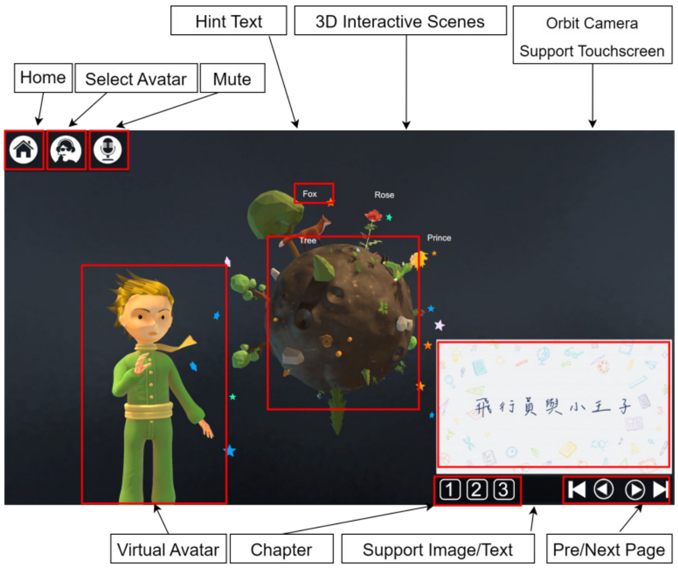


Figure 2. The interactive interface of VAILSS

4 System

We designed and implemented a virtual live interactive storytelling application platform to enable ordinary storytellers to take on role-playing roles to tell story scripts. We selected the Windows operating system as the primary host device, allowing the audience to connect via tablet or PC. We used a camera and AI computer programs to track facial and body movements. The system provided a multimodal capability to meet three essential design requirements: 3D character motion streaming service, 3D scene interaction, and bi-directional communication for better emotional expression. The VAILSS interactive storytelling consists of three components: 1) an Avatar generation framework, 2) an AI motion capture system, and 3) an interactive storytelling engine. This device features bi-direction interaction with joint observation and real-time feedback and supports motion capture. Finally, the system generates a .exe file for the host and an APP for the audience. The following image shows the implementation process of VAILSS in Figure 3.

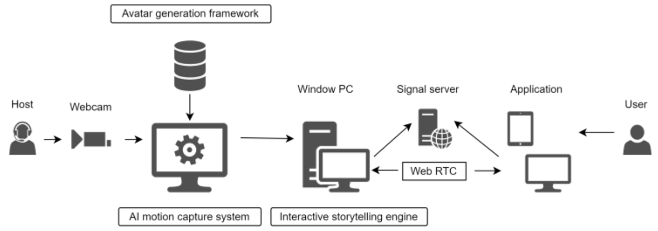


Figure 3. The implementation process of VAILSS

4.1 The Avatar Generation Framework

Due to the limited availability of avatar generation systems that cater to customized stories beyond Japanese-style manga characters, we propose a framework for creating personalized 3D avatars. This framework encompasses various character types, including monsters and non-comic characters, while adhering to the virtual reality modeling (VRM) format [16]. These avatars can be effectively utilized on popular service platforms such as VRChat or VSeeFace, which support VTuber roles.

The procedural framework for customizing 3D avatars is based on visual perception's theoretical foundations by Zell et al. [17]. The design process framework comprises six essential modules: character concepting, character designing, character modeling, character shading, character rigging, and VRM formatting. Refer to Figure 4 for a visual representation of the framework's process for creating a customized 3D avatar.

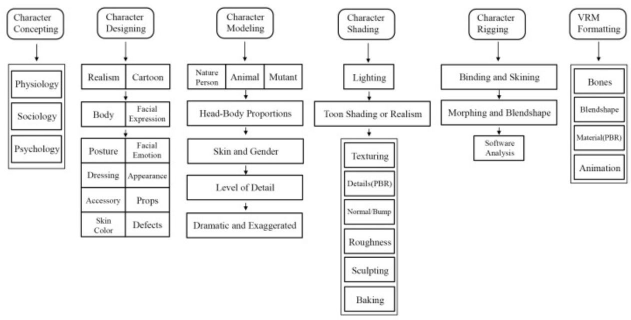


Figure 4. The production framework for creating customized 3D virtual avatars

4.2 The AI Motion Capture System

We use VSeeFace as motion-tracking AI software. VSeeFace is a popular software application that virtual streamers and VTubers use to create and control their virtual avatars in real time. It provides a user-friendly interface and robust facial tracking, lip-syncing, and animation control features. VSeeFace supports various tracking methods, including webcam motion capture, which allows users to animate their avatars using their facial expressions and head movements captured by a webcam.

VSeeFace utilizes webcam motion capture to track users' facial movements and expressions. The software can accurately capture and interpret their real-time facial actions by positioning the webcam in front of the user. It allows the virtual avatar to mimic the user's facial expressions, such as raising eyebrows, smiling, or opening the mouth.

Overall, VSeeFace with webcam motion capture provides an accessible and intuitive solution for users to control and animate their virtual avatars in real-time, adding a layer of interactivity and immersion to their virtual streaming or VTubing experiences.

On the other hand, we employ Webcam Motion Capture software to enhance the precision of capturing performers' hand and body movements. Webcam Motion Capture is a user-friendly software application that utilizes a standard webcam or similar camera device to capture and track body motions. This technology eliminates the need for specialized equipment or markers, transforming an ordinary webcam into a convenient motion capture device.

Webcam Motion Capture software offers an accessible and cost-effective solution for various motion capture applications, making it particularly favored among hobbyists, content creators, and small-scale productions. It empowers users to seamlessly translate their real-world body movements into the digital realm, opening possibilities for creative expression and immersive experiences.

4.3 The Interactive Storytelling Engine

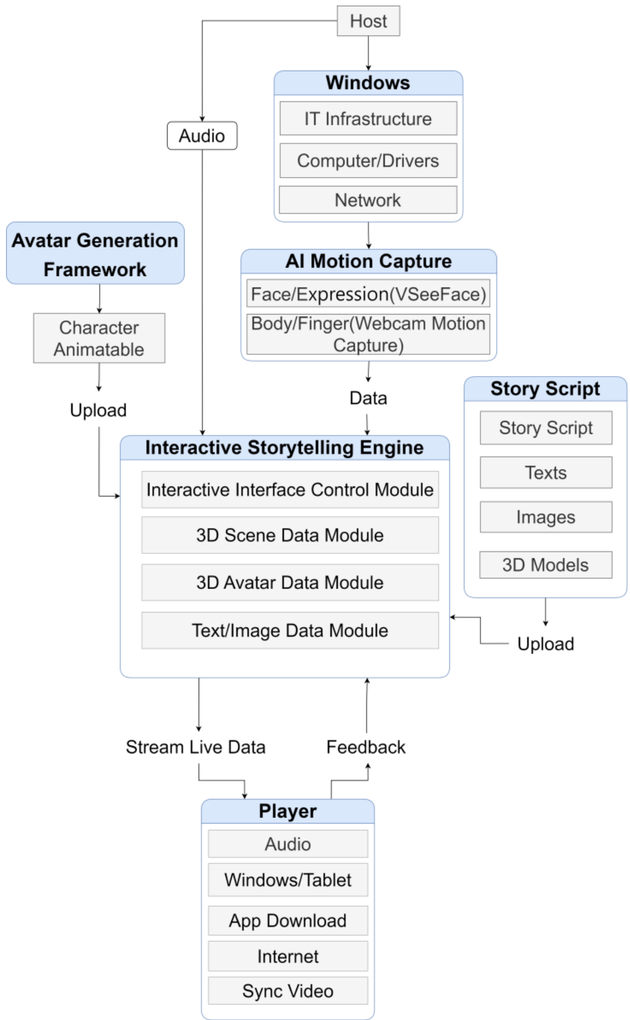


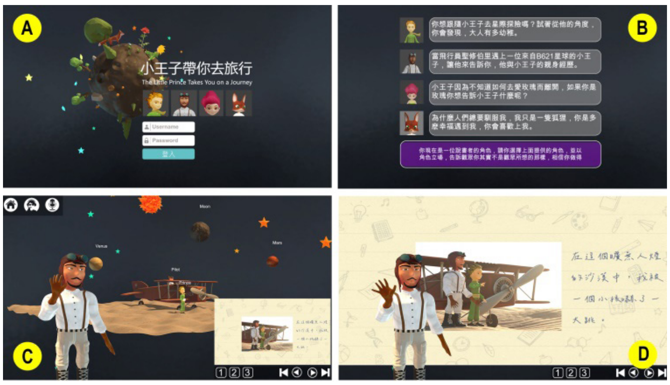
Figure 5. System flow chart

The interactive storytelling engine integrates AI motion capture data to ensure synchronized movements of virtual characters within the 3D scene. It also offers 3D interactive controls, enabling users and hosts to interact with the game by grabbing objects and adjusting the camera zoom. Additionally, it incorporates a PowerPoint presentation interface layout to enhance the assistance provided by graphics and text. The interactive storytelling engine comprises various modules, including 1) the interactive interface control module, 2) the 3D scene data module, 3) the 3D avatar data module, and 4) the graphic/text data module. Figure 5 shows the system flow.

4.3.1 The Interactive Interface Control Module

The interactive design interface arrangement is divided into three scenes: Scene A: The login interface is based on Firebase as the backend system. Scene B: The storyteller role selection page, where users can choose to be storytellers or listeners. Scenes C and D: 3D scenes serve as the storytelling venue. Storytellers can change roles and press the play button to enter the next chapter. Listeners can zoom in or out of the camera, turn the camera, and interact with objects in the scene to explore the story. For a visual representation of the VAILSS UI, please refer to Figure 6, which includes the following items:

A) Main page: The login screen, where data verification is completed using the Firebase SDK.
B) Role options buttons: This interface provides four role options for the presenter, including the little prince, the pilot, the rose, and the little fox.
C) 3D scene: This is the storyteller's stage, where the audience can interact and explore.
D) Presentation scene: The host presents with a graphic/text data background.



A) Main page B) The four role options buttons C) 3D scene D) Presentation scene

Figure 6. A visual representation of the VAILSS UI

We utilize the Unity game engine to integrate several features, including five modules: 1. The interactive control module, enabling player touch interactions. 2. The virtual character module facilitates four characters' extraction and live performance. 3. The 3D scene module, responsible for displaying the Little Prince's planet. 4. The image and text module, allowing custom graphics playback. 5. The audio module enables two-way communication between the user and the live streamer.

To enable real-time streaming of virtual character performances, we utilize WebSocket and WebRTC technologies. WebRTC can bridge the gap between browsers and real-time rendering, allowing us to leverage Unity's evolving graphics capabilities without device performance limitations.

WebRTC is a versatile technology widely used for various real-time communication applications, including video conferencing, live broadcasting, online gaming, and remote collaboration. It is known for its cross-platform support and flexibility. Being an open-source technology, WebRTC enables direct real-time communication between web browsers or compatible applications. It facilitates peer-to-peer communication, allowing tasks like audio and video streaming, file sharing, and data transfer without additional plugins or software installations. We use this technology not only to handle distortions and instabilities caused by virtual character transmission delays but also to accommodate future character changes efficiently, minimizing the effort required to write and update character interaction controls. The following image shows the framework of WebRTC streaming service in Figure 7.

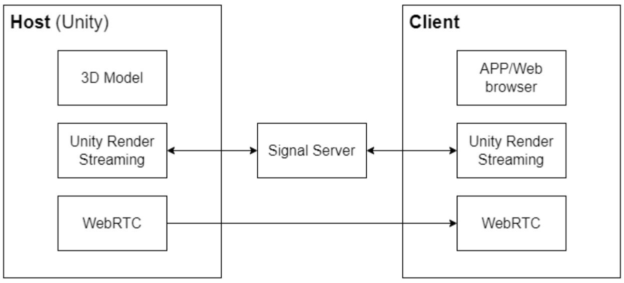


Figure 7. The framework of WebRTC streaming service

4.3.2 The 3D Scene Data Module

The 3D scene provides a 3D interactive narrative environment simulation, where the viewer explores a 3D scene related to the story content through the storyteller's prompts, and the scene includes touch feedback, i.e., animation playback, pop-up messages, etc. The storyteller tells the oral story through real-time voice communication of the storyteller's emotions through Avatar, Support for auxiliary graphics and indicators, including images and diagrams with text descriptions displayed in the lower right corner, providing logical storytelling by the storyteller. The content of relevant 3D scene data, we mentioned in the previous chapter.

Our team has developed a render streaming framework using Unity Render Streaming API, which enables running Unity applications in a browser. Once we add the Render Streaming package to the project, we can easily control it in real-time using standard browsers such as Google Chrome and Safari on desktop and mobile devices. We also use the Normcore plug-in program to facilitate multi-person communication with voice dialogue. In order to enhance the realism and vividness of the storyteller's emotional performance, we used a Windows operating system computer as the essential equipment. We utilized its camera with

VSeeFace software for face and body tracking. In addition, we utilize the Webcam Motion Capture software for hand tracking to ensure the synchronization of the storyteller's movements with the Avatar. Those enable storytellers to convey various facial expressions and gestures in virtual reality realistically. The following image shows the status of interface synchronization via streaming services in Figure 8.

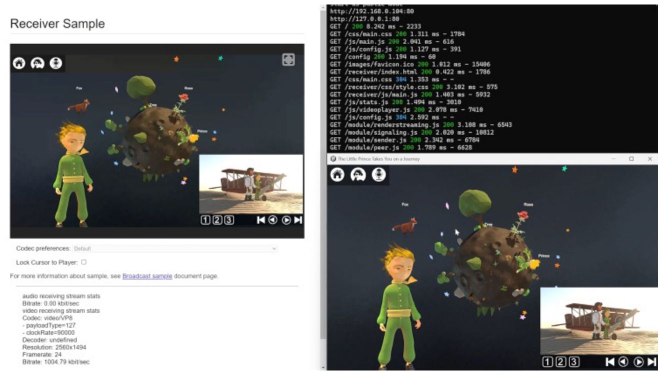


Figure 8. The interface synchronization via streaming services

4.3.3 The 3D Avatar Data Module

We created four virtual characters and provided them to the database, such as the pilot, the little prince, the rose, and the fox. To achieve virtual role-playing motion synchronization between VSeeFace and Unity, we incorporate Open Sound Control (OSC) and Virtual Motion Capture (VMC) components into a 3D avatar model within Unity. By utilizing OSC and VMC APIs plugins for Unity, we can receive, and process data from diverse sources, including VSeeFace, such as motion capture data or other applications. The data can then control and animate objects within in Unity project. We employ the VRM format to customize our storyteller characters for exclusive interactive narratives. This approach distinguishes our characters from the VTuber Japanese-style characters generated through Procedural Content Generation (PCG). Our focus lies in constructing 3D character generation rooted in specific story content.

4.3.4 The Text/Image Data Module

This module includes auxiliary graphics and indicators to support the content presented by the narrator. It allows players and narrators to identify each other through visual information. The system offers zoom-in/out functionality for pictures and texts, enabling users to delve into the details and enhancing readability. Moreover, the system supports mobile app downloads, providing convenience for users to learn and utilize the module.

5 Experiment

In order to gain a deeper understanding of the effectiveness of virtual role-playing in language learning, we conducted a workshop. This workshop examined the interactive system's impact on the user experience through virtual performances. The experiment lasted for three weeks. Using qualitative methods, we observed the influence of interactive systems on user experience from both the presenter's and the audience's perspectives. We assessed the relationship between virtual role-playing and their avatar preference from the presenter's side. From the audience's side, we evaluated user satisfaction regarding software and hardware usage and concrete emotional expression. The following image shows the experimental process in Figure 9. The workshop is divided into four sessions:

1. The modification of the story script (pre-test).
2. Conducting virtual role-play training sessions.
3. Storytelling with emotional embodiment (post-test).
4. The audience experiences of VAILSS (control group).

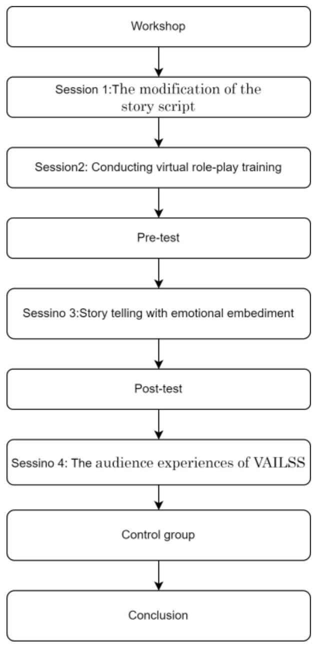


Figure 9. The experimental process

In Session 1, students were asked to give speeches, and professors provided corrections. This session focused on the logical construction of the narrative text.

In Session 2, we introduced the use of virtual characters and the operation of the interactive system and conducted a test drill. Four roles were assigned to performers through random selection. These characters were based on the fairy tales of the Little Prince and were transformed into tour guides with English-speaking skills. During this session, students practiced performance and speech content with the assistance of teaching assistants. Participants conducted a pre-test without engaging in experiential activities.

In Session 3, each student was given 1-3 minutes of story performances. Performers achieved virtual performances through AI motion capture and facial expression capture. Audiences could experience the emotional embodiment through live streaming or recorded videos. After each session, participants filled out a questionnaire to report their user experience.

In Session 4, to gauge the audience's satisfaction with the system, we invited 23 students from the NFU Department of Multimedia Design students taking the game planning and design course to participate in this questionnaire survey. This control group was used to compare the user experiences without participating in the virtual role-playing.

We ran the VAILSS system on a laptop with an Intel i9-

139880HX CPU, 8GB RAM, and Nvidia GeForce RTX 4070 GPU. We opted for a Windows operating system as the essential equipment and utilized the camera and VSeeFace software for face and head tracking. Additionally, we used Webcam Motion Capture for body and hand tracking to ensure the presenter's movements were synchronized with the virtual avatar.

We invited 17 night-school freshmen from the NFU Department of Applied Foreign Languages, taking the English course for the knowledge of tourism, to take part in the workshop.

During the workshop, we provided an overview of the interactive system and demonstrated how to use it, incorporating storytelling in English by the participants. We offered four different roles for the performers to assume. The virtual characters mainly consisted of cartoon shapes, with various characteristics used to convey their personalities and identities. Each participant randomly selected one of the four roles and conducted the briefing. These characters were inspired by the fairy tales of The Little Prince and included the pilot, the little prince, the fox, and the rose. They were divided into three categories: natural people (the pilot & the little prince), animals (the fox), and mutants (the rose). Human characters allow the audience to connect with a realistic portrayal of a person, while animal characters minimize the promotion of prejudices such as racial or body discrimination. In addition, the little prince has more concise cartoon features. Finally, mutant characters possessed metaphorical properties that reflected the relationship between humans and other species. This type of character combined the advantages of the first two, allowing viewers to overcome prejudices and establish a stronger connection with the story.

After the performers finished their experience and testing with the VAILSS, we shared the recorded video with each participant to gather their feedback and assist them in completing the post-test. We conducted a questionnaire survey using a five-level Likert scale to classify the responses. The questions were divided into three dimensions, totaling 17 items. The IBM SPSS was applied to analyze the student's performance in the experiment, including the results of three dimensions. The three dimensions encompass preference analysis of virtual characters, satisfaction analysis of the interactive storytelling system, and satisfaction analysis of its impact on foreign language learning.

6 Results

A questionnaire based on a five-point Likert scale was used to collect responses. The factor analysis technique was also applied using SPSS for dimension reduction and drawing conclusions. The results indicated the extraction of four components: system features, user experiences, language learning achievements, and the perception of virtual characters. Based on statistical data from 40 students and 11 items, Cronbach's value of 0.751 exceeds the threshold of 0.7, indicating that the scale exhibits satisfactory reliability and validity.

We employ exploratory factor analysis (EFA) to assess the validity of the user satisfaction questionnaire. Factor analysis is a valuable tool for assessing questionnaires' construct validity and reliability. We condensed a large number of items from 17 to 11. It aids in establishing construct validity and enables us to categorize users' preferences into distinct psychological constructs, facilitating meaningful inferences.

The sampling size of Kaiser-Meyer-Olkin from the KMO and Bartlett test is 0.661, and the significance level of Bartlett's spherical test is $p = 0.000 < 0.05$, which means there is significance. Table 1 shows the KMO and Bartlett's test of the user satisfaction questionnaire.

Table 1. The KMO and Bartlett's test of the user satisfaction questionnaire

Kaiser-Meyer-Olkin measure of sampling adequacy		0.661
Bartlett's test of Sphericity	Approx Chi-Square	137.366
	df	55
	Sig.	0.000*

$*p < 0.05$

Judging from the four items of system feature analysis, we found that the system can facilitate an immersive experience for the audience while catering to their interest in bi-directional interaction. Additionally, the system assists the presenter in reducing performance stress and building confidence.

6.1 The First Component

In the first component, to assess the equivalence between user and system feature knowledge, we employed one-way analysis of variance (ANOVA) to analyze the results of user satisfaction tests conducted on the pre-test, post-test, and control groups. The results showed no significant difference among the pre-test and post-test and the control group ($p = 0.594, 0.639, 0.878, 0.608, > 0.05$). Table 2 shows the ANOVA results of the four items in the first component. Among the system features, we observed that the average value of the fifth item was higher than the rest. The result suggests that users have greater confidence in the role-playing ability to reduce the performer's pressure. Here are the four questions in the first component:

Table 2. The ANOVA results of the four items in the first component.

		Sum of squares	df	Mean square	F	Sig
Item 2	Between groups	0.497	2	0.249	0.527	0.594
	Within groups	23.578	50	0.472		
	Total	24.075	52			
Item 5	Between groups	0.406	2	0.203	0.452	0.639
	Within groups	22.462	50	0.449		
	Total	22.868	52			
Item 12	Between groups	0.114	2	0.57	0.130	0.878
	Within groups	21.886	50	0.438		
	Total	22.000	52			
Item 17	Between groups	0.427	2	0.214	0.502	0.608
	Within groups	21.271	50	0.425		
	Total	21.698	52			

Item 2: "*IN VAILSS, do you think virtual role-playing can improve self-confidence in language expression?*"

Item 5: "*IN VAILSS, do you think acting with virtual characters is less stressful than speaking directly to a crowd?*"

Item 12: "*IN VAILSS, do you think virtual role-playing can be better integrated into a story?*"

Item 17: "*IN VAILSS, do you think audiences will increase their interaction with storytellers because of virtual role-playing?*"

6.2 The Second Component

In the second component, three items emphasize the benefits of virtual role-playing in enhancing user experiences. These items showcase how virtual role-playing can captivate the audience's interest and attention while allowing presenters to enhance their body language skills. The study results showed no significant difference between the pre-test and post-test and the control group (p = 0.156, 0.704, 0.447 > 0.05). Below is a list of three questions in the second component. Table 3 shows the ANOVA results of the three items in the second component.

Table 3. The ANOVA results of the three items in the second component

		Sum of squares	df	Mean square	F	Sig
Item 1	Between groups	2.477	2	1.239	1.932	0.156
	Within groups	32.051	50	0.641		
	Total	34.528	52			
Item 3	Between groups	0.233	2	0.116	0.352	0.704
	Within groups	16.484	50	0.330		
	Total	16.717	52	0.330	0.353	0.704
Item 4	Between groups	0.978	2	0.489	0.819	0.447
	Within groups	29.852	50	0.597		
	Total	30.830	52			

Item 1: "*Do you think virtual avatar/role-playing would make your presentation more interesting?*"

Item 3: "*Do you think virtual avatar/role-playing can increase audience attention?*"

Item 4: "*Do you think virtual avatar/role-playing makes you pay more attention to your facial expressions when speaking?*"

In this section, we also compare the differences in perceptions between male and female groups. We used independent t-test to analyze whether the means of male or female groups were statistically significantly different. Table 4 displays the gender-specific statistics for user experiences of virtual role-playing in the independent t-test.

Table 4. The gender-specific statistics for user experiences of virtual role-playing

	Gender	N	Mean	Std. deviation	Std. error mean
Item 1	Male	17	3.53	0.943	0.229
	Female	23	4.43	0.590	0.123
Item 3	Male	17	3.94	0.556	0.135
	Female	23	4.39	0.499	0.104
Item 4	Male	17	3.76	0.752	0.182
	Female	23	4.13	0.815	0.170

In item 1, the males' mean is 3.53, and the mean for females is 4.43. The study results indicate significant differences in role-playing interest between the male and female groups. Those results were confirmed by conducting an independent t-test, where the obtained value of sig (2-tailed) was 0.001 ($p < 0.05$). Table 5 displays the independent t-test results between male and female students in item 1.

Table 5. The independent t-test results between male and female students in item 1

Item 1 (Gender)	Levene's test for equality of variances		t-test for equality of means						
	F	Sig	t	df	Sig (2-tailed)	Mean Difference	Std.Error Difference	95% Confidence interval of the difference	
								Lower	Upper
Equal variances assumed	2.077	0.158	-3.730	38	0.001*	-0.905	0.243	-1.397	-0.414
Equal variances not assumed			-3.488	25.061	0.002	-0.905	0.260	-1.440	-0.371

*$p < 0.05$

In item 3, the mean score for males is 3.94, while the mean score for females is 4.39. The study results reveal significant differences in audience attention during role-playing between the male and female groups. This finding was confirmed through an independent t-test, where the obtained value of sig (2-tailed) was 0.011 ($p < 0.05$). Based on the results mentioned above, it was observed that women exhibit a higher level of interest in engaging in virtual role-playing activities. The analysis results in Table 6 show that item 3 significantly differ in independent t-test results between male and female students.

Table 6. The independent t-test results between male and female students in item 3

Item 3 (Gender)	Levene's test for equality of variances		t-test for equality of means						
	F	Sig	t	df	Sig (2-tailed)	Mean difference	Std. error difference	95% Confidence interval of the difference	
								Lower	Upper
Equal variances assumed	2.323	0.136	-2.688	38	0.011*	-0.450	0.167	-0.789	-0.111
Equal variances not assumed			-2.644	32.388	0.013	-0.450	0.170	-0.797	-0.103

*$p < 0.05$

6.3 The Third Component

In the third component, the study's results showed no significant difference among the pre-test, post-test and the control group (p = 0.100, 0.054 > 0.05). Students generally agree with the effectiveness of language learning achievement. Below is a list of two questions in the third component. Table 7 shows the ANOVA results of the two items in the second component.

In item 15: "*Can virtual avatar/character/role-playing replace traditional drama performances?*"

In item 16: "*Does virtual role-playing have value in language learning?*"

Table 7. The ANOVA results of the two items in the second component

		Sum of squares	df	Mean square	F	Sig
Item 15	Between groups	4.415	2	2.208	2.415	0.100
	Within groups	45.698	50	0.914		
	Total	50.113	52			
Item 16	Between groups	3.071	2	1.535	3.103	0.054
	Within groups	24.741	50	0.495		
	Total	27.811	52			

6.4 The Fourth Component

In the fourth component, to explore the influence of the perception of virtual characters on students, a one-way analysis of variance (ANOVA) was used to exclude the difference among the pre-test and post-test and the control group. Below is a list of two questions in the fourth component.

In item 9: "*Do you think cartoon characters are more attractive than realistic ones?*"

In item 11: "*Do you think animal characters are more relatable to the audience?*"

In item 11, the results showed significant differences between the post-test and control groups (audience). This finding was confirmed through an independent *t*-test, where the obtained value of sig (2-tailed) was 0.029 ($p < 0.05$). "*Do you think animal characters are more relatable to the audience?*" Table 8 presents the Post Hoc test results with ANOVA about the perception of virtual characters.

Table 8. The Post Hoc test with ANOVA results about the perception of virtual characters

	(I) Group	(J) Group	Mean difference (I-J)	SD	P	95% Confidence interval	
						Lower bound	Upper bound
Item 11	Post-test	Pre-test	0.308	0.349	0.680	-0.57	1.19
		Control	0.826	0.300	0.029*	0.07	1.58

*$p < 0.05$

The control group, which did not participate in virtual role-playing, obtained a mean score of 3.48. In contrast, the post-test group, which engaged in virtual role-playing, achieved a higher mean score of 4.31. This case highlights that students perceive it easier to cultivate friendly relationships with others by utilizing animal characters when engaging in virtual role-play performances. Table 9 presents the group statistics for the post-test group and control group comparison.

Table 9. The group statistics for the post-test group and control group comparison

	Group	N	Mean	Std. deviation	Std. error mean
Item 11	Post-test	17	4.31	0.751	0.208
	Control	23	3.48	0.975	0.188

Furthermore, our findings indicated significant differences in animal character preferences between male and female groups. The females display a greater interest in selecting animal characters for role-playing. This finding was confirmed through an independent *t*-test, where the obtained value of sig (2-tailed) was 0.001 ($p < 0.05$). Table 10 shows the results of independent *t*-test for animal character preferences between male and female groups.

Table 10. The results of independent *t*-test for animal character preferences between male and female groups

Item 11 (Gender)	Levene's test for equality of variances		t-test for equality of means						
	F	Sig	t	df	Sig (2-tailed)	Mean difference	Std. error difference	95% Confidence interval of the difference	
								Lower	Upper
Equal variances assumed	2.113	0.154	-3.648	38	0.001*	-0.997	0.273	-1.551	-0.444
Equal variances not assumed			-3.789	37.875	0.001	-0.997	0.263	-1.530	-0.464

*$p < 0.05$

7 Conclusion

We introduce VAILSS, a multimodal storytelling system for language learning inspired by virtual role-play storytelling. VAILSS aims to foster a natural speaking situation, alleviate presenter stress, and enhance confidence. The outcomes of the classroom workshop validate the effectiveness of our approach in achieving these objectives while shedding light on various intriguing user experiences and virtual role-playing behaviors. These findings will guide the enhancement of the application and the assessment of user experience. The research findings can be summarized as follows:

1. The system retains control of the original story (Authorial Control)
2. The system provides players to explore story situations and reflections
3. When the storyteller plays different narrative characters, the story is given different interpretations and reflections
4. The system's multimodal display effectively aids in enhancing emotional performance.
5. Virtual character emotions influence character identity and persuasiveness
6. The system's bi-directional interaction feedback makes language learning more natural.

Acknowledgements

Special thanks to the Ministry of Science and Technology, Taiwan, R.O.C. for the grant support and to all personnel involved in developing and experimenting with this project in 2022.Project number: MOST 111-2637-H-150-001.

References

[1] K. Egan, *Teaching as story telling: An alternative approach to teaching and curriculum in the elementary school*, University of Chicago Press, 1989.

[2] R. Barthes, "The Death of the Author", in: S. S. Elliott, M. Waggoner (Eds.), *Readings in the Theory of Religion*, Routledge, 2016.

[3] H. Y. Wu, M. Christie, T. Y. Li, Stories animated: A framework for personalized interactive narratives using Filtering of Story Characteristics, *CASA 2013 - The 26th International Conference on Computer Animation and Social Agents*, Istanbul, Turkey, 2013.

[4] R. Hadidi, C. H. Sung, Pedagogy of online instruction-Can it be as good as face-to-face?, *AMCIS 2000 Proceedings*, Long Beach, CA, 2000, Article No. 288.

[5] J. A. Ollerenshaw, R. Lowery, Storytelling: Eight steps that help you engage your students, *Voices from the Middle*, Vol. 14, No. 1, pp. 30-37, September, 2006.

[6] D. Markouzis, G. Fessakis, Interactive storytelling and mobile augmented reality applications for learning and entertainment—a rapid prototyping perspective, *2015 International Conference on Interactive Mobile Communication Technologies and Learning (IMCL)*, Thessaloniki, Greece, 2015, pp. 4-8.

[7] B. Lufkin, *From Japanese anime characters to barbie, virtual YouTubers talk and act just like people and they could change the way we all interact forever*, BBC Worklife, October, 2018.

[8] Z. Zhao, F. Han, X. Ma, A live storytelling virtual reality system with programmable cartoon-style emotion embodiment, *2019 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, San Diego, CA, USA, 2019, pp. 102-109.

[9] M. C. Costa, P. Santos, J. M. Patrício, A. Manso, An interactive information system that supports an augmented reality game in the context of game-based learning, *Multimodal Technologies and Interaction*, Vol. 5, No. 12, Article No. 82, December, 2021.

[10] Y. T. Lin, J. H. Ye, Development of an Educational Chatbot System for Enhancing Students' Biology Learning Performance, *Journal of Internet Technology,* Vol. 24, No. 2, pp. 275-281, March, 2023.

[11] A. Tse, C. Jennett, J. Moore, Z. Watson, J. Rigby, A. L. Cox, Was I there? Impact of platform and headphones on 360 video immersion, *Proceedings of the 2017 CHI conference extended abstracts on human factors in computing systems*, Denver, Colorado, United States of America, 2017, pp. 2967-2974.

[12] N. De la Pena, P. Weil, J. Llobera, E. Giannopoulos, A. Pomés, B. Spanlang, D. Friedman, M. V. Sanchez-Vives, M. Slater, Immersive journalism: Immersive virtual reality for the first-person experience of news, *Presence: Teleoperators and Virtual Environments*, Vol. 19, No. 4, pp. 291–301, August, 2010.

[13] F. Lu, F. Tian, Y. Jiang, X. Cao, W. Luo, G. Li, X. Zhang, G. Dai, H. Wang, ShadowStory: Creative and collaborative digital storytelling inspired by cultural heritage, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Vancouver, BC, Canada, 2011, pp. 1919-1928.

[14] T. N. Turner, T. Oaks, Stories on spot: Introducing students to impromptu storytelling, *Childhood Education*, Vol. 73, No. 3, pp. 154-157, 1997.

[15] X. Ma, J. Forlizzi, S. Dow, Guidelines for depicting emotions in storyboard scenarios, *Proceedings of 8th International Design and Emotion Conference*, London, England, 2012, pp. 1-12.

[16] M. Forte, *Virtual reality modeling*, The Encyclopedia of Archaeological Sciences, 2018.

[17] E. Zell, K. Zibrek, R. McDonnell, Perception of virtual characters, *ACM Siggraph 2019 Courses*, Los Angeles, CA, USA, 2019, pp. 1-17.

Biographies

Hong-Yi Pai is an assistant professor in the Department of Multimedia Design at National Formosa University, Taiwan. He received his MFA degree in Computer Graphics at Pratt Institute, USA, in 1998. His research interests include multimodal learning, AR/VR displays and interactive 3D.

Ching-Huang Wang is a professor in the Department of Applied Foreign Languages at National Formosa University, Taiwan. He received his Ph.D. degree in Language Education from Indiana University at Bloomington, USA, in 2002. His research interests include TESOL, bibliotherapy, identity, critical literacy, course activities, and cultural studies.

Yi-Chen Lai is a graduate student of the Department of Multimedia Design at National Formosa University, Taiwan. He received his bachelor's degree of Science in Computer Science and Information Engineering at WuFeng University, Taiwan, in 2021. His research interests include interactive design and programming.