Multimodal Data Integration for Interactive and Realistic Avatar Simulation in Augmented Reality

Anchen Sun*, Yudong Tao*, Mei-Ling Shyu*, Shu-Ching Chen†, Angela Blizzard‡,
William Andrew Rothenberg‡, Dainelys Garcia‡, Jason F Jent‡
*Department of Electrical and Computer Engineering
University of Miami, Coral Gables, FL, USA
†Knight Foundation School of Computing and Information Sciences
Florida International University, Miami, FL, USA
‡Mailman Center for Child Development
University of Miami, Coral Gables, FL, USA
Email: a.sun158@umiami.edu, yxt128@miami.edu, shyu@miami.edu, chens@cs.fiu.edu,
amb208@med.miami.edu, war37@miami.edu, ngarcia09@med.miami.edu, jjent@med.miami.edu

***Abstract*—The affordance of Augmented Reality (AR) allows users to view the virtual objects along with real world simultaneously, which provides more realistic and immersive experiences to interact with them. In this paper, we propose to integrate data with different modalities (such as animation, audio, and user behavior data) to generate a realistic child avatar in the AR environment that is responsive to the user's behaviors. The proposed framework has three components: avatar interaction system, avatar action control system, and avatar display system. The avatar interaction system leverages user behavior data to allow for reasonable interactions of the avatar, the avatar action control system integrates actions and audio data to generate realistic avatar actions, and the avatar display system presents the avatar to the user via the AR interface. Furthermore, a child tantrum management training application is implemented based on the proposed system to allow users to experience child tantrum and learn how to respond and manage different child tantrum situations. The application based on our system can run in real-time with an average of 93.41 fps. Based on the qualitative evaluation, the simulated child avatar has shown to be realistic enough and is able to respond to the user's gaze instantly in the AR environment. The action, reaction, treatment, and mitigation of the child avatar behavior among different tantrum levels in the avatar action controller are accurately represented through the clinical evaluation experiences from pediatricians and trained psychologists.**

***Index Terms*—Multimodality, Integration, Augmented Reality (AR), Immersive Media, Child Tantrum**

I. Introduction

Avatar simulation is the backbone of programmatically building intelligent virtual avatars in the Augmented Reality (AR). A realistic and interactive avatar could allow users to quickly step into a virtual environment and gain a sense of immersion [1]. The subtle expressions, movements, sounds, and language can be generated to simulate an avatar as a real person through algorithms [2]. This allows AR environments to turn what was once a flat text interaction into a three-dimensional image or animation to help users have a more immersive and direct interactive experience. At the same time, AI assistants, intelligent voice customer service, and other intelligent tools can be presented in the real world using three-dimensional animations and interacted with the real-world environment through AR technology [3], [4].

Following the development of avatar simulation in an AR environment, more and more attention is being focused on the application of these new technologies in education. As we know, educating and learning how to raise young children has always been a challenge in our society. Caring for children during early childhood is especially important because children in these age ranges are prone to experiencing intense emotions and do not have the language skills to verbalize their needs [5]. Having specialized training to deal with kids is critical. For teachers, dealing with challenging child behaviors is related to turnover in the early childhood setting, so providing teachers with the skills to handle these behaviors has the potential to improve teachers' job satisfaction [6]. Special education teachers may have to work with children with specific disorders that result in unique and frequent tantrums. Similarly, developmental-behavioral pediatricians provide common guidance for parents on child behavior issues. For example, the conditions in children such as Attention-Deficit/Hyperactivity Disorder (ADHD) or Autism Spectrum Disorders (ASD) are often treated with behavioral therapy [7], requiring the parents to undergo training from the doctor on how to react to specific actions from their children, such as a tantrum or throwing objects. However, the training for pediatricians in developmental and behavioral pediatrics is limited, and research indicates a need for improvement in training clinicians in developmental-behavioral areas [8].

Real-world practice is critical to mastering the evidence-based skills associated with managing child behaviors, but it is not feasible to create an environment for this, especially for challenging or unsafe behaviors [9]. Therefore, simulating a child behavior allows for on-demand training opportunities with a diverse array of tense situations that do not exist in real-world settings. Behaviors to be simulated, such as tantrums, often include aggressive and disruptive behaviors [10], and it is not feasible or safe to simulate a real life tantrum between a child and a caretaker.

In recent years, although virtual avatars have been gradually applied in the fields of education [11], healthcare [12], and entertainment [13], the existing applications have significant limitations. Most of the existing AR and VR healthcare and nursing education applications employ pre-designed movements and fixed processes of virtual avatars [11], [14], which fails to provide a good sense of immersion to users [15]. Hence, it is desirable to design and develop a simulation system that can help simulate different behaviors of the virtual avatar in AR so that it can be better used and provides the users realistic experiences.

In this paper, a realistic and interactive avatar simulation based on multimodal data integration is proposed, where animation, audio, and user behavior data are integrated to provide users a realistic experience interacting with the avatar. The spatial immersion tool used in the proposed simulation system is the AR glasses. AR offers a responsive way to leverage technologies to teach the caregivers practical behavior management skills more efficiently than in a behavioral training program. By using AR, we can create a life-like familiar or unfamiliar environment that can mimic a real-world situation while still maintaining consistent control variables through the AR environment. To demonstrate how our proposed system can be utilized, child tantrum management is selected in this study and the simulation exhibits typical behavior symptoms of temper tantrums for young children reacting to the user's gaze (attention). By simulating a young child in a familiar or unfamiliar environment, we are creating a life-like scenario of a child's tantrum, similar to what they would encounter in person. In the simulation, the child can react to the user's gaze with the desired response. Pediatricians and psychologists have professionally evaluated the system's settings for the child and tantrum behavior, movements, sounds, and reactions at different tantrum levels.

The rest of the paper is organized as follows. Section II discusses the related work. In Section III, the design of our proposed multimodal data integration framework for avatar simulation is presented. Section IV describes the evaluation and practical experience of the avatar simulation system. Finally, the limitations of the existing system and the future directions are given in Section V.

II. Related Work

A. Avatar Simulation in Immersive Environments

The virtual avatar is an integral part of the immersive world experience. Virtual avatars extend our sense of self on multiple levels, whether physical, physiological, or psychological [13]. It is no exaggeration to say that the virtual avatar transforms our notions of identity and self-expression into the uncharted territory that we cannot understand or fully comprehend. The virtual avatar is not only a tool for us to enter the world of immersion, but also a guide through which a new generation of intelligent characters will be guided into the world of AR.

Over the past few years, we have witnessed the growth of an entire industry spawned by virtual celebrities such as Brud's Lil Miquela [16] and Activ8's Kizuna AI [17]. Immersive media like AR and VR can provide these virtual personas with the level of interaction they need to augment the virtual persona. An approach combines immersive multimedia and virtual-reality had been proposed in communication system design [18].

B. Augmented Reality (AR) for Education

AR can change the time and place of education and dramatically improve the efficiency of education [19]. AR applications change the way students understand knowledge in a way that traditional teaching cannot. The potential of AR in education is mainly reflected in: (1) visualizing and visualizing abstract learning content; (2) supporting contextual learning in an ubiquitous environment [20]; (3) enhancing learners' presence, intuition, and concentration [21]; (4) interacting with learning objects using natural methods; and (5) combining traditional and new types of learning.

One of the earliest projects in the field of AR is "Handheld Augmented Reality" [22], which was made possible through a grant from the U.S. Department of Education. It has been found that AR technology can improve student performance and increase student interest in learning [23]. Immersive multimedia has also been used in recent studies in the field of education to help students gain experiences that are difficult to accumulate in reality [24].

C. Child Tantrum and Child Tantrum Management

Up to 70% of children between the ages of 18-months and 5 years experience tantrums [25], [26]. While tantrums are common in early childhood, it is difficult for caregivers to manage these behaviors and also difficult to ascertain when tantrum behaviors indicate a serious behavioral problem [9]. Healthy children appear to experience tantrums that are shorter in duration and less severe in expression, compared to children with disruptive behavior disorders [9]. Children with disruptive behavior disorders were reported to show tantrums with more violent, self-injurious, and destructive behaviors [9].

Behavioral parent training programs are considered highly effective for helping caregivers manage challenging childhood behaviors, but these programs are often underutilized due to various logistical, attitudinal, and system-level barriers [27]. Ongoing engagement in behavioral parent training is also a challenge, with estimates of up to 50% of caregivers starting, but not completing, the program [28]. Behavioral parent training delivered via technology is increasing in popularity as a potential means to improve access to evidence-based techniques to manage challenging childhood behaviors [29].

III. Multimodal Data Integration for Avatar Simulation IN AR

In order to create a realistic avatar in AR, it is necessary to leverage data with different modalities [30] and make the avatar behave and interact with the user in a natural manner. To this end, a multimodal data integration framework for avatar simulation in AR is proposed, which consists of three components: 1) Avatar Interaction System (AIS), 2) Avatar

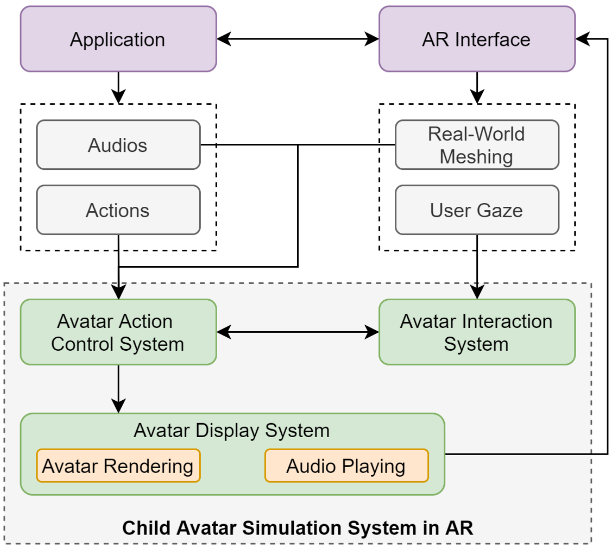


Fig. 1. The proposed multimodal data integration framework for avatar simulation in AR

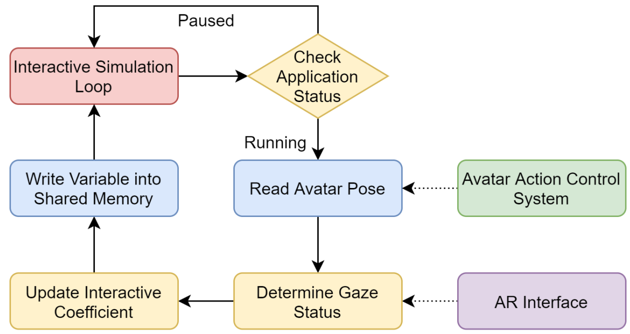


Fig. 2. The flowchart of the Avatar Interaction System

Action Control System (AACS), and 3) Avatar Display System (ADS). The AIS determines the internal states of the child avatar by integrating user behaviors and avatar status. Then, the AACS schedules and simulates the action sequence of the child avatar through integrating avatar's internal state, actions and audios. In the end, the ADS renders the avatar, which are displayed to the user via the AR interface. These components run simultaneously and independently. The AACS also integrates the meshing of the real world from the AR interface to facilitate the realistic action scheduling, accounting for constraints in the real world. Furthermore, the set of available child avatar actions, their animations and sound effects, the set of internal states of the avatar, and the relations between actions and internal states need to be designed based on the applications. Fig. 1 depicts the proposed multimodal data integration framework for avatar simulation in AR.

In the following, we detail and explain the design of the novel components in the proposed child avatar simulation system for a child tantrum management training application. This application is designed to present realistic behaviors of a child in tantrum to the users and simulate the dynamics of tantrum development according to users' behaviors. Thus, it can be utilized to help new parents and healthcare providers to learn how the child tantrum can be managed.

A. *Avatar Interaction System*

The Avatar Interaction System (AIS) is designed to interact with the user and models the relationship between external user factors and the internal status of the simulated avatar. Therefore, the user behavior data needs to be collected and integrated with the avatar's status to determine how the avatar should interact with and react to the user. Specifically, for child tantrum management training, the internal status is modeled as a tantrum coefficient which reflects the level of tantrum of the avatar and impacts the avatar's behavior pattern [10]. Non-contingent gaze has been found to be helpful to manage the tantrum [28], the proposed AIS acquires the position of the user's gaze via the AR interface, obtains the avatar position and pose from AACS, and infers the changes of tantrum coefficient accordingly. The flowchart of the AIS is shown in Fig. 2. When the user's gaze is over-focused or not focused on the child avatar, it means that the user is giving excessive or insufficient attention to the young child's tantrum. This causes the child avatar to either notice that the adult is watching him or feel that there is no way to get the adult's attention. In both cases, the tantrum coefficient increases. Conversely, if the user's attention to the young child avatar is directed at the right place, the tantrum coefficient decreases. Furthermore, since the eye gaze data streams at a different frequeny from the avatar simulation, the tantrum coefficient is updated asynchronously. To allow the AACS to access the most recent internal status and to avoid simultaneous data read and write, a shared memory with the mutual exclusion lock is implemented to store the most updated internal status, which allows for asynchronous communications between AIS and AACS. When a new internal status of the child avatar is obtained, the AIS will write it into the shared memory for the AACS to access.

Additionally, the young child avatar will have a ball icon on his head to indicate whether the user's eyes are in the optimal position to calm the tantrum. The state of the avatar is checked every frame, and the gaze detection module will run when the avatar is running. If the user's gaze is directed at the young child's head, the ball will appear *red* to indicate that the child will perceive that it is receiving undue attention, which is not conducive to controlling the young child's tantrum behavior. If the user's gaze is too intense for the child avatar, the ball will display *yellow* to inform the user that the young child's tantrum cannot be ignored. The *green* color indicates that the user's eyes are in the optimal position for observing the young child's behavior and not acknowledging the child's tantrum.

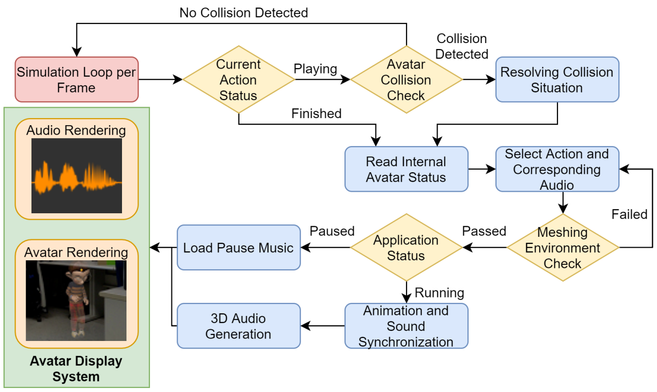


Fig. 3. The flowchart of the avatar action control system

The gaze detection can be summarized as follows.

$$\vec{D}_{Gaze} = P_{Fixation} - P_{Start} \quad (1)$$

$$P_{Gaze} = \begin{cases} P_{hit} & \text{if gaze ray hit} \\ P_{Start} + 2\vec{D}_{Gaze} & \text{otherwise} \end{cases} \quad (2)$$

$$Ray_{Gaze} = Ray(P_{Start}, \vec{D}_{Gaze}) \quad (3)$$

where $P_{Fixation}$ is the point user fixation on, P_{Start} is the point of the main camera (the location of the AR Glass and user), and the P_{Gaze} is the point of gaze ball display in the simulation system. The maximum distance is 10 for hit detection. If the Ray_{Gaze} hits the head or hair of the avatar, the gazing status ball shows *red*. If the Ray_{Gaze} hits other parts or close area of the avatar, the gazing status ball shows *green*. Otherwise, the gazing status ball shows *yellow*.

B. Avatar Action Control System

For the child tantrum simulation, a set of regular actions and tantrum behaviors is included. The regular behaviors include walking, turning, running, jumping, etc., while the tantrum behaviors were determined based on research literature examining tantrum behavior. Temper tantrums are usually defined by physical behaviors (for example, throwing oneself to the floor, hitting) and vocalizations (screaming and crying) as shown in Table I. In temper tantrums, verbal aggression is usually not sophisticated [5].

This AACS is used to control the actions, movements, and sounds of the avatar, where multiple judgements are made and corresponding action scheduling and adjustment are performed at each frame to ensure the smoothness, realism and stability of the simulation. As shown in Fig. 3, the AACS will first determine whether the current action has completed and select the next action and the corresponding audio according to the internal avatar status. Afterwards, the AACS will also check whether the selected action is feasible in the real-world by integrating the mesh data. Based on the infrared light sensor and front-view camera deployed on the AR glasses, a 3D mesh characterizing the real-world space can be generated. Based on the world mesh, it can be determined whether there is enough space for the child avatar to perform the selected action. When

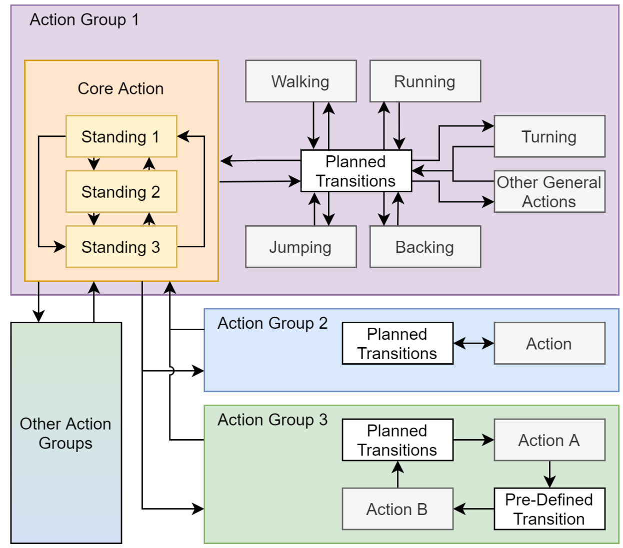


Fig. 4. The designed animator in the avatar action control system

the selected action is feasible, the AACS will synchronize the animation and the audio, generate the audio in the 3D space, and then inform the ADS about the subsequent avatar poses and the corresponding audio to render and display. Also, when the avatar is animating, the AACS performs a collision check based on the real-time 3D mesh data to cope with the dynamic changes in the real world. In the event of a collision, the AACS immediately forces the avatar to pose a standing action to avoid potential collisions and then re-schedules the action to be deployed. During child tantrum management training, the user can easily feel upset due to the continuous crying of the child avatar, so the user can pause the application and a comforting music is played to relieve the emotional tension.

1) Animator: A reasonable and smooth transition between different actions is an important basis for making the overall simulation look realistic. Hence, an animator that generates animation based on the selected action of the child avatar is shown in Fig. 4, where all the actions are fused with the planned transitions. A core action needs to be established in the designed animator to transfer animations between each action loop. The standing movement is set as the core action in our proposed animator. When the AACS needs to stop the animation being played in any situation, the avatar will automatically transition back to the core action. For example, when the avatar is walking forward, an object moves into the avatar's forward path and is detected by the AR glasses. At this time, the avatar will stop the walking action and automatically return to the standing core action through the animator to avoid the potential collision. To make the standing animation look more realistic than a typical simulation, we use three different standing animations played randomly, which makes the presented avatar have some random small movements like a normal child even if it stands in the same place.

Simultaneously, we have also divided the different action loops among the animator into large, small, and general action

TABLE I
CHILD TANTRUM BEHAVIORS

Tantrum Level	**Tantrum Coefficient**	**Behavior Characters**
Pre-Tantrum	0	Smiling, Talking (Neutral statements)
1	1-20	Sighing, Pouting
2	21-40	Holding breath, Covering face, Places head on floor, Crossing arms whining
3	41-60	Whining, Crying (Actual tears, heavy breathing), Throwing themselves on floor
4	61-80	Crying, Screaming (Similar phrases to whining), Flailing around, Throwing themselves on floor
5	81-100	Crying, Screaming, Squatting Down and Weeping

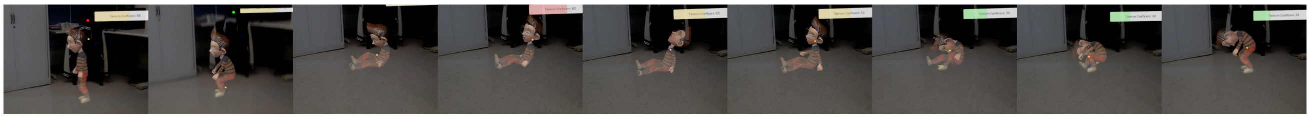


Fig. 5. Example of a series of actions within the same group (sitting on the floor and rolling around). These figures show the step-by-step animation from left to right.

loops. We use the implemented child tantrum avatar simulation application as an example to describe our proposed animator as follows. A small loop is a standard single tantrum action such as crying or getting angry, connected to the core action (standing) by an excessive action. The large loop is a series of tantrum actions such as sitting down after getting angry and rolling around on the floor before standing up (as shown in Fig. 5). The reason for designing large loops is to increase the coherence of the simulation so that the avatar can perform a series of continuous simulation actions. Because of the complexity of the large loop, multiple transitions need to be generated in between to make the whole loop look natural and vivid. Simultaneously, the large loop takes into account the subsequent actions of the child avatar after getting angry, and thus the overall length is longer. As a result, the simulated tantrum of the child avatar is more vivid. Both the large and small loops are controlled by the aforementioned ACS at different tantrum levels and implemented by the designed animator.

Another important part of the animator is the general action loop. When the tantrum level is low, the child avatar will move in the simulated space part of the time and perform some regular actions such as walking, running, jumping, looking to the right and left, talking to itself, etc. These routine actions are played randomly according to the space of the mesh environment. They are connected in a large network in the animator and are connected to all other general actions through the excessive action. The proposed animator enables all general actions to be switched smoothly at any time. This design realistically simulates the regular behavior of children in space and provides a good connection between small and large loops of the tantrum actions.

Before each animation is played, the current environment of the avatar is checked to see if it is suitable for the next animation. If the space, where the avatar is currently located, is not suitable for playing the animation, the system will block the animation and select the next one. At the same time, for the articulation between complex actions, an intermediate standing action was set up to make the articulation look more natural.

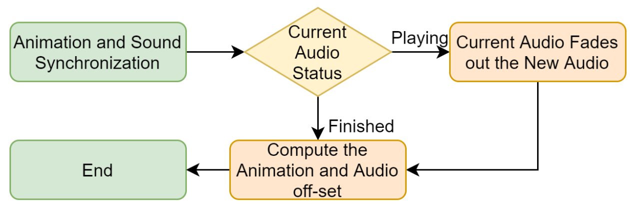


Fig. 6. The flowchart of an animation and sound synchronization scheme, synchronizing audio with animation

2) Animation and Sound Synchronization: After solving the problem between action articulation, the synchronization between sound and animation is also a challenge, which is led by the following reasons: (1) The same action may match different cries in the avatar action system to add a variety and realism aspect; (2) In the simulations, sometimes the young child avatar needs to switch actions frequently to adapt to the existing position and react appropriately because of the terrain. (3) There are two sound players on the avatar's feet and head to simulate the sound more realistically, and these two players need to play different sounds simultaneously or separately when the avatar performs different actions. All of the above reasons make it difficult to synchronize the corresponding sounds during a long time simulation. To solve this problem, a checking and synchronizing scheme called Animation and Sound Synchronization is designed as Fig. 6 in the avatar action system after the integration of animations and sounds. In this scheme, a synchronizer is designed to control the synchronization of the animation and corresponding audio.

3) 3D Audio Generation: We set up a sound player in the avatar's head and one in the avatar's feet to simulate the sound being generated in different parts of the avatar's body. A 3D sound rendering with settings applied proportionally to the Spatial Blend parameter is also applied to both sound players to render the sound close to the real space in real-time according to the user's position [31]. The 3D audio rendering scheme is shown in Fig. 7.

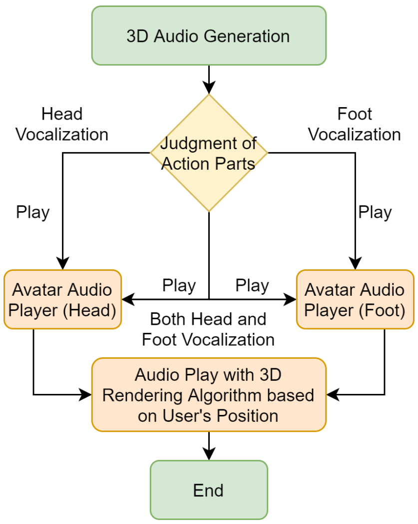


Fig. 7. The 3D Audio Rendering scheme

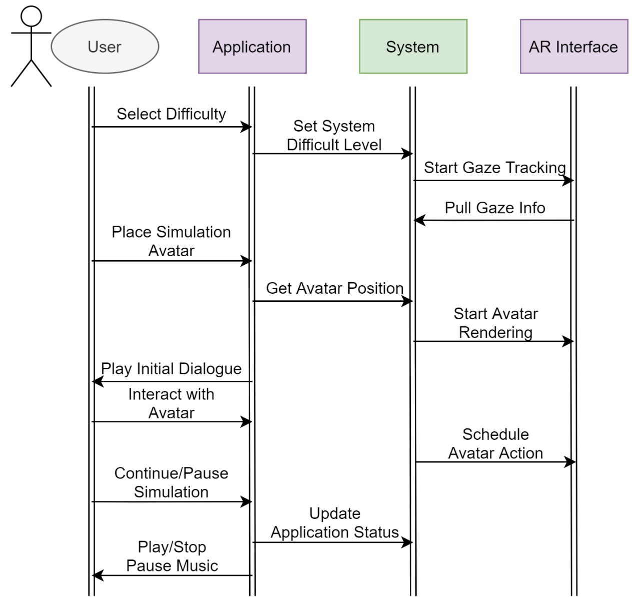


Fig. 8. Sequence Diagram of the Implemented Child Tantrum Simulation Application. Solid lines mean manual processes and dashed lines refer to automated processes.

IV. Evaluation and Analysis

The proposed system was implemented in Unity 3D using C# with OpenGL and Lumin OS v0.98.11 and then was built to the Magic Leap One Creator Edition for testing via TheLab.

A. User Interface Design

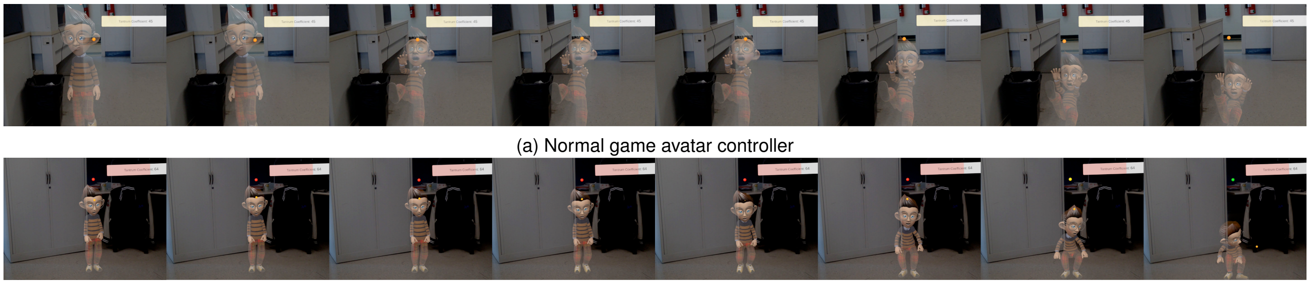
In our implemented application, we design a user interface in the application to communicate with the user. Fig. 8 shows the sequence diagram of our implemented child tantrum simulation application based on our proposed framework. All pages that interact with the user are presented at a fixed angle to the user in the AR interface and follow the user's movement. As shown in Fig. 8, to enable the interaction between the user and the application, a difficulty level selection mechanism is designed and developed. Then the user can place the child avatar through the gaze in the application and listen to the initial dialogue from the child avatar. Once the initial dialogue is finished, the user can interact with the child avatar by the gaze and pause simulation in the application at any time when the user wants with the controller. Also, the user can pause at any time by pressing the home button on the control pad and a soothing music will be played to relax the user. During the simulation, there is a preemptive action that avatar has a 60% chance to rotate in the direction of the user before crying, sitting down, getting angry, or playing any other special animation. This allows the user to better interact with the simulation and observe the child avatar's movements and facial expressions. Once the user reaches any of the end conditions, the simulation will end and the corresponding end text will be displayed. The tantrum progress bar can display the tantrum coefficient and different colors are used to show the current tantrum level (blue for Lv.1, green for Lv.2, yellow for Lv.3, red for Lv.4, and purple for Lv.5 in the current system).

B. Avatar Simulation Evaluation

To demonstrate that our proposed simulation system can control the avatar according to the real environment of the meshing coherently and naturally to make a variety of movements, we implemented two child tantrum simulation applications, one by the proposed avatar action control system, and one by a normal game avatar controller. We compare the reactions and actions of the child avatar simulated by the two applications when different situations are encountered during the simulation and demonstrate that our proposed system solves some of the challenges in the AR avatar simulation. As shown in Fig. 9, it can be clearly seen that the animations in the first row of the actions are relatively rigid with incoherent transitions, lacking a sense of immersion especially between the second and third frames. While the animations in the second row generated by our proposed child avatar simulation system, it is clear that the actions are delicate and vivid articulation, more in line with the human structure of a child, with a good sense of immersion.

C. Real-Time Performance Evaluation

To evaluate the real-time performance of our proposed system, it is compared with an animation controller in the Unity Asset Store (called the baseline) using the Frame Per Seconds (FPS) values. The FPS value is recorded for one minute after the avatar is placed in the simulation environment and starts walking. Five trials for the proposed system and the baseline were conducted, and the results are presented in Table II. As can be seen from Table II, the average frame rate of our proposed system is 19.25% higher than that of the baseline because of our design and optimization. This is due to the facts that our proposed system designs efficient overall frame generation flow and judgment logic in the Avatar Action Control System and the Interactive Action System, and incorporates the Animation and Sound Synchronization to



(a) Normal game avatar controller

(b) **Proposed child avatar simulation system**

Fig. 9. Examples of the fall down action comparison of the two child tantrum simulation applications. The examples in the first row are generated by the child avatar tantrum simulation application implemented by using a normal game avatar controller, and the examples in the second row with bolded title are generated by the child avatar tantrum simulation application implemented by our proposed child avatar simulation system. These figures show the animation by frames from left to right.

TABLE II
FPS PERFORMANCE TEST RESULTS

Version	Proposed System	Baseline
1st Quartile FPS	80	69
99th Quartile FPS	98	86
Average FPS	93.41	78.33
Standard Deviation FPS	2.81	3.49
Median FPS	94	79

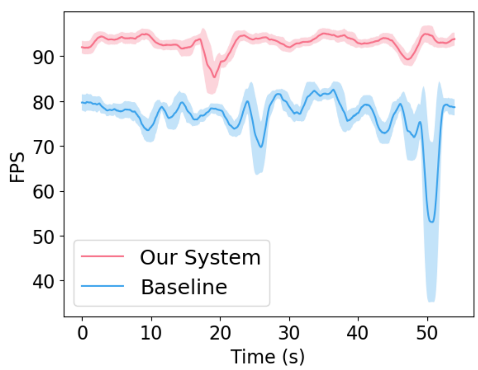


Fig. 10. Real-Time FPS performance evaluation (The line is the average FPS, the upper boundary is the maximum FPS, and the lower boundary is the minimum FPS)

avoid potential out-of-sync, lags, and bugs during animations and audio transitions. It can also be observed that the 1st Quartile FPS and 99th Quartile FPS of the proposed system is 15.94% and 13.95% higher than those of the baseline through the statistical analysis of the FPS test data. The overall frame rate dispersion is also relatively low, signifying that our proposed system can generate frames more smoothly and can significantly reduce the lags in low FPS. Also, the FPS of our proposed and baseline system over an one-minute session is plotted in Fig. 10. It can be observed that our system can achieve higher FPS, as well as lower variance in FPS, compared to the baseline system. This indicates that our proposed multimedia system can efficiently and robustly produce realistic and interactive child avatar simulations.

D. Analysis

This project was conducted in collaboration with pediatricians and psychology professors who were involved in the design and testing of the child avatar tantrum simulation system and provided their expert advice to make the current version more medically sound. The psychology team, including three child-clinical psychologists, reviewed some demos to determine the merits of the simulation system.

Furthermore, in our proposed system, the child avatar accurately demonstrates the behaviors of an escalating tantrum. That is, a tantrum of a lower intensity includes whining and pouting at a lower volume; while a tantrum of a higher intensity includes physically aggressive behaviors and crying at a louder volume. The child avatar demonstrates the ranges and varieties of child tantrum behaviors. All these demonstrate that the user gaze impacts the child avatar's tantrum behavior, in the way that when the user looks at the avatar, the tantrum is escalated. The voice and the tantrum behaviors were realistically aligned to the presentation of child tantrum behaviors.

V. CONCLUSION AND FURTHER WORK

In this paper, a novel multimodal data integration framework is proposed for avatar simulation in AR, which leverages data with different modalities and integrates them to generate realistic and interactive avatar simulations. A user study of child tantrum simulation is conducted based on the proposed system and the child avatar is able to mimic child's behaviors in tantrum condition and interact with the human user, reacting to the gaze. A system performance evaluation has illustrated the effectiveness of the proposed system.

The implications of the developed avatar simulation in AR based on multimodal data integration are vast. This proposed system can be further extended to include more obstacles and symptoms of young children, such as impairments, disabilities, ADHD, ASD, or other complicated factors in child development. Given the immersive AR simulation's utility, it can be incorporated into the curricula of healthcare, teachers, childcare workers, camp counselors, and others who regularly interact with children while more evaluation in training and clinical can be conducted in the future.

VI. Acknowledgments

This work is partially supported by Clinical and Translational Science Institute at the University of Miami.

References

[1] A. Shapiro, A. Feng, R. Wang, H. Li, M. Bolas, G. Medioni, and E. Suma, "Rapid avatar capture and simulation using commodity depth sensors," *Computer Animation and Virtual Worlds*, vol. 25, no. 3-4, pp. 201–211, 2014.

[2] S. Yang and B. Bhanu, "Facial expression recognition using emotion avatar image," in *IEEE International Conference on Automatic Face & Gesture Recognition*, 2011, pp. 866–871.

[3] S. Pouyanfar, Y. Yang, S.-C. Chen, M.-L. Shyu, and S. S. Iyengar, "Multimedia big data analytics: A survey," *ACM Computing Surveys*, vol. 51, no. 1, pp. 10:1–10:34, 2018.

[4] S. Pouyanfar, S. Sadiq, Y. Yan, H. Tian, Y. Tao, M. Presa Reyes, M.-L. Shyu, S.-C. Chen, and S. S. Iyengar, "A survey on deep learning: Algorithms, techniques, and applications," *ACM Computing Surveys*, vol. 51, no. 5, pp. 92:1–92:36, 2019.

[5] K. Österman and K. Björkqvist, "A cross-sectional study of onset, cessation, frequency, and duration of children's temper tantrums in a nonclinical sample," *Psychological Reports*, vol. 106, no. 2, pp. 448–454, 2010.

[6] D. D. Schaack, V.-N. Le, and J. Stedron, "When fulfillment is not enough: Early childhood teacher occupational burnout and turnover intentions from a job demands and resources perspective," *Early Education and Development*, vol. 31, no. 7, pp. 1011–1030, 2020.

[7] R. L. Goldin, J. L. Matson, K. Tureck, P. E. Cervantes, and J. Jang, "A comparison of tantrum behavior profiles in children with ASD, ADHD and comorbid ASD and ADHD," *Research in Developmental Disabilities*, vol. 34, no. 9, pp. 2669–2675, 2013.

[8] R. E. K. Stein, A. Storfer-Isser, B. D. Kerker, A. Garner, M. Szilagyi, K. E. Hoagwood, K. G. O'Connor, C. M. Green, and S. M. Horwitz, "Does length of developmental behavioral pediatrics training matter?" *Academic Pediatrics*, vol. 17, no. 1, pp. 61–67, 2017.

[9] A. C. Belden, N. R. Thomson, and J. L. Luby, "Temper tantrums in healthy versus depressed and disruptive preschoolers: Defining tantrum behaviors associated with clinical problems," *The Journal of Pediatrics*, vol. 152, no. 1, pp. 117–122, 2008.

[10] S. S. Eisbach, F. Cluxton-Keller, J. Harrison, J. R. Krall, M. Hayat, and D. Gross, "Characteristics of temper tantrums in preschoolers with disruptive behavior in a clinical setting," *Journal of Psychosocial Nursing and Mental Health Services*, vol. 52, no. 5, pp. 32–40, 2014.

[11] A. Davis, "Virtual reality simulation: an innovative teaching tool for dietetics experiential education," *The Open Nutrition Journal*, vol. 9, pp. 65–75, 2015.

[12] C. L. Foronda, M. Fernandez-Burgos, C. Nadeau, C. N. Kelley, and M. N. Henry, "Virtual simulation in nursing education: a systematic review spanning 1996 to 2018," *Simulation in Healthcare*, vol. 15, no. 1, pp. 46–54, 2020.

[13] S. Rusdorf, G. Brunnett, M. Lorenz, and T. Winkler, "Real-time interaction with a humanoid avatar in an immersive table tennis simulation," *IEEE Transactions on Visualization and Computer Graphics*, vol. 13, no. 1, pp. 15–25, 2006.

[14] G. Albright, R. Goldman, K. M. Shockley, F. McDevitt, and S. Akabas, "Using an avatar-based simulation to train families to motivate veterans with post-deployment stress to seek help at the va," *Games for Health: Research, Development, and Clinical Applications*, vol. 1, no. 1, pp. 21–28, 2012.

[15] L. I. Kidd, K. I. Morgan, and J. R. Savery, "Development of a mental health nursing simulation: Challenges and solutions." *Journal of Interactive Online Learning*, vol. 11, no. 2, 2012.

[16] R. Blanton and D. Carbajal, "Not a girl, not yet a woman: A critical case study on social media, deception, and lil miquela," in *Handbook of Research on Deception, Fake News, and Misinformation Online*, 2019, pp. 87–103.

[17] X. Zhou, "Virtual youtuber kizuna ai: co-creating human-non-human interaction and celebrity-audience relationship," Master's thesis, Lund University, Lund, Sweden, 2020.

[18] R. Anand and A. Periyannan, "Systems and methods for real-time virtual-reality immersive multimedia communications," U.S. Patent 9 143 729, 9 22, 2015.

[19] K. Lee, "Augmented reality in education and training," *TechTrends*, vol. 56, no. 2, pp. 13–21, 2012.

[20] L. Johnson, A. Levine, R. Smith, and S. Stone, *The 2010 Horizon Report*. ERIC, 2010.

[21] R. Freitas and P. Campos, "Smart: a system of augmented reality for teaching 2nd grade students," *People and Computers XXII Culture, Creativity, Interaction 22*, pp. 27–30, 2008.

[22] D. Wagner and D. Schmalstieg, "First steps towards handheld augmented reality," in *Seventh IEEE International Symposium on Wearable Computers*, 2003, pp. 127–135.

[23] D. Sahin and R. M. Yilmaz, "The effect of augmented reality technology on middle school students' achievements and attitudes towards science education," *Computers & Education*, vol. 144, p. 103710, 2020.

[24] Y. Tao, E. Coltey, T. Wang, M. Alonso, M.-L. Shyu, S.-C. Chen, H. Alhaffar, A. Elias, B. Bogosian, and S. Vassigh, "Confidence estimation using machine learning in immersive learning environments," in *IEEE Conference on Multimedia Information Processing and Retrieval*, 2020, pp. 247–252.

[25] M. S. Bhatia, N. K. Dhar, P. K. Singhal, V. R. Nigam, S. C. Malik, and D. N. Mullick, "Prevalence and etiology: In a non-referral outpatient setting," *Clinical Pediatrics*, vol. 29, no. 6, pp. 311–315, 1990.

[26] M. Potegal, M. R. Kosorok, and R. J. Davidson, "Temper tantrums in young children: 2. tantrum duration and temporal organization," *Journal of Developmental & Behavioral Pediatrics*, vol. 24, no. 3, pp. 148–154, 2003.

[27] A. E. Kazdin and S. M. Rabbitt, "Novel models for delivering mental health services and reducing the burdens of mental illness," *Clinical Psychological Science*, vol. 1, no. 2, pp. 170–191, 2013.

[28] A. Chacko, S. A. Jensen, L. S. Lowry, M. Cornwell, A. Chimklis, E. Chan, D. Lee, and B. Pulgarin, "Engagement in behavioral parent training: Review of the literature and implications for practice," *Clinical Child and Family Psychology Review*, vol. 19, no. 3, pp. 204–215, 2016.

[29] A. Baumel, A. Pawar, J. M. Kane, and C. U. Correll, "Digital parent training for children with disruptive behaviors: Systematic review and meta-analysis of randomized trials," *Journal of Child and Adolescent Psychopharmacology*, vol. 26, no. 8, pp. 740–749, 2016.

[30] X. Yang, M.-L. Shyu, H.-Q. Yu, S.-M. Sun, N.-S. Yin, and W. Chen, "Integrating image and textual information in human-robot interactions for children with autism spectrum disorder," *IEEE Transactions on Multimedia*, vol. 21, no. 3, pp. 746–759, 2019.

[31] D. N. Zotkin, R. Duraiswami, and L. S. Davis, "Rendering localized spatial audio in a virtual auditory space," *IEEE Transactions on multimedia*, vol. 6, no. 4, pp. 553–564, 2004.