Near-contact Person-to-3D Character Dance Training: Comparing AR and VR for Interactive Entertainment

Salva Kirakosian
Technical University of Crete
Chania, Greece
Email: subkiro@gmail.com

Grigoris Daskalogrigorakis
Technical University of Crete
Chania, Greece
Email: gregthask.bt1@gmail.com

Emmanuel Maravelakis
Hellenic Mediterranean University
Chania, Greece
Email: marvel@hmu.gr

Katerina Mania
Technical University of Crete
Chania, Greece
Email: amania@isc.tuc.gr

***Abstract*—This paper presents an innovative, near-contact, person-to-3D character, real-time dance training framework for leaders developed in Augmented Reality (AR) for interactive entertainment. Rather than mimicking pre-recorded dance animations as in previous work, the user is trained to lead a 3D partner while dancing, employing mid-air gestures. Real and virtual hands are joined and dance movements are initiated to which the 3D character responds to, through Unity's Inverse Kinematics solvers, in a non-predetermined manner following Latin dance rules. We evaluate task workload and usability of our AR dance training system compared with the Virtual Reality (VR) equivalent. Perceived training performance, accuracy of actions and fun were lower in AR, evoking increased training effort in AR compared to playfully training in VR. Design recommendations include increasing AR's Field of View, enhancing high quality graphics in AR, removing obtrusive cabling in VR and improving finger tracking in VR.**

Keywords- Augmented Reality, Virtual Reality, Interactive Entertainment, Animation, Dance.

I. Introduction

Virtual reality (VR) and Augmented Reality (AR) Head Mounted Displays (HMDs) are often employed for interactive entertainment, communicating a strong sense of immersion and fun, providing playful multimodal experiences. However, the medium, e.g. either AR or VR is usually prone to usability issues inherent to the technology such as having to deal with the small Field-of-View (FoV) of AR displays as well as motion sickness or movement restrictions for VR [1]. In this paper, we present an innovative, near-contact, person-to-3D character, real-time dance training framework for leaders developed in AR for interactive entertainment, compared for task load and usability to an equivalent implementation in VR summarized in [2]. We result to design recommendations invaluable to AR/VR developers so that usability as well as fun is enhanced.

978-1-6654-3886-5/21/$31.00 ©2021 IEEE

Dance training in AR and VR utilized motion capture technology to capture dance routines and, then, reproduce them, frequently evaluated based on successful system recognition of dancer movements rather than usability. VR dance or sports training systems merely replay captured dance sequences through HMDs in order for the user to learn by mimicking displayed movements without providing any feedback [3], [4] and without live dancing with a 3D character acting as a partner [5], [6], [7]. A recent system has been presented focusing on dance recognition performance and feature extraction rather than usability assessment [8], [9] and without interactivity with the 3D character, limiting dance training effectiveness [10], [11], [12]. A process based on Hidden Markov Models (HMM) for learning the structure of the dance motions of a leader and follower predicted the intention for the following movement [13], [14], [15]. Such methods, as above, are not integrated in a dance training framework.

AR has recently been employed for interactive entertainment and dance, therefore, inherent technical issues of the AR displays such as the limited (FoV) should be addressed. Conceptual models utilize touch, vision and hearing for remote multiple users dancing together, using AR and holography [16]. A prototype mobile AR interface which included a 3D character performing professional folk dance instructed users to imitate the 3D character's movements and captured user dance and body joints through an optical motion capture system. The AR set-up presented did not include interactivity of user and 3D dancer [17]. Dance training systems employed an interactive mirror in AR to provide feedback while users danced without interaction with a digital dancer based on motion capture accuracy [18] or overlaying the movement of the dancer's body over a 3D character, viewed in AR offering limited interaction [19].

Previous work in AR as well as VR interactive entertainment is focusing on system performance rather than interaction affordances or usability. Recent reviews proposed that evalua-

Fig. 1: Dance training in VR in [2]

tions of such AR and VR experiences should focus on concrete hypotheses, involving cognitive aspects, ecological validity [20], analysis of AR acceptance, believability of digital content overlayed on the real-world [21], [22], emotional responses [23] and reduction of motion sickness [24]. New evaluation methods based on exploring multiple aspects of experience are needed [25]. Previous work in VR has shown that the limited FoV significantly affects performance [26], [27]. It is not clear whether the same applies in AR.

In this paper, we propose an innovative near-contact person-to-3D character real-time dance training application for leaders, implemented in AR. A 3D scanned character representing the well-known Latin dancer Alberto Rodriguez follows users' movements and performs dance sequences in real-time, according to the direction the user is leading (See Fig.2). The 3D scanned dancer responds in real-time to the users' (leader) improvised dancing holding virtual with real hands, implemented using Unity's Inverse Kinematics (IK). Motion capture data is applied to the rigged 3D model to create realistic dance animations. We compare task workload in AR and VR [2] employing the NASA TLX [28] and usability survey [29] resulting to design recommendations for VR and AR developers for interactive entertainment. Aspects of the VR/AR technology which hampered user performance and user experience are reported.

II. IMPLEMENTATION IN AR

We employed the same dance mechanics and 3D scanned dancer now displayed in the real-world office space in AR using the Magic Leap One HMD, as in the VR condition in [2]. We selected the Magic Leap One display because it has comparable hand tracking to the Leap Motion utilized in VR, while also performing spatial scanning so that our character could naturally walk on the scanned mesh of the real world's floor. Magic Leap One's FoV is limited compared to the Oculus Rift used in VR (**Figure 2**c). For the AR implementation, we started from the VR scene in [2] and removed scene objects such as floors and walls, leaving only menus and the 3D dancer (**Figure 2**a). We used Magic Leap's spatial mapper, which uses the real-world scan and builds a virtual mesh around the camera. By using Unity's NavMesh for our character, a walkable surface on the real world at runtime was generated. Leap Motion's hand and gesture tracking in VR were replaced by Magic Leap's on-board hand tracking, which has nearly identical position tracking latency and more robust gesture detection than Leap Motion. We added features for re-positioning the menu and the 3D character in case they were spawned inside a wall as well as a re-scanning button for incorrect scanning. A show/hide mesh on demand ensured the room was sufficiently scanned. We, then, hid the mesh before user testing. The built-in gesture recognition works separately from the hand tracking algorithm and can track eight distinct gestures per hand including the "Open Pinch", "Pinch", "Thumbs up" and "OK" gestures, as in VR.

III. USER EVALUATION

A. Evaluation of dance training in AR

We evaluate near contact dance in AR with the equivalent experience in VR, based on measuring perceived workload (NASA TLX [28]) and usability [29] (within subjects) involving real-time dance with the same 3D scanned character as well as employing a think-aloud protocol. Four stages were completed in both AR and VR. (1) Walk towards the virtual character (2) Use gestures to interact with the menus by pinching (3) Watch the 3D dancer and learn the basic steps and (4) Dance while interacting with the 3D dancer. Step 1 determined how confidently users were walking while wearing either a VR or AR headset. Steps 2 and 3 familiarized users with the User Interface (UI) preparing them for real-time person-to-3D character dance in task 4. We propose the following hypotheses:

- (H1) The limited FoV of current AR displays compared to current generation VR displays would negatively impact perceived workload.
- (H2) AR interaction will provoke higher usability ratings as AR does not completely exclude the real-world.

B. Participants

18 participants were recruited (5 female, 13 male, mean age 30, SD=10.5 years, height 1,65m to 1,88m). 5 were near sighted, 3 had slight astigmatism and 1 had cataract. 7 participants had some experience in AR and VR while 8 had limited AR/VR knowledge. 13 users had right and 5 left eye dominance.

C. Procedure

We counterbalanced viewing conditions in AR/VR to avoid any order effects. Users wore either the Oculus or the Magic Leap HMD. We initially trained users to walk towards the 3D

(a) (b) (c)

Fig. 2: (a), (b) Dance training in AR (c) Field of View (FoV) comparison, Magic Leap (AR) vs Oculus Rift HMD (VR)

character, navigate the menu using the ”Thumbs up gesture” and ”Pinch”, follow the ”tutorials” sub-menu for the basic dance moves: Basic step, Side step and turn. When selecting the ”dance” option, the user made contact with the hands of the 3D character and danced along by leading the 3D character for 1 minute. The user followed the same steps in AR/VR.

D. Data Collection and Analysis

The data collected after dance training in AR/VR was based on NASA TLX [28] for perceived workload and an 8-part usability survey [29] on a 7-point Likert scale (1= Very low, 7=Very high). The NASA TLX assesses perceived performance effectiveness on: *Mental demand Physical demand*, *Temporal demand*, *Performance*, *Effort* and *Frustration*. The usability survey assessed training efficiency in: *Learnability*, *Efficiency*, *Memorability*, *Accuracy*, *Satisfaction*, *Intuitiveness* and *Fun*. We compared the NASA TLX and usability ratings between AR and VR by applying a Wilcoxon signed-rank test.

IV. RESULTS

A. NASA TLX

A Wilcoxon signed-rank test yielded a significant difference between dance training in AR and VR in the categories of *Performance* (W=22.5, p=0.01046) and *Effort* (W=20, p=0.04136). In AR, users required more perceived effort to accomplish performance than VR (**Figure 3**). VR users felt more successful in accomplishing what they were asked to do. Dance training in both VR and AR resulted in a similar, low level physical, mental and temporal task demand. Frustration was low in AR/VR, but with higher variance in AR. Users felt they performed the task more efficiently in VR. In VR, the exclusion of real-world surroundings provoked a subjective sense of less effort associated with VR training as well increased self-performance.

B. Usability ratings

A Wilcoxon signed rank test was applied for usability and resulted in a significant difference between AR and VR dance training in the fields of *Accuracy* (W=23.5, p=0.02144) and *Fun* (W=13.5, p=0.0251) where VR proved to result to significantly higher ratings than AR (**Figure 3**). Our second hypothesis (H2) was not supported. *Accuracy* levels, e.g. the level of perceived accuracy and responsiveness of the system were higher in VR. Because of the smaller FoV of the AR display, users felt at times disoriented and hesitant to interact with the 3D character because they could view only part of the 3D character at one time, therefore, they had to move their head repetitively for viewing. Moreover, users expressed higher levels of *Fun* in VR, potentially because of the larger FoV and the stronger sense of immersion in VR than in AR. We observed that in the fields of *Accuracy*, *Intuitiveness*, *Naturalness* and *Fun* AR scores varied greatly amongst users. Same category scores in VR had significantly less variance.

C. General feedback

Limited AR FoV. 14 out of 18 users found the limited FoV of the Magic Leap (40 deg. horizontal, 30 vertical) restrictive for dance training and near-field interactions compared with the 90 deg. FoV of the Oculus display. 5 stated that if the Magic Leap had a wider FoV, they would prefer to train in AR compared to VR. Only 2 users liked AR more than VR even with the limited FoV. This and NASA TLX's reduced performance and low usability accuracy score in AR, support (H1).

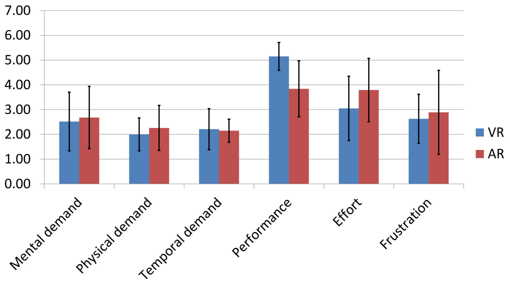
Obtrusive VR cabling. 9 out of 18 users found the Oculus Rift cabling obtrusive. 13 out of 18 were reassured they would not fall or hit walls and objects in the play area. Users were hesitant to conduct a full-body turn when interacting with the 3D character out of fear of pulling the cables. The move to wireless VR for the general user is paramount.

Unstable hand tracking in VR. The Leap Motion was more limited in the covered area for tracking users' hands in VR forcing them to conduct unnatural movements.

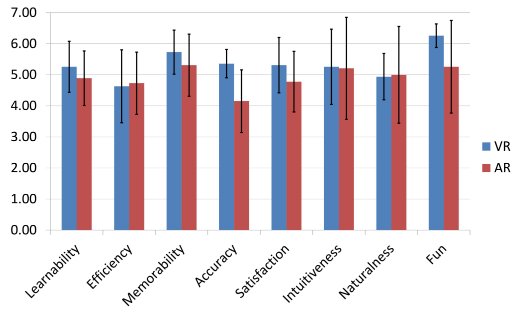
We also asked each participant which would be their preferred platform of choice for dance training. 16 out of 18 agreed that they would prefer to learn how to dance through VR than AR. Even the participants that preferred AR over VR stated that they would prefer to ”dance in a virtual world rather than with a hologram”.

V. DISCUSSION

(1) In terms of task load index, users felt they trained more successfully and with less effort in VR. Because of the limited AR FoV, participants were not viewing the whole AR scene of the 3D dancer coveted in their FoV and, therefore, felt



(a)



(b)

Fig. 3: (a) Nasa TLX in VR/AR (b) Usability ratings in VR/AR

disoriented. They had to perform multiple head movements in AR in order to place the 3D dancer in their FoV.

(2) In terms of usability ratings, users expressed that the VR dance training system was more accurate and responsive as well as more fun. Again, the small FoV of the AR display was a potential reason. Immersive VR played a role in evoking higher ratings of perceived fun than in AR. Certain users preferred the VR dance training system, even if the AR FoV becomes larger in future systems.

(3) Efficient navigation in VR was problematic. Although users were aware of the wireframe that indicated the edges of the safe play area, users were still hesitant in navigating the play area due to fear of falling.

(4) High variance of task load and usability for the AR dance training system was present but not for VR. Certain users were hesitant to wear the VR display preferring the more open and light AR display. Others had negative views about the low quality graphics and small FoV of the AR display.

VI. Conclusion

Both training systems in AR/VR demonstrated minimal to non-perceptible latency, advancing current dance training systems which mostly visualize pre-rendered dance animations the trainees are required to imitate. The user is able to lead a 3D character in Latin dance by joining real and virtual hands and the 3D character can respond, through Unity's IK solvers, to user dance actions in real-time and in a non-predetermined manner. We compared the two systems in regards to their task load and usability when training a user how to dance resulting to design and system recommendations for interactive entertainment in AR/VR. We found that Optical-See-Through AR HMD devices are still not well received for near-field interactions due to their small FoV, even though their tracking and visual integrity are at par with current VR systems. New AR displays should definitely focus on enlarging the available FoV. We also found that the VR system (high quality graphics) is preferred when it comes to training providing a stronger sense of accomplishment as well as system accuracy and responsiveness, also making the process more fun. Cabling of the HMD displays which require PC connection is an issue when the applications are computationally heavy. Finger tracking is also unstable in VR, therefore, interactive entertainment applications should focus on accurate tracking. Future work will include a thorough exploration of the effect of the FoV on task load and usability, employing VR with an artificially restricted FoV to match that of the AR HMD. Game scenarios could also emerge based on real-time 3D character and human interactivity, taking also into account the explosion of AR technologies across sectors.

Acknowledgment

This research has been co-financed by the EU and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH – CREATE – INNOVATE (project: T1EDK-01469).

References

[1] G. A. Koulieris, K. Akşit, M. Stengel, R. K Mantiuk, K. Mania, and C. Richardt. Near-eye display and tracking technologies for virtual and augmented reality. In *Computer Graphics Forum, vol 38, pp 493-519*.

[2] S. Kirakosian, E. Maravelakis, and K. Mania. Immersive simulation and training of person-to-3d character dance in real-time. In *2019 11th Intl. Conf. on Virtual Worlds and Games for Serious Applications, pp 1-4*.

[3] K. Jeonghyun Y. Ungyeon. Implementation and evaluation of "just follow me": An immersive, vr-based, motion-training system. *Presence: Teleoperators and Virtual Environments, vol 11, pp 304-323*.

[4] K. Hachimura, H. Kato, and H. Tamura. A prototype dance training support system with motion capture and mixed reality technologies. In *13th IEEE Intl. Workshop on Robot and Human Interactive Communication pp 217-222*.

[5] D. Alexiadis, P. Kelly, P. Daras, N. O'Connor, T. Boubekeur, and M. B. Moussa. Evaluating a dancer's performance using kinect-based skeleton tracking. In *Proc. 19th ACM International Conference on Multimedia, MM '11, pp 659-662*.

[6] J. C. P. Chan, H. Leung, J. K. T. Tang, and T. Komura. A virtual reality dance training system using motion capture technology. *IEEE Trans. on Learning Technologies, vol 4, pp 187-195*.

[7] M. Raptis, D. Kirovski, and H. Hoppe. Real-time classification of dance gestures from skeleton animation. In *Proc. 2011 ACM SIGGRAPH/Eurographics Symp. on Computer Animation*, pages 147–156.

[8] L. Deng, H. Leung, N. Gu, and Y. Yang. Real-time mocap dance recognition for an interactive dancing game. *Computer Animation and Virtual Worlds*, 22(2-3):229–237, 2011.

[9] Y. Yang, H. Leung, L. Yue, and L. Deng. Automatic dance lesson generation. *IEEE Trans. Learn. Technol.*, 5(3):191–198, January 2012.

[10] M. Kyan, G. Sun, H. Li, L. Zhong, P. Muneesawang, Nan N. Dong, B. Elder, and L. Guan. An approach to ballet dance training through ms kinect and visualization in a cave virtual reality environment. *ACM Trans. Intell. Syst. Technol.*, 6(2):23:1–23:37, 2015.

[11] J. Tang, Jacky J. C. P. Chan, and H. Leung. Interactive dancing game with real-time recognition of continuous dance moves from 3d human motion capture. In *Proc. of 5th ACM Int. Conf. on Ubiquitous Information Management and Communication*, pages 50:1–50:9.

[12] I. Rallis, A. Langis, I. Georgoulas, A. Voulodimos, N. Doulamis, and A. Doulamis. An embodied learning game using kinect and labanotation for analysis and visualization of dance kinesiology. In *10th Int. Conf. on Virtual Worlds and Games for Serious Appl. 2018*, pages 1–8.

[13] C. Mousas. Performance-driven dance motion control of a virtual partner character. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 57–64, March 2018.

[14] T. Takeda, Y. Hirata, and K. Kosuge. Dance step estimation method based on hmm for dance partner robot. *IEEE Transactions on Industrial Electronics*, 54(2):699–706, April 2007.

[15] E. Ho, J. C. P. Chan, T. Komura, and H. Leung. Interactive partner control in close interactions for real-time applications. *ACM Trans. Multimedia Comput. Commun. Appl.*, 9(3):21:1–21:19, July 2013.

[16] V. Saxena, T. Feldt, and M. Goel. Augmented telepresence as a tool for immersive simulated dancing in experience and learning. In *Proc. of ACM India HCI 2014 Conference on Human Computer Interaction*, IndiaHCI '14, pages 86:86–86:89. ACM, 2014.

[17] I. Kico and F. Liarokapis. A mobile augmented reality interface for teaching folk dances. In *25th ACM Symp. on Virtual Reality Software and Technology, 47:1-2*.

[18] F. Anderson, T. Grossman, J. Matejka, and G. Fitzmaurice. Youmove: Enhancing movement training with an augmented reality mirror. In *Proc. 26th Annual ACM Symposium on User Interface Software and Technology, pp 311-320*.

[19] S. Yan, G. Ding, Z. Guan, N. Sun, H. Li, and L. Zhang. Outsideme: Augmenting dancer's external self-image by using a mixed reality system. In *Proc. 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems, CHI EA '15, pp 965-970*.

[20] L. Merino, M. Schwarzl, M. Kraus, M. Sedlmair, D. Schmalstieg, and D. Weiskopf. Evaluating mixed and augmented reality: A systematic literature review (2009-2019). In *2020 IEEE Intl. Symp. on Mixed and Augmented Reality (ISMAR), pp 438-451*.

[21] M. Billinghurst. Grand challenges for augmented reality. *Frontiers in Virtual Reality, vol 2, pp 12*.

[22] A. Dey, M. Billinghurst, R. W. Lindeman, and J Swan. A systematic review of 10 years of augmented reality usability studies: 2005 to 2014. *Frontiers in Robotics and AI*, 5:37.

[23] F. Pallavicini, A. Pepe, and M. E. Minissi. Gaming in virtual reality: what changes in terms of usability, emotional response and sense of presence compared to non-immersive video games? *Simulation & Gaming, vol 50, pp 136-159*.

[24] G. A. Koulieris, G. Drettakis, D. Cunningham, and K. Mania. Gaze prediction using machine learning for dynamic stereo manipulation in games. In *2016 IEEE Virtual Reality (VR), pp 113-120*.

[25] K. Kim, M. Billinghurst, G. Bruder, H. B.-L. Duh, and G. F. Welch. Revisiting trends in augmented reality research: A review of the 2nd decade of ismar (2008–2017). *IEEE trans. on visualization and computer graphics, vol 24, pp 2947-2962*.

[26] JJ-W. Lin, H. B.-L. Duh, D. E. Parker, H. Abi-Rached, and T. A. Furness. Effects of field of view on presence, enjoyment, memory, and simulator sickness in a virtual environment. In *Proc. ieee virtual reality 2002, pp 164-171*.

[27] Y. Ling, H. T. Nefs, W.-P. Brinkman, C. Qu, and I. Heynderickx. The effect of perspective on presence and space perception. *PloS one, vol 8*.

[28] S. G. Hart. Nasa-task load index (nasa-tlx); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting, vol 50, pp 904-908*.

[29] T. Piumsomboon, D. Altimira, H. Kim, A. Clark, G. Lee, and M. Billinghurst. Grasp-shell vs gesture-speech: A comparison of direct and indirect natural interaction techniques in augmented reality. In *2014 IEEE Intl. Symp. on Mixed and Augmented Reality (ISMAR), pp 73-82*.