What's my future: a Multisensory and Multimodal Digital Human Agent Interactive Experience

ANNA SHEREMETIEVA, Softserve, Wroclaw, Poland
IHOR ROMANOVYCH, SAM FRISH, and MYKOLA MAKSYMENKO, SoftServe, Lviv, Ukraine
ORESTIS GEORGIOU, Ultraleap Ltd, Bristol, UK

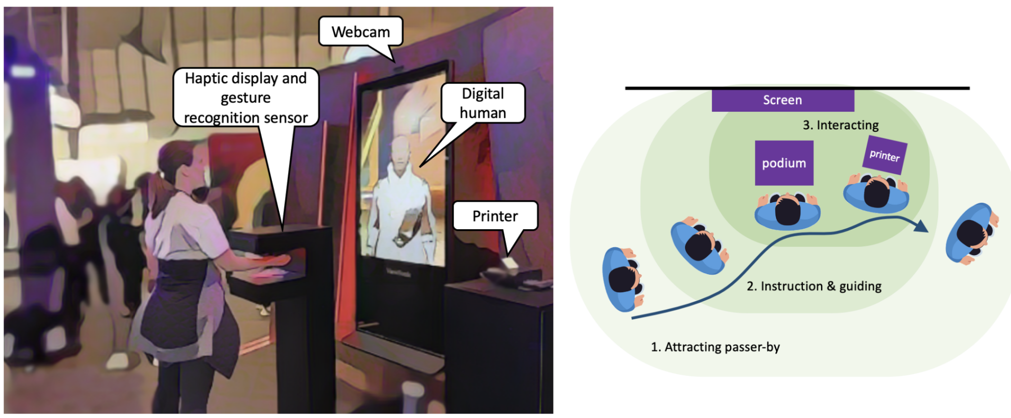

Fig. 1. *Left* - Interactive fortune-telling experience setup and main interface components. A person stands in front of a letterbox-style podium, approx 1.5 m from a screen. The podium houses a mid-air haptic display and hand-tracking camera. The screen displays the digital human agent (DHA). The webcam above the screen is used to detect user presence and to take a picture of their face which is then processed and sent to the printer. *Right* - The three steps of the user journey during the interactive experience. 1. Passers-by are attracted by the experience setup and the DHA on the screen. 2. As they approach they notice the footprints and handprint and also some instructions on the screen. 3. Interaction begins during which users follow a short multisensory (audio-visual-haptics) experience that ends with them collecting a personalized printout giveaway.

This paper describes an interactive multimodal and multisensory fortune-telling experience for digital signage applications that combines digital human agents along with touchless haptic technology and gesture recognition. For the first time, human-to-digital human interaction is mediated through hand gesture input and mid-air haptic feedback, motivating further research into multimodal and multisensory location-based experiences using these and related technologies. We take a phenomenological approach and present our design process, the system architecture, and discuss our gained insights, along with some of the challenges and opportunities we have encountered during this exercise. Finally, we use our singular implementation as a paradigm as a proxy for discussing complex aspects such as privacy, consent, gender neutrality, and the use of digital non-fungible tokens at the phygital border of the metaverse.

CCS Concepts: • **Human-centered computing** → **Interactive systems and tools**; **Haptic devices**; *Mixed / augmented reality*; • **Computing methodologies** → *Image manipulation.*

Additional Key Words and Phrases: Digital Humans, Digital Signage, Multisensory experience, Haptics, NFT

Manuscript submitted to ACM

ACM Reference Format:
Anna Sheremetieva, Ihor Romanovych, Sam Frish, Mykola Maksymenko, and Orestis Georgiou. 2023. What's my future: a Multisensory and Multimodal Digital Human Agent Interactive Experience. In *ACM International Conference on Interactive Media Experiences (IMX '23), June 12–15, 2023, Nantes, France.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3573381.3596161

1 INTRODUCTION

Digital human agents (DHAs) are digital representations of real or imaginary people that are designed to interact with other people and assist them in various ways. They are often used in customer service or other similar roles, where they can provide answers to common questions, guide users through processes, or help them with specific tasks [23]. In each of these cases, the aim of DHAs would be able to understand and respond to user input in a natural and intuitive way, thus providing a more engaging and personalized experience. As such, DHAs have found a plethora of applications in digital signage interactive kiosks located in shopping malls [38], museums [31], airports, and other public spaces as well as in the metaverse.

Implicit in the creation of these interactive experiences is the use of artificial intelligence (AI) and natural language processing (NLP), which allows DHAs to understand and respond to user input in a human-like manner, what is often referred to as conversational agents; a market estimated at around $10 billion in 2023 [5]. Explicit in the deployment and operation of these interactive experiences is a large screen (usually in portrait mode), an internet/server connection, a directional microphone array to detect speech input, and a camera to detect user presence and possibly apply some facial recognition or other spatial computing or sensing. However, the implementation of voice-input DHA kiosks in public spaces is not without its challenges. For example, they are susceptible to environmental audible noise interference that deteriorates microphone reception. Another example is user hesitance to interact due to social acceptability and privacy concerns, particularly when verbally sharing sensitive or personal information in public [11].

Despite the above advancements and challenges, DHAs have had surprisingly little multisensory and multimodal research attention in the digital signage space [9]. Our search did not reveal any works that study human-to-DHA interactions that involve additional input modalities other than voice, or that engage additional senses other than audio-visual. Multisensory digital signage incorporates multiple senses, such as sight, sound, touch, and even smell, to create a more immersive and engaging user experience [7, 36]. Multimodal digital signage incorporates multiple modes of interaction, such as touchscreens, facial expressions, hand or body gestures, and voice input to create a more intuitive and engaging user experience [6, 26, 33].

This paper describes a multisensory and multimodal human-to-DHA interactive experience whereby user input is through hand gestures and facial recognition, and the DHA-to-human communication is through voice and ultrasound mid-air haptics as shown in Figure 1. The experience itself is that of visiting a fortune teller oracle that reads your palm and predicts your future, however, as we discuss later this is just one possible example implementation. Through this singular paradigm, which is orthogonal to most existing human-to-DHA location-based interfaces, we wanted to probe beyond audio-visual and explore the integration of new input modalities and senses. Moreover, we wanted to understand this design space and gain firsthand experience, in order to gather more insights on how multimodal and multisensory DHA-human interactions can be made more natural thus contributing to their advance and development.

To that end, we chose to explore touchless mid-air haptics [3] and gesture recognition [30] since these two are primitives of so-called natural user interfaces [37], and have already been studied in the contexts of digital signage and extender reality (XR). Our aim is therefore to import these two interface technologies and explore how they can be applied to a human-to-DHA location-based interactive experience.

In the following sections, we will provide background and related works (Sec. 2), clarify the paper contribution (Sec. 3), and describe the experience physical setup (Sec. 4). Then, we will describe the experience journey (Sec. 5) followed by the implementation and integration architecture (Sec. 6) along with a discussion on the challenges we faced and decisions we made (Sec. 7).

2 BACKGROUND AND RELATED WORKS

To address our research objective and exploration, the literature is divided into three sub-sections relating to gesture interfaces, mid-air haptics, and DHAs, respectively.

2.1 Gesture interfaces

Hand gesture technology uses proximity and optical sensors along with machine vision algorithms to recognize hand movements and poses in space and use those to control computing systems. Hand gesture interfaces are becoming commonplace in virtual, augmented, and mixed reality (VR/AR/MR) through their native support on headsets like Oculus Quest, Magic Leap, Hololens, Lynx, and Varjo, and thus form a major candidate for interfacing with the metaverse [2]. Gesture interfaces are also increasingly prominent in location-based public interfaces, as well as automotive human-machine interfaces (HMIs). The former has been recently accelerated by the Covid-19 pandemic and concerns about disease transmission through touchscreen interfaces [8, 24]. The latter has been driven by the need to reduce the mental and physical effort of drivers while interfacing with a car's infotainment system since touchscreens tend to demand visual attention thus increasing accident risk [19].

Hand and body gestures have been extensively studied and used in digital signage applications. Chen et al. [4] describe a vision-based gesture recognition approach to interact with digital signage systems and discuss the issues faced by such systems. Hardy et al. [13] provide real-world feedback and insights from their in-the-wild experiments, including the so-called issue of display blindness [21]. Walter et al. proposed methods for revealing mid-air gestures on interactive public displays [36]. Finally, there are several companies that use proximity and optical sensors to offer hand gesture recognition for touchless interaction with digital signage and location-based kiosks such as airport check-in desks and self-service check-outs.

2.2 Mid-air haptics

Mid-air haptic technology uses ultrasound speakers and beam-focusing algorithms to deliver a tactile sensation on a user's hands or fingertips, no wearables or controllers needed (see recent review [12]). The tactile sensations can be dynamically updated to provide haptic feedback, and haptic guidance to the user thus improving the interactivity and performance of gesture interfaces [14, 35]. Corenthy et al. [6] for example gave an early demonstrator of the tech in a movie-poster setting. Importantly, in the context of digital signage, mid-air haptics has been shown to also improve other metrics such as user experience and engagement [17]. Finally, Limerick et al. [16] discussed implementation and deployment insights and best practices, including gamification tactics, applied in a real-world experiment at a cinema in Los Angeles.

2.3 Digital human agents

Creating interactive DHAs has become easier than ever through the use of computer vision, multimodal text-to-speech synthesis, speech-driven face animation, neural network (NN)-based body motion control, and human-digital-human interaction guidelines. For example, Fang et al. describe the release of the MetaHuman Creator by Epic Games in 2021 and claim that 'anyone' can now create a bespoke photorealistic digital human, fully rigged and complete with hair and clothing, in a matter of minutes [10].

As such, interactive DHAs are growing in popularity and use in the media, ranging from providing customer and sales support to a variety of digital signage applications [7, 29, 31], marketing tools, gaming, and teaching aids [23]. Davies et al. provide a full review of the state of the art, the research challenges, and the main drivers behind pervasive and public displays [7]. Silva et al. look at the recent multidimensional uptake of DHAs by the fashion industry during the COVID-19 pandemic which forced fashion businesses to accelerate their digital transformation [29]. Sylaiou et al. review the use of DHAs in museums and places of cultural interest and highlight trends of increasingly sophisticated installations with a focus on mixed reality and behavioral realism. The impact of DHA creation and embodiment on gender diversity was studied by Reyes et al. [25]. The synthesis of a human avatar from a single camera is discussed by Roble et al. [27], while neural network (NN) techniques for creating deepfake face-swaps are discussed by Nguyen et al. [22]. Finally, Nowak et al. give an overview of current debates, methodological approaches, and trends around the use of DHAs from a communication theory perspective [23].

Despite the recent proliferation of conversational agents and platforms such as ChatGPT, and the leaps forward in computer graphics such as the Metahuman Creator, there have been very few examples of multimodal interactions with DHAs. That is, most realizations of DHAs for digital signage rely on voice input, with a touchscreen fall-back option. In contrast, Diez et al. describe a human-to-DHA interface where a gesture component is developed, in addition to the voice interface, which integrates computer vision technologies for face and hand tracking [9]. Specifically, their implementation allowed users to interact directly with the content using their hand gestures, while at the same time allowing the system to estimate the emotional state of the user through his facial expressions and body language.

3 CONTRIBUTION

In this paper, motivated by the works of Diez et al. [9] and Corenthy et al. [6], we explore alternative multisensory and multimodal interfaces for DHAs in a location-based digital signage setting, and aim to 1) amplify the conversational role of hand gesture input, and 2) introduce a suitable tactile component that connects the DHA with the user. To that end, we describe the implementation of an interactive experience that modifies the usual DHA digital signage setup through the addition of a haptic podium (see Figure 1). Further, we put into practice the deployment guidelines for communicating interactivity and contactless haptic affordances from Limerick et al [16] and also seize the opportunity to create a DHA that is gender-neutral, thus making the installation more than just a fortune-telling, fun and engaging experience, but also a discussion point about diversity in the metaverse. Finally, as described in more detail below, the phygital (i.e., where the physical meets the digital) experience we have created synthesizes a human avatar from a picture of the user and mints it on a blockchain, thereby leaving behind a phygital footprint. Taken together, these aspects motivate further research into multisensory and multimodal DHA interactive experiences.

4 EXPERIENCE SETUP

The interface setup is shown in Figure 1 and comprises a large screen in portrait mode, a letterbox-style podium containing an Ultraleap haptic display and hand-tracking camera, a webcam, a powerful PC, and a color printer. The screen displays a gender-neutral DHA called Alnair (Arabic for "the bright one") that calls passersby over to interact with them ("I knew you would come") and then guides them through the fortune-telling experience. The letter-box style pedestal acts as a hand rest while the Ultraleap device projects vibrotactile sensations onto the user's palm and is also used for reading the user's hand gesture input. The webcam detects user presence and takes a picture of the user's

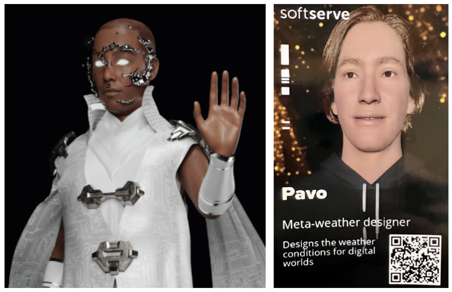

Fig. 2. *Left* - Gender-neutral DHA called Alnair. Their clothing is futuristic, their race is mixed or unclear, and their eyes hollow to indicate their oracle ability. The liquid metal on their face symbolizes the user's fortune which will solidify and shape into a fortune card. *Right* - Example printout giveaway card generated by a deepfake NN and a picture of a user. It contains a synthetic digital human avatar that resembles the user, a randomly selected nickname and job title and description within a hypothetical metaverse, and a QR code that leads to their NFT.

face, and the powerful PC runs a deepfake NN to create a synthetic human 3D avatar that resembles the user's face. A fortune prediction is randomly generated from a look-up-table (LUT) and together with the 3D avatar they are added to the Polygon blockchain (a cost-effective alternative to the Ethereum network) which mints a non-fungible-token (NFT) and a QR code which are sent to the printer (see example printout in Figure 2). The user journey through the physical installation is also shown in Figure 1, while the experience flow is detailed in Figure 3.

The letterbox podium is designed so that an average-height person (165 cm) can comfortably rest their hand (palm facing up just like in a typical fortune teller experience) while standing 1.5 meters away from the screen (see Figure 1). Footprints are affixed to the floor right in front of the podium at the location where the user would need to stand to initiate an interaction. A handprint is also affixed inside the letterbox podium. Together, the footprints and the handprint guide the user toward the right position and pose to initiate the interactive experience. Additional graphical instructions are shown on the main screen. Importantly, the podium communicates interactivity, thereby overcoming possible display blindness shortcomings [21].

The Ultraleap haptic display and hand-tracking device are housed in the upper part of the podium and are facing down, toward the user's palm. The gap between the Ultraleap device and the handprint below is 30 cm, thus allowing for good hand gesture recognition and mid-air haptic operation range. Once the user's palm is detected in position, it is also mirror-displayed on the main screen, a haptic sensation is felt, and the experience begins. At some point during the experience, the user is asked by Alnair for their consent to take their picture using the webcam situated above the screen. Consent is communicated through a hand gesture. At the end of the experience, the user steps to the side and picks up a print-out containing their fortune, a digital human avatar that resembles their face, and a QR code that points to their NFT address that has been inserted into the Polygon blockchain.

5 EXPERIENCE DESCRIPTION

The whole experience takes around 2 minutes to complete and only one user can interact at a time. Therefore, users stand in a queue and wait for their turn. Figure 3 presents a flowchart of the experience.

Alnair is displayed on the screen in idle mode while also inviting passersby over to the haptic podium. Once a user approaches the podium and places their palm inside, the webcam and hand tracking camera recognize the user and their hand, respectively. This initiates the experience which starts off with a tactile *line-scan* sensation on the palm.

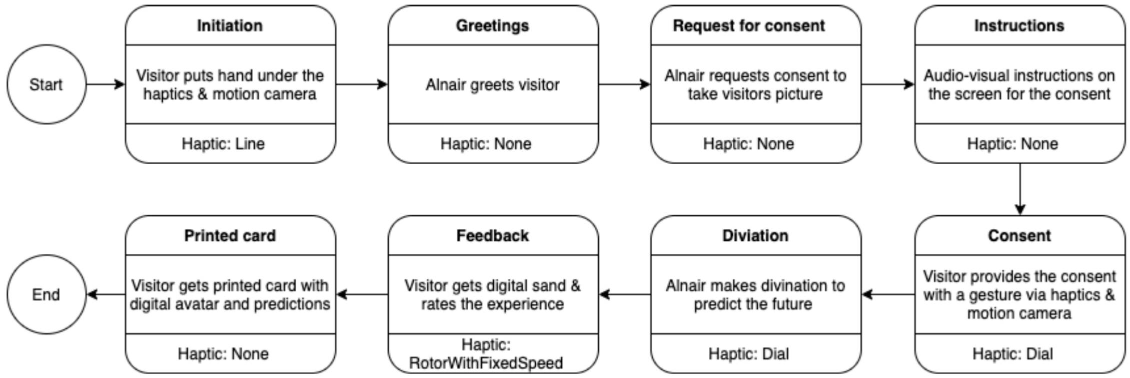

Fig. 3. Flowchart of the interactive experience.

Alnair awakens and greets the visitor. Next, Alnair asks for permission to read the user's face in order to perform the prediction (a GDPR-compliant consent is required for taking and using the user's face image). If the user agrees, the face image will be used to create a personalized digital avatar. If they decline, the user will get a randomly assigned avatar. The user accepts or rejects GDPR consent with a *grab* hand gesture and while doing so also feels the tactile dial sensation. After the consent, Alnair performs a 'divination' dance (prediction process) during which the user feels a tactile *dial* sensation conveying that some kind of magic is being performed. Meanwhile, the user's picture is being processed by a NN deepfake algorithm that runs locally (not online) and generates a digital avatar of the user, and assigns some meta-data containing the nickname of the user and occupation in the metaverse. These are also stored as an NFT. Finally, Alnair reaches out through the podium portal and releases some stardust onto the user's open hand which is felt through a *rotor* tactile sensation. Meanwhile, the printer is printing the card with the prediction.

6 IMPLEMENTATION AND INTEGRATION

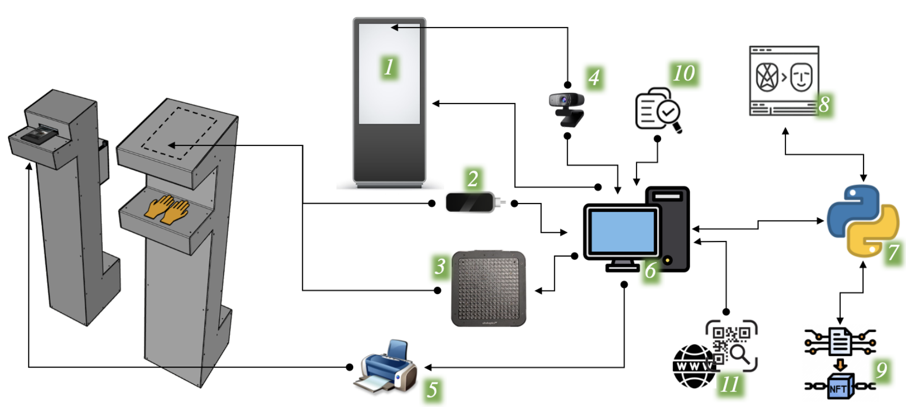

Fig. 4. *Left* - Printer podium (left) and interaction podium (right) housing the hand-tracking camera and the mid-air haptic display. *Right* - System architecture. See the manuscript text for more details on these and how they integrate into the experience.

Architecture: The system architecture is shown in Figure 4. It comprises of a central PC that interfaces with various sensors, actuators, displays, and programs. Specifically, we have a large screen (1), a Leap Motion camera (2), an Ultraleap haptic display (3), a webcam (4), and a printer (5) are all connected to a central PC that runs a game engine (6) application. The screen (1) displays Alnair's DHA and also outputs their voice and other sound effects. The Leap Motion infrared stereo camera (2) tracks the user's hand movements and recognizes any gesture input through the Gemini v5 SDK. The Ultraleap haptic display (3) employs an array of 256 ultrasound transducers to create tactile sensations in mid-air (see Figure 5) through the device SDK v3beta10. Both the Leap Motion camera and the Ultraleap haptic display are housed in the upper part of the interaction podium as seen on the left of Figure 4. OpenCV v4.5 was used to track the user's face through a Logitech C922 Pro webcam (4) which is attached to the top of the large screen (1). On the right-hand side of the PC in Figure 4, the app launches a Python NN server (7) which interfaces with a custom face-swap application for generating 3D avatars (8), and mints an NFT from the Polygon blockchain (9). Finally, the Python server (7) returns the synthetic digital human avatar (8) and NFT (9) back to the PC (6) that randomly pulls information (fortune, meta-occupation, etc) from a LUT (10) and generates a QR code link to a web-page containing the user's fortune and avatar NFT (11). Finally, a Canon Selphy C1300 printer (5) is used to print out the participant's synthetic avatar (8) and their QR code that points to their NFT. Further implementation details are described below.

Alnair's DHA: Alnair's DHA was created and body-animated in the Reallusion Character Creator (CC) tool at a high level of detail (LOD), i.e., geometry and texture complexity. To generate a multiethnic, multiracial, and transgender DHA, a total of 24+24=48 variations of male and female avatars were generated. For each gender, we used three Digital Humans chosen on three levels of character material variation, representing three different ethnicities following Higgins et al. who studied the uncanny valley effect [15]. These were then mixed further within the CC app, and all facial and head hair was removed, resulting in the DHA seen in Figure 2. It should be noted that there are currently no clear guidelines on how to mix or blend ethnicities, races, and gender within the context of a DHA CC app. We also created a voice script, recorded it, and used Nvidia's Omniverse Audio2Face app to lip- and face-sync to Alnair's DHA further customized by hand in the editor.

Synthetic Human Avatar of the User: Using the CC app we pre-generated 48 template avatars (of variable ethnicities, races, genders, and hairstyles) at a low LOD. If consent was obtained from the user, a picture of their face was captured by the webcam and sent to a custom face swap application that runs locally along with a randomly selected avatar template. It takes approximately a minute to generate the synthetic human avatar result and print it.

Hand Gesture Input: A large number of intuitive hand gesture commands have been researched and documented through various gesture elicitation studies (see recent review [34]) as a means for controlling all kinds of interactive human-machine interfaces. In our implementation, we used a Leap Motion camera to detect the *grab-release* template gesture (see Figure 5) to detect consent (right hand) or no consent (left hand). Many other options exist for Human-to-DHA interactions however were not explored here, e.g., pointing, handshake, caress, high-five, or even play a game of rock-paper-scissors.

Haptic Feedback: Mid-air haptic feedback sensations were delivered to the user's palm through an Ultraleap device. The sensations were scripted and triggered through the device API. These included a tactile dial, rotor, and line-scan (see Figure 5). Dial: This sensation involves a high-intensity focal point (HIFP) that moves clockwise around a 2 cm radius circle that is centered at the user's palm. The HIFP is amplitude modulated (AM) at a frequency of 200 Hz and the rotation frequency is 1 Hz. Rotor: This sensation involves a HIFP that moves back and forth along a 4 cm line that is centered at the user's palm. Meanwhile, the line is made to rotate at 1 Hz while the line-traversal frequency is 200 Hz. Line-scan: Similar to the rotor, but instead of rotating the line, we move it up and down the palm at 1 Hz.

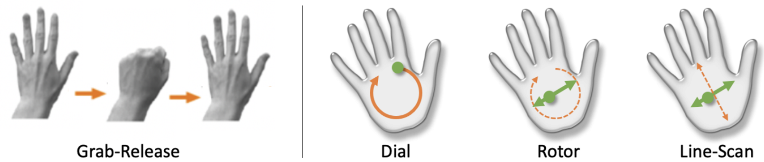

Fig. 5. *Left* - Grab-release gesture used to provide consent. *Right* - The three mid-air haptic sensations used in our implementation.

NFT Printout: First, we generated a list of nicknames and occupations and compiled them into a look up table (LuT) and then used to randomly assign to the user's printout card. The LUT comprised of 300 nicknames that were randomly selected from the published list of the International Astronomical Union (IAU) approved star names. It also had 100 fake occupations that were dreamed up by us as possible jobs in the metaverse after reading various sensational online blogs on this topic. At the end of each experience, a nickname and occupation were chosen at random and printed along with the generated synthetic human avatar. Also printed was a QR code that acts as a Polygon NFT identifier. Specifically, the QR code directs the user to a specially created website, where they are able to see their NFT which is also accessible on the OpenSea marketplace. To claim ownership of their NFT, the participants must transfer it into their own crypto wallet. Instructions are given on the QR code-linked website.

7 DISCUSSION

The experience described in the previous sections has been developed and showcased at WebSummit 2022 in Lisbon. A major aim of this effort was to differentiate our booth from more traditional DHA conversational kiosks through the inclusion of alternative modalities and senses. Moreover, we wanted to highlight the interplay between physical and digital ("phygital") through 1) the use of haptics to interact with a DHA, 2) a printout card that points to an NFT on a blockchain, 3) the use of a bodily gesture that carries agency [20] to communicate consent (rather than simply ticking a box on a screen), 4) a personalized 3D deep-fake avatar synthesis based on the user's real picture. It was also important that the setup interface was ergonomic and intuitive so that anyone could walk-up-and-use [28]. To that end, we followed guidelines from Limerick et al. [16] and Corenthy et al. [6] but also designed and built our own haptic interaction podium. Finally, it was deemed important and timely to bring forth discussion points around gender neutrality in the metaverse [25].

The oracle experience described and presented was overall a great success with hundreds of event participants getting their fortunes read and experiencing the demo from start to finish within 2 minutes. This should be benchmarked against other forms of immersive media consumption such as touchscreens or VR headsets which involve large amounts of setup time (i.e., putting on the headset, connecting cables, changing batteries) and almost always require the presence of a human assistant and frequent disinfecting of the headset and controllers. We thus call for further research on the pros and cons of a touchless multimodal multisensory digital signage podium (e.g., accessibility, hygiene, user experience) that can be applied to human-DHA interactions, but also in other settings such as to create memorable experiences that promote public engagement with brands or content, such as at marketing events, shopping malls, and media art installations.

Our implementation however was not without flaws, and so we share here a list of considerations for improvement, which are however not limited to this singular example. The grab-release hand gesture was not robust enough and not so intuitive to the first-time user. It was initially chosen since it can be used with palm facing up towards the haptic display.

Other gestures like pointing or pinching could have been used, and/or further user tested during internal piloting [30]. The gesture recognition camera was placed facing down intentionally so that it was not exposed to strong lighting which emits infrared radiation as that would interfere and reduce its performance. Despite this, because the surface material of the podium's lower surface was glossy, reflected lights still manage to slightly deteriorate the camera's performance. A second haptic display could have been positioned under the hand to produce a more three-dimensional haptic effect like those seen in papers by Makino et al [18]. Such a haptic apparatus could have enlarged the set of haptic effects beyond the simple sensations shown in Figure 5. Another point for improvement is that Alnair's audio-visual interactivity could have been more spatially aware. For example, the webcam could have been used to adjust Alnair's height, gaze, and spatial audio, relative to the participant, therefore, offering a more personalized experience. Also, Alnair's multiethnic, multiracial, and gender-neutral avatar could have been designed more methodically, instead of the mixed average approach undertaken in our implementation. Finally, the podium distance to the screen could have been smaller to reduce the exhibit's overall footprint. Note that the printer was intentionally placed to the side, and not on the podium to increase participant throughput. Future research scope could therefore include revising the podium design to enable a variety of gesture-input and ultrasound mid-air haptic feedback. The inclusion of other non-contact haptic modalities, e.g., electrostatic or infrared, while also incorporating additional multi-modal interactions (voice/gaze) suitable for this or other applications.

We conclude our discussion by arguing that the 'phygital' world, i.e., where the physical meets the digital, need not be constrained to augmented reality (AR) headsets, and hence our work motivates further research in multimodal and multisensory interactions between humans and DHAs, such as a handshake, a caress, a high-five, or a game of rock-paper-scissors. The oracle demo experience we have described is but a first example where a DHA was able to reach out of their virtual environment and physically touch the user in the real world, and *vice versa* - a kind of cyber-physical portal. This tactile portal concept can potentially be generalized to other interactions, and contexts, and also to non-human digital avatars such as pets, or even animated objects. Finally, one could imagine the existence of additional synergies between voice and gesture input modalities, for example where deictic gestures complement deictic expressions [1] as is the case for novel human-car interfaces [32].

8 CONCLUSION

In this paper, we have presented a metaverse-inspired fortune-telling location-based DHA experience that is multisensory and multimodal. Unlike most current digital signage implementations that use conversational (voice-input, i.e., speech-to-text) methods to interface with DHAs, the interactive experience we have implemented (see Figure 1) and described in the previous sections (physical setup, user journey, and system architecture) combines DHAs with touchless mid-air haptics, and gesture recognition. Also, we presented examples of how these novel technologies could be assembled and integrated to produce engaging new digital signage experiences while also addressing aspects such as consent, privacy, and gender neutrality in the metaverse. Through our example implementation we diverge and discuss opportunities and challenges concerned with multimodal and multisensory DHA experiences, not limited to this specific implementation.

ACKNOWLEDGMENTS

O.G. acknowledges funding from the EU Horizon 2020 research and innovation programme under grant agreement No 101017746, TOUCHLESS.

REFERENCES

[1] Adrian Bangerter and Max M Louwerse. 2005. Focusing attention with deictic gestures and linguistic expressions. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Vol. 27.

[2] Gavin Buckingham. 2021. Hand tracking for immersive virtual reality: opportunities and challenges. *Frontiers in Virtual Reality* (2021), 140.

[3] Tom Carter, Sue Ann Seah, Benjamin Long, Bruce Drinkwater, and Sriram Subramanian. 2013. UltraHaptics: multi-point mid-air haptic feedback for touch surfaces. In *Proceedings of the 26th annual ACM symposium on User interface software and technology*. 505–514.

[4] Qing Chen, François Malric, Yi Zhang, Muhammad Abid, Albino Cordeiro, Emil M Petriu, and Nicolas D Georganas. 2009. Interacting with digital signage using hand gestures. In *Image Analysis and Recognition: 6th International Conference, ICIAR 2009, Halifax, Canada, July 6-8, 2009. Proceedings 6*. Springer, 347–358.

[5] The Business Research Company. 2023. *Conversational AI Global Market Report 2023*. www.thebusinessresearchcompany.com/report/conversational-ai-global-market-report

[6] Loic Corenthy, Marcello Giordano, Richard Hayden, Daniel Griffiths, Craig Jeffrey, Hannah Limerick, Orestis Georgiou, Tom Carter, Jörg Müller, and Sriram Subramanian. 2018. Touchless tactile displays for digital signage: mid-air haptics meets large screens. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–4.

[7] Nigel Davies, Sarah Clinch, and Florian Alt. 2014. Pervasive displays: understanding the future of digital signage. *Synthesis Lectures on Mobile and Pervasive Computing* 8, 1 (2014), 1–128.

[8] Andrew Di Battista, Christos Nicolaides, and Orestis Georgiou. 2021. Modelling disease transmission from touchscreen user interfaces. *Royal Society Open Science* 8, 7 (2021), 210625.

[9] Helen V Diez, Javier Barbadillo, Sara García, Maria del Puy Carretero, Aitor Álvarez, Jairo R Sánchez, and David Oyarzun. 2014. Interactive Multimodal Platform for Digital Signage. In *International Conference on Articulated Motion and Deformable Objects*. Springer, 128–137.

[10] Zhixin Fang, Libai Cai, and Gang Wang. 2021. MetaHuman Creator The starting point of the metaverse. In *2021 International Symposium on Computer Technology and Information Science (ISCTIS)*. IEEE, 154–157.

[11] Teresa Fernandes and Elisabete Oliveira. 2021. Understanding consumers' acceptance of automated technologies in service encounters: Drivers of digital voice assistants adoption. *Journal of Business Research* 122 (2021), 180–191.

[12] Orestis Georgiou, William Frier, and Oliver Schneider. 2022. User Experience and Mid-Air Haptics: Applications, Methods, and Challenges. *Ultrasound Mid-Air Haptics for Touchless Interfaces* (2022), 21–69.

[13] John Hardy, Enrico Rukzio, and Nigel Davies. 2011. Real world responses to interactive gesture based public displays. In *Proceedings of the 10th International Conference on Mobile and Ubiquitous Multimedia*. 33–39.

[14] Kyle Harrington, David R Large, Gary Burnett, and Orestis Georgiou. 2018. Exploring the use of mid-air ultrasonic feedback to enhance automotive user interfaces. In *Proceedings of the 10th international conference on automotive user interfaces and interactive vehicular applications*. 11–20.

[15] Darragh Higgins, Donal Egan, Rebecca Fribourg, Benjamin R Cowan, and Rachel McDonnell. 2021. Ascending from the valley: Can state-of-the-art photorealism avoid the uncanny?. In *SAP*. 7–1.

[16] Hannah Limerick. 2020. Call to interact: Communicating interactivity and affordances for contactless gesture controlled public displays. In *Proceedings of the 9TH ACM International Symposium on Pervasive Displays*. 63–70.

[17] Hannah Limerick, Richard Hayden, David Beattie, Orestis Georgiou, and Jörg Müller. 2019. User engagement for mid-air haptic interactions with digital signage. In *Proceedings of the 8th ACM international symposium on pervasive displays*. 1–7.

[18] Yasutoshi Makino, Yoshikazu Furuyama, Seki Inoue, and Hiroyuki Shinoda. 2016. HaptoClone (Haptic-Optical Clone) for Mutual Tele-Environment by Real-time 3D Image Transfer with Midair Force Feedback.. In *CHI*, Vol. 16. San Jose, CA, 1980–1990.

[19] Keenan R May, Thomas M Gable, and Bruce N Walker. 2017. Designing an in-vehicle air gesture set using elicitation methods. In *Proceedings of the 9th international conference on automotive user interfaces and interactive vehicular applications*. 74–83.

[20] James W Moore. 2016. What is the sense of agency and why does it matter? *Frontiers in psychology* 7 (2016), 1272.

[21] Jörg Müller, Dennis Wilmsmann, Juliane Exeler, Markus Buzeck, Albrecht Schmidt, Tim Jay, and Antonio Krüger. 2009. Display blindness: The effect of expectations on attention towards digital signage. In *Pervasive Computing: 7th International Conference, Pervasive 2009, Nara, Japan, May 11-14, 2009. Proceedings 7*. Springer, 1–8.

[22] Thanh Thi Nguyen, Quoc Viet Hung Nguyen, Dung Tien Nguyen, Duc Thanh Nguyen, Thien Huynh-The, Saeid Nahavandi, Thanh Tam Nguyen, Quoc-Viet Pham, and Cuong M Nguyen. 2022. Deep learning for deepfakes creation and detection: A survey. *Computer Vision and Image Understanding* 223 (2022), 103525.

[23] Kristine L Nowak and Jesse Fox. 2018. Avatars and computer-mediated communication: a review of the definitions, uses, and effects of digital representations. *Review of Communication Research* 6 (2018), 30–53.

[24] Jennifer Pearson, Gavin Bailey, Simon Robinson, Matt Jones, Tom Owen, Chi Zhang, Thomas Reitmaier, Cameron Steer, Anna Carter, Deepak Ranjan Sahoo, et al. 2022. Can't Touch This: Rethinking Public Technology in a COVID-19 Era. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–14.

[25] Zoey Reyes and Joshua Fisher. 2022. The Impacts of Virtual Reality Avatar Creation and Embodiment on Transgender and Genderqueer Individuals in Games. In *Proceedings of the 17th International Conference on the Foundations of Digital Games*. 1–9.

[26] Julie Rico and Stephen Brewster. 2010. Gesture and voice prototyping for early evaluations of social acceptability in multimodal interfaces. In *International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction*. 1–9.

[27] Doug Roble, Darren Hendler, Jeremy Buttell, Melissa Cell, Jason Briggs, Chad Reddick, Lonnie Iannazzo, Deer Li, Mark Williams, Lucio Moser, et al. 2019. Real-time, single camera, digital human development. In *ACM SIGGRAPH 2019 Real-Time Live!* 1–1.

[28] Gustavo Rovelo, Donald Degraen, Davy Vanacken, Kris Luyten, and Karin Coninx. 2015. Gestu-wan-an intelligible mid-air gesture guidance system for walk-up-and-use displays. In *Human-Computer Interaction–INTERACT 2015: 15th IFIP TC 13 International Conference, Bamberg, Germany, September 14-18, 2015, Proceedings, Part II 15*. Springer, 368–386.

[29] Emmanuel Sirimal Silva and Francesca Bonetti. 2021. Digital humans in fashion: Will consumers interact? *Journal of Retailing and Consumer Services* 60 (2021), 102430.

[30] Arthur Sluÿters, Mehdi Ousmer, Paolo Roselli, and Jean Vanderdonckt. 2022. QuantumLeap, a Framework for Engineering Gestural User Interfaces based on the Leap Motion Controller. *Proceedings of the ACM on Human-Computer Interaction* 6, EICS (2022), 1–47.

[31] Stella Sylaiou and Christos Fidas. 2022. Virtual Humans in Museums and Cultural Heritage Sites. *Applied Sciences* 12, 19 (2022), 9913.

[32] Robert Tscharn, Marc Erich Latoschik, Diana Löffler, and Jörn Hurtienne. 2017. "Stop over there": Natural gesture and speech interaction for non-critical spontaneous intervention in autonomous driving. In *Proceedings of the 19th acm international conference on multimodal interaction*. 91–100.

[33] Matthew Turk. 2014. Multimodal interaction: A review. *Pattern recognition letters* 36 (2014), 189–195.

[34] Santiago Villarreal-Narvaez, Jean Vanderdonckt, Radu-Daniel Vatavu, and Jacob O Wobbrock. 2020. A systematic review of gesture elicitation studies: What can we learn from 216 studies?. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference*. 855–872.

[35] Dong-Bach Vo and Stephen A Brewster. 2015. Touching the invisible: Localizing ultrasonic haptic cues. In *2015 IEEE World Haptics Conference (WHC)*. IEEE, 368–373.

[36] Robert Walter, Gilles Bailly, and Jörg Müller. 2013. StrikeAPose: revealing mid-air gestures on public displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 841–850.

[37] Daniel Wigdor and Dennis Wixon. 2011. *Brave NUI world: designing natural user interfaces for touch and gesture*. Elsevier.

[38] Kim Willems, Nanouk Verhulst, and Malaika Brengman. 2021. How COVID-19 could accelerate the adoption of new retail technologies and enhance the (E-) servicescape. *The Future of Service Post-COVID-19 Pandemic, Volume 2: Transformation of Services Marketing* (2021), 103–134.