





Article

Realistic Virtual Humans for Cultural Heritage Applications

Effie Karuzaki ^{1,*}, Nikolaos Partarakis ¹, Nikolaos Patsiouras ¹, Emmanouil Zidianakis ¹,
Antonios Katzourakis ¹, Antreas Pattakos ¹, Danae Kaplanidi ², Evangelia Baka ³, Nedjma Cadi ³,
Nadia Magnenat-Thalmann ³, Chris Ringas ², Eleana Tasiopoulou ² and Xenophon Zabulis ¹

¹ Institute of Computer Science, Foundation for Research and Technology (ICS-FORTH), N. Plastira 100, Vassilika Vouton, 70013 Heraklion, Greece; partarak@ics.forth.gr (N.P.); patsiouras@ics.forth.gr (N.P.); zidian@ics.forth.gr (E.Z.); tonykatz@ics.forth.gr (A.K.); anpattakos@ics.forth.gr (A.P.); zabulis@ics.forth.gr (X.Z.)

² Piraeus Bank Group Cultural Foundation, 6 Ang. Gerontas St., 10558 Athens, Greece; danae.kaplanidi@gmail.com (D.K.); chrringas@gmail.com (C.R.); TasiopoulouE@piraeusbank.gr (E.T.)

³ MIRALab, CUI, University of Geneva Battelle, Building A, 3rd Floor 7, Route de Drize Carouge, 1227 Geneva, Switzerland; ebaka@miralab.ch (E.B.); cadi@miralab.ch (N.C.); thalmann@miralab.ch (N.M.-T.)

* Correspondence: karuzaki@ics.forth.gr

Abstract: Virtual Humans are becoming a commodity in computing technology and lately have been utilized in the context of interactive presentations in Virtual Cultural Heritage environments and exhibitions. To this end, this research work underlines the importance of aligning and fine-tuning Virtual Humans' appearance to their roles and highlights the importance of affective components. Building realistic Virtual Humans was traditionally a great challenge requiring a professional motion capturing studio and heavy resources in 3D animation and design. In this paper, a workflow for their implementation is presented, based on current technological trends in wearable mocap systems and advancements in software technology for their implementation, animation, and visualization. The workflow starts from motion recording and segmentation to avatar implementation, retargeting, animation, lip synchronization, face morphing, and integration to a virtual or physical environment. The testing of the workflow occurs in a use case for the Mastic Museum of Chios and the implementation is validated both in a 3D virtual environment accessed through Virtual Reality and on-site at the museum through an Augmented Reality application. The findings, support the initial hypothesis through a formative evaluation, and lessons learned are transformed into a set of guidelines to support the replication of this work.

Keywords: virtual humans; virtual storytellers; augmented reality; virtual reality



Citation: Karuzaki, E.; Partarakis, N.; Patsiouras, N.; Zidianakis, E.; Katzourakis, A.; Pattakos, A.; Kaplanidi, D.; Baka, E.; Cadi, N.; Magnenat-Thalmann, N.; et al. Realistic Virtual Humans for Cultural Heritage Applications. *Heritage* **2021**, *4*, 4148–4171. <https://doi.org/10.3390/heritage4040228>

Academic Editor: Nicola Masini

Received: 31 August 2021

Accepted: 20 October 2021

Published: 1 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Virtual Humans (VHs) are today considered a commodity in the domain of computer animation, cinema, and games. In a sense, they are changing the way that audio-visual information is presented and in a gaming context, they have contributed to a more human-like interaction metaphor through Non-Playable Characters (NPCs). Furthermore, recent technical advances in computing technologies have made Virtual and Augmented Reality (VR and AR) easily accessible for the wide public through powerful mobile devices and inexpensive VR headsets. In this sense, the evolution of computing technology has made possible access to high-quality audio-visual content including 3D representation from a simple web browser.

AR and VR could benefit from realistic VHs. AR is an interactive experience that uses the real world as a reference and is enhanced by computer-generated sensory stimuli. In the case of visual AR, the goal is to create the pseudo-illusion that the computer-generated content is a seamless part of the environment. In the case of VR, the pursued illusion is that the virtual environment and everything in it are real—including the VHs. In both cases, lack of realism manifested as physical inconsistencies of virtual content with the environment

(physical or digital) are known to be a key factor of observed fatigue and misperception of distance in all types of audiovisual and Extended Reality (XR) displays [1,2]. For example, when VHs are part of the augmented content, lack of realism in very simple and ordinary animations produces dislike for the end outcome, due to the “uncanny valley” effect [3,4]. To avoid such inconsistencies, VHs should be as realistic as possible, including their appearance, animations, and speech.

In the context of digitization and preservation of Cultural Heritage (CH), AR and VR technologies have been widely adopted. Cultural institutions (including museums) all over the world seek compelling ways to reach new audiences and enhance the museum-visiting experience. AR and VR technologies could not be left out of such context, as they enhance the physical and virtual museum-visiting experience. Augmenting interactive exploration essentially allows for users to experience culture without the need to come into contact with the real objects, visit objects that do not have a physical presence or are in maintenance, while also offers additional experiences such as immersive experiences, personalized guidance [5,6] and augmentation of exhibits with several multimedia and storytelling [7]. Additionally, such technologies can direct visitors’ attention by emphasizing and superimposing techniques, which effectively enhance the learning experience [8]. Indeed, digitally mediated personalization and personalized learning are becoming prominent trends in museums in recent years. For example, through mobile apps, museums can provide supplementary information about exhibits or the museum itself [7]. In this context, VHs can be a valuable arrow in the curators’ arch, since they can impersonate persons of the past, storytellers, curators, guards, visitors, personal guides, and so on [9] and undertake the role of guiding visitors while providing extra information in the form of multimedia, offering visitors and enhanced museum-visiting experience. However, little work can be found in the literature about presenting a methodology for creating realistic VHs for presenting tangible and intangible CH aspects, such as presenting an exhibit and explaining its usage (tangible), or narrating stories of the past around the exhibits, conveying the life and work of people of a previous era, as an elder would narrate to their grandchildren (intangible), all according to the visitor’s interests. Such narrators would not only allow visitors to fully understand what they see in a museum’s space but also allow them to mentally travel into the past, providing a deeper understanding of how people lived and work back then.

In light of the above, this paper presents a cost-effective methodology for achieving realistic storyteller VHs creation for CH applications. The proposed methodology covers all steps of creation and presentation of virtual storytellers in various settings including VR and AR, focusing on their looks, movement, and speech. We do that through a case study in the context of the Mingei European project [10], which aims at representing and making accessible both tangible and intangible aspects of crafts as CH [11]. The presented methodology was used in the Mastic pilot of the Mingei project, which aims at preserving the traditional craft of mastic cultivation and processing [12]. This craft is unique in the world and takes place only at the Chios island of Greece. As previous research has shown that emotional responses caused by VHs are heavily affected by whether the VH’s looks adheres to their profession [13], our case study refers to creating realistic interactive VHs that represent actual workers of the Chios Mastic Grower’s Association, which undertakes the mastic tear processing from their initial form into their final form in the market and also into the famous Greek mastic gum. The final VHs will provide information about the machines exhibited at the Chios Mastic Museum and the lives of the people of that age to the visitors through narratives. However, the methodology presented is independent of the case study and can be applied for the creation process of realistic storyteller VHs for any cultural heritage application.

2. Background and Related Work

The usage of VHs in Digital Cultural Heritage (DCH) environments has been a subject of study by several research works [14,15]. For example, in [9] the affective potential,

persuasiveness, and overall emotional impact of VHs with different professional and social characteristics (a curator, a museum guard, and a visitor), in an immersive VM environment has been studied. In the study, persuasiveness relates to VHs' capacity to engage, affect, and stimulate emotional and cognitive responses by employing different narration styles. The authors underline the importance of aligning and fine-tuning narrative styles and contents to VHs, which should correspond in terms of appearance to their roles, and highlight the importance of affective components in their storytelling approach [16]. Testón and Muñoz in [17] undertake the challenge of immersing visitors in the museum in a way that encouraged them to discover the hidden stories of the ancient city of Valencia. To do that, they analyze different interfaces to achieve natural and humanized behaviour in a museum visit, concluding that VH solutions emerge as cost-effective, empathic mediums to engage new audiences and highlighting that narratives "represent a new way to discover hidden treasures from the past". In [17], the early stages of virtual guides for onsite museum experiences are presented. They used several portraits exhibited in the museum to build VHs representing the corresponding personalities, which were then used to present the exhibits and narrate stories of the past to the visitors. In another work [18], VHs were used for guiding users through virtual CH environments. The VHs provide users with the context and background of virtual exhibits through legends, tales, poems, rituals, dances, and customs etc. VHs have been also used for preserving and simulating cultures [19] and teaching crafts. For example, [20] utilize VHs in VE for teaching the craft of printmaking, while [21] utilizes VR as a tool for communicating the craftsmanship of engraving. In [22], Danks et. al. combines interactive television storytelling and gaming technologies to immerse museum visitors with artifacts on exhibition, engaging the user into physical space using virtual stories, while [23] describes the use of a VH as a means for providing interactive storytelling experiences at a living history museum.

Research has shown that VHs can affect the virtual experience and stimulate attention and involvement [24–26] and thus can make the stories presented in VEs more credible and thus influence users positively and constructively. Furthermore, they contribute to the suspension of disbelief and which enables the user to become immersed and follow their story and turn of events. The categorization of areas that should be paid attention to when designing VEs has been defined as i. Information Design, ii. Information Presentation, iii. Navigation Mechanism, and iv. Environment Setting. Information design and presentation, in particular, stress the fact that the users should be able to understand the significance of information providing engagingly through narratives [27]. Narratives, when successfully used for guiding the user through a VM, motivate visitors to stay longer and see more [28]. Besides, visiting a cultural site in the company of a guide who tells fascinating stories about the exhibits becomes a memorable experience, and when human guides are a scarce resource—or doors to a museum or gallery are closed, VHs can bring these experiences to a wider audience as well as provide a welcome invitation to discovery [29].

As previous research has shown, the VH's looks along with their behavior heavily affect the user's response to them. Thus, it is important for the VHs to look, move and sound natural. In this vein, this paper will provide a methodology to address those issues, the tools utilized, and why we have concluded in utilizing them. This section provides a background about the choices made on the tools that helped us achieve each of the three main goals for our characters—looks, motion, and speech.

2.1. Realistic VHs Creation

As we focus on realistic VHs, solutions that aim at creating cartoon-like VHs, such as the Ready Player Me [30] were rejected and instead we examined solutions that offer integrated solutions for building high-resolution realistic ones. Among the latter, Character Creator 3 [31] by Reallusion is a full character creation solution for designers, enabling easy creation and customization of realistic-looking character assets. Another character creation software is DAZ studio [32], which aims at users who are interested in posing human figures for illustrations and animation. MakeHuman [33] is a free, open-source,

interactive modeling tool for creating custom 3D human characters, however, according to their official documentation, some of its tools have not yet been created or are in the early stages of development (poses, animation cycles, managing facial expressions, hair, and clothes). Another software that promises realistic VHs is the Didimo [34], which focuses on the creation of life-like digital representations of real people—however, it currently only supports the generation of human heads. Having reviewed other solutions in addition to the aforementioned ones, we have decided to adopt the Reallusion's Character Creator 3 suite, as it creates high-resolution realistic, whole-body VHs and can perfectly collaborate with the tools chosen for realizing realistic animation, lip-synching, and facial expressions.

2.2. Realistic Presentation of Human Motion

Probably the most sufficient way of achieving believable human-like animation of VHs is motion capture (mo-cap) which refers to the process of recording the movement of objects or people. It is used in military, entertainment, sports, medical applications, and for validation of computer vision [35] and robotics [35,36]. Mo-cap technology was used as early as the 19th century when photographer Eadweard Muybridge studied the motion of humans and animals through stop-motion photography [37]. The basic principles of his study would soon serve filmmakers when Max Fleischer invented the Rotoscope in 1915 (Figure 1). In essence, a camera would project a single frame onto an easel so that the animator could draw over it, frame by frame, capturing realistic movement for the on-screen VHs. Rotoscoping was partially used in 1938's 'Snow White and the Seven Dwarves', 'Star Wars', and others [38].

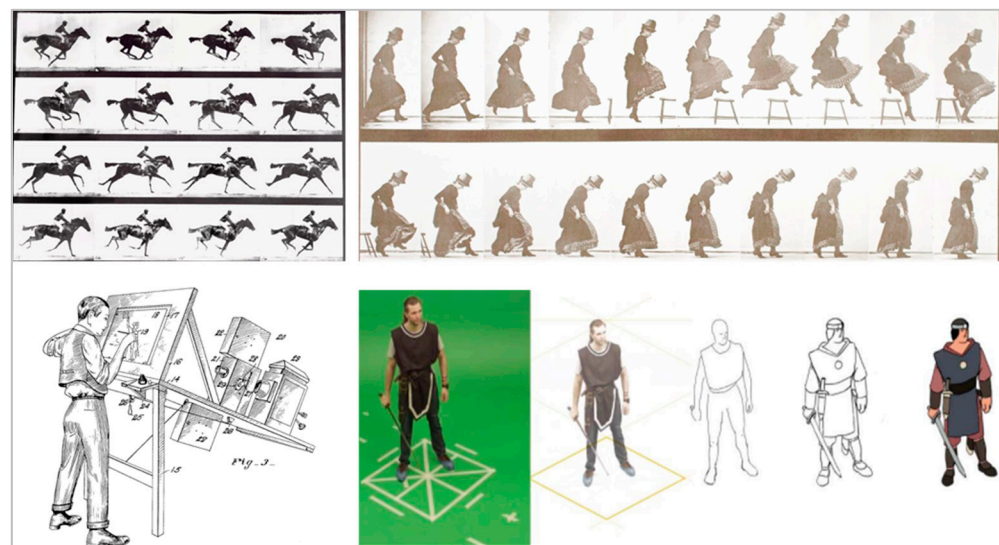


Figure 1. (top) Stop-motion photography; (bottom left) A Rotoscope device; (bottom right) Drawing a character based on the frame captured by the rotoscope device.

During the following years, biomechanic organizations were making strides in monitoring and tracking the human body's motions for medical research, introducing the concepts of degrees of freedom and hierarchical structure of motion control. In the early 1980s, Tom Calvert attached potentiometers to a body to drive computer-animated figures to study movement abnormalities [39], while Ginsberg and Maxwell presented the Graphical Marionette, a system using an early optical motion capture system that featured a bodysuit with the LEDs on anatomical landmarks and two cameras with photodetectors that returned the 2-D position of each LED in their fields of view. The computer then used the position information from the two cameras to drive a 3D stick figure [40]. This technique was evolved through the next years and reached a state where it was used in movies, from "The Polar Express" (2004) to the "Lord of the Rings" (2001–2003) and others (Figure 2).



Figure 2. (top) Tom Hanks performing for ‘The Polar Express’; (bottom) Andy Serkis performing the Gollum’s moves in “Lord of the Rings: The two towers”. Both actors wear suits with reflective markers.

The aforementioned mo-cap techniques that include such retroreflective markers belong to the greater family of “Optical” motion capturing systems. They essentially utilize data captured from image sensors to triangulate the 3D position of a subject between two or more cameras calibrated to provide overlapping projections. One can easily understand that such motion capturing systems are very costly since it requires a lot of cameras to be used (typically a system will consist of around 2 to 48 cameras). Many companies offer complete optical motion capturing solutions, including suits, markers, cameras, and software. Some are Qualisys [41], Vicon [42], NaturalPoint [43], and Motion Analysis [44]. Markerless optical systems have also been developed [44–50]. Such systems use the OpenPose system [51] or multiple sensors like Kinect to track body motion; however, they are less accurate. Hybrid approaches have also been developed, enabling marker-based tracking and markerless tracking with the same camera system [52].

Non-optical motion capturing (mo-cap) systems do not use any sort of physical marker. This is much more affordable since it is mostly software-based and requires much less equipment [35]. Essentially, they track the motion of sensors, e.g., inertial, stretch sensors etc. For example, magnetic systems calculate position and orientation by the relative magnetic flux of three orthogonal coils on the transmitter and each receiver [53]. The relative intensity of the voltage or current allows computers to calculate both range and orientation through mapping the tracking volume. However, magnetic mo-cap systems are susceptible to magnetic and electrical interference from metal objects in the environment, like rebar (steel reinforcing bars in concrete) or electrical sources such as monitors, lights, cables, and computers [53], making them unsuitable for recording motions in most house and office environments. Other mo-cap systems can utilize wearable stretch sensors [54–56], which are not affected by magnetic interference and are free from occlusion. However, the data collected are transmitted via Bluetooth or direct input, severely limiting the freedom of the actors to move. Such systems are used to detect minute changes in body motion [57] and are rarely used for motion capture. Another solution for motion capture in the bibliography is the mechanical motion capture systems, which require the user to wear physical recording devices for each joint, (exoskeleton) such as the Gypsy Mocap

system [58]. Such systems can suffer from motion data drift, noise, and the limitations of mechanical compared to human motion [59].

The most popular non-optical motion capture technology makes use of inertial sensors. Most of the systems that belong to this category make use of Inertial Measurement Units (IMUs), which contain a combination of gyroscope, magnetometer, and accelerometer. The data collected from such units are wirelessly transmitted to a computer where the motion is translated to a VH. This allows for great freedom in movements, excellent accuracy, ease of use, zero occlusion issues, quick and easy setup, and the ability to record in various environments. These benefits, along with the significantly lower cost than the optical-based methods, make inertial systems increasingly popular amongst game developers [60].

2.3. Realistic Presentation of Human Speech

For a VH to speak, two activities have to be synchronized—first, lip moving and face morphing to support facial expressions, and second, audio containing the VH's speech content performed with the VH's voice. Both activities are very important for presenting a realistic virtual speech, especially when there is a need for a very expressive VH. However, in the case of VHs used for narrations in CH sites, more emphasis should be put on the audio part of the speech, as narrators usually do not express very intense emotions (such as surprise, scare, enthusiasm, etc.) that need to be emphasized on the VH's face as it is the case with most cartoonish characters met in games. Most of the time, lip movement and facial expressions of storyteller VHs do not directly contribute to a better user experience and will be hardly noticed by users, as the respective VHs only occupy a small proportion of the user's view window. More specifically, when it comes to AR the VHs will be displayed on a relatively small screen of a tablet or mobile phone, while in VR the VH will be put in a peripheral space, as the main exhibit should be put directly in front of the user's view window to draw their attention. For example, VH narrators for cultural applications can be presented on the side of the user's view window and show mild emotions. Furthermore, in some cases, the character is put in the middle of the user's view window, expressing more vivid emotions [20,61,62].

Face morphing and lip-synching can be automatically generated by software, based on the text or audio that has to be performed, or it can be captured directly from an actor's face. The second option is more accurate and can produce higher-quality facial morphing and lip motion. This instantly inhibits the use of automatic Text-to-Speech (TTS) synthesizers, as it is almost impossible to synchronize automatically-generated speech with the captured lip motion. Moreover, captured facial expressions would express emotions that synthesized voices cannot fully express, leading to an uncanny effect that would break the suspension of disbelief. Capturing facial expressions should always be accompanied by audio recorded by a human actor, preferably at the same time as the facial capturing. However, precise synching of the recorded audio and the captured lip-motion is not an easy task, and additionally, it requires re-capturing of both audio and facial expressions if the narrated story changes (even if the curator needs to add or remove a single word from the narration). On the other hand, automatically generated lip-synching and face morphing is made possible via specialized software that drives the VH's facial anchors in a way that their expression and lips match the given text or audio. TTS software works initially by analyzing the input text, then processed it, and finally convert it to digital audio through concatenating short sound samples from a database [63]. TTS software is cheap, easy to use, and quick to adopt as it requires no voice recordings and the input text can change any time, allowing curators to freely edit narrations. However, as speech is automatically generated, it still faces some drawbacks, for example, inadequate expression of emotions, failing to perform a spontaneous speech in terms of naturalness and intelligibility, natural-sounding, and others [64]. It can be used in cases where natural voice recordings are hard or impossible to take place, such as in screen readers [65], translators, etc. Voice recordings are a much better solution when it comes to performing specific text parts that do not change frequently. They pose a serious time overhead since the recordings should

be performed by actors, usually in a studio, and then processed with audio-processing software. Once recorded, it is difficult to correct wrong pronunciations or change a word, since such changes require time-consuming audio processing or recording a new voice track from scratch. However, voice recordings sound natural, as actors can color their voices and express emotions. They are much more pleasing to hear and are more comprehensible.

Lip-synching is a technical term for matching lip movements with sound. The respective software accepts voice recordings as input, which they analyze to extract individual sounds, known as phonemes. Then, the program uses a built-in dictionary to select the appropriate viseme (mouth shape) for each sound. Much research work has been done on lip-synching [66–68]; for example, Martino et al. in [69] generate speech synchronized 3D facial animation that copes with anticipatory and perseverative coarticulation.

Except for the aforementioned research works, many enterprise solutions promise quality lip-synching for VHs as well. CrazyTalk [70] for example is a facial animation and lip-synching tool mainly targeting the creation of cartoonish effects that use voice and text to vividly animate facial images by defining the facial wireframes in them. Papagayo [71] is another lip-synching program for matching visemes with the actual recorded sound of actors speaking. The lip-synching process is not automated—the developers should provide the text being spoken and drag them on top of the sound’s waveform until they line up with the proper sounds. Salsa LipSync Suite [72] provides automated, high quality, language-agnostic, lip-sync approximation for 2D and 3D characters, offering real-time processing of the input audio files to reduce/eliminate timing lag. It is also capable of controlling eye, eyelid, and head movement and performs random emote expressions, essentially providing a realistic face motion for the target 3D characters. In the context of storyteller VHs for CH applications, we propose lip-synching over facial capturing, as it requires no synchronization between audio and face motion, and it is easier for curators to revise narrated texts, either via sound-editing or via re-recording the target speech with no need for face re-capturing.

2.4. AR and VR Presentation

The presentation of virtual content within visual representations of a scene captured through a camera has been often achieved using physical markers in designated locations and arrangements. After the proliferation of unique keypoint features in Computer Vision, markers are substituted by (almost) unique key point features occurring at multiple scales in the environment, mainly in the form of visual textures. In the latter case, marker placement is substituted by a prior reconstruction of the environment to find and spatially index key points and their arrangement in the 3D environment. When the system is later presented with an image of the environment, it recognizes these points and estimates the pose of the camera. Once the camera position is estimated a virtual camera is (hypothetically) placed at the same point and content is ray-traced, visually predicting its (hypothetical) appearance on the surface that is imaged by the camera.

Since the advancement of holographic technology, AR headsets are evolving including interactive features like gesture and voice recognition, as well as improvements in resolution and field of view. In addition, untethered AR headsets paved the way for mobile experiences without the need for external processing power from a PC. Such embedded systems, facilitate great tools to represent Virtual Museums (VMs) [73] due to their lack of cables and enhanced interactive capabilities. VMs are institutional centers in the service of society, open to the public for acquiring and exhibiting the tangible and intangible heritage of humanity for education, study, and enjoyment. In addition, True AR has recently been defined to be a modification of the user’s perception of their surroundings that cannot be detected by the user [74] due to their realism. VHs and objects should blend with their surroundings, supporting the “suspension of disbelief”.

In recent years, many approaches to holographic CH applications emerged, each one focusing on a different aspect of representing the holographic exhibits within the real environment. A published survey [75] investigated the impact of VR and AR on the

overall visitor experience in museums, highlighting the social presence of AR environments. Papaefthymiou et al. [76] presented a comparison of the latest methods for rapid reconstruction of real humans using as input RGB and RGB-D images. They also introduce a complete pipeline to produce highly realistic reconstructions of VHs and digital asserts suitable for VR and AR applications. The InvisibleMuseum project contributed with an authoring platform for collaborative authoring of Virtual Museums with VR support [77]. Another project [78] integrates ARCore and ARCore to implement a portal-based AR virtual museum along with gamified tour guidance and exploration of the museum's interior. Storytelling, presence, and gamification are three very important fields that should be considered when creating an MR application for CH. Papagiannakis et al. [79] presented a comparison of existing MR methods for virtual museums and pointed out the importance of these three fields for applications that contribute to the preservation of CH. In [80] fundamental elements for MR applications alongside examples are presented. Another recent example [81] presented two Mixed Reality Serious Games in VR and AR comparing the two technologies over their capabilities and design principles. Both applications showcased antiquities through interactive mini-games and a virtual/holographic tour of the archaeological site using Meta AR glasses. Abate et al. [82] successfully published an AR application for visualizing restored ancient artifacts based on an algorithm that addresses geometric constraints of fragments to rebuild the object from the available parts.

3. Proposed Methodology

The primary hypothesis of this research work regarding achieving realistic storytelling animation for VHs is that it is important that they look, move, and sound natural. In this vein, and considering the above analysis, we have decided to use ultra-high-resolution VHs to make them look realistic, mo-cap technologies to ensure that animations look natural, and software that can use voice recordings, lip-synching and facial expressions. The overall workflow proposed is shown in Figure 3. First, high-resolution VHs have to be created. Then, a motion-capturing suit should be used to record the VH's movement for each of the stories we want them to narrate. Although any mo-cap system could be adequate for this task, we propose the mo-cap suit as a cost-efficient yet effective mo-cap system in comparison to any other solution (cost of 2000 euros per suit in conjunction to ~200K for a vicon room). The suit provides accurate recording and can cope with visual occlusions which is the main drawback of optical systems. Furthermore, we propose that the corresponding voice clips of each narration should be recorded in parallel with the motion capturing, as this will allow for perfect synchronization between the recorded animation and the audio. Then, to retarget the captured animation to the VH, a game engine is facilitated (in our case Unity). Finally, human voice recordings are proposed to be used for the narrations instead of automatically generated ones via TTS. Finally, we propose that the face and lips of the VH should be automatically controlled via software and that face-capturing solutions should be avoided in the context discussed. We justify these propositions in Sections 3.1–3.6.

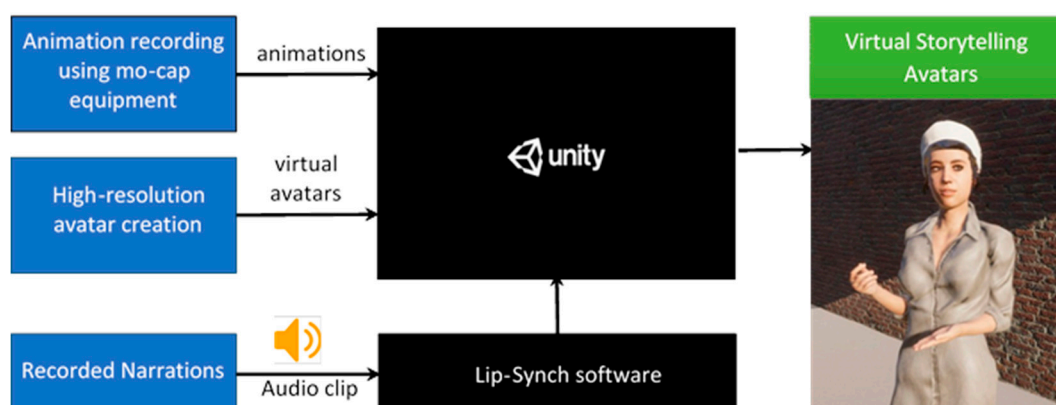


Figure 3. The overall workflow we adopt to bring narrator VHs to life.

3.1. Implementation of VHs

The VHs' bodies and clothes were created to obtain one unified and optimized model, enhancing the visual impact of the characters with texture mapping and material editing. The 3D generation of the virtual bodies has also to take into consideration the total number of polygons used to create the meshes to keep a balance between the 3D real-time simulation restrictions and the skin deformation accuracy of the models.

For VHs acting as conversational agents are meant to be part of a storytelling scenario. The requirement to use a blend-shape system for the facial animation meant working with software that supports from one hand the external BVH files for the animation of the body and from the other hand gives tools for controlling the facial animation. Reallusion pipeline fits all these requirements since it is a 2D and 3D character creation and animation software with tools for digital humans' creation and animation pipelines, that provides a motion controller for the face, body, and hands (see Figures 4 and 5).

Once the character is created in CC3, it can be directly exported to iClone [83] which is the animation module where the external motion file can be tested after its conversion inside 3Dxchange [84].

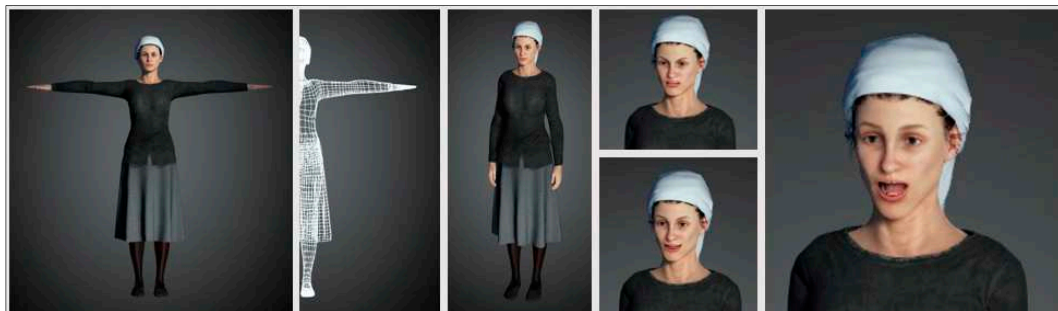


Figure 4. Fully rigged female VH with face and body control.

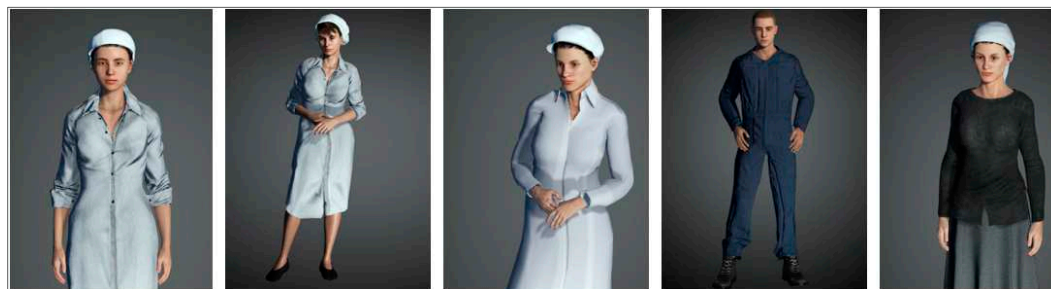


Figure 5. VHs of conversational agent type.

3.2. Animation Recording

After carefully analyzing the pros and cons of each of these systems, always considering their cost, we chose the Rokoko suit due to its high accuracy, portability, easiness of setup and use, and its overall price to quality ratio. The SmartSuite uses 19 Inertial Measurement Unit (IMU) sensors to track full-body, minus the fingers, which are tracked by the SmartGloves. The sensors have a wireless range of up to 100 m and require no external parts. The recordings, acquired using the suit, can be trimmed within the Rokoko Studio and exported under various mainstream formats (FBX, BVH, CSV, C3D) using various skeletons (HumanIK, Mixamo, Biped, and Newton).

Using the Rokoko equipment and software, we put on the suit and the gloves and recorded unique animations for each narration. For the narration moves to be more realistic, we also narrated the stories during the recordings and used a voice recording program to capture our voice. In this way, the synchronization of voice and movement in the narration

was a lot easier, and it guaranteed a more natural narration. Moreover, to further enhance the realism of the animation we used male and female “actors” for this process according to the scenario of the narration and the gender of the expected VH by the scenario narrator (see Figure 6).



Figure 6. (left) Female “actor” (right) Male “actor”.

Once the narration animations were recorded, we segment then [85] and we exported them in .fbx format, using the Newton skeleton. This action essentially creates a series of bones, body joints, and muscles, and defines their rotations in the 3D space over time.

3.3. Retargeting Recorded Animations to Virtual Humans

The next step is to import the narration animations and the VHs into the Unity game engine. After that, we can add the VH to our scene, and define an animator component to control their animations. The controller defines which animations the VH can perform, as well as when to perform them. Essentially, the controller is a diagram, which defines the animation states and the transitions among them. In Figure 7, an animation controller is shown where the VH initially performs an idle animation, and it can transit (arrows) to a state where the character introduces herself (“Self_introduction”) or to a state where she narrates a specific process (“Narration about Sifting Process”).



Figure 7. Simple animator controller for a narrator VH.

3.4. Recording Natural Human Voice in Virtual Narrators

Recent studies on the usage of state-of-the-art TTS synthesizers instead of human voices recordings [86] have indicated that a human voice is still preferable: both in terms of exhibited listener facial expressions indicating emotions during storytelling, as well as in terms of non-verbal gestures involving head and arms. Indeed, the main purpose of storyteller VHs is to complement a user’s visit to a CH site with audio stories; the VHs remain on the side of the user’s view window and help users to understand the exhibit in front of them or to discover more information around it. Thus, the narrator’s voice should sound clear and natural, while slight fluctuations in the narrator’s tone, speed, and volume can arise users’ interest and draw their attention to what’s important. This unfortunately comes with the cost of having to process or re-record the audio clip every time the narration script changes. In this vein, we propose to record natural human voice for every narration

part that the humans have to reproduce, in separate audio clips so that the re-recording (if needed) should be easier. In this work, we recorded human voice and then all audio clips were trimmed and lightly processed to remove excess parts and ensure the audio level will be equal among all the clips.

3.5. Lip Synchronization and Face Morphing

When it comes to VH's facial expressions, we propose to avoid using face capturing techniques for building storyteller VHs for CH applications. That is because, in practice, storyteller VHs do not make very vivid facial expressions (such as surprise, fear, over-excitement, etc.), thus mild facial morphs automatically controlled by software will cover our needs. Secondly, building virtual storytellers in the context of CH applications implies that curators will be in charge of providing the scripts that the VHs should narrate, and they should be able to slightly alter those scripts without needing to re-capture facial animations from scratch; they should only need to process or re-record the respective audio clips. In the case of face capturing, the slightest change in the scripts would make the whole face look out-of-synch, as the VHs would move their lips in a different way than one would expect during narrations. The same would happen for the different languages supported—each language would require a new face capture to look and feel natural. Auto facial morphing and lip-synching provide the advantage of automatically controlling the VH's face based on the provided audio. In terms of quality, automatic face morphing and lip-synching results are inferior to facial capturing; but as the VH's voice and body language remain the main focus of the users this is no problem.

In the light of the above, in this work, we used software for controlling both facial morphs and lip-synching, as these two should comply with each other. We have used the Crazy Minnow Studio's Salsa lip-sync suite [87], as it creates face morphs and lip synchronization from any given audio input, produces realistic results, and is fully compatible with the Unity game engine and the software used for the creation of the VHs. Such compatibility is important because the lip-synch algorithms mix the existing blend-shapes of the VHs, and such blend-shapes differ depending on the software used to create the VHs. Following the aforementioned steps, we have applied Salsa lip-sync to the VHs provided by the Miralab and we assigned to each VH the corresponding recorded animations. The narration stories refer to the mastic chewing gum creation process (Mastic Pilot), the carafe creation process (Glass Pilot), and the ecclesiastical garments creation process (Silk Pilot).

3.6. Putting Them All Together

The unity game engine was used to compile VHs, animations, lip-synching, and voice recordings together. The animations and the VHs were imported using the Filmbox (.fbx) format [88], and a humanoid rig was applied to them. Then, animation controllers were built. Each character was bound to one animation controller which defines which animations he/she will be able to perform and controls the transitions among them using transition parameters. Such parameters can be then triggered via code each time we need a specific animation to be played. Colliders were also used to define proximity areas around the VHs so that specific parameters would be triggered upon the trespassing of them. An example is that, when a player's VH approaches a virtual narrator, the latter greets and introduces itself. Such a collider is shown with green lines Figure 8 (left). On the right part of the image, the collider's properties are shown. Notice that the collider is set to be a trigger because we don't want to stop other objects from entering the collider area, but we need the VH's introduction animation to be triggered upon a character's trespassing the collider's borders.

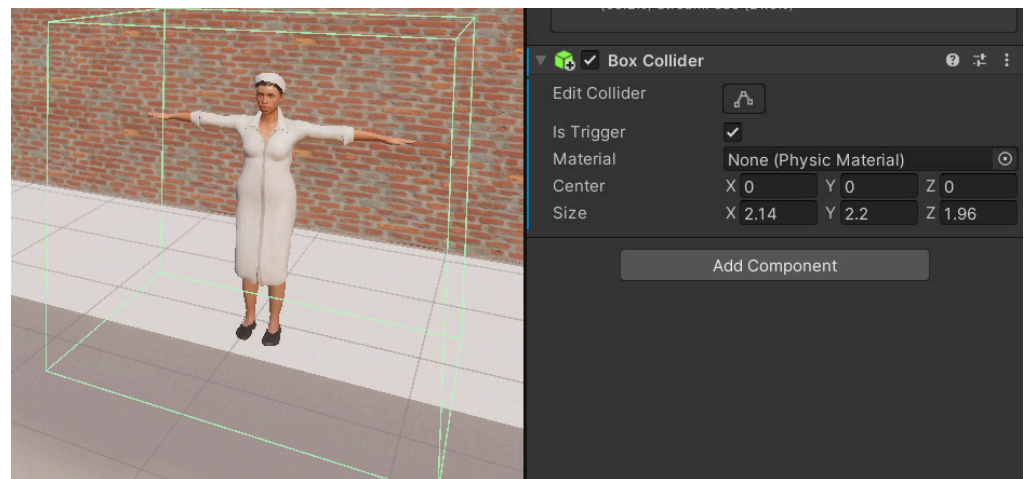


Figure 8. (left) green lines that delimit character's collider; (right) collider's properties.

4. Use Case

The Chios Gum Mastic Growers Association is agricultural cooperation established in 1937 in Chora of the island of Chios, after a long period of crisis that the mastic market faced. While its establishment helped mastic growers have better work conditions concerning their compensation. The use case study in the context of this research work regards the traditional mastic processing that takes place in the mastic growers association, on the island of Chios, Greece. VHs have been created once (following the aforementioned propositions) and have been deployed in two CH applications. In all applications, the VHs represent actual workers of the association, and they are narrating stories from their personal life, their work-life, and their duties at the factory. Those narrations are created based on real testimonies as explained in 0. Through them, the museum visitors can virtually travel back in time to that era and learn how people lived, and also about the mastic processing stages, the machines' functionality, and more. In this vein, the AR application regards the augmentation of the physical museum exhibits with narrator VHs. The second application, which enables a virtual visit to the old factory of the traditional Chios mastic chicle. The idea is that users can explore the 3D models of the museum's machines via 3D while the VH standing in front of each one will provide extra information about the machine, the respective step of the process line, and their personal lives. The production line, the VHs, and the scripts that they will narrate have been defined after careful examination of the requirements that the museum curators provided and analyzing former and present-day workers of the Association (research of the late 2000s) kept in the PIOP archive.

4.1. Requirements Analysis and Data Collection

4.1.1. Production Line at the Warehouse and the Factory

The production line at the Chios Mastic Grower's association is illustrated in Figure 9. The top part of the figure regards the process of collection, cleaning, and sorting mastic tears based on their quality resulting in the following byproducts (a) mastic oil, (b) mastic teardrops, and small tears used for gum production. In the case of chewing gum, the production takes place in the factory. There, workers make the chicle mixture by adding natural mastic, sugar (optional), butter, and cornflour in a blending machine (h). When the chicle dough is ready (i), the workers transfer it to a marble counter on top of which they have sprinkled icing sugar and knead the chicle dough to form 'pies'. The 'pies' are then placed on wooden shelves to cool before being transferred to an automated machine called 'cutting machine' (j). The cutting machine shapes the pies into sheets and cuts them into gum dragées (k). The dragées are left to cool again on wooden shelves. When they are ready, workers break the sheets to separate the formed dragées (l). If the dragées are not well shaped, they are sent back for heating in the blending machine. If they are well

shaped, they are loaded into the candy machine to be coated with sugar (m). After the coating has finished, the dragées are left to cool down once again, and then are polished using a revolving cylinder (n). When the chicle dragées are ready, they are packaged and packed in boxes (o) [89].

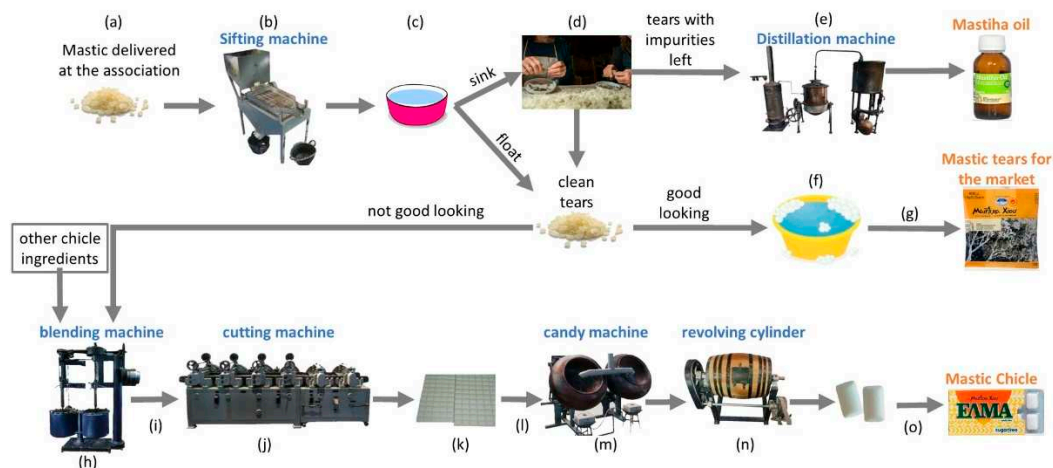


Figure 9. The production line of the Chios Mastic Grower's Association. The steps are explained within the text.

4.1.2. Stories of the Workers for the Virtual Warehouse and Factory of the Chios Gum Mastic Growers Association

The virtual factory of the Chios Gum Mastic Growers Association is based on the exhibition space of the Chios Mastic Museum at the island of Chios in Greece. The museum's space is designed to showcase the machinery that the Association first used during the 1950s and 1960s. The machines for the exhibition were a donation of the Association to the Piraeus Bank Group Cultural Foundation (PIOP) to include a section in the Chios Mastic Museum regarding the industrial processing of mastic. The museum space aims to introduce visitors to the industrial aspect of mastic processing during the 1960s through its exhibition, and offer a unique experience of interacting with some of the machines.

In the context of the Mingei project, after completing a series of co-creation meetings, it was decided to enhance the museum-visiting experience through storytellers. Each of these VHS represent a worker of the association who is in charge of the process step carried out by the respective machine. The goal of using VHS is two-fold: Firstly, they can provide information about their personal lives and their work lives, virtually transporting the museum visitors to another era, allowing them to get a deeper understanding of those times. Secondly, it has been observed that the visitors do not follow or understand the order of processing mastic to produce mastic chewing gum. VHS can fill in this gap by providing information about each exhibit around them and by explaining the whole process to the visitors. To create the stories that those VH workers would narrate, it was necessary to go through the PIOP archive, and more specifically, the archive part containing oral testimonies of former and present-day workers of the Association (research of the late 2000s). From these testimonies, we have been able to extract information regarding the gender of the workers in each process, their age, their family background, as well as other information concerning the Chian society, which we then used to create the stories that each VR would narrate. One of the most significant observations was that the majority of workers at the warehouse and factory of the Association were women. The age of the interviewees spanned from forty to eighty years old in twenty-three participants. As not all of the interviewees worked at the Association at the same time, it is interesting to spot the differences and similarities in their testimonies to unveil the developments made at those times regarding the working environment and/or in their personal lives.

4.2. Implementation

4.2.1. Creation of Stories

The profiles and stories of the VHs are a mix-and-match of the material in the oral testimonies. Eight VH have been created, seven of them representing people coming from different villages of southern Chios (also known as mastic villages) and one VH represents a woman coming from Sidirounta (a northern village). The represented villages were chosen after they repeated appearance in the archive, either in singularity or by wider locality (e.g., the village of Kini might have been referenced once, so the profile of the participant will correspond better with those of participants coming from villages of the same area, for example from Kalamoti). The age of the VHs was also defined as the middle age of the participants coming from villages of the same area.

In creating the content of the profiles and the stories of the VH workers, it was sought to represent how life at the villages was, how the worker grew up in the village (i.e., education, agricultural life, leisure time, adolescence, and married life), what led them to seek work at the Association in Chora of Chios, how their working life in the Association was, and in which process(es) they worked in. All this information is divided into sections according to (a) family background and early and adult years of life, (b) work-life in the Association, and (c) explanation of the processes. Through this process, eight personas were created. Each persona has a different name, age, work experience, and family life background, which were then imprinted into eight story scenarios. Then, eight VHs were created based on these personas.

4.2.2. Creation of Virtual Humans

As explained previously, based on the requirements analysis and the stories that are to be narrated, eight VHs were created; seven females and one male. This decision reflects the disproportion of the workers' sex at the Chios Gum Mastic Growers' Association, where women workers were preferred over men in most steps of the production line.

Life and work conditions back then. Figure 10 shows an example of a VH in different poses, while Figure 11 shows the total of 8 VHs built for the Chios Mastic Museum.



Figure 10. An example of a female VH (Stamatia).

The Reallusion's Character Creation3 (CC3) software was used for creating the VHs. Their outfit has been designed to match the actual clothes that workers wore at the factory—mostly a white robe and a white cap, while their facial and body characteristics were designed to match those of an average Greek woman living at Chios island. This step was essential, since, as explained previously, making the characters look like actual workers of that time can travel users back in time and prompt them to learn more about.



Figure 11. All 8 VHs built for the Chios Mastic Museum.

4.2.3. AR Augmentation of the Museum's Machines

In the context of the Mingei project, an AR application has been built to augment exhibits of the Chios Mastic Museum with VHs as described in Section 4.2.2. VHs were created following the methodology proposed in this paper. Viewing the machines through the museum's tablets, the visitors will be able to see VHs standing next to them, ready to share their stories and explain the functionality of the respective machines. The exact VH's location will be initially defined by a museum's curator, and visitors will select the story they are interested in by selecting the story from the left part of the tablet's screen. Due to COVID-19, the app has not yet been installed in the museum (initially planned for early 2021 but now it is scheduled for October 2021). In Figure 12, a screenshot of the application running in our lab is shown.

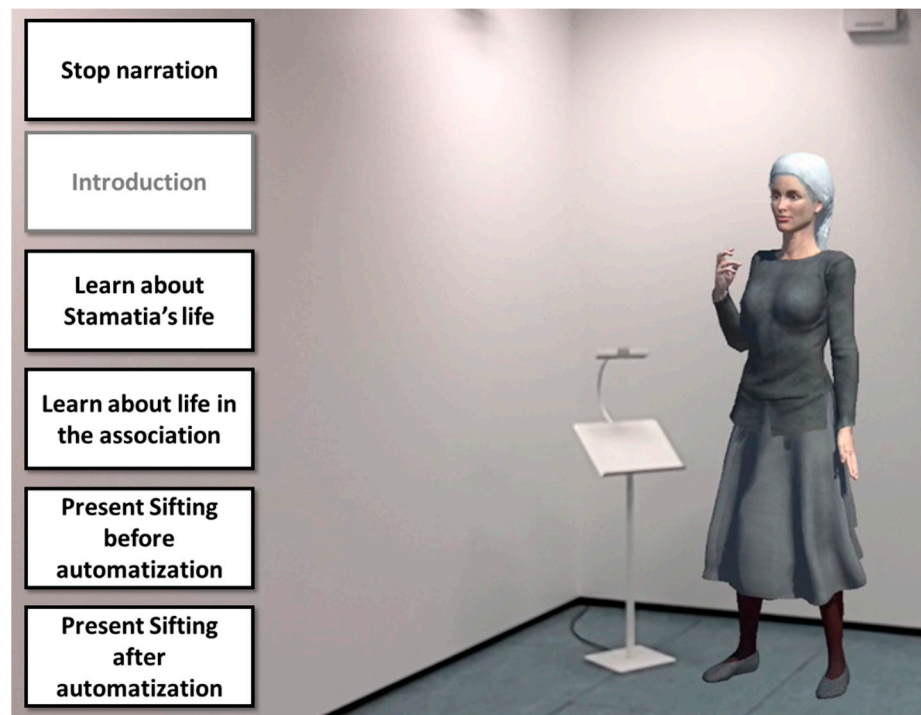


Figure 12. A VH narrating a story.

4.2.4. A Tour inside a Virtual Mastic Factory

The second application created in this use case regards a 3D model of an old mastic factory, where visitors can discover the machines met in the chicle and mastic oil production line and interact with VHs standing before them. Each machine carries out a specific task in the mastic chicle/oil production line. They have been reconstructed from the machines that are exhibited in the Chios museum; the machines were thoroughly scanned using a handheld trinocular scanner. Finally, the 3D was further processed using the Blender 3D creation and editing software [90]. The 3D reconstruction technique works well for objects that comprise flat surfaces that reflect light in a straight manner. However, when it comes to scanning and reconstructing curvy or hollow objects it is really difficult to achieve a decent result. Such objects were post-processed with 3D editing software. Figure 13 displays the reconstruction of the Candy Machine, which features both curvy and hollow surfaces. The machine as it was photographed at the museum is displayed on the left-hand side of the image. In the middle, one can witness the automatic reconstruction of the machine and on the right the machine after post-processing which is very close to the original machine.

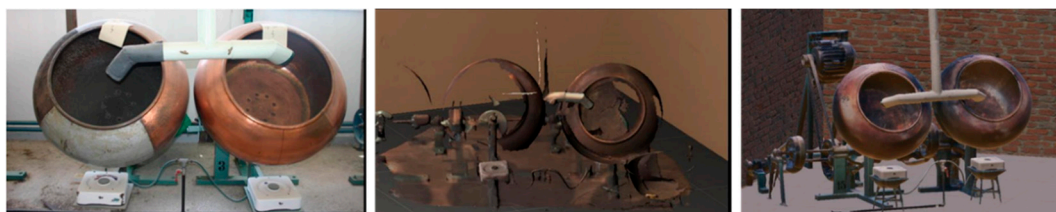


Figure 13. Recreation of the Candy Machine. **(left)** a photo of the machine at the museum; **(middle)** the reconstructed 3D model based on the scans; **(right)** the reconstructed 3D model in its final form (after processed in Blender).

Overall, seven machines were reconstructed, namely, (i) the sifting machine, (ii) the blending machine, (iii) the cutting machine, (iv) the candy machine, (v) the revolving cylinder, (vi) the printing machine, and (vii) the distillation machine. All of them are met

in the Chicle production line, except the distillation machine that is used for producing mastic oil. The final 3D models of these machines are illustrated in Figure 14.



Figure 14. All the machines that were scanned and reconstructed as 3D models for the virtual factory. At the top, from left to right: Sifting machine, blending machine, and cutting machine. At the bottom, from left to right: Candy machine, revolving cylinder, printing machine, and distillation machine.

The 3D model of the factory building was initially created in Unity and then imported in Blender for further processing (e.g., windows and doors were cut and a High Dynamic Range (HDR) environment texture was used, to provide ambient light in the scene). We have used a “stone tile” texture for the walls and plain “cement” texture for the floor, as these materials were predominately used in factory buildings in Greece at that age. Then the 3D model was imported back into the Unity game engine and used as the virtual factory’s building. We have used Unity 2020 with the High Definition Rendering Pipeline (HDRP) enabled to achieve a high-quality, realistic lighting result. Inside the virtual factory building, we have placed the 3D machine models in the order that are met in the Chicle production line. To guide the users to visit the machines in the correct order, virtual carpets have been placed on the floor in a way that creates corridors for the visitors to follow, starting from the factory’s front door as shown in Figure 15.

We have used the same VHs as the ones in the AR application. Each VH is placed next to their respective machine and is ready to explain the functionality of the respective machine, explain how the respective process step was performed at the factory before and after the machine acquisition, and narrate stories about their personal and work lives. When a VH is approached by the camera controlled by the user, he/she starts talking to introduce themselves. Then, the available stories are presented to the users in the form of buttons, that visitors can press to listen to the respective narrations (Figure 16).



Figure 15. The interior space of the virtual factory, with and without the VHs. Carpet corridors are placed on the floor to guide users to follow the mastic chicle production line. VHs are placed next to each machine, ready to share their stories.



Figure 16. VHs narrating Mingei stories and demonstrating the relative processes.

5. Replication and Evaluation

5.1. Guidelines for Lessons Learned by This Research Work

The process of defining the proposed methodology has provided insights on the facilitated technologies summarized in a collection of guidelines to be considered by people replicating this research work:

Guideline #1: Review the plethora of available mo-cap systems and select the one most capable to support your scenario. For example, if multiple users are to be tracked optical mo-cap systems may be preferable. Please consider the possibility of visual occlusions and the effect on your use case. Examine whether the usage of tools or handheld equipment is required to be tracked. Carefully consider the option of a solution for hand tracking if this is the case or a solution for inferring tool state from the recorded animation [91,92].

- Guideline #2:** Consider scale when transferring motion for the tracked system to an VH. Calibrate both the mo-cap system and the VH dimensions accurately to reduce retargeting issues
- Guideline #3:** Invest in a high-quality VH to enhance the visual appearance of textures and support for manipulation of facial morphology
- Guideline #4:** Carefully consider TTS vs speech recording based on the requirements of your scenario. The first is cost-efficient but less realistic, the second is more realistic but needs re-capturing sound for each change in the script.

5.2. Formative Evaluation

Both applications were planned to be installed in the Chios Mastic Museum in early 2021. However, they were delayed due to the COVID-19 pandemic and are currently planned for October 2021. Nevertheless, the applications have been completed and we expect the results to be very promising. As a first step towards evaluating the outcome of this research work, an expert-based evaluation was conducted. In this approach, the inspection is conducted by a user experience usability expert. Such evaluation is based on prior experiences and knowledge of common human factors and ergonomics guidelines, principles, and standards. In this type of inspection, the evaluator looks at the application through the “eyes” of the user, performing common tasks in the application or system while noting any areas in the design which may cause problems to the user. For the case of the two applications, the following issues were reported by the usability expert.

- AR application
 - Make sure that the VH is looking at the user browsing the application
 - Include subtitles on the bottom side of the screen since in noisy environments the information may not be audible.
 - In some cases, there are some deficiencies in the animations such as non-physical hand poses and arms overlapping with parts of the body
- 3D application
 - Improve camera handling by locking the z-axis practically not required by the session
 - Make sure that the VH is looking to the location of the “user” accessing the application (center of the screen)

The above usability issues were all considered critical in terms of user experience and were corrected before installation in the pilot site.

6. Conclusions and Future Work

To sum up, this research work contributed a cost-effective methodology for the creation of realistic VHs for CH applications. The VH models have been created using the Character Creation3 (CC3) software. The animations for the VHs have been recorded using the Rokoko suit and gloves, which utilize inertial sensors. The decision to use the specific suite was made after thoroughly analyzing mo-cap systems on the market in terms of their performance, usability, comfort, cost, ease of use, etc. Each animation recorded using the Rokoko suit and gloves features natural, expressive body moves and gestures corresponding to the story that the VH will narrate, while we have used an audio-recording application to simultaneously record natural voice narrating the desired stories. Then, the VH models along with the animations recorded and the recorded audio clips have then been imported into the Unity game engine. To make the narrator VHs narrate the desired stories, we make use of the Salsa Lip-Synch software, which can produce realistic lip-synch animations and facial expressions, and is compatible with both the Unity game engine and the CC3 software used to create the VHs. An automated module is applied to the VHs to get them ready for lip-synching, and for each narration part, we are loading the corresponding audio clip upon the start of each animation. This way we achieve on-demand realistic VHs,

capable of narrating the desired stories, thus offering an enhanced user experience to the Mingei platform users.

The final solution has been tested in two settings: the first is a true AR application that can present VH as storytellers next to the actual factory machines and the second is a virtual environment of a mastic factory experienced as a Windows application and can be also be experienced in VR.

The validation until now of the proposed workflow and solutions supports our initial hypothesis and results in the production of realistic avatars. The selection of technologies has been proven suitable for the needs of the project and minor adjustments to the methodology can further enhance the output. For example, the careful calibration of the tracking suit and software concerning the anthropometric characteristics of the user recorded and those of the created VH can save labor on retargeting and animation tuning. Furthermore, the combination of motion and voice recordings has been proven to be beneficial both for the pursued realism and development time. It is thus expected that further experimentation is required when replicating the proposed solution to define the required fine tunings for each hardware and software setup.

Regarding future research directions, the above-mentioned knowledge is expected to be enriched through valuable feedback received from the installation and evaluation with end-users of the two variations in the context of the mastic pilot of the Mingei project. More specifically evaluation will target information quality, education value, realism, and perceived quality of experience. More specifically, end-users will access the implemented applications and will be requested to fill in a user experience evaluation questionnaire, participate in targeted interviews, and be monitored while interacting with the system (observation sessions). It is expected that the analysis of the data acquired by the aforementioned methods will provide further input for the improvement of the implemented prototypes.

Author Contributions: Conceptualization, E.K., N.P. (Nikolaos Partarakis), D.K. and X.Z.; Data curation, D.K.; Funding acquisition, N.P. (Nikolaos Partarakis) and X.Z.; Investigation, D.K.; Methodology, E.K., N.P. (Nikolaos Patsiouras), N.P. (Nikolaos Partarakis) and X.Z.; Project administration, N.P. (Nikolaos Partarakis) and X.Z.; Resources, D.K.; Software, E.K., N.P. (Nikolaos Partarakis), E.Z., A.K., A.P., E.B. and N.C.; Supervision, N.P. (Nikolaos Partarakis), E.Z., N.M.-T. and X.Z.; Validation, D.K., C.R. and E.T.; Visualization, E.K., N.P. (Nikolaos Patsiouras) and A.P.; Writing—original draft, E.K. and N.P. (Nikolaos Partarakis); Writing—review & editing, E.K., N.P. (Nikolaos Partarakis), A.P. and D.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work has been conducted in the context of the Mingei project that has received funding from the European Union’s Horizon 2020 research and innovation program under grant agreement No 822336.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data are available upon request.

Acknowledgments: The authors would like to thank the Chios Mastic Museum for its contribution to this work.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Huang, Y.P. Visual perception and fatigue in AR/VR head-mounted displays. *Inf. Disp.* **2019**, *35*, 4–5. [[CrossRef](#)]
2. Hoffman, D.M.; Girshick, A.R.; Akeley, K.; Banks, M.S. Vergence–accommodation conflicts hinder visual performance and cause visual fatigue. *J. Vis.* **2008**, *8*, 33. [[CrossRef](#)] [[PubMed](#)]
3. MacDorman, K.F.; Chattopadhyay, D. Reducing consistency in human realism increases the uncanny valley effect; increasing category uncertainty does not. *Cognition* **2016**, *146*, 190–205. [[CrossRef](#)]
4. Ishiguro, H. The uncanny advantage of using androids in social and cognitive science research. *Interact. Stud.* **2006**, *7*, 297–337.

5. Ada and Grace—Virtual Human Museum Guides from 2009 until Today. Available online: https://ict.usc.edu/wp-content/uploads/overviews/Ada%20and%20Grace_Overview.pdf (accessed on 5 May 2021).
6. Swartout, W.; Traum, D.; Artstein, R.; Noren, D.; Debevec, P.; Bronnenkant, K.; White, K. Ada and Grace: Toward realistic and engaging virtual museum guides. In Proceedings of the International Conference on Intelligent Virtual Agents, Philadelphia, PA, USA, 20–22 September 2010; Springer: Berlin/Heidelberg, Germany, 2010; pp. 286–300.
7. Ding, M. Augmented reality in museums. In *Museums Augmented Reality—A Collection of Essays from the Arts Management and Technology Laboratory*; Carnegie Mellon University: Pittsburgh, PA, USA, 2017; pp. 1–15.
8. Ghouaiel, N.; Garbaya, S.; Cieutat, J.M.; Jessel, J.P. Mobile augmented reality in museums: Towards enhancing visitor's learning experience. *Int. J. virtual Real.* **2017**, *17*, 21–31. [\[CrossRef\]](#)
9. Sylaiou, S.; Kasapakis, V.; Gavalas, D.; Dzardanova, E. Avatars as storytellers: Affective narratives in virtual museums. *Pers. Ubiquitous Comput.* **2020**, *24*, 829–841. [\[CrossRef\]](#)
10. Mingei Project's Website. Available online: <https://www.mingei-project.eu/> (accessed on 5 May 2021).
11. Zabulis, X.; Meghini, C.; Partarakis, N.; Beisswenger, C.; Dubois, A.; Fasoula, M.; Galanakis, G. Representation and preservation of Heritage Crafts. *Sustainability* **2020**, *12*, 1461. [\[CrossRef\]](#)
12. Zabulis, X.; Meghini, C.; Partarakis, N.; Kaplanidi, D.; Doulgeraki, P.; Karuzaki, E.; Beisswenger, C. What is needed to digitise knowledge on Heritage Crafts? *Mem. Rev.* **2019**, *4*, 1.
13. Parmar, D.; Olafsson, S.; Utami, D.; Bickmore, T. Looking the part: The effect of attire and setting on perceptions of a virtual health counselor. In Proceedings of the 18th International Conference on Intelligent Virtual Agents, Sydney, Australia, 5–8 November 2018; pp. 301–306.
14. Machidon, O.M.; Duguleana, M.; Carrozzino, M. Virtual humans in cultural heritage ICT applications: A review. *J. Cult. Herit.* **2018**, *33*, 249–260. [\[CrossRef\]](#)
15. Addison, A.C. Emerging trends in virtual heritage. *IEEE Multimed.* **2000**, *7*, 22–25. [\[CrossRef\]](#)
16. Partarakis, N.; Doulgeraki, P.; Karuzaki, E.; Adami, I.; Ntoa, S.; Metilli, D.; Bartalesi, V.; Meghini, C.; Marketakis, Y.; Kaplanidi, D.; et al. Representation of socio-historical context to support the authoring and presentation of multimodal narratives: The Mingei Online Platform. *J. Comput. Cult. Herit* **2021**, in press.
17. Ana, M.T.; Adolfo, M. Digital Avatars as Humanized Museum Guides in the Convergence of Extended Reality. MW21: MW 2021. 1 February 2021. Consulted 24 October 2021. Available online: <https://mw21.museweb.net/paper/digital-avatars-as-humanized-museum-guides-in-the-convergence-of-extended-reality/>. (accessed on 15 June 2021).
18. Decker, J.; Doherty, A.; Geigel, J.; Jacobs, G. Blending Disciplines for a Blended Reality: Virtual Guides for A Living History Museum. *J. Interact. Technol. Pedagogy* **2020**, *17*. Available online: <https://jitp.commons.gc.cuny.edu/blending-disciplines-for-a-blended-reality-virtual-guides-for-a-living-history-museum/> (accessed on 15 June 2021).
19. Huseinovic, M.; Turcinhodzic, R. Interactive animated storytelling in presenting intangible cultural heritage. In Proceedings of the CESC 2013: The 17th Central European Seminar on Computer Graphics, Smolenice, Slovakia, 28–30 April 2013.
20. Bogdanovych, A.; Rodriguez, J.A.; Simoff, S.; Cohen, A. Virtual agents and 3D virtual worlds for preserving and simulating cultures. In Proceedings of the International Workshop on Intelligent Virtual Agents, Amsterdam, The Netherlands, 14–16 September 2009; Springer: Berlin, Heidelberg, 2009; pp. 257–271.
21. Carrozzino, M.; Lorenzini, C.; Duguleana, M.; Evangelista, C.; Brondi, R.; Tecchia, F.; Bergamasco, M. An immersive vr experience to learn the craft of printmaking. In Proceedings of the International Conference on Augmented Reality, Virtual Reality and Computer Graphics, Otranto, Italy, 15–18 June 2016; Springer: Cham, Switzerland, 2016; pp. 378–389.
22. Carrozzino, M.; Lorenzini, C.; Evangelista, C.; Tecchia, F.; Bergamasco, M. AMICA: Virtual reality as a tool for learning and communicating the craftsmanship of engraving. *Digit. Herit.* **2015**, *2*, 187–188.
23. Danks, M.; Goodchild, M.; Rodriguez-Echavarria, K.; Arnold, D.B.; Griffiths, R. Interactive storytelling and gaming environments for museums: The interactive storytelling exhibition project. In Proceedings of the International Conference on Technologies for E-Learning and Digital Entertainment, Hong Kong, China, 11–13 June 2007; Springer: Berlin/Heidelberg, Germany, 2007; pp. 104–115.
24. Geigel, J.; Shitut, K.S.; Decker, J.; Doherty, A.; Jacobs, G. The digital docent: Xr storytelling for a living history museum. In Proceedings of the 26th ACM Symposium on Virtual Reality Software and Technology, Ottawa, ON, Canada, 1–4 November 2020; pp. 1–3.
25. Dzardanova, E.; Kasapakis, V.; Gavalas, D.; Sylaiou, S. Exploring aspects of obedience in VR-mediated communication. In Proceedings of the 2019 Eleventh International Conference on Quality of Multimedia Experience (QoMEX), Berlin, Germany, 5–7 June 2009; pp. 1–3.
26. Carrozzino, M.; Colombo, M.; Tecchia, F.; Evangelista, C.; Bergamasco, M. Comparing different storytelling approaches for virtual guides in digital immersive museums. In Proceedings of the International Conference on Augmented Reality, Virtual Reality and Computer Graphics, Otranto, Italy, 24–27 June 2018; Springer: Cham, Switzerland, 2018; pp. 292–302.
27. Gratch, J.; Wang, N.; Okhmatovskaia, A.; Lamothe, F.; Morales, M.; van der Werf, R.J.; Morency, L.P. Can virtual humans be more engaging than real ones? In Proceedings of the International Conference on Human-Computer Interaction, Beijing, China, 22–27 July 2007; Springer: Berlin/Heidelberg, Germany, 2007; pp. 286–297.
28. Ibrahim, N.; Mohamad Ali, N.; Mohd Yatim, N.F. Factors facilitating cultural learning in virtual architectural heritage environments: End user perspective. *J. Comput. Cult. Herit.* **2015**, *8*, 1–20. [\[CrossRef\]](#)

29. Rizvic, S. Story guided virtual cultural heritage applications. *J. Interact. Humanit.* **2014**, *2*, 2. [CrossRef]
30. Ready Player, Me. Available online: <https://readyplayer.me/> (accessed on 15 March 2021).
31. Character Creator. Available online: <https://www.reallusion.com/character-creator/> (accessed on 15 March 2021).
32. DAZ Studio. Available online: <https://www.daz3d.com> (accessed on 15 March 2021).
33. MakeHuman. Available online: <http://www.makehumancommunity.org/> (accessed on 12 May 2021).
34. Didimo. Available online: <https://www.didimo.co/> (accessed on 9 February 2021).
35. A Brief History of Motion Tracking Technology. Available online: <https://medium.com/@lumrachele/a-brief-history-of-motion-tracking-technology-and-how-it-is-used-today-44923087ef4c> (accessed on 9 June 2021).
36. Yamane, K.; Hodgins, J. Simultaneous tracking and balancing of humanoid robots for imitating human motion capture data. In Proceedings of the 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems, St. Louis, MO, USA, 10–15 October 2009; pp. 2510–2517.
37. Motion Capture. Available online: https://en.wikipedia.org/wiki/Motion_capture (accessed on 15 March 2021).
38. The Technology Behind Avatar Movie. Available online: <https://www.junauza.com/2010/01/technology-behind-avatar-movie.html> (accessed on 15 March 2021).
39. Calvert, T.W.; Chapman, J.; Patla, A. Aspects of the kinematic simulation of human movement. *IEEE Comput. Graph. Appl.* **1982**, *2*, 41–50. [CrossRef]
40. Ginsberg, C.M.; Maxwell, D. Graphical marionette. *ACM SIGGRAPH Comput. Graph.* **1984**, *18*, 26–27. [CrossRef]
41. Qualisys. Available online: <https://www.qualisys.com/> (accessed on 1 February 2021).
42. Vicon. Available online: <https://www.vicon.com/> (accessed on 19 May 2021).
43. NaturalPoint. Available online: <https://www.naturalpoint.com/> (accessed on 12 June 2021).
44. Motion Analysis. Available online: <https://www.vay.ai/> (accessed on 17 February 2021).
45. Nakano, N.; Sakura, T.; Ueda, K.; Omura, L.; Kimura, A.; Iino, Y.; Yoshioka, S. Evaluation of 3D markerless motion capture accuracy using OpenPose with multiple video cameras. *Front. Sports Act. Living* **2020**, *2*, 50. [CrossRef]
46. Zago, M.; Luzzago, M.; Marangoni, T.; De Cecco, M.; Tarabini, M.; Galli, M. 3D tracking of human motion using visual skeletonization and stereoscopic vision. *Front. Bioeng. Biotechnol.* **2020**, *8*, 181. [CrossRef]
47. Corazza, S.; Mündermann, L.; Gambaretto, E.; Ferrigno, G.; Andriacchi, T.P. Markerless motion capture through visual hull, articulated icp and subject specific model generation. *Int. J. Comput. Vis.* **2010**, *87*, 156–169. [CrossRef]
48. Mehta, D.; Sridhar, S.; Sotnychenko, O.; Rhodin, H.; Shafiei, M.; Seidel, H.P.; Theobalt, C. Vnect: Real-time 3D human pose estimation with a single rgb camera. *ACM Trans. Graph.* **2017**, *36*, 1–14. [CrossRef]
49. Mizumoto, T.; Fornaser, A.; Suwa, H.; Yasumoto, K.; de Cecco, M. Kinect-based micro-behavior sensing system for learning the smart assistance with human subjects inside their homes. In Proceedings of the 2018 Workshop on Metrology for Industry 4.0 and IoT, Brescia, Italy, 16–18 April 2018; pp. 1–6.
50. Shotton, J.; Fitzgibbon, A.; Cook, M.; Sharp, T.; Finocchio, M.; Moore, R.; Blake, A. Real-time human pose recognition in parts from single depth images. In Proceedings of the CVPR, Colorado Springs, CO, USA, 20–25 June 2011; pp. 1297–1304.
51. Cao, Z.; Simon, T.; Wei, S.E.; Sheikh, Y. Realtime multi-person 2d pose estimation using part affinity fields. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 7291–7299.
52. Markerless Marker-Based Motion Capture. Available online: <https://www.qualisys.com/applications/human-biomechanics/markerless-motion-capture/> (accessed on 22 July 2021).
53. Motion Capture: Magnetic Systems. Next Generation. *Imagine Media* (10): 51. October 1995. Available online: <https://archive.org/details/nextgen-issue-010>, (accessed on 1 May 2021).
54. Kramer, R.K.; Majidi, C.; Sahai, R.; Wood, R.J. Soft curvature sensors for joint angle proprioception. In Proceedings of the 2011 IEEE/RSJ International Conference on Intelligent Robots and Systems, San Francisco, CA, USA, 25–30 September 2011; pp. 1919–1926.
55. Shen, Z.; Yi, J.; Li, X.; Lo, M.H.P.; Chen, M.Z.; Hu, Y.; Wang, Z. A soft stretchable bending sensor and data glove applications. *Robot. Biomim.* **2016**, *3*, 1–8. [CrossRef] [PubMed]
56. Huang, B.; Li, M.; Mei, T.; McCoul, D.; Qin, S.; Zhao, Z.; Zhao, J. Wearable stretch sensors for motion measurement of the wrist joint based on dielectric elastomers. *Sensors* **2017**, *17*, 2708. [CrossRef]
57. Al-Nasri, I.; Price, A.D.; Trejos, A.L.; Walton, D.M. A commercially available capacitive stretch-sensitive sensor for measurement of rotational neck movement in healthy people: Proof of concept. In Proceedings of the 2019 IEEE 16th International Conference on Rehabilitation Robotics (ICORR), Toronto, ON, Canada, 24–28 June 2019; pp. 163–168.
58. Gypsy Mocap System. Available online: <https://metamotion.com/gypsy/gypsy-motion-capture-system.htm> (accessed on 20 August 2021).
59. Wearable Stretch Sensors. Available online: <https://encyclopedia.pub/1487> (accessed on 20 March 2021).
60. Meador, W.S.; Rogers, T.J.; O’Neal, K.; Kurt, E.; Cunningham, C. Mixing dance realities: Collaborative development of live-motion capture in a performing arts environment. *Comput. Entertain.* **2004**, *2*, 12. [CrossRef]
61. Tour the World from Home. Available online: <https://promocommunications.com/press-release-sherpa-tours-tourtheworldfromhome/> (accessed on 27 July 2021).

62. Virtual Tutankhamun Is Giving Mixed Reality Tours of The Egyptian Museum until Next Weekend. Available online: <https://cairoscene.com/Buzz/Virtual-Tutankhamun-is-Giving-Mixed-Reality-Tours-of-The-Egyptian-Museum-Until-Next-Weekend> (accessed on 19 August 2021).
63. Trivedi, A.; Pant, N.; Shah, P.; Sonik, S.; Agrawal, S. Speech to text and text to speech recognition systems-Areview. *IOSR J. Comput. Eng.* **2018**, *20*, 36–43.
64. Kuligowska, K.; Kisielewicz, P.; Włodarz, A. Speech synthesis systems: Disadvantages and limitations. *Int. J. Res. Eng. Technol.* **2018**, *7*, 234–239. [\[CrossRef\]](#)
65. Text-to-Speech Technology: What It Is and How It Works. Available online: <https://www.readingrockets.org/article/text-speech-technology-what-it-and-how-it-works> (accessed on 19 July 2021).
66. Edwards, P.; Landreth, C.; Fiume, E.; Singh, K. JALI: An animator-centric viseme model for expressive lip synchronization. *ACM Trans. Graph.* **2016**, *35*, 1–11. [\[CrossRef\]](#)
67. Xu, Y.; Feng, A.W.; Marsella, S.; Shapiro, A. A practical and configurable lip sync method for games. In Proceedings of the Motion on Games, Dublin, Ireland, 6–8 November 2013; pp. 131–140.
68. Hoon, L.N.; Chai, W.Y.; Rahman, K.A.A.A. Development of real-time lip sync animation framework based on viseme human speech. *Arch. Des. Res.* **2014**, *27*, 19–28.
69. De Martino, J.M.; Magalhães, L.P.; Violaro, F. Facial animation based on context-dependent visemes. *Comput. Graph.* **2006**, *30*, 971–980. [\[CrossRef\]](#)
70. CrazyTalk. Available online: <https://www.reallusion.com/crazytalk7/videogallery.aspx> (accessed on 12 June 2021).
71. Papagayo. Available online: <http://lostmarble.com/papagayo/> (accessed on 12 June 2021).
72. Salsa LipSync Suite. Available online: <https://crazyminnowstudio.com/unity-3d/lip-sync-SALSA/>, (accessed on 12 June 2021).
73. Liarokapis, F.; Sylaiou, S.; Basu, A.; Mourkoussis, N.; White, M.; Lister, P.F. An interactive visualisation interface for virtual museums. In Proceedings of the 5th International conference on Virtual Reality, Archaeology and Intelligent Cultural Heritage VAST'04, Oudenaarde, Belgium, 7–10 December 2004; pp. 47–56.
74. Sandor, C.; Fuchs, M.; Cassinelli, A.; Li, H.; Newcombe, R.; Yamamoto, G.; Feiner, S. Breaking the barriers to true augmented reality. *arXiv* **2015**, arXiv:1512.05471.
75. Jung, T.; tom Dieck, M.C.; Lee, H.; Chung, N. Effects of virtual reality and augmented reality on visitor experiences in museum. In *Information and Communication Technologies in Tourism*; Springer: Cham, Switzerland, 2016; pp. 621–635.
76. Papaefthymiou, M.; Kanakis, M.E.; Geronikolakis, E.; Nochos, A.; Zikas, P.; Papagiannakis, G. Rapid reconstruction and simulation of real characters in mixed reality environments. In *Digital Cultural Heritage*; Springer: Cham, Switzerland, 2018; pp. 267–276.
77. Zidianakis, E.; Partarakis, N.; Ntoa, S.; Dimopoulos, A.; Kopidaki, S.; Ntagianta, A.; Stephanidis, C. The invisible museum: A User-centric platform for creating virtual 3D exhibitions with VR support. *Electronics* **2021**, *10*, 363. [\[CrossRef\]](#)
78. Efstratios, G.; Michael, T.; Stephanie, B.; Athanasios, L.; Paul, Z.; George, P. New cross/augmented reality experiences for the virtual museums of the future. In Proceedings of the Euro-Mediterranean Conference, Nicosia, Cyprus, 29 October–3 November 2018; Springer: Cham, Switzerland, 2018; pp. 518–527.
79. Papagiannakis, G.; Partarakis, N.; Stephanidis, C.; Vassiliadi, M.; Huebner, N.; Grammalidis, N.; Margetis, G. *Mixed Reality Gamified Presence and Storytelling for Virtual Museums*; No. IKEEBOOKCH-2019–233; Springer: Cham, Switzerland, 2018.
80. Ioannides, M.; Magnenat-Thalmann, N.; Papagiannakis, G. (Eds.) *Mixed Reality and Gamification for Cultural Heritage*; Springer: Berlin, Germany, 2017; Volume 2.
81. Zikas, P.; Bachlitzanakis, V.; Papaefthymiou, M.; Kateros, S.; Georgiou, S.; Lydatakis, N.; Papagiannakis, G. Mixed reality serious games and gamification for smart education. In Proceedings of the European Conference on Games Based Learning, Paisley, UK, 6–7 October 2016; Academic Conferences International Ltd.: Reading, UK, 2016; p. 805.
82. Abate, A.F.; Barra, S.; Galeotafiore, G.; Díaz, C.; Aura, E.; Sánchez, M.; Vendrell, E. An augmented reality mobile app for museums: Virtual restoration of a plate of glass. In Proceedings of the Euro-Mediterranean Conference, Nicosia, Cyprus, 29 October–3 November 2018; Springer: Cham, Switzerland, 2018; pp. 539–547.
83. iClone. Available online: <https://www.reallusion.com/iclone/> (accessed on 12 June 2021).
84. 3Dxchange. Available online: <https://www.reallusion.com/iclone/pipeline.html> (accessed on 17 July 2021).
85. Partarakis, N.; Zabulis, X.; Chatziantoniou, A.; Patsiouras, N.; Adami, I. An approach to the creation and presentation of reference gesture datasets, for the preservation of traditional crafts. *Appl. Sci.* **2020**, *10*, 7325. [\[CrossRef\]](#)
86. Costa, S.; Brunete, A.; Bae, B.C.; Mavridis, N. Emotional storytelling using virtual and robotic agents. *Int. J. Hum. Robot.* **2018**, *15*, 1850006. [\[CrossRef\]](#)
87. Crazy Minnow Studio's Salsa Lip-Sync Suite. Available online: <https://crazyminnowstudio.com/docs/salsa-lip-sync/modules/overview/> (accessed on 19 May 2021).
88. FBX. Available online: <https://en.wikipedia.org/wiki/FBX> (accessed on 17 July 2021).
89. Plytas. *Detailed Description of Various Production Lines of Mastic at the Premises of the Chios Gum Mastic Growers Association*; Archive of Piraeus Bank Group Cultural Foundation: Athens, Greece, 2010.
90. Blender. Available online: <https://www.blender.org/> (accessed on 19 May 2021).

-
91. Stefanidi, E.; Partarakis, N.; Zabulis, X.; Papagiannakis, G. An approach for the visualization of crafts and machine usage in virtual environments. In Proceedings of the 13th International Conference on Advances in Computer-Human Interactions, Valencia, Spain, 21–25 November 2020; pp. 21–25.
 92. Stefanidi, E.; Partarakis, N.; Zabulis, X.; Zikas, P.; Papagiannakis, G.; Thalmann, N.M. TooltY: An approach for the combination of motion capture and 3D reconstruction to present tool usage in 3D environments. In *Intelligent Scene Modeling and Human-Computer Interaction*; Springer: Cham, Switzerland, 2021; pp. 165–180.