Artwork Accessibility for People with Low Vision through Augmented Reality Mediated by Mobile Devices

DRAGAN AHMETOVIC, Università degli Studi di Milano, Dipartimento di Informatica, Italy
CRISTIAN BERNAREGGI, Università degli Studi di Torino, Dipartimento di Matematica, Italy
SERGIO MASCETTI, Università degli Studi di Milano, Dipartimento di Informatica, Italy

In this paper we explore the use of Augmented Reality as a means to provide more widespread and equitable access to art venues and artworks, in particular for people with low vision. We investigate how people with low vision frame, access and interact with artworks in Augmented Reality, using a mobile app specifically designed for accessible and inclusive museum visits. Through a user study with 10 participants in real museum settings, we explore the specific challenges related to the accessibility of different artwork types: a medium size painting, a large tapestry, a statue, and a historical keyboard instrument. Results show that participants were able to access all artworks through an Augmented Reality mobile app running on the user's own device. The system is also perceived as useful and usable. Additionally, we uncover human and environment factors that influence the way users access different artwork types, in particular considering Augmented Reality interactions.

CCS Concepts: • **Human-centered computing** → *Accessibility systems and tools*; • **Applied computing** → *Arts and humanities.*

Additional Key Words and Phrases: Art accessibility; Audio-tactile interfaces; Visual impairment.

ACM Reference Format:
Dragan Ahmetovic, Cristian Bernareggi, and Sergio Mascetti. 2025. Artwork Accessibility for People with Low Vision through Augmented Reality Mediated by Mobile Devices. *Proc. ACM Hum.-Comput. Interact.* 9, 5, Article pn8991 (September 2025), 20 pages. https://doi.org/10.1145/3743713

1 Introduction

People who are blind or have Low Vision[1] (LV) are interested in enjoying products of culture (*e.g.*, literature, artworks) and accessing places of cultural interests (*e.g.*, historical sites, art venues, galleries and museums) [9, 18, 33, 34]. The need for equitable access to products of culture and places of cultural interest by people with disabilities is also recognized as a right in many countries, including Article 30 of the United Nations Convention on the Rights of People with Disabilities [84].

In this paper, we address the problem of equitable access to artworks by people with LV. Specifically, we explore Augmented Reality (AR), as a possible solution to achieve inclusive access to artworks for people with LV. In this context, one prior work proposes *Musa* [3], an AR mobile application designed to recognize artworks framed by the device camera and provide interactive audio-visual descriptions of those artworks, aimed at making the museum experience more accessible and engaging also to people with LV. Results of a preliminary study conducted with seven

[1] People with LV have residual vision but their vision loss cannot be improved by standard treatments [70].

Authors' Contact Information: Dragan Ahmetovic, dragan.ahmetovic@unimi.it, Università degli Studi di Milano, Dipartimento di Informatica, Milan, Italy; Cristian Bernareggi, cristian.bernareggi@gmail.com, Università degli Studi di Torino, Dipartimento di Matematica, Milan, Italy; Sergio Mascetti, sergio.mascetti@unimi.it, Università degli Studi di Milano, Dipartimento di Informatica, Milan, Italy.

ACM 2573-0142/2025/9-ARTpn8991
https://doi.org/10.1145/3743713

representative participants show that AR applications like *Musa* are more effective than traditional audio guides in supporting people with LV for accessing and appreciating artworks. However, this evaluation was conducted in a simulated scenario and not during an actual museum visit. Hence, it did not provide evidence of the applicability of the approach in a real-world museum. Furthermore, *Musa* was designed to recognize average-size bi-dimensional artworks (*e.g.*, paintings) only. AR exploration of other artworks frequently exhibited in museums, such as large-size bi-dimensional ones (*e.g.*, tapestries), as well as 3D objects (*e.g.*, statues) [33, 35, 45] were not addressed in this previous work.

To address these limitations, we conduct a new user study with ten LV participants, with the goal of assessing the applicability of AR access to artworks of different types in a real-world museum scenario. We analyze the interaction between visitors with LV, the system, and different artwork types, with the aim of identifying factors which affect the use of an AR system designed for artwork accessibility by people with LV in art venues. In particular, we tackle the following new research questions:

Rq1 How effective, useful, and easy to use is an AR-based approach in supporting people with LV with artwork exploration in a real museum environment?

Rq2 Which specific challenges emerge when framing, visualizing and interacting with various artwork formats, including large 2D artworks or 3D objects such as statues?

This paper contributes to the state of the art by showing advantages and limitations of the use of AR to support museum exploration by visitors with LV. In particular, results show that a mobile AR app like *Musa* allows independent access to artworks and is therefore perceived as useful by people with LV. The app usability was evaluated positively by the participants as well, and the AR functionalities for framing, visualizing, and interacting with artworks, were found to be useful and easy to use. Still, some challenges emerged for specific types of artworks and situations. In particular, 3D artworks lacking salient visual features were harder to frame from different sides and to interact with using the touchscreen. Another issue, which impacted 3D artworks reflecting light, was that the clarity of the overlay AR elements in presence of light glare. Also, large 2D artworks with many elements took more time to explore than the other artworks, but this did not negatively impact the exploration experience. Unexpectedly, synthetic speech was preferred to audio recordings because it allowed to customize the speech rate according to the user needs, thus resulting in a lower perceived fatigue.

2 Related Work

Section 2.1 reports prior work on artwork accessibility for blind and LV people. Section 2.2 explores AR accessibility and its use for assistive technologies, in particular for LV people. Section 2.3 describes prior work on *Musa*.

2.1 Artwork Accessibility

Blind and LV people are interested in visiting art venues and appreciating artworks [9, 18, 34], despite the many accessibility barriers [85, 87]. Four key challenges that impede equitable access to art venues and artworks by blind and LV people were previously identified [33]: retrieving information on art venues, accessing art venues, obtaining support from art venue staff, and accessing visual artworks. In this paper we focus on the last challenge. Thus, in the following, we only address the relevant literature for this specific problem.

Tactile representations are one of the main solutions for visual artwork accessibility. While extensively studied in prior research [6, 8, 37, 45, 71], they are usually available for few selected artworks only [33], as they are hard to design [19], create [66, 76] and maintain [61]. They are also mostly aimed at blind people, while being less accessible to people with LV who are not acquainted with tactile representations [35] and braille labels [88].

Audio guides [23, 30, 75, 77] are available in most museums. However, they are designed for sighted users and mostly provide a historical and artistic perspective rather than a morphological one [61, 75, 77]. As a result, they can be difficult to access and are not effective for blind and LV people [8]. Descriptions specifically designed for blind and LV people exist [1, 11], but they are rarely available and included in audio guides. Instead, they are commonly adopted during specialized tours that focus on the needs of blind and LV visitors [13, 54]. However, these tours are rarely inclusive to other art venue visitors [9] and only involve a few selected artworks [9, 87].

Remote approaches such as online artwork descriptions [53] are also appreciated, in particular by users with residual vision or acquired blindness [57]. Home-printed audio-tactile representations [2, 38, 68], and audio-touchscreen exploration [4, 50] have also been proposed. Virtual tours are becoming popular as well [25, 37], in particular following the COVID pandemic [20]. However, these remote approaches lack the social aspect of art venue visits and direct interaction with artworks is absent [16, 20, 44]. In this context, *Musa* [3] is a solution for visual artwork accessibility that overcomes the limitations of prior approaches. Indeed, it uses morphological artwork descriptions [1], presented through AR interaction [48] to provide immersive and accessible experiences.

2.2 Augmented Reality for People with LV

AR is broadly defined as the augmentation of natural feedback with simulated cues [63]. Most often, AR is visual [69], but other senses may be used as well [59]. For blind and LV people, two research aspects related to AR can be identified: the accessibility of AR (Section 2.2.1) and the use of AR to create assistive technologies (Section 2.2.2). *Musa* relates to both these aspects as it is a novel assistive technology in AR and also proposes accessible AR visualizations and interactions for users with LV.

2.2.1 AR Accessibility for People with LV. AR content is often purely visual and has poor accessibility for blind and LV users [67]. Despite this limitation, prior literature reports the desire of people with LV to access AR using their residual vision [52]. Similar to Virtual Reality (VR) [65], presentation and interaction with virtual content are two key challenges for AR accessibility [24]. Additionally, in AR there is also the need to make the scanning of real world elements accessible [36]. For people with LV, being able to perceive a virtual element or text can depend on its position, size, color, contrast, and font [93]. These aspects can be improved using visual correction. For example, by adapting luminosity and contrast, using magnification, highlighting edges, or repositioning parts of the scene [92]. For people with severe LV or blindness, it is also possible to access the scene elements using a screen reader-based approach [36]: the virtual elements in the scene can be announced when framed with the camera, or they can be selected from a list to receive instructions on how to turn around to frame them.

AR interaction in *Musa* is based on these findings. Specifically, artwork audio descriptions are associated to visual highlighting of the described elements [36]. Similarly, artwork elements on the touchscreen are highlighted when the user interacts with them. In this work, since *Musa* is extended to support 3D artworks, specific challenges related to AR interactions with 3D objects are examined (**Rq2**). Finally, one improvement to *Musa* used in our experiments is a customizable filter that allows adapting luminosity, contrast and saturation of the artworks [92].

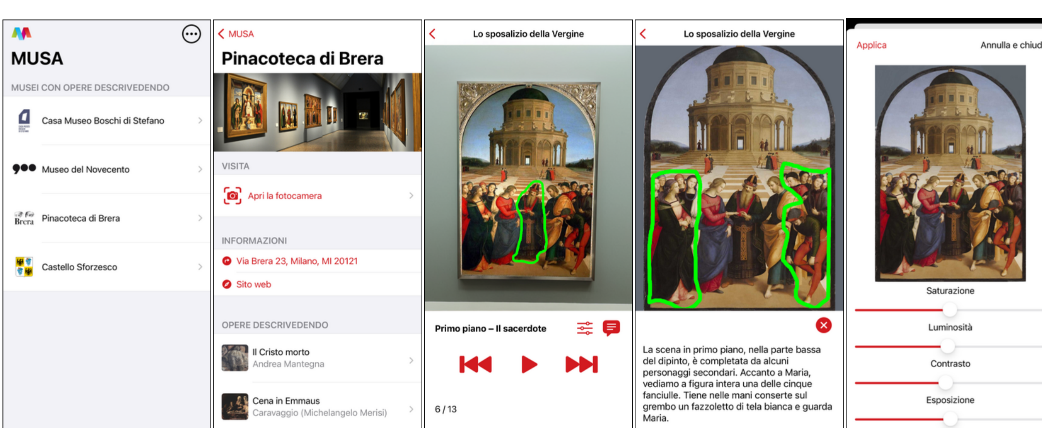
2.2.2 Assistive Technologies for People with LV Using AR. Various AR applications have been proposed to support people with LV. Some are tools that aim to mitigate vision problems using headsets or smart glasses [42, 92]. Among these, some improve visual acuity through magnification [31, 80], contrast [39] or edge enhancement [40], while others correct field of view defects by remapping the scene onto the usable portion of the view area [78, 82].

Other works focus on mobility assistance, using visual [64, 93] or audio [46, 89] guidance. In particular, visual AR guidance was shown to be more accurate than audio guidance for people with LV, with also a lower perceived cognitive load [94]. AR obstacle avoidance systems have also been investigated, using sonification [73], or visual feedback [26]. Some approaches specifically cater to different visual impairments, such as tunnel vision [86], retinitis pigmentosa [7] or peripheral vision loss [90], reducing collisions up to 50%.

Finally, some solutions process the scene to extract high-level knowledge and provide it in AR. For example, some solutions detect and read text present in the scene [39], others segment and highlight nearby objects [49, 91], and some recognize and convey facial expressions of nearby people [51]. Interaction assistance can also be provided to identify objects [22] or guide the user to reach them [5, 83]. In this context, *Musa* advances the state of the art by providing AR support to blind and LV people for accessing visual artworks present in the scene.

2.3 *Musa*

Musa [3] is an iOS app, designed to support users with LV in accessing and appreciating artworks. When the app starts, the user can select a museum from those supported by the system (Figure 1a) to display its information and a list of available artworks. The user can scan the environment with the device camera to detect the supported artworks and access their descriptions. Morphological artwork descriptions, created with the *DescriVedendo* methodology [1], are read using speech synthesis or a human-recorded version is played, if available. The descriptions are partitioned into *chapters*, each describing one artwork area. While the description is read, the corresponding area is contoured on the screen through AR overlays (Figure 1c). The user can also tap an area of the displayed artwork to play the associated chapter. A similar interaction was previously proposed for web-based access to artwork images [4]. Alternatively, the user can select an artwork from the list. In this case, instead of localizing the artwork with the camera, its image is displayed on the screen (Figure 1d), and can be accessed in the same way as the AR exploration described above. *Musa* is compatible with various accessibility tools, including screen reader, magnifier, and enlarged fonts.



(a) Museum selection (b) Artwork selection (c) Audio player - AR (d) Text mode - virtual (e) Filter settings

Fig. 1. Screens of the *Musa* app

Musa was evaluated through a preliminary study with 7 LV participants, conducted in a simulated scenario during which the participants accessed one image of a medium-sized painting shown on an external screen. The evaluation compared *Musa* to an audio guide baseline, showing that *Musa* was considered more useful and effective. This paper extends the work presented in [3] along two main directions. First, we investigate the validity of the approach in a real-world scenario, thus clarifying how real-world usage would influence the usage and appreciation of the system (**Rq1**). Second, this paper evaluates Musa with different types of artworks, including large 2D artworks and 3D objects, which are common in art venues but rarely accessible to people with LV [33, 35, 45], thus identifying specific challenges that may emerge for these artwork types (**Rq2**).

3 Evaluation

Through a within-subject study with ten participants with LV, we evaluated the effectiveness and appreciation of the *Musa* system (**Rq1**) for accessing different kinds of 2D artworks, including large-sized ones, as well as 3D artworks (**Rq2**) in a real-world museum setting. The study was approved by our institution's ethics committee. It lasted about one hour and a half for each participant and was supervised by one researcher.

3.1 Evaluation Setting and Stimuli

The study was conducted, during opening times, in two museums located in Milan, Italy: *Castello Sforzesco*[2] and *Pinacoteca di Brera*[3]. These museums were chosen because they contain various 2D and 3D artworks for which *DescriVedendo* descriptions are available. We selected four different artwork types, with various sizes and shapes, that would require the users to position themselves in various ways to frame and interact with them using *Musa*. The first artwork was from Pinacoteca di Brera while the other three were from Castello Sforzesco:

- "Lo Sposalizio della Vergine" - "**Marriage**[4] of the Virgin" (Figure 2a [74]) is a painting by Raphael. It is placed in a corridor 5m wide and its dimensions (width 1.86m, height 2.34m) are comparable to "Cena in Emmaus" - "Supper at Emmaus" (width 1.41m, height 1.962m), which was used in the previous study [3].
- "Il Virginale Ruckers" - "The Ruckers **Virginal**" (Figure 2b [41]) is a laquered wooden keyboard instrument with a painting decoration on its lid. This 3D artwork (width 4.92m, depth 1.708m, height 2.52m) can be observed standing in front of it in a long corridor about 4m large;
- "L'Arazzo del Mese di Maggio" - "May **Tapestry**" (Figure 2c [14]) is a large tapestry (width 4.75m, height 4.96m), one of the twelve "Trivulzio Tapestries" corresponding to the months of the year. It is placed on the middle of the long side of a room 15m large and 20m long, and can be observed from various positions;
- "La Pietà Rondanini" - "Rondanini **Pietà**" (Figure 2d [62]) is an unfinished white marble sculpture made by Michelangelo. The statue, 1.95m high, is placed on a plinth (width 2m, depth 3m, height 1m), on a 3m wide platform in the center of a large hall. Thus, it can be observed from all the sides;

3.2 Apparatus

Participants accessed the artworks using *Musa* running on an *iPhone 12 Pro* device provided by the supervisor. Audio feedback was conveyed through smartphone loudspeakers at a low volume to avoid disturbing other visitors but still allowing the researcher to relate participants' behavior to

[2]https://www.milanocastello.it/en
[3]https://pinacotecabrera.org/
[4]We refer to the artworks using the term in bold

(a) Marriage of the Virgin (b) The Ruckers Virginal (c) May Tapestry (d) Rondanini Pietà

Fig. 2. Artworks used in the study

their interaction with the system. Besides **Pietà**, which had voiced audio descriptions, the other artworks had speech-synthesized audio descriptions.

Since we were interested in understanding how AR-based access to artworks applies to artwork types different than medium-sized paintings, our experimental apparatus extends the original *Musa* system [3] to support the recognition of 3D artworks as well. Technical extension of the *Musa* system capabilities is outside of the scope of this work. However, for clarity, we add details about how the system was extended in Appendix A.

We also introduce two additional improvements, motivated by the prior results [3]. First, the new app provides customizable visual filters for the displayed artworks. Specifically, the user can tune luminosity, contrast, saturation, and exposure (Figure 1e). This improvement was directly requested by the participants of the prior study [3], and it is similar to one of the AR accessibility tools proposed in [92]. The functionality replaces the negative color filter that was present in the older version, which was not considered to be useful by the participants. The second improvement is the possibility to display artwork descriptions as text, instead of having them read aloud (Figure 1d). This functionality is aimed at users who can use their residual vision to read text displayed on the screen and prefer accessing information this way, rather than listening to audio descriptions [52].

3.3 Metrics and Data Analysis

Participants' interactions with *Musa* were automatically logged. In particular, the following metrics were collected:

(1) The duration of the following activities:
- **Framing Time** required for the system to recognize the artworks
- **Exploration Time** dedicated to each artwork

(2) The number of times that the following functionalities were used:
- **Player Interaction** (play/pause description, previous/next chapter)
- **Touch Exploration** of the artwork displayed on the screen
- **Visual Filters** to change the aspect of the artwork on the screen
- **Text Descriptions** coupled with audio descriptions of the artworks

For each artwork we also asked an 8-item Likert-like scale [58] questionnaire. Since no standardized questionnaires exist on this topic, we designed a new one focusing on ease of use, usefulness and clarity of app functionalities (Figure 4). Answers ranged between 1 (very little) and 5 (very much).

We analyzed whether these metrics varied significantly among different artworks (**Rq2**) using Friedman's test [28], with post-hoc pairwise comparisons using Conover procedure [21] adjusted by

Benjamini-Hochberg False Discovery Rate correction for multiple comparisons [15]. A significance threshold of $\alpha = .05$ was used. Effect sizes were reported using Kendall's coefficient of concordance (Kendall's W) [81].

A final questionnaire, including System Usability Scale (SUS) [17] and additional five-level Likert-like scale items, examined the general disposition of the participants with respect to the system. The answers were descriptively analyzed, and comparisons with benchmark values [56] in the case of SUS were performed. We also report relevant comments that were spontaneously provided by the participants during the study.

3.4 Evaluation Protocol

Participants are first briefed about the scope of the research and study tasks, and can ask questions about the study to the supervisor. After filing the consent form, participant's demographic data is collected (Section 3.5). A brief training on a sample painting ("Il bacio" [3, 27]) is conducted to acquaint the participant with *Musa*. The supervisor verifies that the participant has learned to use the system before proceeding to the next step.

After the training, the participant is asked to access the four artworks using the system. The sequence in which the artworks are accessed could not be completely counter-balanced as **Marriage** is displayed in a different museum than the others. Therefore, this artwork was always accessed last, after a ten-minutes walk between museums. For each artwork, the task proceeds as follows: the supervisor accompanies the participant in the vicinity of the artwork; the participant scans the surroundings using the camera until the artwork is correctly framed and recognized by *Musa*; the participant uses the system to explore the artwork and listen to its description.

Once the participant finishes exploring an artwork, the supervisor administers the ease of use, clarity and usefulness questionnaire (Section 3.3). After visiting all the artworks, the final questionnaire is administered. Finally, the supervisor collects the participant's comments and suggestions for improvement.

3.5 Participants

10 participants (3 female, 7 male) with LV were recruited through convenience sampling, involving associations of people with LV (Table 1). Participants had between 21 and 54 years of age. All had some residual vision, with visual acuity between 1/10 and 3/10, and visual field between 10% and 50%. $P1$, $P2$, $P5$ had visual impairment since birth or first years of life, while others acquired it between 15 and 30 years of age.

Table 1. Participants' data (SR: Screen reader, LF: Large font, IC: Inverted color, GS: Grey scale, Z: Zoom)

P#	Age	Gen.	Visual disability			Museum visit		Assistive technology used
			acuity	field	onset	freq.	support	
$P1$	27	F	2/10	30%	age 5	3/year	friends, family	SR LF — GS Z
$P2$	26	F	1/10	50%	birth	1/year	-	SR LF IC GS Z
$P3$	41	M	2/10	30%	age 20	3/year	friends	— LF —— Z
$P4$	46	M	3/10	50%	age 26	1/year	-	— LF —— Z
$P5$	38	M	2/10	30%	birth	1/year	friends	SR LF — GS Z
$P6$	44	M	2/10	20%	age 15	1/year	friends	SR LF —— Z
$P7$	21	M	1/10	10%	age 20	1/year	friends, family	SR LF —— Z
$P8$	30	M	1/10	20%	age 20	1/year	friends	SR LF IC — Z
$P9$	41	F	1/10	30%	age 20	1/year	friends, venue staff	SR ——— Z
$P10$	54	M	1/10	20%	age 30	1/year	friends, family	SR LF —— Z

All participants reported to be experienced in operating a mobile device and used smartphones for more than 5 years (6 iOS and 4 Android). Mobile devices were mostly used with assistive technologies leveraging participants' residual sight: all used magnification, 9 used large font, 2 inverted color, and 2 grey scale. 8 participants used screen readers (*i.e.*, TalkBack or VoiceOver). All visited museums or galleries between 1 and 3 times per year. During visits, 8 were usually aided by friends, 3 by family members, and 1 by venue staff. The support involved both access to venues and artwork appreciation. All besides *P*8 participated to guided tours (often during school trips), but only 5 experienced guided tours for people with LV (*P*1, *P*2, *P*5, *P*7, *P*9). Only *P*7 experienced tactile replicas, while all but *P*8 and *P*10 have experienced audio guides. None had ever used assistive tools specifically designed for people with LV to explore artworks.

4 Experimental Results

One major result of our study is that all participants were able to explore the four artworks independently, without requiring assistance from the supervisor (**Rq1**). Here we report a detailed analysis of app usage data (Section 4.1), questionnaire scores (Section 4.2), and comments provided by the participants (Section 4.3).

4.1 App Usage Measurements

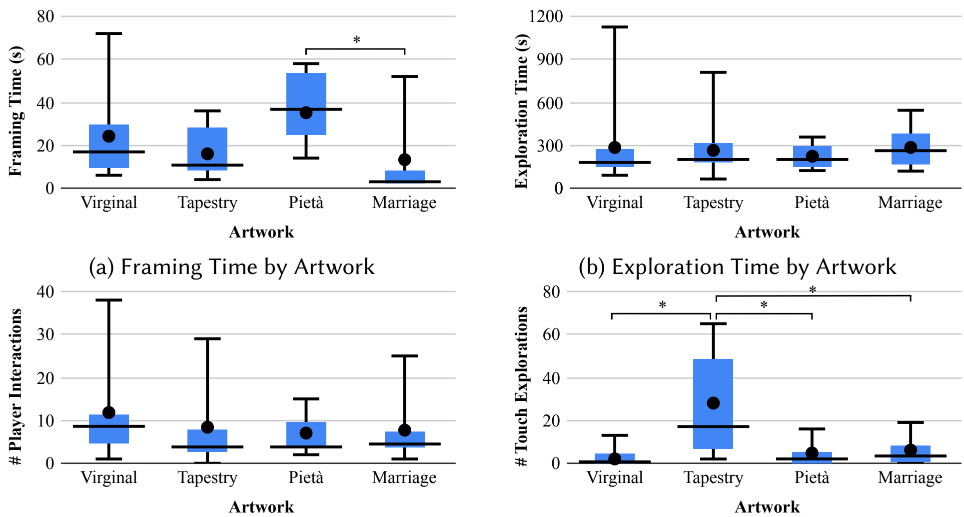
The app usage data analysis focused on three aspects. First, on the time required by the participants to frame different artworks and have them recognized by the system (Section 4.1.1). Second, on the time dedicated to exploring different artworks (Section 4.1.2). Third, on the usage of different app functionalities among the artworks. Specifically, we focused on the interactions with the audio guide player (Section 4.1.3), touch screen exploration of the artwork images (Section 4.1.4), visual filters (Section 4.1.5) and text descriptions (Section 4.1.6).

4.1.1 Framing Time. On average, the participants needed 22.73 s (SD = 19.16 s) to locate, frame and have the artwork recognized by the system (Figure 3a). This value varied significantly (p = .019, χ^2 = 9.98, W = .21) among different artworks, with the highest value recorded for **Pietà** (M = 37.3 s, SD = 17.14 s). This was expected, as the system uses salient visual features to identify artworks and **Pietà** lacks such features. Pairwise comparisons detected a significant difference (p = .007) between **Pietà** and **Marriage** (M = 11.7 s, SD = 17.33 s), but not for **Tapestry** (M = 17.1 s, SD = 12.2 s) and **Virginal** (M = 24.8 s, SD = 20.9 s).

4.1.2 Exploration Time. Exploration time did not significant differ across the artworks (Figure 3b). It took the participants about four and a half minutes (M = 272.53 s, SD = 194.8 s) to explore an artwork, with **Marriage** taking most time (M = 294.7 s, SD = 141.42 s), followed by **Virginal** (M = 285.4 s, SD = 303.23 s) and **Tapestry** (M = 284 s, SD = 203.95s). **Pietà** took the least time, with 226 s (SD = 87.24 s) on average.

4.1.3 Player Usage. Usage of the player functionalities did not differ significantly across the artworks (Figure 3c). On average, participants used these functionalities 8.575 (SD = 8.76) times per artwork. For **Virginal**, they were used on average 11.6 (SD = 12.04) times; for **Tapestry** 8.5 (SD = 10.3) times; for **Marriage** 7.6 (SD = 6.87) times; and for **Pietà** 6.6 (SD = 4.33) times. The most common functionality (55.7%) was to go to the next chapter of the description ("Forward"). "Play/Pause" was used 26.5% of times and "Back" was used 17.8% of times.

4.1.4 Touch Exploration. AR touch screen exploration was performed on average 10.27 (SD = 16.24) times per artwork. Differences among the artworks were statistically significant (p = .006, χ^2 = 12.53, W = .35). Pairwise comparisons reveal that touch exploration was used more often for **Tapestry** (M = 27.7, SD = 24.5) than for **Marriage** (M = 5.7, SD = 5.81, p = .001), **Pietà** (M = 4.4,



(a) Framing Time by Artwork

(b) Exploration Time by Artwork

(c) Player Functionalities usage by Artwork

(d) Touch Exploration usage by artwork

Fig. 3. App usage results (• Mean, — Median, ∗ Significant difference)

SD = 6.06, p = .007), or **Virginal** (M = 3.3, SD = 4.37, p = .002), as shown in Figure 3d. A reason could be that **Tapestry**, being a large artwork with many small features, required moving around in search for details more than the other (smaller) artworks.

4.1.5 Visual Filters Usage. All participants tried visual filters at least a few times, but usage was generally short. Two (*P*1, *P*8) used filters for less than a minute, on two artworks each. Four (*P*3, *P*4, *P*6, *P*7) used them on three artworks each, for 5 to 9 minutes. Four (*P*2, *P*5, *P*9, *P*10) tried them on all the artworks for about 6 to 12 minutes. For **Pietà**, multiple participants with a higher residual vision tried 2 (*P*3) or 3 (*P*4, *P*5) different filter configurations, possibly because this artwork's features are harder to sense by sight compared to the others.

4.1.6 Text Descriptions Usage. Text descriptions were tried by all the participants but not consistently used. Indeed, two participants (*P*5, *P*9) tried them on one artwork only, three (*P*1, *P*2, *P*7) activated it for two different artworks, and five (*P*3, *P*4, *P*6, *P*8, *P*10) on three artworks each. However, most participants accessed only a single chapter for the artworks on which they used this functionality. Only *P*2, *P*4, and *P*7 accessed multiple chapters for **Sposalizio**, **Pietà** and **Virginale**, respectively.

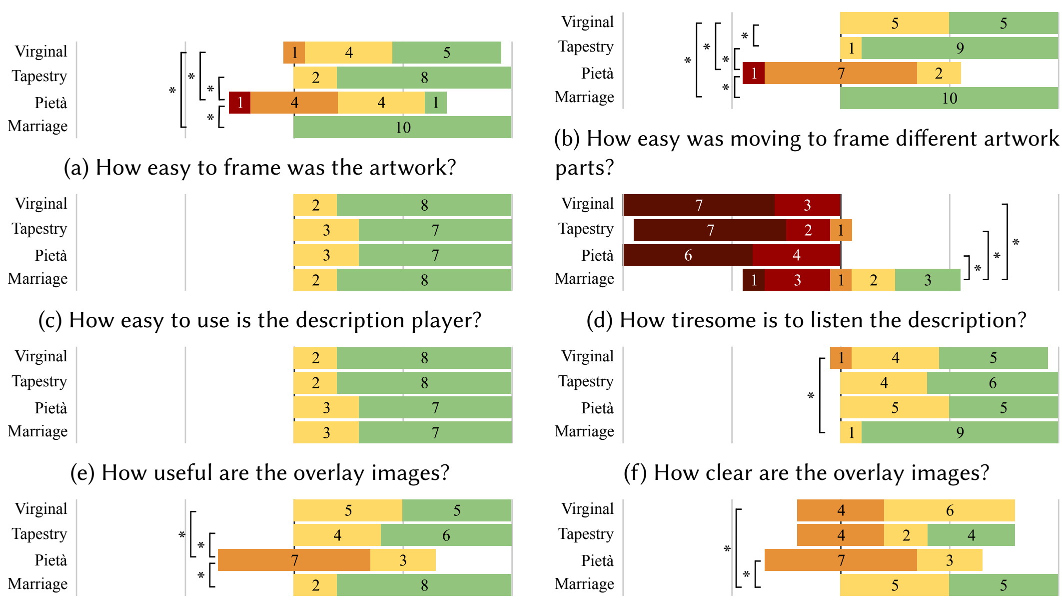
4.2 Subjective Feedback

We investigated artwork framing (Section 4.2.1), audio descriptions (Section 4.2.2), overlay images (Section 4.2.3), and artworks touch exploration (Section 4.2.4). Finally, in Section 4.2.5 we assess system usefulness and usability.

4.2.1 Artwork Framing. Considering the ease of initially framing an artwork to have it described (Figure 4a), significant differences ($p < .001$, $\chi^2 = 20.24$, $W = .34$) were found. Pairwise comparisons revealed that **Pietà** ($M = 3.5$, $SD = 0.85$)[5] was perceived to be harder to frame than all the other artworks ($p < .001$ for all). These results were expected due to the visual characteristics of the

[5]We report means as they better indicate multipoint-item central tendency than medians [55]

artwork. **Virginal** was also harder to frame than **Marriage** ($p = 0.07$), possibly due to light reflections on its surface. Significant differences ($p < .001$, $\chi^2 = 26.52$, $W = .35$) were also found regarding the ease of finding suitable positions from which to frame different parts of the artworks while they are being described (Figure 4b). All participants found this activity very or moderately easy for **Marriage** ($M = 5$, $SD = 0$), **Tapestry** ($M = 4.9$, $SD = 0.32$), and **Virginal** ($M = 4.5$, $SD = 0.53$) , but significantly more difficult for **Pietà** ($M = 3.1$, $SD = 0.57$, $p < .001$ for all). A reason may be that the statue can be viewed from all sides unlike the other artworks. Despite the positive scores, **Virginal** was also more difficult than **Mariage** ($p < .001$) and **Tapestry** ($p = .003$), possibly due to reflections on its surface, as noted above.



(a) How easy to frame was the artwork?
(b) How easy was moving to frame different artwork parts?
(c) How easy to use is the description player?
(d) How tiresome is to listen the description?
(e) How useful are the overlay images?
(f) How clear are the overlay images?
(g) How useful is touchscreen artwork access?
(h) How clear is touchscreen artwork access?

Fig. 4. Per-artwork questionnaire results (∗ Significant difference, Scale: 1 2 3 4 5)

4.2.2 *Audio Descriptions.*

The audio description player was found very or moderately easy to use by all participants (Figure 4c), without significant differences across the artworks. Indeed, the average scores were very similar: 4.8 ($SD = 0.42$) for **Virginal** and **Marriage**, 4.7 ($SD = 0.48$) for **Tapestry** and **Pietà**. Instead, significant differences ($p = .002$, $\chi^2 = 14.36$, $W = .25$) were found considering the perceived fatigue while listening to artwork descriptions (Figure 4d). Pairwise comparisons found **Marriage** ($M = 3.3$, $SD = 1.49$) to be more tiresome to listen to than **Virginal** ($M = 1.3$, $SD = 0.48$, $p = .011$), **Tapestry** ($M = 1.4$, $SD = 0.7$, $p = .011$), and **Pietà** ($M = 1.4$, $SD = 0.5$, $p = .021$). As we discuss in Section 5.1, this may be due to the fact that this artwork was the only one with voice-recorded descriptions and not synthesized speech.

4.2.3 *Overlay Images.*

AR overlays were found to be either very or moderately useful (Figure 4e) by all the participants, with no significant differences across the artworks. On average, for **Virginal** and **Tapestry**, the score was 4.8 ($SD = 0.42$), and for **Marriage** and **Pietà** it was 4.7 ($SD = 0.48$).

The perceived clarity of the AR overlays was high for all the artworks (Figure 4f). It was 4.9 ($SD = 0.32$) on average for **Marriage**, 4.6 ($SD = 0.52$) for **Tapestry**, 4.5 ($SD = 0.53$) for **Pietà**, and 4.4 ($SD = 0.7$) for **Virginal**. However this metric differed significantly across the artworks ($p = .037$, $\chi^2 = 8.45$, $W = .67$). Specifically, **Marriage** obtained significantly better scores than **Virginal** ($p = .031$). This may be because, as reported by the participants (Section 4.3.4), light reflections on the lacquered wood of **Virginal** impaired the overlay visibility.

4.2.4 Touch Exploration. This functionality was generally found to be useful (Figure 4g). Significant differences were found among different artwork types ($p < .001$, $\chi^2 = 19.01$, $W = .22$). Scores were significantly ($p < .001$) higher for **Virginal** ($M = 4.5$, $SD = 0.53$), **Tapestry** ($M = 4.6$, $SD = 0.52$), and **Marriage** ($M = 4.8$, $SD = 0.53$) than for **Pietà** ($M = 3.3$, $SD = 0.48$). A reason may be that parts of **Pietà** are visible from specific points of view only. Thus, it is required to walk around the artwork to interact with parts that are hidden from a given position.

Regarding touchscreen exploration clarity, the scores were overall lower (Figure 4h). This may indicate that, for LV people, this interaction is uncommon and therefore less clear. **Pietà** had an average score of 3.3 ($SD = 0.48$), **Virginal** 3.6 ($SD = 0.52$), **Tapestry** 4 ($SD = 0.94$), and **Marriage** 4.5 ($SD = 0.53$). Significant differences ($p = .07$, $\chi^2 = 12.18$, $W = .23$) were detected, specifically between **Pietà** and **Marriage** ($p = .002$), and **Virginal** and **Mariage** ($p = .02$). An interpretation of this may be that for the artworks that have less visually contrasted elements, in particular **Pietà**, it was not clear to the participants which were the explorable artwork parts.

4.2.5 System Usefulness and Usability. As reported in Figure 5, *Musa* was unanimously found to be very useful for accessing artworks (**Rq1**). All participants appreciated the idea of using their own smartphone as a museum guide. Using *Musa* for accessing artwork descriptions at home was still perceived as useful, but with less enthusiasm.

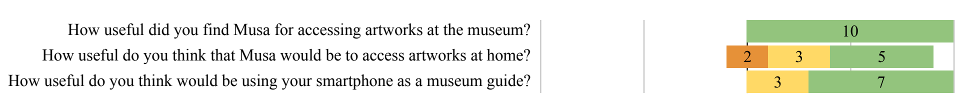


Fig. 5. Final questionnaire results (Scale: 1 2 3 4 5)

System usability was also rated highly (Figure 6), with an average score of 94.3 ($SD = 3.13$), regarded as "Best Imaginable" based on prior literature [12]. All the specific questions achieved scores considered “good” in prior benchmarks [56]. In particular, the participants were eager to use the system frequently (**Q1**, $M = 4.6$, $SD = 0.52$) and did not find it unnecessarily complex (**Q2**, $M = 1$, $SD = 0$). They found the system easy to use (**Q3**, $M = 4.7$, $SD = 0.48$) and did not feel they needed technical support to use it (**Q4**, $M = 1$, $SD = 0$). The functionalities were perceived as well integrated (**Q5**, $M = 4.9$, $SD = 0.32$) and were not found to be inconsistent (**Q6**, $M = 1.2$, $SD = 0.42$). *Musa* was also found to be quick to learn (**Q7**, $M = 4.4$, $SD = 0.7$) and wasn’t perceived to be cumbersome to use (**Q8**, $M = 1.2$, $SD = 0.42$). Finally, participants felt confident when using the system (**Q9**, $M = 4.8$, $SD = 0.42$) and did not need to learn a lot of things before using it (**Q10**, $M = 1.3$, $SD = 0.48$).

4.3 Comments and Suggestions

Participants commented on various system functionalities and proposed several suggestions for improvement. In particular, they suggested addressing the problem of mobility inside art venues. Specifically, *P*1 proposed providing navigation assistance in the venue, while *P*3 suggested providing framing and navigation support for the exploration of 3D artworks in crowded contexts.

4.3.1 Framing of Large 2D Artworks. *P*1, *P*3, and *P*9 commented that **Tapestry** was easy to frame, even from a long distance or from its side. This confirms the results shown in Sections 4.1.1

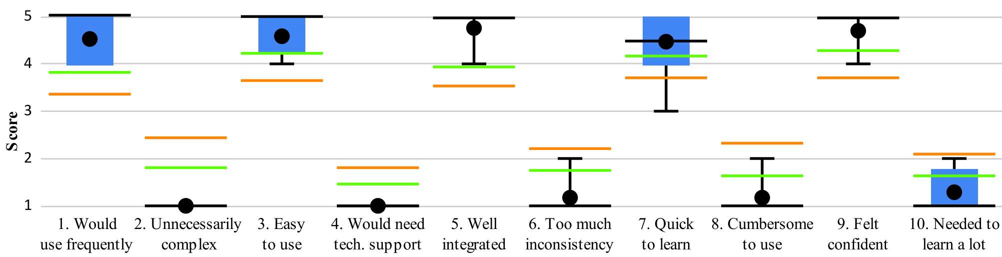


Fig. 6. SUS results (• Mean, — Median, — Benchmark Average, — Benchmark Good)

and 4.2.1. However, with other artworks in the vicinity, this could potentially confuse the user (*P*3, *P*9). Regarding this, *P*3 remarks[6]:

> *"I framed it from the opposite side [of the room]... It was easy to frame, but if another painting was close to the tapestry, it would not be clear which artwork I have to observe"*

4.3.2 Framing of 3D Artworks. All participants were able to correctly frame **Virginal**. However, *P*1, *P*3, and *P*7 observed that it had to be framed from a specific angle and distance to account for light reflections, as noted in Section 4.2.3. For **Pietà**, the framing problems highlighted in Sections 4.1.1 and 4.2.1 were explained by *P*3 and *P*5 who reported that the statue is not well recognized if framed from the right or from behind. Moreover, as *P*3 reports, while walking around the artwork, other visitors often partially cover the view, blocking the recognition:

> *"If someone stands in the way between me and the sculpture, Musa does not recognize it. However, if someone only covers the lower part, if I frame from the belt upward, it recognizes it correctly"*

4.3.3 Audio Descriptions. While all the participants found it easy to listen to audio descriptions and navigate through them, some improvement suggestions were provided. For example, regarding **Pietà**, *P*3 observes that quantitative information about the size of the sculpture elements (*e.g.*, length of legs and arms) would facilitate the understanding of its morphological features. *P*2 and *P*3 instead suggested adding other capabilities to the player such as bookmarks and paragraph skipping. Participants also observed that, when VoiceOver is switched off, the descriptions' speech rate is set to a default value, which is often slower than their speed setting (*P*3, *P*5, *P*6, *P*7). Indeed, participants would sometimes switch VoiceOver off (*e.g.*, to zoom in/out), and found this change of speech rate confusing. Furthermore, not being able to change the speech rate was found to be tiresome (*P*3, *P*9). This was particularly evident for **Marriage**, which was voice-recorded (Section 4.2.2). On this, *P*3 remarks:

> *"VoiceOver doesn't read this description so I can't listen it at a faster speed rate"*

4.3.4 Overlays and Filters. Filters, in particular, "contrast" and "brightness", were considered useful (*P*1, *P*2, *P*3, *P*6, *P*7). Instead, the effects of the "saturation" filter were not clear to *P*4 and *P*5. We highlight that, since saturation depends on the presence of color, it was of little use for **Pietà** and also had limited effect on **Tapestry** and **Virginal**. Participants expressed high appreciation for the overlays as well. However, as previously noted (Section 4.2.3), light reflection limited the visibility of the overlays for **Virginal**. Indeed, *P*3 observed:

> *"The image is very bright when I magnify it. I cannot see well the river mentioned in the description"*

[6]Comments are translated from Italian

4.3.5 Touchscreen Interactions.

Touchscreen interaction was generally appreciated, but it had some issues with large 2D artworks (Section 4.1.4) and 3D artworks (Section 4.2.4). For **Pietà**, participants found it hard to identify the interactive artwork parts (*P*2, *P*3, *P*4, *P*5) which sometimes resulted in touching an already selected part, triggering a repeat in the description (*P*2, *P*4). Instead, for **Tapestry**, due to the high number of elements, *P*1 and *P*3 had difficulty finding the right position to tap on elements of interest. For example, *P*3 noted:

> *"When I move away and try to tap the Gemini sign it sometimes describes a different part"*

5 Discussion

We discuss human (Section 5.1) and contextual (Section 5.2) factors whose interplay influences system usage and appreciation. We also acknowledge the limitations of the study and the adopted methodology (Section 5.4).

5.1 Human Factors

Participants' ability to access artworks autonomously confirms the potential of AR-based technology, thus reinforcing previous findings [3]. It also resulted in a consistently high appreciation of *Musa* (system usefulness was unanimously praised by the participants), despite some potential issues, in particular regarding the time required to frame an artwork. We believe that the excitement toward this new enabling technology may have positively influenced their perception of the system, for instance by emphasizing its usability.

One result that could suggest a negative user experience is the framing time. Indeed, participants required 22 seconds on average to frame an artwork, with a maximum time of more than a minute. However, we note that in most cases, the recognition was almost instantaneous once the artwork was correctly framed, but still it required a few seconds for the participant to correctly position themselves so that no other visitor was occluding the view. In few occasions only we observed that, although the artwork was correctly framed, it was not immediately detected. This happened for **Pietà** and, less frequently, for **Virginal**, as discussed in Section 5.2.2. Indeed, we highlight that if we exclude **Pietà**, in 96.6% of the cases framing was perceived as easy or moderately easy.

Participants appreciated the AR interaction provided by *Musa* and found it to be easy to use. However, AR interaction was sometimes unclear for the participants. We believe that LV users are not used to such forms of interaction since AR applications are usually designed without considering their accessibility needs. Despite this, participants found it easy learning to use the system and felt they would not need support in using it.

While most participants appreciated accessing artwork descriptions through synthesized speech, some also used text descriptions. This indicates that they would sometimes like to use their residual vision and avoid accessing descriptions through audio. Finally, the results and the participants' comments highlight that voice recording is more tiresome to listen to than synthesized speech. This is because, unlike synthesized speech, voice recordings cannot be customized and, in particular, cannot be quickened. Thus, we recommend that verbal descriptions of artworks should be provided through synthesized speech, affording personalization of speech rate and other preferences.

5.2 Contextual Factors

Two contextual factors influenced the user experience most: the environment and the artwork type.

5.2.1 Environment Properties.

Experimenting in a real-world environment unveiled two undetected challenges. First, several participants have mentioned that they would need support in detecting distant artworks [49] and reaching them [75], confirming the need for navigation assistance in museum settings [10, 29, 43]. The artworks themselves can be near to one another, making it difficult for a blind and LV user to discriminate between them during AR access. Thus, navigation assistance should be accurate enough to support users in locating different artworks without ambiguity.

Second, the presence of other visitors can also impact the ability to navigate inside art venues and access artwork for blind and LV people. Indeed, as our participants reported, crowded areas may be difficult to travel and other visitors may occlude artworks, making it more difficult to interact with them in AR. Thus, navigation support should also account for the presence of other visitors, guiding the user around them [47] and avoiding occlusion when accessing artworks in AR in the vicinity of other people.

5.2.2 *Artwork Type.* Specific challenges also emerged in relation to AR interaction with artworks of different types and characteristics (**Rq2**). First, artworks lacking salient visual features, such as **Pietà**, were hard to detect in AR from some directions, which resulted in significantly longer framing times. A similar problem emerged for **Virginal** where light reflections sporadically resulted in non-immediate artwork detection. The lack of strong visual features also made it difficult for the participants to understand which elements of **Pietà** can be explored in AR. Indeed, many participants have attempted to use filters to improve visual access, confirming the usefulness of this feature. As an additional support, AR overlays could also be used to display boundaries between artwork parts, hinting what can be touched.

A second challenge for 3D artworks, such as **Pietà**, and large 2D artworks, such as **Tapestry**, is that different artwork parts may not all be accessible from a single point of view. Due to this, participants needed to move around while interacting with the artwork, which was found to be difficult. Thus, supporting users' movements while interacting with artwork parts should also be considered in the design of the navigation system.

5.3 Technological Challenges

Three main technological challenges emerged. First, navigation was mentioned by several participants, confirming a need already reported in the literature [9, 72]. Navigation support should be provided to enable independent mobility in art venues, and it should similarly guide the users to facilitate artwork framing. Navigation assistance should also provide guidance during the exploration of artworks that can be observed from multiple points of view (*e.g.*, statues).

A second technological challenge is to detect other visitors [32, 95]. In general, detecting other people is a relevant feature that facilitates users' navigation. However, in this case, it is also needed prevent artwork occlusion, by informing the user of occluding visitors and, possibly, to also guide the user to avoid these occlusions.

Finally, the evaluation also highlighted the importance of a reliable solution to detect the artworks, in particular when they lack salient visual features. This technological limitation was also noted in prior literature [48]. One solution would be to improve artwork detection robustness by complementing visual recognition with the use of 3D mapping of the entire scene, leveraging LIDAR sensors available on modern mobile devices. An alternative solution would be to rely of fiducial markers. While this solution is technologically simpler, fiducial markers would have an aesthetic impact that may be unsuitable for art venue settings [60]. In particular, curators of the museums where we conducted the experiments categorically excluded the possibility to place fiducial markers.

5.4 Limitations

Conducting the study in a museum setting allowed us to acquire reliable results on the applicability of the approach to real-world environments. However, it introduced several limitations in the recruitment of the participants that need to be acknowledged. Indeed, we only had access to the available local population and therefore could recruit only 10 participants, all from the same geographical region, and a limited age range (21-54). While we note that such challenges are common in the accessibility field and that this number of participants is consistent with other works [79], we highlight that additional studies are needed to ensure that the results generalize.

Another limitation is the potential presence of the effects of order in the experiment. Indeed, because two museums were involved in the study and participants had to walk between them, we always had one artwork (**Marriage**) accessed last. Although we strongly believe that the increase in the fatigue of listening the description for **Marriage** can be attributed to the use of voice recordings, as also highlighted by the participants, we cannot exclude a confounding effect due to the order of the experiments. We also highlight that there could be a possible novelty bias in the subjective evaluation of the system, as the participants have never used a similar system before. Indeed, all the results of the SUS questionnaire were much higher than the benchmark scores. However, we also note that these scores are in line with prior results that compared *Musa* with an audio guide baseline.

6 Conclusions and Future Work

This paper investigates how people with LV can access different types of artworks through the *Musa* application. All participants were able to autonomously access the content for all artworks and found the application useful and usable. However, specific challenges related to different artwork types were also identified. In particular, for artwork lacking salient visual features, visual recognition during framing remains a challenge.

These results can provide guidance for future developments in this field. Indeed, while on one side this paper shows the advantages an AR-based application like *Musa*, it also illustrates possible limitations. For example, from the user interaction point of view, synthetic speech is preferable to recorded voice. As another example, considering contextual factors, it is preferable to avoid artworks that lack salient visual features or present glares.

This paper also identifies possible future improvements to the system, including the integration of a navigation support, and enhanced recognition for artworks lacking salient visual features through the mapping of the entire scene using a LIDAR sensor. As a future work, we also believe that it could be possible to better investigate how LV people interact with an application like *Musa*, possibly using methodologies like interaction analysis. Finally, while *Musa* was previously compared with an audio guide in a simulated scenario [3], we also want to repeat this experiment in a real world scenario to confirm the validity of the approach with respect to a traditional solution.

References

[1] [n. d.]. Descrivedendo Guidelines. https://www.descrivedendo.it/ [accessed: 2024-03-01].

[2] [n. d.]. IMap audio tactile accessible artworks. http://www2.tate.org.uk/imap/index.shtml [accessed: 2024-03-05].

[3] Dragan Ahmetovic, Cristian Bernareggi, Kristian Keller, and Sergio Mascetti. 2021. MusA: artwork accessibility through augmented reality for people with low vision. In *International Web for All Conference*. ACM.

[4] Dragan Ahmetovic, Nahyun Kwon, Uran Oh, Cristian Bernareggi, and Sergio Mascetti. 2021. Touch Screen Exploration of Visual Artwork for Blind People. In *The Web Conference*. ACM.

[5] Dragan Ahmetovic, Daisuke Sato, Uran Oh, Tatsuya Ishihara, Kris Kitani, and Chieko Asakawa. 2020. ReCog: Supporting Blind People in Recognizing Personal Objects. In *Conference on Human Factors in Computing Systems*. ACM.

[6] Giorgos Anagnostakis, Michalis Antoniou, Elena Kardamitsi, Thodoris Sachinidis, Panayiotis Koutsabasis, Modestos Stavrakis, Spyros Vosinakis, and Dimitris Zissis. 2016. Accessible museum collections for the visually impaired: combining tactile exploration, audio descriptions and mobile gestures. In *Conference on Human-Computer Interaction with Mobile Devices and Services - Adjunct*.

[7] Anastasios Nikolas Angelopoulos, Hossein Ameri, Debbie Mitra, and Mark Humayun. 2019. Enhanced depth navigation through augmented reality depth mapping in patients with low vision. *Scientific reports* (2019).

[8] Vassilios S Argyropoulos and Charikleia Kanari. 2015. Re-imagining the museum through “touch”: Reflections of individuals with visual disability on their experience of museum-visiting in Greece. *Alter* (2015).

[9] Saki Asakawa, João Guerreiro, Dragan Ahmetovic, Kris M Kitani, and Chieko Asakawa. 2018. The present and future of museum accessibility for people with visual impairments. In *SIGACCESS Conference on Computers and Accessibility - Adjunct*. ACM.

[10] Saki Asakawa, João Guerreiro, Daisuke Sato, Hironobu Takagi, Dragan Ahmetovic, Desi Gonzalez, Kris M Kitani, and Chieko Asakawa. 2019. An Independent and Interactive Museum Experience for Blind People. In *International Web For All Conference*. ACM.

[11] Elisabeth Salzhauer Axel and Nina Sobol Levent. 2003. *Art beyond sight: A resource guide to art, creativity, and visual impairment.* AFB.

[12] Aaron Bangor, Philip Kortum, and James Miller. 2009. Determining what individual SUS scores mean: Adding an adjective rating scale. *Journal of usability studies* (2009).

[13] Francesco Bellotti, Riccardo Berta, Alessandro De Gloria, and Massimiliano Margarone. 2006. Guiding visually impaired people in the exhibition. In *Mobile Guide Workshop*. University of Turin.

[14] Benedetto da Milano e collaboratori su disegni di Bramantino. ca. 1503-1508. Arazzi Trivulzio: Mese di Maggio. Wikimedia Commons. https://commons.wikimedia.org/wiki/File:Benedetto_da_milano_e_collaboratori_su_disegno_del_bramantino,_arazzi_trivulzio,_maggio,_1504_ca.-1509,_01.JPG [User: Sailko; CC-BY-SA-3.0; accessed 2024-01-26].

[15] Yoav Benjamini and Yosef Hochberg. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)* (1995).

[16] Galina Bogdanova, Negoslav Sabev, Zhivko Tomov, and Mirena Ekmekci. 2021. Physical and Digital Accessibility in Museums in the New Reality. In *International Symposium on Multidisciplinary Studies and Innovative Technologies*. IEEE.

[17] John Brooke. 1996. SUS-A quick and dirty usability scale. In *Usability evaluation in industry*. CRC Press.

[18] Serap Buyurgan. 2009. The Expectations of the Visually Impaired University Students from Museums. *Educational Sciences: Theory and Practice* (2009).

[19] Luis Cavazos Quero, Jorge Iranzo Bartolomé, Seonggu Lee, En Han, Sunhee Kim, and Jundong Cho. 2018. An Interactive Multimodal Guide to Improve Art Accessibility for Blind People. In *SIGACCESS Conference on Computers and Accessibility*. ACM.

[20] Rafie R Cecilia. 2021. COVID-19 Pandemic: threat or opportunity for blind and partially sighted museum visitors? *Journal of conservation and Museum Studies* (2021).

[21] William Jay Conover. 1999. *Practical nonparametric statistics*. Wiley.

[22] James M Coughlan and Joshua Miele. 2017. AR4VI: AR as an accessibility tool for people with visual impairments. In *IEEE International Symposium on Mixed and Augmented Reality - Adjunct*. IEEE.

[23] Karin De Coster and Volkmar Mühleis. 2007. Intersensorial translation: visual art made up by words. In *Media for all*. Brill Rodopi.

[24] John Dudley, Lulu Yin, Vanja Garaj, and Per Ola Kristensson. 2023. Inclusive Immersion: a review of efforts to improve accessibility in virtual reality, augmented reality and the metaverse. *Virtual Reality* (2023).

[25] Aram Dulyan and Ernest Edmonds. 2010. AUXie: Initial evaluation of a blind-accessible virtual museum tour. In *Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction*. ACM.

[26] Dylan R Fox, Ahmad Ahmadzada, Clara Tenia Wang, Shiri Azenkot, Marlena A Chu, Roberto Manduchi, and Emily A Cooper. 2023. Using augmented reality to cue obstacles for people with low vision. *Optics Express* (2023).

[27] Francesco Hayez. 1859. Il Bacio. Wikimedia Commons. https://commons.wikimedia.org/wiki/File:Hayez,_Il_bacio_(1859).png [accessed 2024-01-26].

[28] Milton Friedman. 1937. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the american statistical association* (1937).

[29] Giuseppe Ghiani, Barbara Leporini, and Fabio Paternò. 2009. Vibrotactile feedback to aid blind users of mobile guides. *Journal of Visual Languages & Computing* (2009).

[30] Desi Gonzalez. 2017. A Path With Choice: What We Learned From Designing An Inclusive Audio Guide. In *Museums and the Web*. Archives & Museum Informatics.

[31] Sarika Gopalakrishnan, Sanjana Chouhan Suwalal, Gnanapoonkodi Bhaskaran, and Rajiv Raman. 2020. Use of augmented reality technology for improving visual acuity of individuals with low vision. *Indian Journal of Ophthalmology* (2020).

[32] Martin Grayson, Anja Thieme, Rita Marques, Daniela Massiceti, Ed Cutrell, and Cecily Morrison. 2020. A dynamic AI system for extending the capabilities of blind people. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–4.

[33] Kozue Handa, Hitoshi Dairoku, and Yoshiko Toriyama. 2010. Investigation of priority needs in terms of museum service accessibility for visually impaired visitors. *British journal of visual impairment* (2010).
[34] Simon Hayhoe. 2013. Expanding our vision of museum education and perception: An analysis of three case studies of independent blind arts learners. *Harvard Educational Review* (2013).
[35] Morton A Heller and Edouard Gentaz. 2013. *Psychology of touch and blindness.* Psychology press.
[36] Jaylin Herskovitz, Jason Wu, Samuel White, Amy Pavel, Gabriel Reyes, Anhong Guo, and Jeffrey P Bigham. 2020. Making mobile augmented reality applications accessible. In *SIGACCESS Conference on Computers and Accessibility*. ACM.
[37] Leona Holloway, Kim Marriott, Matthew Butler, and Alan Borning. 2019. Making sense of art: Access for gallery visitors with vision impairments. In *Conference on Human Factors in Computing Systems*. ACM.
[38] Caro Howell and Dan Porter. 2003. *Re-Assessing Practice: Visual Art, Visually Impaired People and the Web.* Technical Report. Archives & Museum Informatics.
[39] Jonathan Huang, Max Kinateder, Matt J Dunn, Wojciech Jarosz, Xing-Dong Yang, and Emily A Cooper. 2019. An augmented reality sign-reading assistant for users with reduced vision. *PloS one* (2019).
[40] Alex D Hwang and Eli Peli. 2014. An augmented-reality edge enhancement application for Google Glass. *Optometry and Vision Science* (2014).
[41] Ioannes Ruckers. ca. 1600. Virginale doppio (con ottavino). Wikimedia Commons. https://commons.wikimedia.org/wiki/File:Doppio_vjrginale_Castello_Sforzesco.jpg [User: Saggitarius A; CC-BY-SA-4.0; accessed 2024-01-26].
[42] Mina Iskander, Titilola Ogunsola, Rithambara Ramachandran, Richard McGowan, and Lama A Al-Aswad. 2021. Virtual reality and augmented reality in ophthalmology: A contemporary prospective. *The Asia-Pacific Journal of Ophthalmology* (2021).
[43] Dhruv Jain. 2014. Pilot evaluation of a path-guided indoor navigation system for visually impaired in a public museum. In *SIGACCESS conference on Computers & accessibility*. ACM.
[44] Katerina Kabassi, Alessia Amelio, Vasileios Komianos, and Konstantinos Oikonomou. 2019. Evaluating museum virtual tours: the case study of Italy. *Information* (2019).
[45] Hıdır Karaduman, Ümran Alan, and E Özlem Yiğit. 2023. Beyond "do not touch": the experience of a three-dimensional printed artifacts museum as an alternative to traditional museums for visitors who are blind and partially sighted. *Universal Access in the Information Society* 22, 3 (2023), 811–824.
[46] Brian FG Katz, Slim Kammoun, Gaëtan Parseihian, Olivier Gutierrez, Adrien Brilhault, Malika Auvray, Philippe Truillet, Michel Denis, Simon Thorpe, and Christophe Jouffrais. 2012. NAVIG: Augmented reality guidance system for the visually impaired: Combining object localization, GNSS, and spatial audio. *Virtual Reality* (2012).
[47] Seita Kayukawa, Tatsuya Ishihara, Hironobu Takagi, Shigeo Morishima, and Chieko Asakawa. 2020. Guiding blind pedestrians in public spaces by understanding walking behavior of nearby pedestrians. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* (2020).
[48] Jens Keil, Laia Pujol, Maria Roussou, Timo Engelke, Michael Schmitt, Ulrich Bockholt, and Stamatia Eleftheratou. 2013. A digital look at physical museum exhibits: Designing personalized stories with handheld Augmented Reality in museums. In *2013 Digital Heritage International Congress (DigitalHeritage)*, Vol. 2. IEEE, 685–688.
[49] Max Kinateder, Justin Gualtieri, Matt J Dunn, Wojciech Jarosz, Xing-Dong Yang, and Emily A Cooper. 2018. Using an augmented reality device as a distance-based vision aid—promise and limitations. *Optometry and Vision Science* (2018).
[50] Nahyun Kwon, Yunjung Lee, and Uran Oh. 2022. Supporting a crowd-powered accessible online art gallery for people with visual impairments: a feasibility study. *Universal Access in the Information Society* (2022).
[51] Florian Lang, Albrecht Schmidt, and Tonja Machulla. 2020. Augmented reality for people with low vision: symbolic and alphanumeric representation of information. In *International Conference on Computers Helping People with Special Needs.* Springer.
[52] Florian Lang, Albrecht Schmidt, and Tonja Machulla. 2022. Mixed reality as assistive technology: Guidelines based on an assessment of residual functional vision in persons with low vision. In *International Conference on Computers Helping People with Special Needs.* Springer, 484–493.
[53] Barbara Leporini, Ivan Norscia, et al. 2008. "Fine Tuning" Image Accessibility for Museum Web Sites. *Journal of Universal Computer Science* (2008).
[54] Nina Levent and Christine Reich. 2013. Museum accessibility: Combining audience research and staff training. *Journal of Museum Education* (2013).
[55] James R Lewis. 1993. Multipoint scales: Mean and median differences and observed significance levels. *International Journal of Human-Computer Interaction* (1993).
[56] James R Lewis and Jeff Sauro. 2018. Item benchmarks for the system usability scale. *Journal of Usability Studies* (2018).
[57] Franklin Mingzhe Li, Lotus Zhang, Maryam Bandukda, Abigale Stangl, Kristen Shinohara, Leah Findlater, and Patrick Carrington. 2023. Understanding Visual Arts Experiences of Blind People. In *Conference on Human Factors in Computing Systems.* ACM.

[58] Rensis Likert. 1932. A technique for the measurement of attitudes. *Archives of psychology* (1932).

[59] Robert W Lindeman and Haruo Noma. 2007. A classification scheme for multi-sensory augmented reality. In *Symposium on Virtual reality software and technology*. ACM.

[60] Kaj Mäkelä, Sara Belt, Dan Greenblatt, and Jonna Häkkilä. 2007. Mobile interaction with visual and RFID tags: a field study on user perceptions. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 991–994.

[61] Susana Mesquita and Maria João Carneiro. 2016. Accessibility of European museums to visitors with visual impairments. *Disability & Society* (2016).

[62] Michelangelo Buonarroti. 1552-1554. Pietà Rondanini. Wikimedia Commons. https://commons.wikimedia.org/wiki/File:Pieta_Rondanini.jpg [User: Bahar; CC-BY-SA-3.0; accessed 2024-01-26].

[63] Paul Milgram, Haruo Takemura, Akira Utsumi, and Fumio Kishino. 1995. Augmented reality: A class of displays on the reality-virtuality continuum. In *Telemanipulator and telepresence technologies*. Spie.

[64] Hein Min Htike, Tom H. Margrain, Yu-Kun Lai, and Parisa Eslambolchilar. 2021. Augmented reality glasses as an orientation and mobility aid for people with low vision: a feasibility study of experiences and requirements. In *Conference on Human Factors in Computing Systems*. ACM.

[65] Martez Mott, Edward Cutrell, Mar Gonzalez Franco, Christian Holz, Eyal Ofek, Richard Stoakley, and Meredith Ringel Morris. 2019. Accessible by design: An opportunity for virtual reality. In *International Symposium on Mixed and Augmented Reality - Adjunct*. IEEE.

[66] Mukhriddin Mukhiddinov and Soon-Young Kim. 2021. A Systematic Literature Review on the Automatic Creation of Tactile Graphics for the Blind and Visually Impaired. *Processes* (2021).

[67] Sergio Naranjo-Puentes, Camilo Escobar-Velásquez, Christopher Vendome, and Mario Linares-Vásquez. 2022. A preliminary study on accessibility of augmented reality features in mobile apps. In *International Conference on Software Analysis, Evolution and Reengineering*. IEEE.

[68] Andrea Narcisi, Huiying Shen, Dragan Ahmetovic, Sergio Mascetti, and James M Coughlan. 2024. Accessible point-and-tap interaction for acquiring detailed information about tactile graphics and 3D models. In *International Conference on Computers Helping People with Special Needs*. Springer, 252–259.

[69] Theofilos Papadopoulos, Konstantinos Evangelidis, Theodore H Kaskalis, Georgios Evangelidis, and Stella Sylaiou. 2021. Interactions in augmented and mixed reality: an overview. *Applied Sciences* (2021).

[70] Ramachandra Pararajasegaram. 2004. Low vision care: the need to maximise visual potential. *Community eye health* (2004).

[71] Petros Pistofidis, George Ioannakis, Fotis Arnaoutoglou, Natasa Michailidou, Melpomeni Karta, Chairi Kiourt, George Pavlidis, Spyridon G Mouroutsos, Despoina Tsiafaki, and Anestis Koutsoudis. 2021. Composing smart museum exhibit specifications for the visually impaired. *Journal of Cultural Heritage* (2021).

[72] Yaniv Poria, Arie Reichel, and Yael Brandt. 2009. People with disabilities visit art museums: an exploratory study of obstacles and difficulties. *Journal of Heritage Tourism* (2009).

[73] Giorgio Presti, Dragan Ahmetovic, Mattia Ducci, Cristian Bernareggi, Luca Ludovico, Adriano Baratè, Federico Avanzini, and Sergio Mascetti. 2021. Iterative Design of Sonification Techniques to Support People with Visual Impairments in Obstacle Avoidance. *Transactions on Accessible Computing* (2021).

[74] Raffaello Sanzio. 1504. Sposalizio della Vergine. Wikimedia Commons. https://commons.wikimedia.org/wiki/File:Raffaello_-_Spozalizio_-_Web_Gallery_of_Art.jpg [accessed 2024-01-26].

[75] Kyle Rector, Keith Salmon, Dan Thornton, Neel Joshi, and Meredith Ringel Morris. 2017. Eyes-free art: exploring proxemic audio interfaces for blind and low vision art engagement. *PACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* (2017).

[76] Andreas Reichinger, Stefan Maierhofer, and Werner Purgathofer. 2011. High-quality tactile paintings. *Journal on Computing and Cultural Heritage* (2011).

[77] Carmen Santoro, Fabio Paterno, Giulia Ricci, and Barbara Leporini. 2007. A multimodal mobile museum guide for all. In *Mobile Interaction with the Real World Workshop*. University of Munich.

[78] Ahmed M Sayed, Mohamed Abou Shousha, MD Baharul Islam, Taher K Eleiwa, Rashed Kashem, Mostafa Abdel-Mottaleb, Eyup Ozcan, Mohamed Tolba, Jane C Cook, and Richard K Parrish. 2020. Mobility improvement of patients with peripheral visual field losses using novel see-through digital spectacles. *PloS one* (2020).

[79] Andrew Sears and Vicki Hanson. 2011. Representing users in accessibility research. In *Conference on Human factors in computing systems*. ACM.

[80] Lee Stearns, Victor DeSouza, Jessica Yin, Leah Findlater, and Jon E Froehlich. 2017. Augmented reality magnification for low vision users with the microsoft hololens and a finger-worn camera. In *SIGACCESS Conference on Computers and Accessibility*. ACM.

[81] Maciej Tomczak and Ewa Tomczak. 2014. The need to report effect size estimates revisited. An overview of some recommended measures of effect size. (2014).

[82] Matthew GJ Trese, Naheed W Khan, Kari Branham, Erin Brown Conroy, and Sayoko E Moroi. 2016. Expansion of severely constricted visual field using Google Glass. *Ophthalmic Surgery, Lasers and Imaging Retina* (2016).

[83] Nelson Daniel Troncoso Aldas, Sooyeon Lee, Chonghan Lee, Mary Beth Rosson, John M Carroll, and Vijaykrishnan Narayanan. 2020. AIGuide: An augmented reality hand guidance application for people with visual impairments. In *SIGACCESS Conference on Computers and Accessibility*. ACM.

[84] United Nations. 2006. Convention on the Rights of Persons with Disabilities. https://www.un.org/disabilities/documents/convention/convention_accessible_pdf.pdf

[85] Raša Urbas, Matej Pivar, and Urška Stankovič Elesini. 2016. Development of tactile floor plan for the blind and the visually impaired by 3D printing technique. *Journal of graphic engineering and design* (2016).

[86] Fernando Vargas-Martin, Eli Peli, et al. 2002. Augmented-view for restricted visual field: multiple device implementations. *Optometry and Vision Science* (2002).

[87] Marcus Weisen. 2020. How accessible are museums today? In *Touch in Museums*. Routledge.

[88] Jaroslaw Wiazowski. 2014. Can Braille be revived? A possible impact of high-end Braille and mainstream technology on the revival of tactile literacy medium. *Assistive Technology* (2014).

[89] Chris Yoon, Ryan Louie, Jeremy Ryan, MinhKhang Vu, Hyegi Bang, William Derksen, and Paul Ruvolo. 2019. Leveraging augmented reality to create apps for people with visual disabilities: A case study in indoor navigation. In *SIGACCESS Conference on Computers and Accessibility*. ACM.

[90] Ola Younis, Waleed Al-Nuaimy, Fiona Rowe, et al. 2019. A hazard detection and tracking system for people with peripheral vision loss using smart glasses and augmented reality. *International Journal of Advanced Computer Science and Applications* (2019).

[91] Kaixing Zhao, Rui Lai, Bin Guo, Le Liu, Liang He, and Yuhang Zhao. 2024. AI-Vision: A Three-Layer Accessible Image Exploration System for People with Visual Impairments in China. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* (2024).

[92] Yuhang Zhao, Edward Cutrell, Christian Holz, Meredith Ringel Morris, Eyal Ofek, and Andrew D Wilson. 2019. SeeingVR: A set of tools to make virtual reality more accessible to people with low vision. In *Conference on human factors in computing systems*. ACM.

[93] Yuhang Zhao, Michele Hu, Shafeka Hashash, and Shiri Azenkot. 2017. Understanding low vision people's visual perception on commercial augmented reality glasses. In *Conference on Human Factors in Computing Systems*. ACM.

[94] Yuhang Zhao, Elizabeth Kupferstein, Hathaitorn Rojnirun, Leah Findlater, and Shiri Azenkot. 2020. The effectiveness of visual and audio wayfinding guidance on smartglasses for people with low vision. In *Conference on human factors in computing systems*. ACM.

[95] Yuhang Zhao, Shaomei Wu, Lindsay Reynolds, and Shiri Azenkot. 2018. A face recognition application for people with visual impairments: Understanding use beyond the lab. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–14.

A Implementation of 3D Artwork Recognition in *Musa*

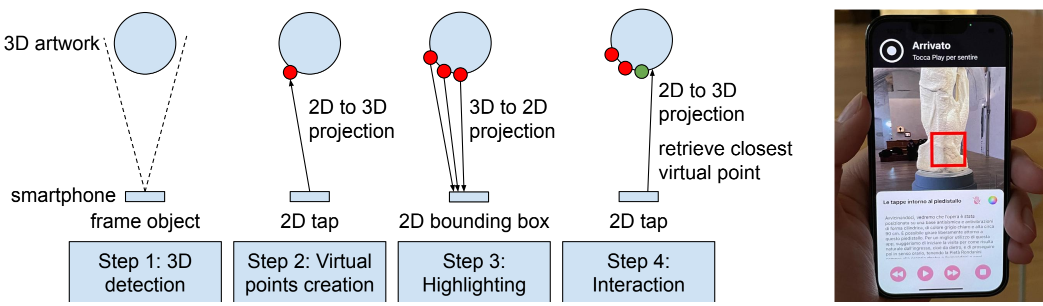
The detection capability is implemented using the ARKit object detection API[7]. It requires that the artwork is first scanned to store its 3D structure, which is then used for the detection. The artwork scanning uses a publicly available application provided by Apple as part of their 3D object detection framework. This app requires manually specifying the object bounding boxes and framing the object from multiple positions. The result is a file representing the object 3D structure (point cloud) and textures.

To support the highlighting and touchscreen interaction with artwork parts, it is also necessary to define the artwork parts and to map them to the corresponding description chapters. The procedure consists of four steps, depicted in Figure 7a, and makes use of a custom app that we developed to be used used by museum staff when adding a new 3D artwork to the system. As the first step, the custom app is used to detect the artwork using the previously created 3D scan. As the second step, the user chooses a description chapter corresponding to an artwork part and selects some points on the object that belong to that part by tapping on the screen. At each tap, the app projects the 2D screen coordinates to the surface of the artwork with the raycast function[8]. The resulting 3D points, which we call *virtual points*, are stored associated with the corresponding chapter. As the

[7] https://developer.apple.com/documentation/arkit/arkit_in_ios/content_anchors/scanning_and_detecting_3d_objects

[8] https://developer.apple.com/documentation/arkit/arraycastquery

third step, while a museum visitor uses *Musa* to interact with the 3D artwork, the virtual points are projected onto the 2D screen reference system. The bounding box containing all the points of an artwork part is highlighted when the corresponding chapter is read (Figure 7b). Finally, as the fourth step, during interaction, when the user touches the artwork on the touchscreen, the 2D screen coordinate is projected on the artwork 3D model (again, using raycast), the closest virtual point is computed, and its associated chapter is played.



(a) Highlighting and touchscreen interaction (b) Result

Fig. 7. *Musa* with 3D artworks