

**Envisioning the Museum Voice:
Gaze and Speech as Modes of Interacting with Art**

Tate Strickland

A thesis
submitted in partial fulfillment of the
requirements for the degree of

Master of Design

University of Washington
2017

Committee:
Axel Roesler
Karen Cheng
Christopher Ozubko

Program authorized to offer degree:
Design

© Copyright 2017
Tate Strickland

University of Washington

Abstract

Envisioning the Museum Voice:
Gaze and Speech as Modes of Interacting with Art
Tate Strickland

Chair of the Supervisory Committee:
Associate Professor Axel Roesler
School of Art

This thesis presents an exploratory vision of how augmented reality could be used in an art gallery setting in a way that is educational yet still appropriate to context. Inspired by the traditional audio guide, the system — prototyped on the Microsoft HoloLens — uses spoken narration as the primary output, with voice and gaze as methods of input to control the experience.

MASTER OF DESIGN THESIS DOCUMENTATION

Envisioning the Museum Voice

Gaze and Speech as Methods of Interacting with Art

Tate Strickland

Contents

Introduction	4
<i>Abstract</i>	
<i>Preface</i>	
Background and Approach	8
<i>Motivation</i>	
<i>What is Augmented Reality?</i>	
<i>Sound, Augmented Reality, and the Museum</i>	
<i>The Microsoft HoloLens</i>	
The Arnolfini Portrait	12
Early Prototypes	16
<i>"Wizard of Oz" Head-Mounted Camera Test</i>	
<i>API.AI Chatbot</i>	
<i>Observations and Results</i>	
HoloLens Prototype	24
<i>The Aesthetics of Voice</i>	
<i>Observations and Results</i>	
Final Design	32
Next Steps	40
Acknowledgments	42
Bibliography	43
Appendix: Narration Scripts	44

1

Introduction

ABSTRACT

This thesis presents an exploratory vision of how augmented reality could be used in an art gallery setting in a way that is interactive and educational, yet still appropriate to context. Inspired by the traditional audio guide, the system — prototyped on the Microsoft HoloLens — uses spoken narration as the primary output, with voice and gaze as methods of input to control the experience.

PREFACE

The origin of this project lies in work that I began seven years ago, in 2010. Mobile app design was a booming area of interest for designers — the iPhone was less than three years old, and the App Store presented a new genre of specialization for interaction designers. I was an undergraduate student at American University, doing my senior-level coursework in user experience design.

I created an augmented reality concept for the iPhone that would, hypothetically, use the camera to recognize paintings in an art gallery and serve text or audio narration about them on the device — either general information, or specific facts tied to tappable targets. The idea, I supposed, was that AR could allow museum curators to populate the gallery with *virtual* content while preserving the aesthetics of the physical space. I illustrated this concept in a video that was less than two minutes long, filmed in the National Gallery of Art in Washington.

The next year, the piece was accepted into the International Symposium on Mixed and Augmented Reality in Basel, Switzerland, and displayed at their exhibit in the new media art festival, SHIFT. I attended, and while surreptitiously lurking next to my video in the gallery to observe reactions, I overheard a few people talking about it. To paraphrase, one of the viewers said something to the effect of, “I’m not sure if I’d want to walk around a museum holding my phone out like that.”

Screen captures from *NGA Experience* (2010).



It was an important critique of my design. The ability to view art on the Internet or in printed reproductions did not destroy the art museum; visitors clearly still find value in experiencing original works of art in person. In Walter Benjamin's terms, they go to see the "unique existence" of the actual work of art, and to experience the "aura" that only it, as the original work, exudes.¹ It's why an estimated six million people visit the *Mona Lisa* in the Musée du Louvre in Paris every year², crowding around the small portrait that is arguably seen more clearly in a reproduction. Viewed from this perspective, it is unsurprising that visitors like the one from the art festival might be hesitant to use a screen-based device to mediate between themselves and the work.

1 Walter Benjamin, "The Work of Art in the Age of Mechanical Reproduction," in *Literary Theory: An Anthology*, 2nd, ed., ed. Julie Rivkin and Michael Ryan (Maiden: Blackwell Publishing, 2004), 1236.

2 Bob Chaundy, "Faces of the Week," *BBC News*, September 29, 2006, accessed June 7, 2017, http://news.bbc.co.uk/2/hi/uk_news/magazine/5392000.stm.

The concept of this project has since become a reality. Most notably, ArtLens, a mobile app released by the Cleveland Museum of Art in 2013, scans two-dimensional artwork in its collection and provides additional information in a similar way. (The app also includes navigational features and non-AR content.³) The core of this project idea also appeared in an app called LayAR — tappable targets tracked to a painting and which triggered information on the screen. After trying LayAR myself, I concluded that the experience was indeed odd. Rather than viewing the painting at full scale on the wall, one fiddles with a small representation on a screen in a way that sapped the unique experience of standing in front of the original work of art.

One might argue that the shortcomings of the experience were technology-related; certainly, the tracking was not perfect and load times were slow. But fundamentally, it felt to me that this particular application of AR was near its functional limit. Using a smartphone as an AR device practically demands the sort of interaction that LayAR used, and that I had proposed in my undergraduate work — using a screen to mediate between viewer and art.

This thesis project is, loosely, a re-imagining of the original premise I conceived seven years ago. In designing it, I've moved away from the screen and attempted to more closely match the type of digital mediation to the context of the museum. In doing so, I've considered other devices and forms of interaction, some of which were unavailable before now.

I also specifically set out to create a functional prototype in this project, rather than stopping at a concept video. Not only could a prototype be tested and evaluated, but the act of creating the prototype would reveal the current capabilities and limitations of current AR tools — and as any designer knows, a design process is nothing without limitations.

Like the earlier project, the prototypes in this thesis certainly have their own drawbacks, which I will attempt to address along the way and in the conclusion. Consequently, I acknowledge that this project is a work in progress. As augmented reality technology evolves, my

³ "ArtLens App," *The Cleveland Museum of Art*, accessed June 8, 2017, <http://www.clevelandart.org/gallery-one/artlens>.

hope is that the conclusions I've derived in this thesis will contribute to our understanding of its function in the small but fascinating domain of the art museum.



Background and Approach

MOTIVATION

The latter half of the 20th century has seen a radical shift in the role of museums in society. Once seen as places that existed to serve the objects in their collections, their focus has turned outward and they now are broadly seen as existing to serve to the public instead.⁴ In this new role of public service, the visitor, too, is considered differently — now treated as a consumer to be educated and engaged. For some museum directors and curators, this engagement can involve the timely and relevant implementation of technology.

These technological solutions sometimes adopt novel formats. Replacing the traditional audio tour at the San Francisco Museum of Modern Art, for instance, is a new smartphone app that combines synchronized audio and indoor positioning to create an experience driven solely by the visitor's path through the museum.⁵ The Cooper Hewitt in New York now provides visitors with a digital pen that can save items in the collection for later viewing, and which enables other kinds of interaction at certain exhibits.⁶ But even beyond these ambitious projects, interactive digital content — from apps to audio guides — is now commonplace in museums seeking new ways to connect with visitors.

The museum as a laboratory for educational applications of new technology is the background upon which this exploration into augmented reality is situated. *How can augmented reality be used to make the art museum guide more engaging, personalized, and interactive, but still appropriate to the gallery context?*

4 Kenneth Hudson, "The Museum Refuses to Stand Still," *Museum International* 66, no. 1-4 (2014): 136.

5 Rene Chun, "The SFMoMA's App Will Forever Change How You Enjoy Museums," *WIRED*, May 5, 2016, <https://www.wired.com/2016/05/sfmoma-audio-tour-app/>.

6 "Designing the Pen," *Cooper Hewitt*, accessed June 8, 2017, <https://www.cooperhewitt.org/new-experience/designing-pen/>.

WHAT IS AUGMENTED REALITY?

Augmented reality rests on the Reality-Virtuality Continuum established by Milgram et. al. in 1995, as a subset of mixed reality. At one extreme of this continuum are environments that are entirely real — a scene viewed via video, for example — and at the other extreme are environments that are entirely virtual, as are currently encountered when using VR devices like the Oculus Rift. Mixed reality is any environment between these two extremes of reality and virtuality, in which both real and virtual objects are presented together. Augmented reality is a type of mixed reality that is more real than virtual — “augmenting natural feedback to the operator with simulated cues.”⁷

It is especially relevant to think of an AR art museum interface in terms of this continuum between reality and virtuality, considering the special nature of reality in this space — containing the “aura” of original objects. The original object and the real, physical environment are vitally important to the experience of seeing art in a museum, as I have previously discussed. Any experience designed for this space must supplement, not supplant. Thus, in selecting augmented reality, I have purposefully privileged the real world over any virtual supplementation in my implementation of this thesis.

SOUND, AUGMENTED REALITY, AND THE MUSEUM

AR is almost always considered to be primarily, if not exclusively, a visual medium, usually through some kind of transparent layer. Notably, however, the definition provided by Milgram does *not* specifically specify that the “simulated cues” must be visual, even though in almost all practical applications, they are.

As we are considering the museum, perhaps the museum audio guide could be a kind of AR — one that augments with sound rather than with visuals. If we accept this assumption, it is a form of AR that is very appropriate for the context. Voice as a medium makes sense

7 Milgram, Paul, Haruo Takemura, Akira Utsumi, and Fumio Kishino. “Augmented reality: A class of displays on the reality-virtuality continuum.” In *Photonics for industrial applications*, pp. 282-292. International Society for Optics and Photonics, 1995.

in the gallery. It does not interfere with the work visually, and in the case of the audio guide, can be enjoyed privately with earphones. Thinking about an audio experience in the museum as a kind of AR is perhaps more convincing when considering a “smarter” application like SFMoMA’s previously-mentioned audio guide. If that system were to provide its information visually, mapped out into the gallery space as the visitor walked through it, we would certainly consider it to be AR. Instead, however, it responds to and augments the gallery experience using tracks of spoken audio.

This thesis seizes on the idea of sound-based AR and takes it a step further, attempting to fully integrate elements of AR into the museum audio guide to make it more interactive. This has a number of proposed benefits. At present, an audio recording must be reduced to a short segment in order to retain visitors’ attention and not overwhelm them. Enabling users to interact with the device not only opens the possibility of hearing more content (should visitors wish to hear it), but also allows for the presentation of multiple narratives about the work of art beyond a singular curatorial voice.

THE MICROSOFT HOLOLENS

As previously mentioned, I aimed to create a testable prototype in this thesis rather than a concept video. In choosing a prototyping tool, the Microsoft HoloLens stood out as having many desirable features for this project — audio output, robust support for voice recognition, and image recognition (through the Vuforia plug-in), in addition to being wireless and portable.

The HoloLens also includes support for gaze interaction, which presents another exciting opportunity: simply looking at elements of a painting to select them. As a form of interaction, if speech makes sense in the context of a gallery, so, too, does sight.

The HoloLens also has significant drawbacks. At time of writing, it is in developer preview. Although it presents an impressive step forward in wearable AR, it is expensive, cumbersome to wear, and requires instruction to operate properly. In its current form, it is certainly not the type of device one envisions wearing around a gallery. As such, the prototype developed in this project is not a proposal for an immediate, implementable solution. However, the act



of prototyping it with real-world technology grounds this thesis in the capabilities *and* limitations of wearable AR as it currently exists. Simply postulating about this project idea running on ideal future technology prevents discussion of the project and its conclusions in practical terms. At the same time, after using the HoloLens for this project, it is possible to speculate with some degree of confidence about the future of this class of technology — certainly less obtrusive and more immersive, which bodes well for this type of project in the museum space.

The Microsoft HoloLens.

Source: "Microsoft HoloLens: Press Materials," *Microsoft News Center*.
<https://news.microsoft.com/microsoft-hololens-press-materials/>.



The Arnolfini Portrait

Prototyping the experience required that I choose a work of art on which to deploy the AR experience. An ideal subject would be rich with hidden information, and would be the subject of a large amount of scholarship or interesting historical information. With this in mind, I considered Pablo Picasso's *Les Femmes d'Alger*, Édouard Manet's *A Bar at the Folies-Bergère*, and Edvard Munch's *The Dance of Life*. The experience could have been particularly compelling with a colossal work in situ — for instance, Michaelangelo's *Last Judgment* or Raphael's *School of Athens* — but I eliminated these for practical reasons. I also wished to work at scale, which eliminated large canvases such as Théodore Géricault's *Raft of the Medusa*, also for practical reasons.

I ultimately chose *The Arnolfini Portrait*, painted by Jan Van Eyck in 1434. It is dense with symbolism and scholarship, and at 84.5 cm × 62.5 cm (though my copy was slightly smaller), it could easily be kept in the studio. Critical to understanding the portrait is the existence of several conflicting scholarly interpretations, none of which have been definitively proven. I chose three to discuss in this project.

Erwin Panofsky provided the most well-known interpretation in 1934, proposing that the scene depicts the couple's wedding and that the painting itself served as a legal certification of the marriage. Under this interpretation, the gestures are part of a marital oath; the single burning candle in the chandelier is a traditional marriage candle; and the signature turns the portrait into a legal document.⁸ In the years since, many art history students (myself included) were introduced to the painting as the *Arnolfini Wedding Portrait*. Other interpretations have been proposed as well; the idea that the painting is a memorial to the deceased wife, for instance, was proposed by Margaret Koster in 2003.⁹ The fact that the painting is an exquisite and innovative double portrait of a wealthy couple has never been questioned, but the official position of the National Gallery London (where the painting is currently located) is that this simple explanation is the

Opposite page: *The Arnolfini Portrait* by Jan Van Eyck (1434), National Gallery, London.

Source: "The Arnolfini Portrait," *The National Gallery*, <https://www.nationalgallery.org.uk/paintings/jan-van-eyck-the-arnolfini-portrait>.

8 Erwin Panofsky. "Jan van Eyck's Arnolfini Portrait," *The Burlington Magazine for Connoisseurs* 64, no. 372 (1934): 123, 124, 126.

9 Margaret L. Koster, "The Arnolfini Double Portrait: A Simple Solution," *Apollo* 158, no 499 (2003): 6.



correct one.¹⁰ Under each interpretation, different objects take on different meanings. For instance, the dog could be a symbol of marital fidelity, in the wedding theory¹¹; a reference to the dogs depicted on women's tombs, in the memorial theory¹²; or another symbol of wealth, in the "just a portrait" theory.¹³

10 "The Arnolfini Portrait," *The National Gallery*, accessed June 7, 2017, <https://www.nationalgallery.org.uk/paintings/jan-van-eyck-the-arnolfini-portrait>.

11 Panofsky, 125.

12 Koster, 11.

13 Wendy Beckett, *Sister Wendy's Story of Painting*, "A Hero Steps Forth," (1996; Beverly Hills: CBS/Fox Video, 1997), VHS.



Early Prototypes

The first prototypes were centered on a simple idea — if the museum guide could talk to the visitor, what if the visitor could talk back? Ask questions? I created two low-fidelity prototypes to test this idea: a “Wizard of Oz” test (where a human acts as the computer AI — named for the wizard behind the curtain in the eponymous film), and a voice-driven chatbot driven by API.AI. At their core, both used the same simple, two-stage user flow:

1. Painting Introduction



2. Open-Ended Q&A

In the first stage, the computer would introduce the *Arnolfini Portrait* with brief information about the artist, the scene, and the painting’s enigmatic nature (much like an actual museum docent might). The computer would then invite the user to ask questions for which it would attempt to provide appropriate answers. When the design succeeded, it would provide a response to most reasonable, relevant questions without falling back on a default error message (e.g. “I’m sorry, I don’t know anything about that.”).

“WIZARD OF OZ” HEAD-MOUNTED CAMERA TEST

To simulate the gaze cursor and voice interaction in a low-fidelity prototype, I attached a laser bore sighter to a head mounted camera in such a way that the red laser would always stay pointed in the direction of the user’s gaze; the camera would allow me to record what tests user saw and how they used the cursor. For the voice interaction component, I sat behind the user and acted as the voice

Opposite page: The head-mounted 4K camera with a laser bore sighter that simulated the gaze cursor.





Top: The test in progress.

Bottom: Screenshot of the video recorded by the line-of-sight camera. Text prompts are superimposed on the image.

of the AR experience in real time. I printed out a set of pre-written responses to read out loud, as well as a few generic error messages that I would use if the user's question didn't match any of the responses.

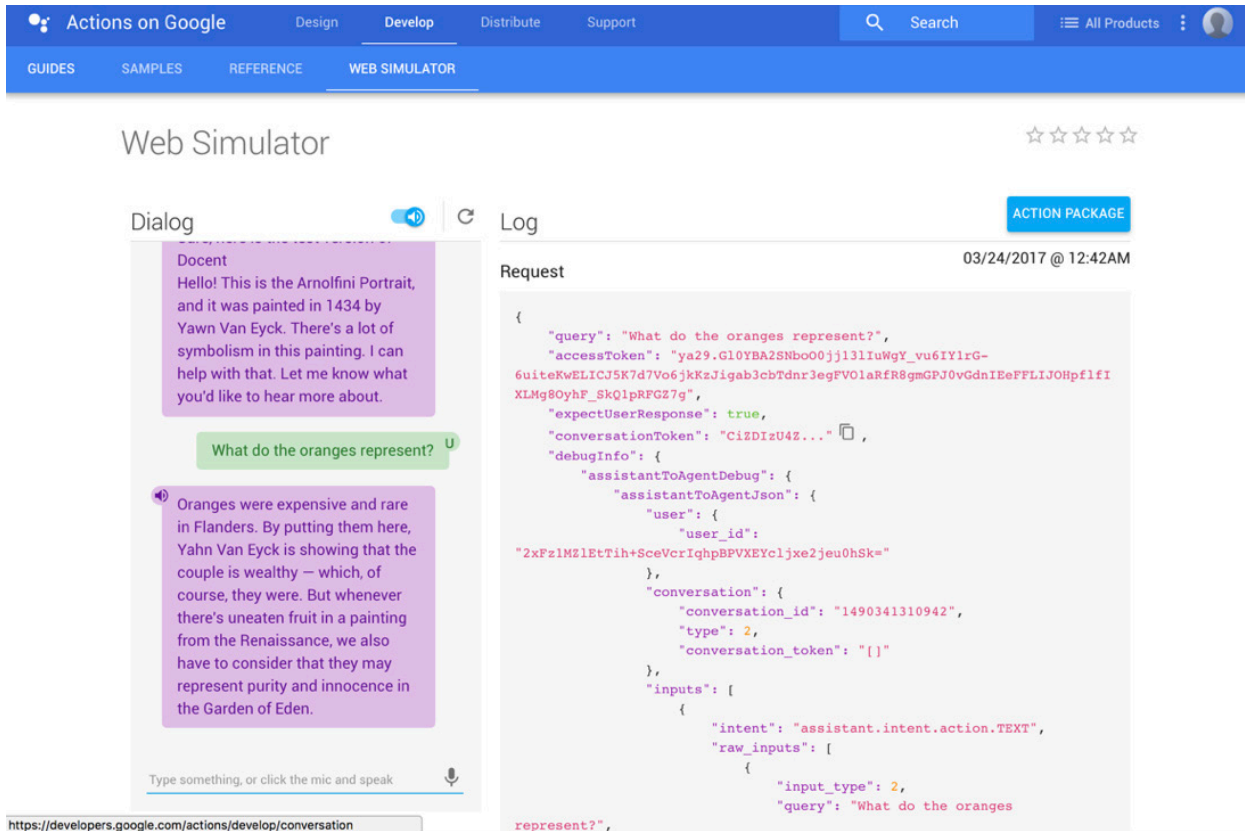
The responses were triggered by the utterances that related to different objects in the painting, such as "bed," "woman," and "dog," and to general attributes like "artist," "theories," and "medium." The user did not need to say the word alone in order for the response to trigger; for instance, he or she could say "Tell me about the dog," or "Why is the dog there?" and get the same response about the dog at the bottom of the painting.

Finally, in order to help guide the user on some of the possible questions he or she could ask, I lightly superimposed keywords onto the painting, which was displayed on a large screen connected to a computer. They would appear if the user said something like "what can I ask?" (which the introductory narration indicated was an available option). As the user moved closer and farther away from the on-screen painting, the words would get more or less specific at three different distances. General words like "theories" or "artist" appeared at a far distance, major objects like "woman" or "man" appeared at a medium distance, and detail words like "reflection" or "rosary" appeared at the closest distance. As the person operating the test, I triggered these manually by pressing keys on the keyboard.

API.AI CHATBOT

API.AI is a platform for conversational user interfaces that can be used to build text-based chatbots and, as of recently, voice-based user experiences for Google Assistant, Amazon Alexa, and Microsoft Cortana. Fundamentally, it interprets natural language inputs, matches them to an "intent", and outputs data — a text or speech response, for instance.

I used API.AI to create "Docent," an entirely speech-based prototype that would allow me to gauge the difficulty and complexity involved in programming different intents that would successfully match and respond to a user's questions about the Arnolfini Portrait. It would also allow me to evaluate the effectiveness of an artificial text-to-



Running simulation of “Docent”, a prototype of a conversational user interface that answered questions related to the *Arnolfini Portrait*. Notice the spelling of Jan Van Eyck as “Yawn Van Eyck,” an attempt to force the artificial voice to pronounce the name correctly.

speech (TTS) voice in this context. I ran the prototype on the Google Actions Web Simulator and on the Google Home hardware device.

OBSERVATIONS AND RESULTS

The API.AI “Docent” chatbot and the head-mounted camera test both yielded critical findings that would affect future iterations of the design.

Lost narratives

When put into an open question/answer mode, users tended to ask questions about individual objects in the painting, e.g. the dog, the window, or the oranges. As a result, when asked about what they learned from the experience, users tended to retain some information about these details — *but they missed overarching narratives, such as the different interpretations of the painting as a wedding or a memorial*. They came away with a fragmented understanding of the work of art. I regarded this as an important

failure, because the painting is best regarded not as a random assortment of objects, but as an overall story to which those objects contribute via symbolic meaning or historical significance. In this particular work, those stories are multiple — however, users were lucky if they understood even one. This was the most important takeaway from these early prototypes.

Accommodating intents

It was not hard to identify intents for basic objects in the painting — for instance, in both tests, I used an intent for Jan Van Eyck’s signature in the center of the composition that would trigger with sentences containing “signature,” “writing,” “autograph,” or “written.” A user could speak a number of utterances in different permutations to trigger the same response, e.g. “What is that signature?,” “Why is that signature on the wall?,” “Tell me about the signature,” etc., and the prototype would respond with a few pertinent pre-written sentences.

However, it quickly became apparent that it would be difficult or impossible to create intents for *every* possible question related to the painting. I recalled two interesting questions that came up in the head-mounted camera test — “does this house still exist today?” and “what is that hanging thing?” (referring to the red hanging fabric on the bed) — examples of the many questions I would have found difficult to anticipate and accommodate.

I also realized that, even with the difficulty I was having in accommodating a large number of intents for this prototype, the objects in the *Arnolfini Portrait* constituted a best-case scenario. As a representational painting, common vocabulary exists for every object depicted in the scene: mirror, window, dress, dog, etc. If I was having difficulty with this Jan Van Eyck painting, what might it be like with a surrealist painting by Salvador Dalí, in which objects cannot easily be identified verbally? Or an abstract expressionism painting by Jackson Pollock? In my prototyped system, users could be handicapped by their inability to name an object.

Synthesized vs. human voices

The API.AI “Docent” test made it immediately apparent that a synthesized text-to-speech voice would not be appropriate in this context. While the sound produced by the concatenated speech of

TTS might be acceptable for purely functional applications (reading out the weather or the time, for instance), telling stories about art would require a human voice in order to be engaging.

Errors

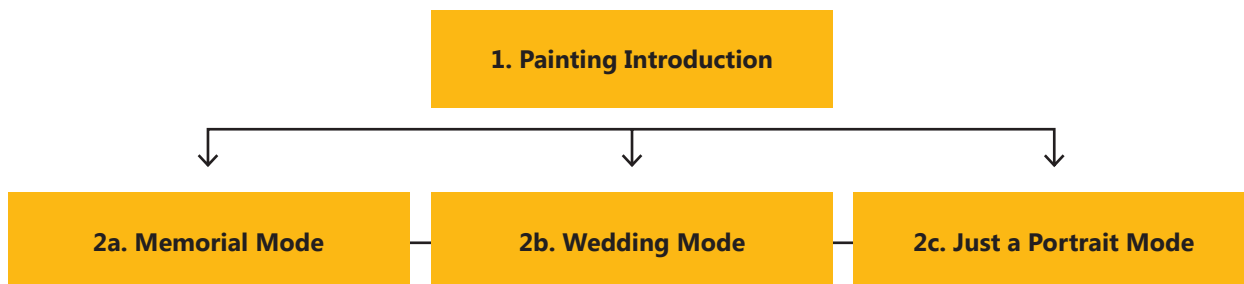
Errors are seldom an intended or desirable outcome in an experience such as this one. But when viewing a painting and asking questions about it, generic error messages ("Sorry, I don't know anything about that") felt especially out of place. Even the worst museum docent could come up with a better, more human-sounding answer. If the system instructs the user to ask any question he or she wishes about the painting, but then returns several errors when unexpected questions are asked, why not just tell the user what the system *does* know?

Several of these conclusions lead me to the conclusion that *the open-ended questioning system was not the best approach*. In the next prototype, I decided to make the experience more guided.



HoloLens Prototype

The prototype on the HoloLens was developed with Unity and C#, in collaboration with a developer partner. As a result of previous tests, I changed the underlying structure to focus on the multiple narratives that existed surrounding the *Arnolfini Portrait*, and eliminated the open-ended verbal questioning component in favor of gaze interaction.



After the global narrator introduced the painting (including the fact that multiple interpretations are debated), the viewer was prompted to choose an interpretation by saying either “It’s a memorial,” “It’s a wedding,” or “It’s just a portrait.” These options also appeared as holograms on the right side of the painting. The HoloLens would recognize the phrase, and the viewer entered the corresponding mode.

In each mode, a narrator introduced the general rationale for the painting being interpreted as a wedding, a memorial, or just a portrait. The gaze cursor then activated in the center of the viewer’s line of sight, and the viewer was able to search the painting for hotspots; the circular cursor would get larger if there was information available. The viewer could then select the hotspot either by using a separate handheld clicker, using the HoloLens’s “air tap” gesture, or saying “select.”¹⁴ The viewer could switch theories at any time simply by saying a different key phrase.

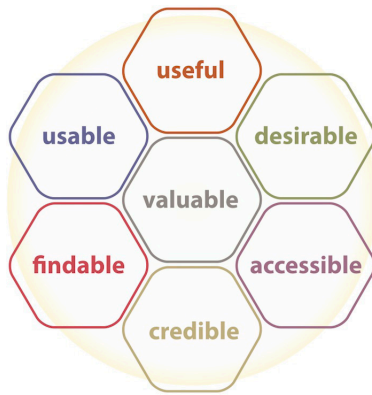
As narrative took a primary role in this prototype, the style of narration also became important to consider.

¹⁴ In testing, the handheld clicker was clearly the easiest to use.



THE AESTHETICS OF VOICE

The earlier prototypes made it clear, among other things, that a TTS voice would not be acceptable for this application. Thus, for the HoloLens prototype, I turned my attention toward human voice talent for each of the three theories and for the global narration — and in doing so, I reflected on the way that qualities of voice can be consciously considered in a piece of interaction design.



Peter Morville's User Experience Honeycomb (2004).

Source: Peter Morville, "User Experience Honeycomb," *Intertwined*, October 11, 2016, <http://intertwined.org/user-experience-honeycomb/>.

When designing a voice interface, most principles of general UX design remain applicable.¹⁵ As an example, consider Peter Morville's "User Experience Honeycomb," which he created in 2004.¹⁶ It is not difficult to imagine that while a website or app should be *useful*, *desirable*, *accessible*, *credible*, etc., then a voice interface should be the same. How this is accomplished requires more specific thinking in some cases, but it seems safe to say that the goals and high-level principles of UX design can be translated to the realm of voice interfaces in a fairly straightforward fashion. Indeed, much research regarding the design of conversational UI centers around principles of UX, and the role of UX designers in voice interaction has been well-established.

Of course, when designing a user interface, one traditionally considers not only the UX design, but also principles of visual design — color, composition, etc. And by contrast, the principles of visual design are very much subverted by the *non-visual* nature of voice UI design — excepting, of course, any visual interface that supports the voice interface. There is no color or composition inherent in a voice UI like there is in a GUI. But the design of a voice persona can be compared to a visual style guide in its effect on the user's perception of the experience.¹⁷ We know that vocal qualities — gender and emotion, for instance — have an effect on the experience of the VUI. Clifford Nass and Scott Brave provide many examples in their 2005 book *Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship*. If the experience of a VUI is heard, rather than seen, there is still an "aesthetic" quality to voice that must be understood — and once it is understood, it must be designed.

¹⁵ Laura Klein, *Design for Voice Interfaces*, Sebastopol: O'Reilly Media: 6.

¹⁶ Peter Morville, "User Experience Design," *Semantic Studios*, June 12, 2004, http://semanticstudios.com/user_experience_design/.

¹⁷ Klein, 8.

It is beyond the scope of this thesis to fully explore the ways in which vocal qualities can be manipulated to specific effect in a VUI design. But there are connections between speech and visuals that are useful to consider. For example, it has long since been established that the sounds of speech can be non-arbitrarily associated with visual qualities, independent of language and culture. In Wolfgang Köhler's 1929 experiment, when participants are shown two simple shapes, one spiked and the other rounded, and asked to name one "*takete*" and the other "*baluma*," they will almost always name the rounded shape "*baluma*" and the spiked shape "*takete*". In 2001, this effect was reproduced with 95% accuracy by Vilayanur Ramachandran and Edward Hubbard, using the names "*kiki*" and "*bouba*".¹⁸ The fact that this experiment has been reproduced not only across languages, but also in children under one year of age, suggests that this phenomenon ("sound symbolism") may be innate.¹⁹ Other studies have attempted to expand this phenomenon to other sounds and shapes; for instance, shapes with sharp angles tend to be matched with sounds that are louder and articulated with less rounding of the lips.²⁰

For a designer's purposes, however, a voice must be described in terms acknowledging that it is more than a collection of articulated consonants and vowels, just as gestaltism would suggest that an interface is more than a collection of colors, shapes, and words.

In researching how a visual designer might evaluate and choose a voice for a certain interface, I spoke with voice actors and casting directors — people who are intimately familiar with the qualities of voice and how they fit (or do not fit) a certain context. In my conversations, I observed that many of the words used to describe voices were identical to words designers use to describe visuals. A voice could be "warm," it could have "texture," it could be "rough" or "smooth," "feminine" or "masculine," "dark" or "light." It was tempting to imagine matching a piece of design, or a GUI, to a voice that uses the same descriptors. Additionally, it was clear from speaking with

18 Vilayanur S. Ramachandran and Edward M. Hubbard, "Synaesthesia — a window into perception, thought and language," *Journal of Consciousness Studies* 8, no. 12 (2001): 19.

19 Annette D'Onofrio, "Phonetic Detail and Dimensionality in Sound-shape Correspondences: Refining the *Bouba-Kiki* Paradigm," *Language and Speech* 57, no. 3 (2013): 368.

20 D'Onofrio, 369.



TO CHANGE THEORIES, SAY...

It's a memorial.

It's just a portrait.

It's a wedding.

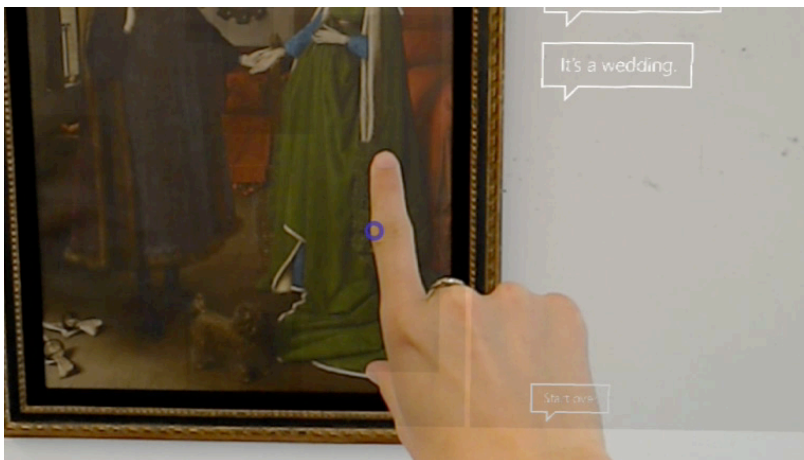
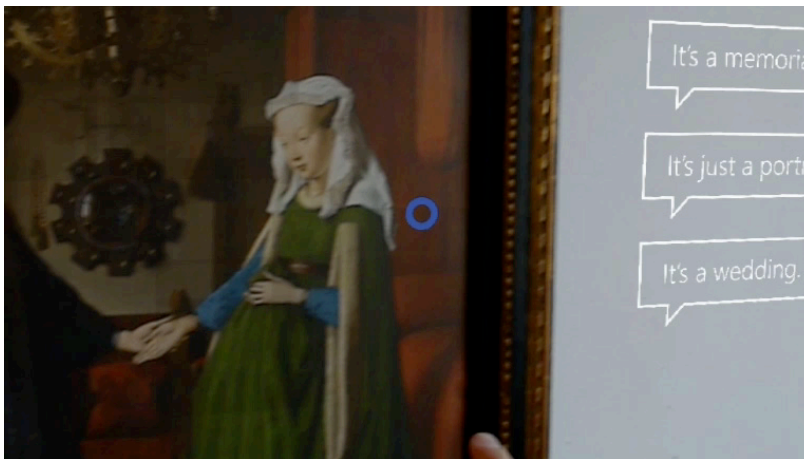
Mockup of the HoloLens prototype experience with gaze cursor (blue circle), holographic prompts on the right, and selectable hotspots (which were invisible in the actual prototype).

these professionals that vocal qualities played an integral role in the perception of whatever context the voice appeared in, potentially functioning as a kind of brand identity. Voice actress Jessica Bogart, who had done work for Subaru, memorably demonstrated for me on a phone call the subtle variations she could apply while saying “The all-new Subaru Forester” to create different effects.

It was important to me that each theory — wedding, memorial, and just a portrait — was experientially distinct from the others. Variety would create more interest, I felt, and it would be easier to distinguish between modes if the audio experience noticeably differed between theories. I also envisioned a kind of debate between the different advocates for each theory — scholars trying to convince the visitor about his or her theory of choice. For these reasons, I chose three different actors who could lend each theory the vocal equivalent of the correct “look



Video stills from the prototype recorded on the HoloLens, depicting the holographic cues to the right of the painting, the gaze cursor, and the "air tap" gesture. Note that the darker box surrounding the painting was not visible in the actual prototype, only in the video recording.



and feel," based on the topic. For the wedding theory, I chose a female voice with a smooth, elegant tone. For the memorial theory, I chose an older male voice that was deeper with more texture. And for the "it's just a portrait" theory, I chose a more energetic, confident, clear female voice.

OBSERVATIONS AND RESULTS

Communicating the narratives

In testing, the multiple narratives successfully came through in this prototype. Because they constituted the core of the experience, all testers were able to identify them afterward and understood the controversial nature of the painting. I received positive feedback on the selection of music and voice to differentiate modes from one another. Some users responded that they were better able to recall which facts corresponded with one another because they were delivered by different voices.

User interest

User tolerance for listening to the experience varied widely, and depended on prior interest and disposition. Some users listened to a few recorded clips before declaring that they were finished, and even expressed a desire to interrupt the voice. Others listened to every single clip, returned to theories they had previously explored, and searched around using the gaze cursor for more information even when they had already exhausted the options. For these highly-interested users, the experience proved much more engaging and informative than the standard museum audio guide, if only for the fact that they received far more information and learned about multiple perspectives, not just one. It is unknown how this experience could better engage the users who were less interested, though I speculate that this audio-based learning style might have simply not been a good fit for these users.

Gaze cursor design

The appearance of the gaze cursor was the most problematic part of the design. It took the form of a circle that became larger when it was over an active hotspot in the painting. However, most users found it difficult to perceive this scale change until I intervened and told them about it.

Problems with the hardware

The creation, deployment, and testing of this prototype was predictably made more difficult by the in-production nature of the HoloLens. While the prototype was fully functional, there were

elements of it — particularly around tracking — that could have been smoother.

Most significantly, however, wearing and using the HoloLens proved unintuitive to viewers, and they required assistance and a brief tutorial. (Of course, on first activation, the HoloLens provides such a tutorial to the user.) This supported my strong early assumption that the HoloLens, while an impressive and highly promising feat of technology, would not currently be appropriate for broad deployment in an art museum.



Final Design

Informed by the previous functioning prototypes, I created a video mockup to showcase the final design vision for this project. This was partly out of necessity, as high quality video could not be recorded from the HoloLens device.

The experience shown in the video very closely resembles the HoloLens prototype, with a number of key differences:

- The global narrator was improved and changed from an American male to a British male. Slight modifications to the script were made to improve clarity.
- A short narrated instruction plays when the gaze cursor first appears. In the HoloLens prototype, many users asked for help at this point in the test.
- The gaze cursor was redesigned to be more noticeable and more precise; it changed from a circle that changed sizes to a cross that changed colors. Additionally, when the gaze cursor first appears, it calls attention to itself with a brief animation.
- When an object is selected, a brief glow is superimposed on the painting to better identify the object.
- The holographic cues for changing theories were often mistook for clickable buttons in the HoloLens prototype. In the final design, they've been changed to text in quotation marks.
- A brief glow shows which theory was selected after the viewer chooses one.
- Several narration clips were shortened for brevity in the video.

Opposite page: Final design of the gaze cursor animation.





0:07

The viewer puts the HoloLens device on while standing in front of the painting.



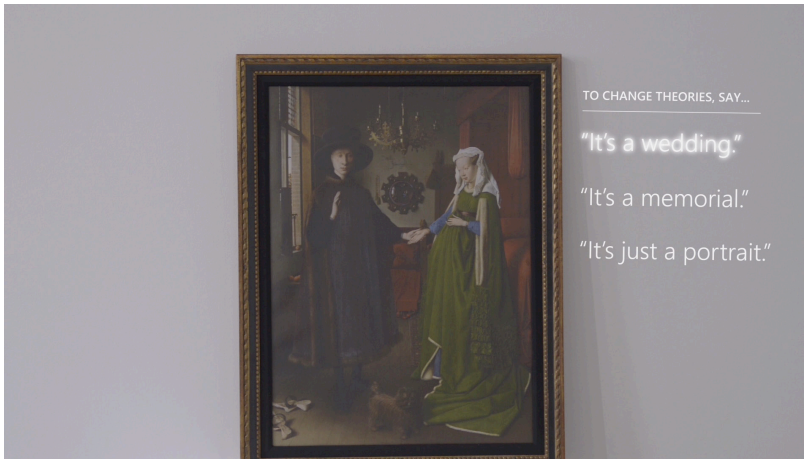
0:19

A sound effect indicates the painting is identified, and the global narrator introduces the painting with basic information.



0:35

As the narrator asks the viewer to choose a theory of interpretation, the options appear as holograms to the right of the painting.



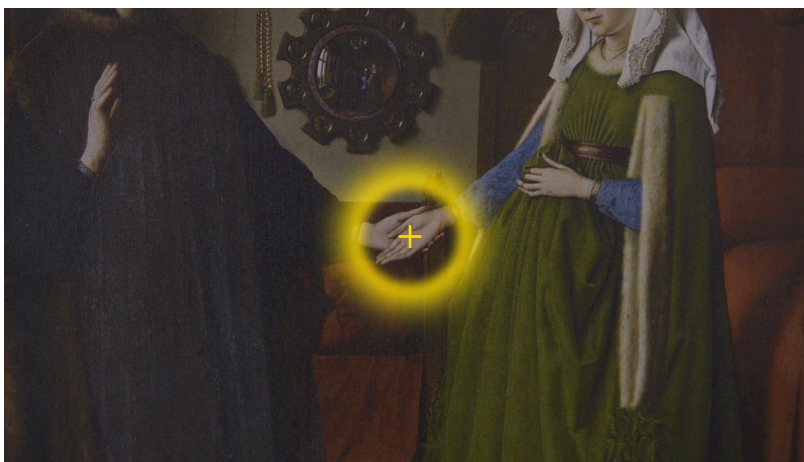
0:37

When the viewer says "It's a wedding," the corresponding choice flashes and a sound effect plays to indicate that the key phrase was heard. The global narrator notes that the viewer can change theories again anytime.



0:55

A new narrator introduces the painting as a depiction of the couple's wedding, as music begins playing in the background.



1:00

The first time the gaze cursor appears in a mode, a burst-like animation plays to call attention to it. This time only, the global narrator instructs the viewer to "look around using the gaze cursor, and tap to learn more."



1:09

The viewer is holding a clicker device, which he uses in conjunction with the gaze cursor to select the chandelier.



1:11

A momentary glow highlights the chandelier once it has been clicked. The gaze cursor disappears as the narrator begins to describe the significance of the chandelier in this theory.



1:26

The gaze cursor reappears when the narrator finishes. It turns yellow when there is information underneath it, and gray when there is not.



1:32

The viewer says "It's a memorial" to switch theories. The music stops, and after the global narrator confirms the choice, new music begins.



1:51

A different narrator introduces the painting as a memorial. The gaze cursor reappears when he is finished, but blue this time.



1:53

The cursor turns blue over the chandelier, and the viewer selects it again. However, in the memorial theory, the chandelier has different significance.



2:09

The viewer changes theories a final time, saying "It's just a portrait." The holographic text glows in response, a sound effect plays, and the global narrator confirms the selection.



2:29

A third narrator introduces the painting simply as a double portrait of a wealthy couple.



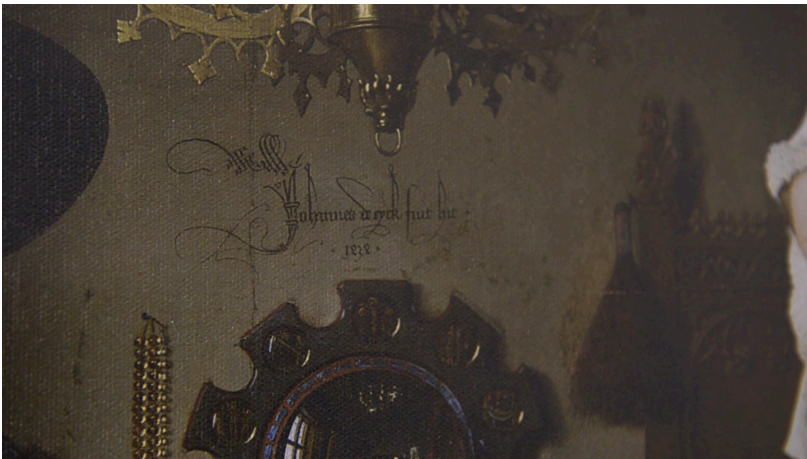
2:38

The viewer uses the gaze cursor to select the signature in the center of the painting, which flashes green in this mode.



2:46

The narrator tells the viewer about the significance of the signature. The viewer leans in for a closer look.



2:53

Because the system doesn't require the viewer to stand in a particular location relative to the painting, stepping closer and seeing the signature in detail is possible.

Next Steps

The direct scope of this thesis is limited to a single work of art and three scholarly perspectives on its interpretation. However, the restrained approach taken in this design nevertheless has relevance for a number of other contexts and permutations.

Other paintings

The *Arnolfini Portrait* is famous for its numerous possible interpretations, as has been previously discussed; these interpretations served as the conceptual foundation for the HoloLens prototype and the final design. However, even paintings that are not the subject of such vigorous, well-known debate can benefit from the exposition of multiple narratives.

For example, many works of art can be studied and explained through different methodologies of art history, with different results. *Formalism*, for instance, might call attention to the striking contrast and flattened picture plane in Manet's *Olympia*; *feminism* might focus on the way the woman actively challenges the male viewer; and *Marxism* might reflect on the painting's "spectacle of sexual capitalism" made confrontational.²¹



Édouard Manet, *Olympia* (1863).

Source: "Olympia," *Google Arts & Culture*, accessed June 9, 2017, <https://www.google.com/culturalinstitute/beta/asset/ywFEI4rxgCSO1Q>.

Other voices

Implicit in the multiple narrative approach is a challenge to the singular, authoritative voice of the museum curator. Were this guide to be installed in the National Gallery of Art in London, it would already undermine the museum's singular assertion that the portrait is not a depiction of their wedding. I see this design as a part of what Eilean Hooper-Greenhill describes in *The Educational Role of the Museum* as, "the cultural move from modernism, with its monolithic metanarratives... to postmodernism with its more fragmented and diverse approach."²² At times, the user's voice joins the curator's to create a shared meaning. But the curator's voice can also be

²¹ Laurie Schneider Adams, *The Methodologies of Art*, Boulder: Westview Press, 1996, 17, 77, 80.

²² Eilean Hooper-Greenhill, preface to *The Educational Role of the Museum*, 2nd ed., London: Routledge, 1999, xi.

challenged directly. In 2005, *Slate* began their “unauthorized” audio tours, including tours of the Met and of the National Mall, promising listeners “the commentary museums don’t want you to hear,” in podcast form.²³ What if, instead of different scholarly theories of interpretation, this design offered different “unauthorized” opinions and perspectives on the work of art? Or if it communicated the personal significance that this work of art has held for any number of lay viewers?

Contextual consideration

This design does not take into consideration the painting’s physical location; indeed, over the course of the design process, I developed and tested it in a number of available classrooms, studios, and galleries out of necessity, with no effect on how the design functioned.

Of course, real works of art are situated in contexts that are important to take into account. In this regard, the clearest next step would be to refer the viewer to other paintings that are installed in the immediate surroundings. In the National Gallery, the *Arnolfini Portrait* hangs directly to the right of Van Eyck’s *Portrait of a Man*, painted one year earlier. It is speculated that one of the two figures reflected in the mirror in the *Arnolfini Portrait* is Jan Van Eyck himself; the *Portrait of a Man*, too, is often thought to be a self portrait. The ability to see this comparison directly certainly captures one of the great advantages of viewing art in person.

23 Andy Bowers, “Introducing SlateAudio Tours,” *Slate*, July 26, 2005, http://www.slate.com/articles/news_and_politics/slate_audio_tours/2005/07/introducing_slateaudio_tours.html.

Acknowledgments

The advice of my committee — Axel Roesler, Karen Cheng, and Christopher Ozubko — was instrumental in shaping this effort.

I am appreciative of the many professionals who lent their talents, knowledge, and time to this project: Steve Bunn, Jessica Bogart, Wendy K Gray, Robert Branom, Kim Butler Wingfield, Yana Sakellion, Sumedha Kshirsagar, Shweta Grampurohit, Richard Ladner, Kelly Moscinski, Sarah Takahashi, the staff of Allison & Ross Fine Art Services, Emily Sappington, Cody Irizarry, and Darren Tibbits.

The HoloLens prototype would simply have not been possible without the very talented Bryan Oltman, and I am grateful that he agreed to create it with me.

I am also very grateful for the support given to me by my sister, Stephanie, and my parents, Steve and Tammy.



Clockwise from upper left: Scott Tsukamaki, Tate Strickland, Sarah Reitz, and Richelle Dumond.

Finally, this project — and my tenure as a graduate student — was fundamentally defined by the support, encouragement, feedback, and occasional commiseration from the other students in my cohort, Richelle Dumond, Sarah Reitz, and Scott Tsukamaki. Going through this experience with such a diverse and talented group of designers has forever broadened my understanding of design and its ubiquity in our world.

Bibliography

- Adams, Laurie Schneider. *The Methodologies of Art: An Introduction*. Boulder: Westview Press, 1996.
- "The Arnolfini Portrait. *The National Gallery*. Accessed June 7, 2017, <https://www.nationalgallery.org.uk/paintings/jan-van-eyck-the-arnolfini-portrait>.
- "ArtLens App." *The Cleveland Museum of Art*. Accessed June 8, 2017. <http://www.clevelandart.org/gallery-one/artlens>.
- Beckett, Wendy. *Sister Wendy's Story of Painting*. "A Hero Steps Forth." 1996. Beverly Hills: CBS/Fox Video, 1997. VHS.
- Benjamin, Walter. "The Work of Art in the Age of Mechanical Reproduction." In *Literary Theory: An Anthology*, 2nd ed., edited by Julie Rivkin and Michael Ryan, 1235–1241. Maiden: Blackwell Publishing, 2004.
- Bowers, Andy. "Introducing SlateAudio Tours." *Slate*. July 26, 2005. http://www.slate.com/articles/news_and_politics/slate_audio_tours/2005/07/introducing_slateaudio_tours.html.
- Chaundy, Bob. "Faces of the Week." *BBC News*. September 29, 2006. Accessed June 7, 2017. http://news.bbc.co.uk/2/hi/uk_news/magazine/5392000.stm.
- Chun, Rene. "The SFMoMA's App Will Forever Change How You Enjoy Museums." *WIRED*. May 5, 2016. <https://www.wired.com/2016/05/sfmoma-audio-tour-app/>.
- "Designing the Pen." *Cooper Hewitt*. Accessed June 8, 2017. <https://www.cooperhewitt.org/new-experience/designing-pen/>.
- D'Onofrio, Annette. "Phonetic Detail and Dimensionality in Sound-shape Correspondences: Refining the Buba-Kiki Paradigm." *Language and Speech* 57, no. 3 (2013): 367–393.
- Hooper-Greenhill, Eilean. Preface to *The Educational Role of the Museum*, 2nd ed., London: Routledge, 1999, x–xiii.
- Hudson, Kenneth. "The Museum Refuses to Stand Still." *Museum International* 66, no. 1-4 (2014): 136–143.
- Klein, Laura. *Design for Voice Interfaces*. Sebastopol: O'Reilly Media, 2016.
- Koster, Margaret L. "The Arnolfini Double Portrait: A Simple Solution." *Apollo* 158, no. 499 (2003): 3–14.
- Morville, Peter. "User Experience Design." *Semantic Studios*. June 12, 2004. http://semanticstudios.com/user_experience_design/.
- Nass, Clifford and Scott Brave. *Wired for Speech: How Voice Activates and Advances the Human-Computer Relationship*. Cambridge: MIT Press, 2005.
- Panofsky, Erwin. "Jan van Eyck's Arnolfini Portrait." *The Burlington Magazine for Connoisseurs* 64, no. 372 (1934): 117–127.
- Ramachandran, Vilayanur S., and Edward M. Hubbard. "Synaesthesia — a window into perception, thought and language." *Journal of Consciousness Studies* 8, no. 12 (2001): 3–34.

Appendix: Narration Scripts

Global narrator

You're looking at a portrait of Giovanni Arnolfini, an Italian merchant, standing with his wife in their lavishly-furnished home in Bruges. It's called the Arnolfini Portrait, and it was painted by Netherlandish painter Jan Van Eyck in 1434. But beyond these basic details, it's hard to know what exactly we're seeing here. Does it depict the couple's wedding? Is it a memorial to the woman on the right? Or is it just a portrait that they commissioned as a conversation piece for their home? Everything you see can be interpreted in different ways, depending on what you believe the painting represents.

Try out a few of the theories for yourself. To get started, tell me whether you'd like to explore it as a wedding, a memorial, or just a portrait.

—

OK. Let's explore this as a wedding portrait. You can try another theory anytime.

—

OK. Let's explore this simply as a portrait of a wealthy couple. You can try another theory anytime.

—

OK. Let's explore this as a memorial portrait. You can try another theory anytime.

—

Look around with the gaze cursor, and tap to learn more.

Wedding narrator

The theory that this painting depicts a wedding has been dominant for the last century — so much so, that it's often called the *Arnolfini Wedding Portrait*. What's interesting is that not only does it show the marriage of the Arnolfini couple, but the painting itself provides legal proof that it happened. The charm of this painting lies in the way it encodes the evidence for this theory into the objects furnishing the beautiful interior of their home.

A medieval spectator would have been keen to read symbols in a painting like this, and would have recognized this dog as a symbol of fidelity between the bride and groom.

There's obviously no practical purpose for this one candle burning in the daytime. But a Flemish wedding at the time wouldn't have been complete without a traditional marriage candle symbolizing the all-seeing eye of God — which is why it's been included here.

A Catholic marriage of the time could take place without a priest, but it required certain actions and words. The raised hand seen here was one of them, part of a marital oath. The traditional "joining of hands" also formed part of that oath, and Van Eyck has made sure to depict it prominently as well.

The scene outside shows that this is a spring wedding. The cherry tree just outside the window serves as a symbol of love and innocence, or possibly of fertility.

Legal cases from the Middle Ages are full of disputed marriages — even cases where someone claimed to be married to a person who denied it. That's why witnesses were so important. Reflected in the

mirror are the two witnesses to this marriage — an unknown woman, and the artist himself. To remove any further doubt, he signed the portrait in the middle like an affidavit: “Jan Van Eyck was here.”

Just a portrait narrator

We can speculate otherwise, but there’s no definitive proof that this painting is anything but a beautiful double portrait of a wealthy couple in Bruges. But that doesn’t make it any less interesting. Imagine the Arnolfini couple showing off this portrait to guests in their home — perhaps unveiling it from behind a curtain. It’s a conversation piece — a tour de force of innovative techniques that would have dazzled and amazed contemporary viewers. It would have invited them to look and wonder at all of the fine detail before them, just as it invites you to do the same.

—

It’s easy to take this for granted, but you’re looking at the first known portrait of its kind in Western art. People had never seen anything like it! Standing portraits of non-religious, non-royal people were unheard of at the time. You’d typically see just their upper body in profile.

—

Is she pregnant, or not? She may appear to be an expectant mother, but that’s actually just the latest fashion of the time — a full-skirted dress with fabric cascading down the front, which she gathers up at her waist.

—

With this mirror, Van Eyck challenges the limitations of painting itself. You can’t walk around the scene in a painting, like you can with a sculpture. But Van Eyck finds a clever way to show us the whole room in this distorted reflection. Suddenly, this room is fully-realized space. We can even see the other side of the couple standing in the middle. The mirror impressed viewers then, just as it does now, and it remains one of the most famous items in all of art history.

—

Imagine that you're wealthy enough to hire the greatest living painter to paint your portrait, in your house. The Arnolfini couple was, and they wanted people to know. So many objects in this painting are there to show off the wealth of the couple, and this signature, proving the presence of this famous artist, is the centerpiece — "Jan Van Eyck was here." Any visitor would have been impressed.

Jan Van Eyck was a master of the new oil medium, which allowed him to do things in painting that no one had ever seen before. Notice, for example, the detail in these glass rosary beads. Oil is translucent, and he used it to recreate the light and color of this interior with stunning accuracy that continues to amaze viewers today.

Today, an orange is nothing special. But in Bruges in 1434, oranges would have been a very expensive import from the south. Yet here they are — another wink to the viewer that signifies just how rich this couple is.

Memorial narrator

Imagine a grief-stricken Giovanni Arnolfini, the man on the left, commissioning this painting to memorialize his wife Costanza, who has just passed away at a young age. In the past, Costanza been dismissed as the identity of the woman because she died before this was painted. And so it was assumed that this was, perhaps, a second wife. But close examination of the symbols in this painting provides evidence for a new theory, that the late Costanza is the woman in this painting — an unusual portrait of both living and deceased.

The colors of their clothing are crucial. At the time, green symbolized young love, blue symbolized faithfulness — and both are present on Costanza's dress. But more interesting is the black fur that Giovanni is wearing. It's not so unusual... until you consider that in 1434, black was not yet popular in clothing, except in mourning.

It was common on women's tombs of the period to have a full-length representation of the deceased, with a dog at her feet — similar to what you're looking at here. Perhaps the dog was to accompany the dead in eternity.

Why light a single candle in the chandelier in the middle of the day? It's one of this painting's biggest mysteries, but the memorial theory solves it. Look closely, and you'll see a burnt out candle on the right side. The lit candle is on the side of the living Giovanni. The burnt out candle, on the side of the deceased Costanza.

Look at the tiny scenes painted around the mirror. They're all scenes from the life of Christ. But their placement is no accident. All of the scenes on the left — Giovanni's side — are scenes depicting a living Christ. And the ones on Costanza's side? All scenes from after Christ's death.

If we use infrared light to see Van Eyck's original plans for the work, we can see that this gesture was originally more firm. Perhaps Costanza died while the painting was being completed, prompting a different sort of gesture. In the original plan, Giovanni had a better grip on his wife's hand. Now it seems to "slip through his fingers, as perhaps did Costanza herself."²⁴

