


**Please cite the Published Version**

Courneya, Marsha, Jackson, David  and McKinley, Roger (2022) Talking with a Virtual Human: Transdisciplinary Design Considerations for Speech-based Interaction in the Audience with a Hero Project. Leonardo Electronic Almanac, 23 (1). ISSN 1071-4391

**Publisher:** Leonardo/ International Society for the Arts, Sciences and Technology

**Version:** Accepted Version

**Downloaded from:** <https://e-space.mmu.ac.uk/629396/>

**Additional Information:** This is an Author Accepted Manuscript of an article published in the Leonardo Electronic Almanac by Leonardo/ International Society for the Arts, Sciences and Technology.

**Enquiries:**

If you have questions about this document, contact [openresearch@mmu.ac.uk](mailto:openresearch@mmu.ac.uk). Please include the URL of the record in e-space. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from <https://www.mmu.ac.uk/library/using-the-library/policies-and-guidelines>)

## Talking with a virtual human: transdisciplinary design considerations for speech-based interaction in the Audience with a Hero project

### Address:

Marsha Courneya

Senior Research Assistant, School of Digital Arts, Manchester Metropolitan University, UK

Email: [m.courneya@mmu.ac.uk](mailto:m.courneya@mmu.ac.uk)

Google Scholar: \*

ORCID: <https://orcid.org/0000-0002-2927-3606>

Twitter: <https://twitter.com/marshacourneya>

Dr David Jackson

Lecturer in Digital Visualisation, School of Digital Arts, Manchester Metropolitan University, UK

Email: [d.j.jackson@mmu.ac.uk](mailto:d.j.jackson@mmu.ac.uk)

Google Scholar:

<https://scholar.google.com/citations?user=wAw6uNAAAAAJ&hl=en&citsig=AMD79opQjnMtDUxN4xmdn3pNMiQAbC48CA>

ORCID: <https://orcid.org/0000-0002-7985-2104>

Roger McKinley

Research Associate, School of Digital Arts, Manchester Metropolitan University, UK

Email: [r.mckinley@mmu.ac.uk](mailto:r.mckinley@mmu.ac.uk)

Google Scholar:

[https://scholar.google.com/citations?view\\_op=list\\_works&hl=en&user=g7yLgdUAAAAJ](https://scholar.google.com/citations?view_op=list_works&hl=en&user=g7yLgdUAAAAJ)

ORCID: 0000-0002-0501-5653

Twitter: @RogerMcKinley1

### Abstract

Exchanges with digital assistants, such as Apple's Siri, have a functional aesthetic related to asking informational questions, such as for weather updates. However, when these human-machine conversations include recorded testimony by real-life people, the identity of our conversational partner changes our relationship with these talking machines. Such complexities require transdisciplinary perspectives to identify the conversational dynamics at play. The Innovate UK funded *Audience with a Hero* research project undertaken by Forever Holdings LLC, Bright White Ltd. and Manchester Metropolitan University sought to allow audiences to speak with a celebrity figure in a virtual reality environment. Research from that project presented here provides insights into the psychological, sociological, narratological and linguistic design considerations that arose during development. It finds that when we reposition the speaking and listening computer as a virtual human celebrity, our language is constrained by many of the anxieties of a real-life encounter. Existing personal relationships with the subject lend a sense of *aliveness* to the conversational agent as we enter the exchange. Once in conversation, a sense of fidelity in both human prosody and convincing instances of self-

disclosure help maintain the necessary social presence required to conduct a worthwhile conversation with a machine.

**Keywords:** Machine learning, parasocial relationships, art, AI, prosody

## **Heading: Introduction**

[INDENT QUOTE]

*“I am sitting in a room different from the one you are in now.”* [1]

[END QUOTE]

Our research and insights in the following essay are based on the *Audience with a Hero* project, still in progress at time of writing. The project was a partnership between researchers at Manchester Metropolitan University (UK) and three UK-based partners: Forever Holdings LLC, Bright White ltd. and Pollen Studio. It was funded by the Creative Industries support body, Innovate UK, as part of its Audiences of the Future call. The project combines research into conversational narrative content design with new film-based production processes in virtual and extended reality to allow a fan to meet and interact with their hero in a virtual environment. One of the purposes of public funding for the project has been to produce new knowledge on conversational agent-based experience design for the benefit of creative industries in the UK and globally, which this paper seeks in part to provide.

Creating a conversational experience that conveys more than just information necessitates a transdisciplinary design approach. Challenges more familiar in creative media production arose along with those already explored in the field of HCI. [2] Without integrated disciplinary knowledge, the creative and technical teams collaborating on the project required new research input to inform human-machine conversational interactions. Following a user-centered rather than discipline-based approach, [3] transdisciplinary design researchers are immersed across relevant disciplines, informing solutions to emergent and more complex design problems where existing disciplinary knowledge is not yet established. [4] The paper documents the findings from transdisciplinary research into the user’s journey, both before and during the conversation.

The experience was due to develop knowledge from interactions with audiences both at SXSW in March 2020 and through the launch of a prototype experience in Autumn 2020. Covid-19 has delayed the rollout of an out-of-home virtual reality experience. As a result, in this paper we describe the user journey identified through the design-informing transdisciplinary research and novel aspects that provide new insights into the production of language models with a living media figure as an interactive subject. The paper begins by describing the lead-up to a dialogue with a hero in VR, using AI as a tool for enabling a conversation; many elements in the journey

before the actual conversation with the digitized hero influence it. Once inside the experience, they adapt to having an exchange with a virtual digital persona, the AI who mediates the experience, as well as the hero. The user for this journey is described in the text as ‘we.’ In this context, the collective pronoun is meant to refer to us the authors, academics from a UK university, as participants in the imagined experience. The approach continues to be relevant beyond the normal design cycle because the delivery of the experience itself has been postponed. We have taken a subjective approach to presenting the research to reflect the experiential nature of the experience. [5]

## **Heading: 1. The Morning of the Event**

### **Subheading: Reflections on the ubiquity of voice-based interaction**

The digital assistant uses our name when it wakes us. It also states the time, appries us of the weather outside and asks whether we would like to know our schedule. *Yes, we say. Please, we add.* A female-sounding voice responds:

*“At 12.30 this afternoon you have your Audience with a Hero appointment. You have no other appointments today.”*

The human voice is one of the fundamental tools at our disposal for interacting with the world and others in it. Through synthesis of the voice, computers can provide naturalistic interaction, learning from what we say both through algorithmic processes and through hundreds of thousands of human hours spent listening and transcribing recordings from our devices to improve its software. [6] The *Audience with a Hero* project is a novel articulation of the very same speech-recognition technology in our digital assistant.

The experience we are going to have was created by the team that developed the *Forever Project* (2017), a mixed reality experience commissioned by the National Holocaust Centre in the UK which began in 2013. That project was designed to safeguard the testimonies of Jewish Holocaust survivors and allow audiences to witness and ask questions of the survivors in perpetuity. A paper by Ma, Coward and Walker, describes how processes focused on both creating an experience that felt like audiences speaking to a survivor and on legitimate archival processes for high quality interactive content that would ensure the legacy of documentary footage captured. [7] The *Audience with a Hero* project further develops those processes but also imagines a more intimate kind of encounter, a one-to-one experience with a living celebrity.

The experience will be voice responsive in a way similar to the interfaces that already cohabit millions of homes: smart Internet of Things (IoT) devices such as Google Home, Amazon’s Echo, and others. However, interactions with household digital assistants are typically ‘telic’, in the sense that the interactions are goal orientated (i.e., to find out the temperature, change the music or the lighting). In marked contrast, the *Audience with a Hero* experience is ‘autotelic’ [8]: we interact with language technology *for the sake of the experience*. There is no obvious extrinsic goal in the idea of having a conversation with our hero, except to have a conversation with our hero. Such an autotelic experience creates a new role for the audience: engaging in a

more ambiguous, conversational encounter that emulates more closely our daily human interactions. [9]

## **Heading: 2. Onboarding**

### **Subheading: Experiencing parasocial anxiety whilst we wait**

On arrival at the venue, there is enough time before our appointment to realize that we are anxious. It is an 'unreal' and virtual meeting, but we have been fans of the hero for some time. He has been pretty open about his personal struggles and seems like a genuine person but suddenly the prospect of meeting him face to face is intimidating. We remind ourselves again that this is a virtual encounter.

Face-to-face relationships help us model our interactions with figures from the media, such as newscasters or actors and the characters they play on TV. They each constitute a parasocial relationship (PSR), a concept introduced by theorists Horton and Wohl in their 1956 paper, "Mass Communication and Para-Social Interaction: Observations on Intimacy at a Distance." In the paper, Horton and Wohl use the terms "para-social relationship" and "para-social interaction" somewhat interchangeably but describe a parasocial relationship as one made up of a "seeming face-to-face relationship with a performer." [10] These virtual relationships develop through parasocial interaction (watching, reading, listening to perceive "relational development" [11]) and as we intertwine our lives with our technology, we do so while attentive to the "nuance and appearance of gesture to which ordinary social perception is attentive and to which interaction is cued." [12]

This has an impact on our upcoming conversation. We might as well be meeting a computer in 'person,' for the effect this audience with our hero will have on our relationship with the hero himself. He won't remember meeting us at all and the version of him we're about to encounter is a collage of his utterances. Having learned the formation of relationships primarily with other humans, we subject our interactions with technology to many of the same standards. It has never been the condition of a parasocial relationship that the mediated figure is 'real' in the way that a human is 'real.' When a crucial dimension of the parasocial relationship is that its participants have not met, the mediated persona is also free to be fictional.

When encountering a person with whom we have developed a parasocial relationship, the collision of the 'ordinary' frame of the individual and the 'extraordinary' frame of the figure can result in a variety of unpredictable outcomes related to the different rules of conduct and rituals attached to each. [13] During a conversational encounter, intense "emotion can be generated in this collision of frames: excitement, disappointment, exhilaration, risk, superiority, and shame." [14] If we misspeak or have difficulty navigating the technology of the experience, will it be 'alive enough' to be embarrassing for us?

### **Subheading: 'Alive enough' to speak to?**

The idea of ‘alive enough’ identities is derived from Sherry Turkle’s studies into children’s interactions with “sociable robots” [15] through which she observed new *ad hoc* and pragmatic categories of aliveness. When interacting with computer-controlled toy animals, for example, children considered them “alive enough for this purpose.” [16] In doing so, they go beyond movement-based notions of what constitutes aliveness (if it can move on its own it is alive) to a psychological model of aliveness (if it can think then it is alive). It is a *conditional* aliveness based on an assessment of its living qualities in the context of a purpose. Part of our nervousness is in trying to identify what our purpose is for our meeting with this hero: what will be enough in terms of the intelligent features of a conversation?

The project uses machine learning so we can directly ask our hero a question and get an appropriate response. What we don’t know is how ‘real’ it will feel once we’re inside the virtual room.

### **Heading: 3. Our Virtual Encounter**

We are sitting in a dark, virtual room similar to the one we left behind at the venue, our senses encased in a VR headset. There are familiar objects in this space that appear to be obeying the laws of physics. There is a chair, an amplifier, and guitar stand. We reach out and brush empty air; the objects are here and not here simultaneously. Someone approaches, emerging out of the black surroundings, playing a guitar. He takes the seat in front of us, settling.

“Hello,” he says. “What’s your name?”

#### **Subheading: The prosodic dimensions of virtual conversations**

When the conversation begins there is a quality to his voice, the shadow of the original interaction, the “paradoxical simultaneity of presence and absence” [17] of the speaker whom we have come to recognize through other recorded media. However, in this instance he waits for a response, and when we do speak, he replies, if not always with the answer we were looking for.

The well-learned grammar of recorded media – for example, interviewees responding to an unheard question or spontaneous leaps in narrative time and space common to film and television grammar – is disrupted by our own voice speaking. The voice denotes the body, both ours and his. [18] Each utterance is the body’s auditory signature and situates the body “as a kind of receiver for stimuli given by the world and generator of appropriate responses to it.” [19] So, by speaking to our mediated hero we confer on him a body, we become attuned to his presence. But listening more closely, it is more than his words that we are attuned to. The way we say something demands a particular and nuanced response from the digital speaker’s voice. In their review of current studies of prosody in linguistics, Heinz and Moroni note that prosodic features (such as the tone, volume, speed, and pitch) in language are considered to be essential contextualization cues involved in important conversational tasks, such as signaling focus, clarifying grammatical constructions and providing rhythmic indicators for turn taking between speakers. [20] The way we say things is contextual to the way our interlocutor has spoken. It is probably this, a sort of deep listening, that we have an ear for when our hero responds to our questions and niceties. There is a slight but perceptible dissonance, the rhythmic indicators of

another conversation are present in our interactions, even when his message is congruent with our prompts and questions. In Emma Rodero's 2017 study comparing stories told by humans and by machines, she found that the expressive prosodic qualities of the human voice are very important for listeners when understanding and remembering a story. Human voices were "better processed" in terms of "effectiveness, attention, concentration and recall" of the stories. [21] Human use of prosody, through normal patterns of stress, pitch and pace clarified the meaning of what was being spoken for the listener, allowing them to "focus on the story." [22] Rodero notes that people "evaluate a synthetic voice interaction as they do a human interaction." [23] Given the multiple functions of prosody in facilitating meaningful speech, it is easy to appreciate the difficulty that humans have in decoding speech without normal prosodic features. It also makes clear the scale of the challenge faced by machine learning schemes made to synthesize prosodic patterns in speech.

One of the expressive qualities of the voice is its emotional quality. Prosodic features, such as vowel features, are instrumental in expressing the various emotional states of speakers. [24] Emotional arousal in speech is so irrepressible that, when they are in conflict with the message that is spoken, prosodic features "can automatically reveal the effect we are trying to suppress." [25] In a study by Spinelli, Fasolo, and Aureli, it was observed that when the subjects of a study spoke about a topic from their childhood that upset them but attempted to dismiss its effect on them, there was a consistency "between the valence of verbal content and prosody," representing a lack of correspondence between "narration of [their] experiences and the real affective relevance of these experiences." [26] That there is sometimes a perceivable gap between what is said and the way it is said, is one of the vocal cues that "a good conversationalist needs to pick up on" says Trevor Cox. [27] Yet, whilst we must be sensitive to these cues, it is also the case that such features are difficult for humans to interpret systematically. Both trained humans as well as machines that detect micro-tremors in the voice for conflicts between what is said and the way it is said are generally as bad or worse than random when detecting liars, for example. [28] One of the reasons for this is the *Othello error*: although we can detect stress in the voice easily, it is difficult to identify the cause of stress. [29] With agents that possess clear emotional features such as the one who sits in front of us now, we automatically model possible psychological motivations for changes in the emotional quality of the agent's speech regardless of its validity. So, we can adapt to slight anomalies in the emotional patterns of our hero's utterances because we are used to ambiguity of motivation when conducting other conversations.

The fact that the agent is a mediated entity speaking to us through video clips adds additional complexity to our conversation. After all, the act of recording and editing the voice renders every media object a deliberate artefact even in the context of a conversational interactive work such as *Audience with a Hero*. Whilst in a real-life context, evidence collected by Cox suggests it is practically impossible to reliably detect the motivations of suppressed emotional arousal, in the context of film and other media, it is one of the ways in which storytellers indicate a knowable subtext in character utterances. [30] To narratologist Robert McKee, the presentation of these conflicts through language choices "convey[s] [a character's] inner life, conscious and subconscious, without announcing it": the audience member's knowledge of the dramatic conventions of prosodic dissonance, their 'story-trained sonar' allows them to become a 'mind reader' of the on-screen character, often understanding their internal life better than they know themselves. [31] When building models of conversational narrative such as the one developed as

part of *Audience with a Hero*, we must consider the role of emotional arousal and its suppression both in real-world relations between people and through the conventions used by media storytellers to portray the inner world of a character.

### **Subheading: Increasing social presence through self disclosure**

After a heartfelt response about working on a song with one of his longtime collaborators, the hero looks at us expectantly, awaiting the next question. We feel closer to him based on his answers so far but can't think of anything to say to move the exchange forward.

We are prompted to ask something else and are relieved to choose from suggestions. The hero wants us to feel comfortable, so he tells the stories from his career that people always respond to. Feeling the gentle pressure of social presence deepens our engagement with the AI because we must constantly remind ourselves that the real man cannot hear us, even though he responds to our questions. Feeling socially present is also supported by the consistency of the virtual space, that familiar gravity and orientation, wherein "the reduced cognitive load and decreased disbelief...may make it easier for users to become deeply engrossed in the virtual environment and increase feelings of social presence." [32] By providing the user with a familiar social framework in the setting, greeting, and rhythm of conversation, the attachment to nonhuman, digital entities is eased. The AI borrows a mask from the recorded testimony and the recorded testimony borrows a body from the AI's conversational framework. By mimicking the contingent responses of human conversation (such as giving responses that call back to things previously mentioned), the perceived interactivity and effectiveness of the overall experience can be enhanced. [33]

Due to this enhanced sense of social presence, we are less inclined to view the exchange as narratological exposition of character, in the way Robert McKee describes the function of cinematic character. Filmed in 3D and rendered as an immersive environment, the clips of the hero's responses do not appear on a flat screen like the images we are used to from narrative entertainment. These factors reduce the number of familiar cues for the audience member to interpret elements of prosody like they might in trying to 'read' the face of a character in a film. The subtext of our hero's utterances is affected by the ordering and selection of answers to questions that we pose. If there is a narrative arc to the experience, it is one of co-construction. The hero is used to managing his image in the media, but the unpredictability of the order and selection of clips means it is safer to just limit the dynamic range of overall expressiveness.

This positive mood, reflected in the upbeat emotional prosody, and which appears genuine, is something we're familiar with from his social media channels. The live streams on Instagram and real-time updates on Twitter make those platforms a reliable place for feeling socially present with the hero and for developing parasocial relationships with other celebrities. The act of self-disclosure, which occurs across breadth (variety) and depth (degree of discourse) of information, involves sharing personal details that users wouldn't otherwise learn. [34] A celebrity's post about the illness of a family member, or a comparison of lunches across a week, helps Twitter become a more socially present space for users. Upon learning about a celebrity, "the part of the brain that processes these messages often doesn't seem to be able to make [the] distinction between readily available real-life mates and less accessible potential partners from media." [35] When the hero's self-disclosure occurs in a virtual space of relative intimacy, social



presence, the combination of this self-disclosure with consistent prosody on the part of the hero helps the AI retrieving answers appear more in tune with natural, conversational rhythm. If the audience member already feels an attachment stemming from a parasocial relationship, the proximity sought in a virtual encounter “can substitute for the actual physical presence of the desired figure.” [36]

The conversation has taken around ten minutes. At the finish, we are told to remove our head mounted display and proceed out into the lobby. We begin to reflect on what we have experienced and what it might mean in our everyday lives.

### **Heading: Conclusion**

In this project, there were many novel design aspects that had not previously been considered due to the limited scope of voice interactive tools in our culture. Predominantly, voice-based interactions with technology are limited to telic or functional purposes such as online shopping or providing weather reports. The agent in this sense is limited to the role of passive, attentive servant or tool. In the case of our audience with a hero we encounter a different phenomenon where the purpose is autotelic or experiential; where the information gained through the exchange is accompanied by social nuance that has value of its own. Such an experience therefore requires new design considerations grounded in transdisciplinary research.

The transdisciplinary design research that has contributed to the Audience with a Hero experience has brought to light a number of theoretical considerations implicit in immersive experiences featuring a human-like conversational agent. A principle from psychology that we have considered is Turkle’s observation of social robots as needing to be ‘alive enough’: we do not need to be convinced that the agent is a fully responsive living person to have a conversation, we only need it to display some of the conversational attributes of a person.

In the experience documented here, the agent occupies a greater sense of aliveness in its capacity to respond with instances of self-disclosure. Identifying with the agent through human testimony builds on concepts from sociology and parasocial studies: the interactions with our celebrity hero figure are informed by the user’s existing parasocial relationship. The audience member brings along their previous impressions of the hero based on their presence in the media, populating the virtual encounter with an established, one-sided relationship dynamic. As a result, the conversation and language therein is constrained by some of the normal anxieties of an encounter with a celebrity. This encounter could offer an ‘alive enough’ proximity to deepen the parasocial relationship without transgressing its essential face-to-face boundary.

From linguistics, we note that the extent to which we grant aliveness to the media figure then depends on other sensitive registers, such as to what extent the agent reflects and responds to the prosodic features of our own speech during conversation. Speakers use prosody to communicate both explicit, implicit and hidden intentions belying speech between people both in real life and in narrative media. These features of speech are hard for an agent to duplicate. However, in both real life and in narrative theory, it is accepted that we will not understand all the intentions of the people we speak with. Therefore, we expect that users will be more forgiving of the strangenesses of mismatched prosodic features and might even imagine narrative arcs in the ebbs and flows of enthusiasm, passion and attention that prosody denotes. In other words, we credit

the agent with contiguous intentionality and memory during our conversation because we are used to doing this to fill in gaps in understanding while consuming narrative media and during our conversations with other people.

The context in which we currently communicate with voice-based agents is heavily influenced by our interactions with other people, celebrity and narrative media, and virtual assistants. As voice interactive agents proliferate, will we develop a new literacy for machine language that changes how we understand other people?

### **Acknowledgements**

We are grateful for the support and finance from UK Research and Innovation (UKRI) as part of the Industrial Strategy Challenge Fund, specifically the Audiences of the Future programme in supporting and financing the project. We are also pleased to be working with project partners Forever Holdings PLC, Bright White Ltd. and Pollen Studio on this work.

### **Author Biographies**

Marsha Courneya is a Canadian writer, editor and open licensing specialist who works as a Senior Research Assistant at Manchester Metropolitan University. Her work centers on copyright law reform and using open licensing to address income inequality through sustainable economic models of collective authorship.

David Jackson is a lecturer exploring relationships between artificial intelligence, narrative media, and its audiences. In the project Audience with a Hero, he is researching the purposes, constraints and effects of narrative in VR conversational experiences and the changing production processes required to develop VR conversational experiences. Other projects include Unhealthy Bias, a project examining raced and gendered influences on machine learning (ML) approaches to detecting bias in public health messaging and a transmedia project which reflects on the history of acoustic surveillance in the age of always-on mobile and home listening agents such as Siri and Alexa with members of Columbia University Digital Storytelling Lab.

Roger McKinley has many years of experience in research and production in the cultural sector. In his previous role as Head of Innovation at FACT, Liverpool, he set up the Research and Innovation Department to deliver strategic development through new value chains between arts organizations, Creative and Digital Industries (CDI) businesses, FE and HE institutions regionally, nationally and internationally. He is Research Associate in the Audience with a Hero project, focusing on market intelligence and evaluation.

### **Heading: Notes and References**

[1] Christopher Sciacca, "I Am Sitting in 4 Rooms: Presence and Absence in the Work of Alvin Lucier and Jacob Kierkegaard," Guggenheim Museum, New York, 1970, accessed 29 March 2020, [https://www.academia.edu/12380463/I\\_am\\_Sitting\\_in\\_4\\_Rooms\\_Presence\\_and\\_absence\\_in\\_the\\_work\\_of\\_Alvin\\_Lucier\\_and\\_Jacob\\_Kierkegaard](https://www.academia.edu/12380463/I_am_Sitting_in_4_Rooms_Presence_and_absence_in_the_work_of_Alvin_Lucier_and_Jacob_Kierkegaard).

[2] For example, see Rafael Valencia-García, Francisco García-Sánchez, “Natural Language Processing and Human–Computer Interaction,” *Computer Standards & Interfaces* 35, no. 5, (2013): 415-416.

[3] [Kyle Wm Hall](#), [Adam J Bradley](#), [Uta Hinrichs](#), [Samuel Huron](#), [Jo Wood](#), [Christopher Collins](#), [Sheelagh Carpendale](#), “Design by immersion: A transdisciplinary approach to problem-driven visualizations,” *IEEE Transactions on Visualization and Computer Graphics* 26, no. 1, (2020): 109-118.

doi: 10.1109/TVCG.2019.2934790

[4] Leonardo Moreno & Erika Rogel, “Transdisciplinary Design: Tamed complexity through new collaboration,” *Strategic Design Research Journal*, 1 (2018): 42-50.  
10.4013/sdrj.2018.111.07.

[5] For practical reasons we cannot go into depth on the new AI tools, methods and application of these under development in the project. Restrictions on sharing technical details is an emerging condition within arts and cultural research where it crosses over with the commercial world of the creative industries and requires new approaches in cultural studies. At time of writing, we could not specifically mention the subject due to commercial sensitivity but during review it has become possible to share the identity of the subject: musician, producer and collaborator Nile Rodgers.

[6] Jing Cao and Dina Bass, "Why Google, Microsoft and Amazon Love the Sound of Your Voice – Bloomberg," accessed 19 March 2020, <https://www.bloomberg.com/news/articles/2016-12-13/why-google-microsoft-and-amazon-love-the-sound-of-your-voice>.

[7] Minhua Ma, Sarah Coward, and Chris Walker, "Question-Answering Virtual Humans Based on Pre-Recorded Testimonies for Holocaust Education," in *Serious Games and Edutainment Applications*, ed. Minhua Ma and A Oikonomou (Springer, Cham 2017), 391–409, [https://doi.org/10.1007/978-3-319-51645-5\\_18](https://doi.org/10.1007/978-3-319-51645-5_18).

[8] Mihaly Csikszentmihalyi, *Creativity: Flow and the Psychology of Discovery and Invention*, 1st ed (New York: HarperCollinsPublishers, 1996), 1.

[9] It is worth noting that there are inadequate descriptors for this kind of user experience itself. The experience of such a process is not well served by familiar terms in arts practice such as ‘interactive’ or ‘immersive’. Alternative descriptors would be ‘participant’, ‘collaboration’, ‘play’ and ‘exchange’, because ‘participant’ implies active engagement, ‘collaboration’ is necessarily co-creative, ‘play’ is understood as a co-produced activity that is for enjoyment and ‘exchange’ is transactional. The audience member or ‘player’ is a ‘participant’ in the ‘exchange’. We could describe the audience taking part in a ‘playful collaboration with a Hero’ or a ‘collaborative exchange with a Hero’ or experiencing a ‘collaborative and playful exchange with a Hero’ or ‘participating in a real-time question and answer session with a Hero’.

[10] Donald Horton and R. Richard Wohl, "Mass Communication and Para-Social Interaction," *Psychiatry* 19, no. 3 (1 August 1956): 215, <https://doi.org/10.1080/00332747.1956.11023049>.

[11] William J. Brown, "Examining Four Processes of Audience Involvement With Media Personae: Transportation, Parasocial Interaction, Identification, and Worship: Examining Four Processes of Audience Involvement With Media Personae," *Communication Theory* 25, no. 3 (August 2015): 275, <https://doi.org/10.1111/comt.12053>.

[12] Horton and Wohl, "Mass Communication," 215.

[13] Kerry O. Ferris, "Seeing and Being Seen: The Moral Order of Celebrity Sightings," *Journal of Contemporary Ethnography*, (25 July 2016), 241, <https://doi.org/10.1177/0891241604263585>.

[14] O. Ferris, "Seeing," 241.

[15] Sherry Turkle, *Alone Together*, First Trade Paper Edition (New York, NY: Basic Books, 2013), 1.

[16] Sherry Turkle, "Sherry Turkle — Alive Enough? Reflecting On Our Technology," interview by Krista Tippet, On Being Studios, November 15, 2012. Audio, 15:25 accessed 26 March 2020, <https://soundcloud.com/onbeing/sets/sherry-turkle-on-alive-enough>.

[17] Norie Neumark, Ross Gibson and Theo Van Leeuwen, *Voice: Vocal Aesthetics in Digital Arts and Media* (Cambridge, Mass.: MIT Press, 2010), xvii.

[18] Neumark, Gibson, and Van Leeuwen, *Voice: Vocal Aesthetics in Digital Arts and Media*.

[19] Lucy Suchman, *Human-Machine Reconfigurations: Plans and Situated Actions*, 2nd Edition (Cambridge University Press, 2006), 230.

[20] Matthias Heinz and Manuela Caterina Moroni, "Prosody: Information Structure, Grammar, Interaction," *Linguistik Online* 88, no. 1 (1 January 2018): 3–11.

[21] Emma Roderer, "Effectiveness, Attention, and Recall of Human and Artificial Voices in an Advertising Story. Prosody Influence and Functions of Voices" (Elsevier Science Publishers B. V., 1 December 2017): 344, <https://doi.org/10.1016/j.chb.2017.08.044>.

[22] Roderer, "Effectiveness," 344.

[23] Roderer, "Effectiveness," 345.

[24] Kanu Boku et al., "Speech Synthesis of Emotions Using Vowel Features," *Artificial Life and Robotics* 19 (1 February 2014): 27–32, <https://doi.org/10.1007/s10015-013-0126-9>.

- [25] Maria Spinelli et al., "It Is a Matter of How You Say It: Verbal Content and Prosody Matching as an Index of Emotion Regulation Strategies during the Adult Attachment Interview," *International Journal of Psychology* 54, no. 1 (2019): 102, <https://doi.org/10.1002/ijop.12415>.
- [26] Spinelli et al., "It is a Matter," 102.
- [27] Trevor Cox, *Now You're Talking* (London: Penguin, 2019), 197.
- [28] Cox, *Now You're Talking*, 208.
- [29] Cox, *Now You're Talking*, 212.
- [30] Robert McKee, *Dialogue : The Art of Verbal Action for the Page, Stage and Screen* (London, UK: Methuen, 2016).
- [31] McKee, *Dialogue*, 48–49.
- [32] Kwan-Min Lee and Clifford Nass, "Social-Psychological Origins of Feelings of Presence: Creating Social Presence With Machine-Generated Voices," *Media Psychology* 7, no. 1 (2005): 35.
- [33] Ji Hee Song and George M. Zinkham, "Determinants of Perceived Web Site Interactivity," *Journal of Marketing* 72 (2008): 102.
- [34] I. Altman and D.A. Taylor, *Social Penetration: The Development of Interpersonal Relationships* (New York: Holt, Rinehard & Winston, 1973).
- [35] Gayle S. Stever, "Evolutionary Theory and Reactions to Mass Media: Understanding Parasocial Attachment," *Psychology of Popular Media Culture* 6, no. 2 (2017): 96.
- [36] Stever, "Evolutionary Theory," 97.

## GLOSSARY

**Telic** – Motivated by goal-orientated activity, where the purpose of engaging in an activity is to achieve a defined goal or outcome or complete a task. The term was coined by linguist Howard B Garey in 1957 in his article "Verbal Aspects" in French *Language* (Howard B. Garey, Vol. 33, No. 2 Apr. - Jun., 1957, pp. 91–110) and later applied to motivation theory by psychologist Michael J Apter (Murgatroyd, Rushton, Apter & Ray; The development of the Telic Dominance Scale; *Journal of Personality Assessment* 1978 Oct 4 (5): 519-28).

**Autotelic** – “An autotelic activity is one we do for its own sake because to experience it is the main goal. Applied to personality, autotelic denotes an individual who generally does things for their own sake, rather than in order to achieve some later external goal.” (Csikszentmihalyi, 1996. *Creativity: Flow and the Psychology of Discovery and Invention*. 1st ed. New York: Harper Collins Publishers. p. 117).

## Bibliography

Altman, I., and D.A. Taylor. 1973. *Social Penetration: The Development of Interpersonal Relationships*. New York: Holt, Rinehard & Winston.

Armstrong, Anne-Marie, ed. 2004. *Instructional Design in the Real World: A View from the Trenches*. Hershey Pa: Information Science Pub

Bernabei, Roberta. 2017. “Wearable Words: A Case Study Applying Jewellery Theory and Practice to the Education of Fine Art, Textiles Innovation and Design, Graphic Communication and Illustration Students.” *The Design Journal* 20 (sup1): S1503–10.  
<https://doi.org/10.1080/14606925.2017.1352674>.

Boku, Kanu, Taro Asada, Yasunari Yoshitomi, and Masayoshi Tabuse. 2014. “Speech Synthesis of Emotions Using Vowel Features.” *Artificial Life and Robotics* 19 (February): 27–32.  
<https://doi.org/10.1007/s10015-013-0126-9>.

Brown, William J. 2015. “Examining Four Processes of Audience Involvement With Media Personae: Transportation, Parasocial Interaction, Identification, and Worship: Examining Four Processes of Audience Involvement With Media Personae.” *Communication Theory* 25 (3): 259–83. <https://doi.org/10.1111/comt.12053>.

Cao, Jing, and Dina Bass. “Why Google, Microsoft and Amazon Love the Sound of Your Voice.” *Bloomberg*, 13 December 2016, Accessed March 19, 2020.  
<https://www.bloomberg.com/news/articles/2016-12-13/why-google-microsoft-and-amazon-love-the-sound-of-your-voice>.

Cox, Trevor. 2019. *Now You're Talking*. London: Penguin. [/books/1111924/now-you-re-talking/9781784705220](https://books/1111924/now-you-re-talking/9781784705220).

Csikszentmihalyi, Mihaly. 1996. *Creativity: Flow and the Psychology of Discovery and Invention*. 1st ed. New York: Harper Collins Publishers.

Ferris, Kerry O. 2016. "Seeing and Being Seen: The Moral Order of Celebrity Sightings." *Journal of Contemporary Ethnography*, July. <https://doi.org/10.1177/0891241604263585>.

Garvis, Susanne. 2009. "Establishing the Theoretical Construct of Pre-Service Teacher Self-Efficacy for Arts Education." *Australian Journal of Music Education*. <https://eric.ed.gov/?id=EJ912408>.

Hall, K., Bradley, A., Hinrichs, U., Huron, S., Wood, J., Collins, C. and Carpendale, S. 2020. "Design by immersion: A transdisciplinary approach to problem-driven visualizations." *IEEE Transactions on Visualization and Computer Graphics* 26 (1): 109–118. <https://doi.org/10.1109/TVCG.2019.2934790>

Heinz, Matthias, and Manuela Caterina Moroni. 2018. "Prosody: Information Structure, Grammar, Interaction." *Linguistik Online* 88 (1): 3–11.

Horton, Donald, and R. Richard Wohl. 1956. "Mass Communication and Para-Social Interaction." *Psychiatry* 19 (3): 215–29. <https://doi.org/10.1080/00332747.1956.11023049>.

Lee, Kwan-Min, and Clifford Nass. 2005. "Social-Psychological Origins of Feelings of Presence: Creating Social Presence with Machine-Generated Voices." *Media Psychology* 7 (1): 31–45.

Leydesdorff, Loet, and Liwen Vaughan. 2006. "Co-Occurrence Matrices and Their Applications in Information Science: Extending ACA to the Web Environment." *Journal of the American Society for Information Science and Technology* 57 (12): 1616–28. <https://doi.org/10.1002/asi.20335>.

Luhmann, Niklas. 1982. "THE WORLD SOCIETY AS A SOCIAL SYSTEM." *International Journal of General Systems* 8 (3): 131–38. <https://doi.org/10.1080/03081078208547442>.

Ma, Minhua, Sarah Coward, and Chris Walker. 2017. "Question-Answering Virtual Humans Based on Pre-Recorded Testimonies for Holocaust Education." In *Serious Games and Edutainment Applications*, edited by Minhua Ma and Andreas Oikonomou, 391–409. Springer Link. [https://doi.org/10.1007/978-3-319-51645-5\\_18](https://doi.org/10.1007/978-3-319-51645-5_18).

McKee, Robert. 2016. *Dialogue : The Art of Verbal Action for the Page, Stage and Screen*. London, UK: Methuen.

Moreno, Leonardo & Rogel, Erika. 2018. "Transdisciplinary Design: Tamed complexity through new collaboration." *Strategic Design Research Journal* 1:42–50. <https://doi.org/10.4013/sdrj.2018.111.07>.

Neumark, Norie, Ross Gibson, and Theo Van Leeuwen. 2010. *Voice: Vocal Aesthetics in Digital Arts and Media*. Cambridge, Mass.: MIT Press.

Rodero, Emma. 2017. "Effectiveness, Attention, and Recall of Human and Artificial Voices in an Advertising Story. Prosody Influence and Functions of Voices." Elsevier Science Publishers B. V. <https://doi.org/10.1016/j.chb.2017.08.044>.

Sciacca, Christopher. n.d. "I Am Sitting in 4 Rooms: Presence and Absence in the Work of Alvin Lucier and Jacob Kierkegaard." Accessed March 29, 2020. [https://www.academia.edu/12380463/I\\_am\\_Sitting\\_in\\_4\\_Rooms\\_Presence\\_and\\_absence\\_in\\_the\\_work\\_of\\_Alvin\\_Lucier\\_and\\_Jacob\\_Kierkegaard](https://www.academia.edu/12380463/I_am_Sitting_in_4_Rooms_Presence_and_absence_in_the_work_of_Alvin_Lucier_and_Jacob_Kierkegaard).

Shannon, C. E. 1948. "A Mathematical Theory of Communication." *Bell System Technical Journal* 27 (3): 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>.

Song, Ji Hee, and George M. Zinkham. 2008. "Determinants of Perceived Web Site Interactivity." *Journal of Marketing* 72: 99–113.

Spinelli, Maria, Mirco Fasolo, Gabrielle Coppola, and Tiziana Aureli. 2019. "It Is a Matter of How You Say It: Verbal Content and Prosody Matching as an Index of Emotion Regulation Strategies during the Adult Attachment Interview." *International Journal of Psychology* 54 (1): 102–7. <https://doi.org/10.1002/ijop.12415>.

Stever, Gayle S. 2017. "Evolutionary Theory and Reactions to Mass Media: Understanding Parasocial Attachment." *Psychology of Popular Media Culture* 6 (2): 95–102.

Suchman, Lucy. *Human-Machine Reconfigurations: Plans and Situated Actions*. 2nd Edition. Cambridge University Press, 2006.

Turkle, Sherry. 2013 *Alone Together*, First Trade Paper Edition, New York, NY: Basic Books.

———. 2012. "Sherry Turkle — Alive Enough? Reflecting On Our Technology," interview by Krista Tippett, On Being Studios, November 15, 2012. Audio, accessed 26 March 2020, <https://soundcloud.com/onbeing/sets/sherry-turkle-on-alive-enough>.

Valencia-García, Rafael, Francisco García-Sánchez. 2013. "Natural Language Processing and Human-Computer Interaction." *Computer Standards & Interfaces* 35 (5): 415–416. <https://doi.org/10.1016/j.csi.2013.03.001>.



