

Audio Matters Too: How Audial Avatar Customization Enhances Visual Avatar Customization

Dominic Kao
kaod@purdue.edu
Purdue University
USA

Rabindra Ratan
rar@msu.edu
Michigan State University
USA

Christos Mousas
cmousas@purdue.edu
Purdue University
USA

Amogh Joshi
joshi134@purdue.edu
Purdue University
USA

Edward F. Melcer
eddie.melcer@ucsc.edu
UC Santa Cruz
USA

ABSTRACT

Avatar customization is known to positively affect crucial outcomes in numerous domains. However, it is unknown whether audial customization can confer the same benefits as visual customization. We conducted a preregistered 2 x 2 (visual choice vs. visual assignment x audial choice vs. audial assignment) study in a Java programming game. Participants with visual choice experienced higher avatar identification and autonomy. Participants with audial choice experienced higher avatar identification and autonomy, but only within the group of participants who had visual choice available. Visual choice led to an increase in time spent, and indirectly led to increases in intrinsic motivation, immersion, time spent, future play motivation, and likelihood of game recommendation. Audial choice moderated the majority of these effects. Our results suggest that audial customization plays an important enhancing role vis-à-vis visual customization. However, audial customization appears to have a weaker effect compared to visual customization. We discuss the implications for avatar customization more generally across digital applications.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**.

KEYWORDS

Games; Avatar; Audio; Voice; Customization; Identification; Player Experience

ACM Reference Format:

Dominic Kao, Rabindra Ratan, Christos Mousas, Amogh Joshi, and Edward F. Melcer. 2022. Audio Matters Too: How Audial Avatar Customization Enhances Visual Avatar Customization. In *CHI Conference on Human Factors in Computing Systems (CHI '22)*, April 29-May 5, 2022, New Orleans, LA, USA. ACM, New York, NY, USA, 27 pages. <https://doi.org/10.1145/3491102.3501848>



This work is licensed under a Creative Commons Attribution-Share Alike International 4.0 License.

CHI '22, April 29-May 5, 2022, New Orleans, LA, USA
© 2022 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-9157-3/22/04.
<https://doi.org/10.1145/3491102.3501848>

1 INTRODUCTION

Avatars are ubiquitous across digital applications. Using avatars as representations of ourselves, we socialize, play, and work. Increasingly, researchers have become interested in *avatar customization* [152]. Avatar customization, or the ability to modify one's avatar, increases outcomes including intrinsic motivation [25], helping behavior [59], user retention [26], learning [175], flow [136], and of especial importance, avatar identification [221]. Avatar identification, or *the temporary alteration in self-perception of the player induced by the mental association with their game character* [224], leads to increased motivation [25, 26, 220], creative thinking [34, 56, 87], enjoyment [134, 175, 217], performance [113], player experience [107], flow [207], and trust in others [117]. Despite the large corpora of literature on avatar customization, studies have focused almost exclusively on *visual* aspects of the avatar. Limited adoption of *audial* aspects in avatar customization is potentially because avatar audio is perceived as non-critical and has substantial overhead (e.g., multiple voice actors, region localization) [237]. Recent advances in artificial intelligence (e.g., neural networks) have vastly improved text-to-speech engines and voice cloning software, however, and these programs are able to produce artificial voices nearly indistinguishable from real ones. Voice cloning software was used in a study in which participants played a game with avatars that had either a similar voice or a dissimilar voice (as compared to the player) [114]. Results showed that participants in the similar voice condition had increased performance, time spent, similarity identification, competence, relatedness, and immersion. Prior research adds further support that the importance of avatar audio may be underappreciated. Audio in games is linked to increased physiological responses [90], emotional realism [20, 66], performance [103], and immersion [67, 115, 131, 168, 199]. A meta-analysis of 83 studies in virtual environments found that adding audio has a small- to medium-sized effect on presence [55]. Given prior work demonstrating the importance of avatar customization and audio separately, allowing players to *audially* customize their avatars may have beneficial effects.

Customizing one's avatar is often viewed as inherently enjoyable [105]. This customization is now part of a lucrative "skin" market in online games [102]. Game skins can be used to customize an avatar's appearance, and research estimates the skin market to be worth \$40 billion (USD) per year [226]. While a few ventures have begun to explore customization of the player's voice, these

efforts have been limited to external tools (e.g., voice-changing software [163, 229]). A small number of games *do* offer the option of customizing avatar audio. Final Fantasy XIV [208], Saints Row IV [230], and Monster Hunter: World [36] allow the user to choose between different sets of voices. Black Desert Online [178], Red Dead Redemption 2 [192], and The Sims 4 [68] allow the user to customize pitch. More generally, avatar customization interfaces are understood to vary greatly between games with regards to both quantity and quality of customization options [152, 156]. For the purposes of the present study, we created four character models and four character voices. We then created four character customization interfaces that varied (1) whether the character model was chosen or randomly assigned and (2) whether the character voice was chosen or randomly assigned. These customization interfaces were explicitly designed to test whether aural customization would have any effect on outcomes vis-à-vis visual customization.

We conducted an online study on Amazon’s Mechanical Turk (MTurk) in which participants were randomly assigned to one of the four character customization interfaces. Participants then played a Java programming game for 10 minutes. After 10 minutes had passed, an in-game survey collected measures of avatar identification, autonomy, intrinsic motivation, immersion, motivated behavior, motivation for future play, and likelihood of game recommendation.¹ After completing the survey, participants could quit or continue playing for as long as they liked, reflecting motivated behavior.

Our results show that visual customization leads to higher avatar identification and autonomy. Aural customization leads to higher avatar identification and autonomy, but only within the grouping of participants in which visual customization was available. In the grouping of participants without visual customization, aural customization had no effect on avatar identification or autonomy. Visual customization leads to higher time spent playing, and indirectly (through the mediators of avatar identification and autonomy), it leads to higher intrinsic motivation, immersion, time spent playing, motivation for future play, and likelihood of game recommendation. Aural customization moderated the direct effect of visual customization on time spent playing, as well as the indirect effects of visual customization on intrinsic motivation, immersion, motivation for future play, and likelihood of game recommendation. The moderation effect was such that the effect was non-significant when aural customization was unavailable but significant when aural customization was available. Our results show that aural customization, although having an overall weaker effect than visual customization, can strengthen existing effects of visual customization on outcomes. This suggests that avatar customization systems in games can be improved by adding aural customization options. Moreover, our study provides motivation to extend this research to other domains as potential beneficiaries of aural avatar customization (e.g., virtual reality, digital learning, health applications). In the highly understudied area of avatar audio, we contribute baseline results in a large-scale preregistered study that can spur further work in this domain.

¹Study hypotheses, analyses, experiment design, data collection, sample size, and measures were all preregistered. Preregistration: <https://osf.io/dbvbkp/>. Raw Data: <https://osf.io/mnpsd/>.

2 RELATED WORK

2.1 Avatar Customization

Avatar customization is the process of changing aspects of a video game character. Players customize their avatars’ physical (e.g., body shape), demographic (e.g., age, race, gender), and transient (e.g., clothes, ornaments) aspects. The avatar customization process can also include choosing roles (e.g., playing as a warrior, archer, mage, or a healer), attributes (e.g., luck, intelligence), and group membership (e.g., playing as horde or alliance) [53, 219]. Customizing one’s avatar can lead to direct and indirect effects on gameplay [100, 219]. For example, choosing a role of a warrior affords different game mechanics and play strategies (i.e., favoring close combat) compared to playing as an archer. Similarly, customizing skill attributes can also affect gameplay—e.g., favoring increased charisma gives lower prices on game items in *Fallout 4* [24]. Customizing avatars’ physical appearance or the name of the avatar, on the other hand, usually does not affect gameplay (directly) but can have a psychological effect on the players [25, 138, 201]. To understand these psychological effects, many studies have used off-the-shelf games (e.g., Massively Multiplayer Online Games, or MMOs) that offer a comprehensive avatar customization process, such as changing physical, demographic, and transient aspects; as well as choosing roles, group membership, and attributes. Lim and Reeves used a popular MMORPG (*World of Warcraft*, or *WoW* [53]) where participants were randomly assigned to play the game with avatar customization or to play with a premade avatar [138]. The study found that players who customized their avatar experienced greater physiological arousal [138]. Similarly, players reported greater physiological arousal and subjective feelings of presence when playing advergames that offered avatar customization options, suggesting greater game enjoyment [14]. It has also been shown that players remember more game features—such as spatial features of landmarks and characteristics of NPCs—when playing with customized avatars [175]. Teng [215] examined how customizing avatars’ transient aspects in MMORPGs impact identification and loyalty with the game. The study found that customizing these items (e.g., clothes, shoes, etc.) positively impacted identification with the avatar, which subsequently increased gamer loyalty. Other studies have also explored how customizing non-human objects (e.g., race cars) influences player experience [185, 201]. One study used the game *Need for Speed: ProStreet* [64] to understand if customizing a racecar affects players’ enjoyment of the game [201]. Players customized their cars’ visual appearance, such as changing the car’s shape, aftermarket components (spoilers, rims), color, and skins. Players who customized their cars experienced greater identification, leading to higher game enjoyment, than those who played with pre-made customized cars. One key limitation of these studies is the time duration of their investigation. Many studies have only investigated the effect of avatar customization on short playing time (~1 hour) [220]. MMOs are long-term games, with players’ gameplay experience and expertise evolving with time. Previous studies have found that players playing these games spend approximately 10 hours playing each week [63]. Turkey and Kinzer investigated how players’ identification and empathy towards their avatar evolved over ten hours of playing *Lord of the Rings Online (LotRO)* [209] [220]. The study found that players who customized their avatars had

a stronger identification and expressed greater empathy towards them than those who played the game with premade avatars.

Studies have also used bespoke games to understand the effects of avatar customization [25, 126, 140]. Birk, Atkins, Bowey, and Mandryk [25] investigated if players who customized their avatars experienced greater intrinsic motivation compared to those who used premade avatars. The researchers leveraged *Unity Multipurpose Avatar* [1] to develop a character creator which allowed players to customize their game characters' appearance (e.g., skin tone, clothing), personality (e.g., extraversion), and attributes (e.g., intelligence, stamina, willpower). Players who customized their game character experienced greater identification with their avatars, which led to greater autonomy, immersion, invested effort, enjoyment, positive affect, and time spent playing in an infinite runner [25]. In a subsequent paper, Birk and Mandryk investigated the effect of avatar customization on attrition and sustained engagement while playing a mental health game over three weeks [26]. The study found a reduced attrition rate for the players who customized their avatar compared to those who played with a generic avatar [26]. In another study, playing an exergame with autonomy-supportive features (which included customizing an avatar) led to increased effort, autonomy, motivation to play the game again, and greater likelihood to recommend the game to peers compared to participants who played the game without autonomy-supportive features [180]. Similarly, in a virtual reality exergame, players customized their avatars using an off-the-shelf software tool (*Autodesk Character Creator* [11]) to create an avatar similar to themselves. Players could customize their avatars (e.g., skin tone, hair and eye color, clothes, shoes). The study found that players who competed against their customized self-similar avatars performed significantly better compared to the players who competed with generic avatars [126]. The effect of customization has also been observed in learning environments. Students engaged with a computational learning game (over seven sessions lasting an hour each) with a customized avatar of their choosing [140]. Customization options included skin tone, hairstyle, and eye-color options. The study found that players who customized their avatars remembered and understood greater computational concepts than those who played the game with a premade avatar. Kao and Harrell [113] investigated how avatar identification influenced players in a computational learning environment (*MazeStar* [112]). Players customized their avatars using a freely available Mii creator. The study found that avatar identification promoted outcomes including player experience, intrinsic motivation, and time spent playing [113].

These studies suggest that avatar customization affects player experience in a wide variety of settings (e.g., games for entertainment or learning), virtual environments (e.g., desktop, VR) and timespans (both one-off play sessions and longitudinal) [25, 114, 126, 138, 201, 220]. More importantly, a subset of these studies highlight that avatar customization generates attachment and identification with their game character [26, 113, 201, 215, 220], which consequently affects a wide range of variables: intrinsic motivation [25, 180], autonomy [25, 26, 114], empathy [220], performance [26, 114, 126], game enjoyment [217], loyalty [215] and player experience [25, 114, 138, 201, 220].

2.1.1 Avatar Identification. Identification is a mechanism wherein media experiences—such as reading a story or watching a movie—are interpreted and experienced by audiences as if “the events were happening to them” [46]. The mechanism of identification differs in interactive and non-interactive media experiences. In a typical media experience (e.g., movie or a late-night talk show), the relationship between the audience and media-character is often categorized as a *self* versus *other* (often referred to as a dyadic relationship) [44, 61]. Within games, the distance between the self and the other is said to be diminished due to games' affording direct control over the game character and their interactions in the virtual world [91, 122]. Players control, customize, and interact with their game character and the game world using an avatar. Consequently, the player-avatar relationship is often said to be “a merging of [the player] and the game protagonist” [44].

Avatar identification is thought to be a shift in self-perception [123]. Players can temporarily adopt salient characteristics of the avatar [224] or channel their expectations into the avatar creation, thereby facilitating avatar identification [220]. Many factors influence the nature of identification that can take place with the avatar. Flanagan [73] asserts that player identification with a game character is complicated by the various roles embodied by the player (such as being a subject, spectator, participant, etc.) during game-play. Murphy [167] elaborates on how players' abilities, player characters' abilities, game events, and other players influence the player's sense of agency in virtual environments. While many authors agree that identification takes place between a player and the game character, the nature of identification remains understudied [220].

One avenue of understanding identification is through understanding the avatar customization process. When players customize their avatar, they cycle through many “possible selves” [148] as they experiment and adopt the game characters' attributes for themselves. In two separate studies by different researchers, there are a few common trends regarding players' avatar creation and customization experiences [62, 105]. In one of the studies, researchers investigated reasons for avatar customization and creation in three virtual worlds: *World of Warcraft* [53], *Second Life* [141] and *Maple Story* [238]. Researchers found that players in these virtual worlds created and customized their avatars for various reasons, including to project an ideal self, follow a trend, or stand out from others [62]. Another study examined the avatar creation and customization process for players in *Whyville* [105]. Players customized their avatars for aesthetic reasons, to follow a popular trend, and to express themselves (e.g., show some aspect of their authentic selves). Moreover, they also found that players customize their avatars with a functional intention, such as to experiment with gender or to play different roles [105].

These findings have led researchers to consider avatar identification as a multi-faceted construct [61], which has been operationalized into three distinct dimensions: *similarity identification*, *wishful identification*, and *embodied identification* [224]. Similarity identification refers to players identifying with an avatar that looks like them [61]. Avatars that look similar to players can facilitate feelings of familiarity and stronger empathetic experience [224]. Research shows that similarity identification can play an important role in the player's motivations for playing [224], learning outcomes [114],

player experience [107], and behaviors [126, 134]. Players can also identify with their game characters and see them as role models for future action or identity development [224]. Players desiring to align their personal attitudes, aesthetics, and attributes with those of their game character is referred to as wishful identification [61, 224]. For example, previous research has documented that older players often create avatars younger than themselves [62]. Lastly, players also identify with their avatars when manipulating avatars' bodies as their own. Perceiving to be present in a virtual environment through one's avatar, or so-called "body container" [61, 224], heightens embodied identification [220].

The process of avatar customization is often a precursor for generating greater avatar identification. For example, players wanting to create an avatar that has similar attributes (e.g., physical appearance, hair style, hair color) may generate greater similarity identification [220]. On the other hand, players customizing their avatars according to their ideal self may increase their wishful identification [220]. Players typically interact with a user interface that allows players to fluidly cycle through choices to allow players to constitute their desired digital body. As such, the design and options presented to the players can play a crucial role in helping (or hindering) players to create their desired avatar [153, 155].

2.1.2 Avatar Customization Interface. The interface that the players use to create and customize their avatars—sometimes referred to as a character customization interface (CCI) [156]—represents a "space of liminality" [234] where players spend a significant amount of time intentionally creating their desired avatar [62, 156]. McArthur states that these interfaces generate action possibilities for avatar creation and customization [156]. Players cycle through many possible customization options to create their desired avatar. Avatar customization interfaces are not only important in terms of usability, but also in how they communicate cultural ideologies [153, 156].

For instance, the design of "default" options in avatar customization interfaces and the order (hierarchy) of body customization options oftentimes implicitly reinforces existing hegemonic structures in society [156, 170]. Avatar customization interfaces are known to constrain user choices, in part due to their oftentimes exclusionary design [155]. Previous research has found a limited number of options for players belonging to diverse ethnic groups and gender, suggesting that customization favors the creation of light-skinned male avatars [51, 156, 177]. While our focus in the present study is on understanding if aural avatar customization can confer similar benefits to visual avatar customization, the exclusionary potential of aural avatar customization options should be studied closely in future research.

Research has emphasized the role played by other aspects including game world aesthetics, co-situated players, social context, and avatars of other characters in influencing the avatar customization process [105, 153]. Kafai found that new players felt out of place with their generic avatars when interacting with avatars with detailed customization [105]. Players also reported customizing their avatars to avoid being bullied in online settings by other players [105]. Players customize their avatars differently depending on the context of the virtual environment, such as changing clothes and accessories when the social context switched between "game" and "job" [218]. Players also adhere to group norms while creating

and customizing their character [101]. User characteristics, such as age, gender and self-esteem, play a role in the avatar creation and customization process. Individuals with higher self (and body) esteem represent their avatars with a greater number of body details and emphasis on sexual characteristics that identified their gender [227]. Adolescent boys customized their avatars to create a more stereotypical masculine body compared to girls who focused on customizing transient aspects of the avatar, such as clothing and accessories [227].

Although the process of avatar customization has been extensively investigated, research has largely ignored the effect of voice options on avatar creation and customization. Contemporary games seldom offer voice customization options; however, there do exist some examples. Some games offer a "voice template" that can be chosen during avatar customization, such as in *Black Desert Online* [178]. *Sims 4* [68] allows characters' voices to be customized according to three voice qualities: "sweet," "melodic," and "lilted" for women, and "clear," "warm," and "brash" for men. Other games allow players to customize a given voice by directly changing specific aspects of the voice, such as pitch. The games *Saints Row IV* [230] and *Cyberpunk 2077* [38] offer the ability to modify pitch. This project investigates the effect of providing aural avatar customization options on a variety of player outcomes.

2.2 Audio in Games

Game audio performs many functions, such as emphasizing visuals [174], contextualizing a place [67], highlighting emotions and thoughts of the game-character [174], and immersing the player in the game world [193]. To understand the design of audio in games, researchers have defined audio typologies. One typology classifies sound based on the source [19]. Sound is referred to as "diegetic" if it originates from the game world (e.g., game sound [86, 169]), and sounds that have origins different than the game world (e.g., interface sounds) are called "non-diegetic" [86, 169]. Liljedahl [137] classifies sounds into three categories: speech and dialogue, sound effects (e.g., ambient noise, avatar sounds, object and ornamental sounds), and music.

Research shows that players appreciate the inclusion of audio elements in the game. Klimmt et al. [124] investigated the role of background music on gameplay experiences of players. Players experienced greater enjoyment while playing a game (*Assassins Creed: Black Flag* [222]) with background music included. Background music can also affect performance in a game—participants who played a role-playing adventure game (*The Legend of Zelda: Twilight Princess* [176]) performed better with background music present [205]. Some games incorporate background music that changes according to events in games. An adaptive soundtrack has also been shown to improve player experience. Researchers designed a game with a soundtrack that increased in tension depending on the chance of success or failure of players in the game [182]. Participants who played the game with an adaptive soundtrack experienced greater tension, suggesting a more engaging experience. Players playing a first-person shooter game reported higher game experience (immersion, flow, positive affect) with the presence of sound effects (e.g., ornamental and character sounds) [169]. Audio may also influence motivated behaviors such as time played [110] and actions

performed [108]. Lack of thematic fit between audio and visuals (also known as game atmosphere) can affect player experience. In a study, players played a survival horror game (*Bloodborne* [76]) either with background and voiceover audio relevant for the game (built-in game audio) or with experimenter-induced music and voiceovers [83]. Players experienced a lower degree of perceived game atmosphere when the audial elements did not fit the game’s visual elements.

Avatar sounds are sounds related to the avatar activity, such as breathing and footstep sounds [137]. These sounds help immerse the player into the game world [85], provide feedback for avatar movement [67], and play a crucial role in localizing the player in audio games [75, 78]. For example, Adkins et al. [3] developed an audio game wherein the players selected an animal as a game character—a cow, dog, cat, and frog—to navigate through a maze. The four animals also had representative animal sounds that provided essential user feedback for nearby obstacles and intersections. Providing sound cues for the movement of an avatar helps the virtual world conform to the players’ expectations [75] and induce immersion into the game world [85].

2.2.1 Avatar Voice. Avatar voice includes linguistic (e.g., dialogue and voiceovers) and non-linguistic vocalizations such as emotes (e.g., effort grunts, screams, sighs) [95]. Avatar voice can be used to control actions of the game character [8], converse with NPCs [60], and converse with other players in the game world [231]. While conversation with NPCs is usually supported through prerecorded dialogues [95], games also facilitate avatar control and player-to-player communication through voice interaction [37].

Voice dialogue in games supports storytelling, the development of a rich and believable world, and setting emotional tone [95]. As players explore and interact with a novel game world, conversing with NPCs can reveal important information regarding historical events and new quests that can ultimately help in the narrative progression. A common feature in many open-world games is the presence of a social space (e.g., local tavern) containing music and ambient sounds that are concurrent and continuous [206]. The social space also contains jumbled, indistinct conversations (Walla) among social actors (NPCs) [95]. Therefore, a sonic environment comprising of music, sounds, and voices helps in several ways: creating a game-feel, setting the mood, and making the game world believable [49, 95]. Game characters also use emotionally-laden dialogues to engender emotions in a player that can forge a deeper connection between the game character and the player [211]. For instance, an urgent request for help can arouse the player to take action.

Voice interaction focuses on using players’ voices as input in the game [8, 37]. Beyond using voice interaction to converse with other players [231], recent advances in software and hardware technology [7] have made it possible to use voice interaction to control avatar actions and in-game events [8]. Two popular approaches exist here: “voice-as-sound” [88, 99] and “voice-as-speech” [8, 37]. Voice-as-sound uses players’ voice characteristics such as pitch and tone [88, 99]. Hämäläinen et al. [88] describes the design of two games that used the voice-as-sound approach. The players navigate a hedgehog through a maze in the first game by singing at the correct pitch. The authors also developed a turn-based ping-pong

game where the players had to navigate their paddle at appropriate positions using the correct pitch. Voice-as-speech uses speech recognition technology to interpret players’ commands in games [7, 8, 88]. Players can use their voice to navigate menus [37], engage in unscripted conversations with a virtual pet (a fish in the game *Seaman* [228]) [8], and cast spells using voice commands in *Skyrim* [7, 23].

Carter suggests that voice interaction can facilitate a deeper connection with the players’ game characters [37]. The voice of an avatar is a part of game characters’ identity, and providing a way to use players’ voices for avatar actions can lead to a merging of identity (player-avatar convergence). Embodied identification, that is, the degree of control over the game characters’ movement and action, can imbue players with a greater sense of agency and identification [220]. Players playing *Tomb Raider* [54] can use voice commands to initiate player actions, such as attack and defend [37], while simultaneously performing (other) actions with the game controller. In this sense, voice interaction may facilitate embodied identification by affording greater control over game characters’ actions. Voice interaction may also facilitate wishful identification by affording associations between players’ voice and the game characters’ voice. *Splinter Cell Blacklist* [223] allows users to distract enemy NPCs by using a specific speech phrase (“Hey, you!”), which is repeated by the voice of the game character in the virtual world. Players in *FIFA 14* [65] embody the role of manager and perform actions such as selecting players for the tournament and giving advice on the field. Players can voice specific commands that change the behavior of their chosen team to adopt a defensive or attacking mindset [37]—a typical action that coaches and managers perform. Lastly, voice interactions can facilitate similarity identification by allowing users to interact with the game characters using their voices. For example, avatar representation in karaoke games is almost entirely through the voice of the player [37].

2.2.2 Avatar Voice and Learning Environments. Studies have investigated how engagement and learning outcomes are influenced by voice characteristics of the instructional agents [133, 150]. Learners rate voices more likeable when voice characteristics of instructional agents are similar to themselves in perceived gender [132] or personality [171, 172]. Research also documents persistent stereotypes in the design of instructional agents’ voices. Deutschmann [58] evaluated how students perceive a male and female avatar delivering a lecture. Students perceived the male avatar as more knowledgeable, and the female character as more likable. Along a similar line, authors designed three avatars—the instructor’s face, male-anime, and female-anime—to understand how students perceive and perform in an online course. Students showed higher likeability for the female-anime avatar but performed higher when instructors’ own face delivered lectures. Although these studies show that the voice of an avatar plays a role in students’ perception and performance, a general limitation is the poor quality of voice morphing in these studies [58, 97].

More recently, research has also sought to understand how an avatar’s voice can affect self-presentation in digital environments. Zhang et al. characterized users’ voice customization preferences on social media websites [242]. The study highlighted gender, personality, age, accent, pitch, and emotions as key factors that users wanted

to customize to represent their avatar in digital spaces [242]. The study also highlighted the need to provide customization options to modulate pitch and voice depending on the context—e.g., sounding serious and formal for professional websites such as LinkedIn. A common trend in studies leveraging personalized avatar voice in virtual environments is the beneficial effects of using a self-similar avatar voice [12, 114]. In a public speaking experiment, participants stood in front of a virtual classroom to give a speech [12]. Participants either used their own voice to give the speech or had another participant’s speech played back. Participants who used their own voice showed significantly higher social presence [12]. Kao, Ratan, Mousas, and Magana leveraged recent advances in voice cloning and found that learners using a more self-similar voice (as opposed to a self-dissimilar voice) in a game-based learning environment had higher performance, time spent, similarity identification, competence, relatedness, and immersion. Additionally, they found that similarity identification was a significant mediator between voice similarity and all measured outcomes [114].

While research provides strong support for avatar voice influencing avatar identification, no study (to the best of our knowledge) has investigated the effects of providing avatar aural customization options. We present a study that provides aural (voice) avatar customization options alongside visual avatar customization options in a Java programming game. Our goal is to understand how providing aural avatar customization options affect measured outcomes.

2.3 Hypotheses

We had seven overarching hypotheses (each broken down into three sub-hypotheses) in this study. All hypotheses and research questions were part of the study preregistration.² Because prior work has shown that avatar customization leads to an increase in avatar identification (similarity identification, embodied identification, and wishful identification) [25, 26, 57], we hypothesized that visual customization would lead to an increase in avatar identification. Research has shown that game audio is important to player experience (PX) [66, 67, 168] and that avatar audio can influence avatar identification [114]. Therefore, we hypothesized that aural customization would lead to an increase in avatar identification. Additionally, we hypothesized a lack of an interaction effect between visual and aural customization because existing work gives us no reason to believe their effects would depend on one another.

H1.1: Visual customization will lead to higher avatar identification.

H1.2: Aural customization will lead to higher avatar identification.

H1.3: No interaction effect between visual and aural customization for avatar identification.

Prior studies have shown that character customization leads to greater autonomy [25, 180]. Therefore, we hypothesized that visual customization would lead to greater autonomy. Similar to H1.2, we hypothesized that aural customization will play a similar role to visual customization and will also increase autonomy. We again hypothesized a lack of an interaction effect for the same reason as H1.3.

H2.1: Visual customization will lead to higher autonomy.

H2.2: Aural customization will lead to higher autonomy.

H2.3: No interaction effect between visual and aural customization for autonomy.

Prior work has shown that avatar customization is linked to intrinsic motivation [25], immersion [25], time spent playing [25], motivation for future play [180], and likelihood of game recommendation [180]. Furthermore, avatar identification and autonomy are increased through avatar customization (e.g., [25, 180]), and also affect intrinsic motivation, immersion, time spent playing, motivation for future play, and likelihood of game recommendation [25, 113, 180, 183, 198]. Therefore, we hypothesized a model in which visual customization directly, and indirectly through avatar identification and autonomy, influences intrinsic motivation, immersion, time spent playing, motivation for future play, and likelihood of game recommendation. Lastly, given the lack of prior work on aural customization, we posed as research questions (without any formal hypotheses) whether aural customization moderated any of these effects.

H3.1: Visual customization will lead to higher intrinsic motivation.

H3.2: Avatar identification will mediate H3.1.

H3.3: Autonomy will mediate H3.1.

H4.1: Visual customization will lead to higher immersion.

H4.2: Avatar identification will mediate H4.1.

H4.3: Autonomy will mediate H4.1.

H5.1: Visual customization will lead to higher time spent playing.

H5.2: Avatar identification will mediate H5.1.

H5.3: Autonomy will mediate H5.1.

H6.1: Visual customization will lead to higher motivation for future play.

H6.2: Avatar identification will mediate H6.1.

H6.3: Autonomy will mediate H6.1.

H7.1: Visual customization will lead to higher likelihood of game recommendation.

H7.2: Avatar identification will mediate H7.1.

H7.3: Autonomy will mediate H7.1.

Research Question: Does aural customization moderate H3–H7?

3 EXPERIMENTAL TESTBED

Our experimental testbed is CodeBreakers⁴ [109], which was created for conducting avatar-based studies. CodeBreakers is a Java programming game in which players solve increasingly difficult problems by throwing snippets of code. See Figure 1. CodeBreakers was iteratively created with feedback from professional game developers, game designers, and Java developers, and it included informal play testing over an eighteen-month span with playtesters. There were 14 total puzzles, spanning 6 levels. CodeBreakers was designed to incorporate best practices on effective learning curves [142]. Programming topics include data types, conditionals and control flow, classes and objects, inheritance and interfaces, loops and recursion, and data structures. Each puzzle had up to 3 hints, which are increasingly detailed. Players controlled their character using the keyboard and mouse. CodeBreakers was originally developed for Microsoft Windows and macOS. However, for the purposes of

³Note that the avatar model color was changed to gray for this study. See Section 4 for details.

⁴Gameplay video: <https://youtu.be/x5U-Jd6tKXA>.

²<https://osf.io/dbvbk/>.



Figure 1: Data type puzzle (L). Curing a wounded knight (R). Placeholders indicate where code snippets can be thrown.³

this experiment, CodeBreakers was converted to WebGL and was therefore playable on any PC inside of the browser (e.g., Chrome, Firefox, Safari). See Section 4.4.1 for details. In total, there were 30 possible voice lines that could have been triggered. Other than the first voice line (*What am I doing here? Did my ship crash? How long have I been lying here for? I guess I should get up and look around.*), audio lines typically come before and after each puzzle. For example, prior to puzzle #7: *The castle is under siege!*. And after completing puzzle #7: *It worked! I neutralized all of the bugs by using the staff.* These voice lines were accompanied by speech bubbles (see Figure 2).

4 METHODS

For this study, we explicitly aimed to create stereotypically-appearing (and sounding) “male” and “female” avatars. We created four avatar appearances (two male and two female) and four avatar voices (two male and two female). We made these design decisions with an understanding that a binary view of gender is problematic, but we did so for ecological validity with the majority of existing games. While it would have been possible to create a more inclusive set of gender choices, this might present as a possible confound as such choices are not currently available in most of today’s games. Our goal is to develop a baseline understanding of the presence of customization choices that mirror current games. Such baseline understandings can inform future avatar customization research and implementation, in which we hope that more inclusive design choices become the norm. Finally, our rationale for creating two visual choices and two audial choices for each gender was to add a (minimal) degree of visual and audial choice.

4.1 Model Development

All four models used in this experiment were designed and created from scratch by a professional 3D game artist. The models were purposefully designed to avoid known color effects (e.g., the color red is known to reduce mood, affect, and performance in cognitive-oriented tasks [84, 98, 111, 127, 158, 159]). We chose gray because it matched the aesthetic of the game and is not associated with negative physiological effects on cognition and heart rate variability

(HF-HRV) [69]. All four models shared the same identical skeleton and joints, and therefore all animations (i.e., idle, walking, picking up code, throwing code, using weapons, falling, dying, stopped in front of a wall, etc.) were identical across the four models. Only visual appearance differed. See Figure 3.

4.2 Voice Development

4.2.1 Voice Development Goal. Our goal was to create four avatar voices (two stereotypical male and two stereotypical female). We wanted each voice to be appropriate for the game *and* to be appropriate for either of the two models from the same gender. Additionally, we wanted each male voice to have a “matching” female voice as rated on a scale of perceived vocal dimensions—e.g., strong vs. weak, smooth vs. rough, resonant vs. shrill [82].⁵ In other words, we wanted these matched voices to sound as similar as possible. The reason this matching was done was to mitigate confounds from large differences between voices. High variance between voices would add an additional dimension to the manipulation which could influence the study results. Nevertheless, we wanted both male voices to be distinct from one another and both female voices to be distinct from one another. If this were not the case (e.g., both male voices sounded the same), then our manipulation of giving users a *choice* of voice would only be illusory.

4.2.2 Creating Voices. We hired two professional voice actors with over ten years of experience in character voice acting. Both voice actors were screened through their portfolios, which contained samples of their work. Both voice actors provided sample voice clips for CodeBreakers prior to being hired. We decided on hiring two voice actors instead of four because: (1) we could ensure greater overall consistency across voices, helping to bound the variance across voices and (2) both voice actors had demonstrated evidence of being able to perform a multitude of different voices and characters, assurance that each voice actor could produce two unique-sounding voices. Both voice actors self-identified as white and have lived in the U.S. for their entire lives. One voice actor self-identified as male and was 49 years old. The other voice actor self-identified as female and was 38 years old. The two voice actors were instructed

⁵We discuss this scale in more detail in the validation section below.

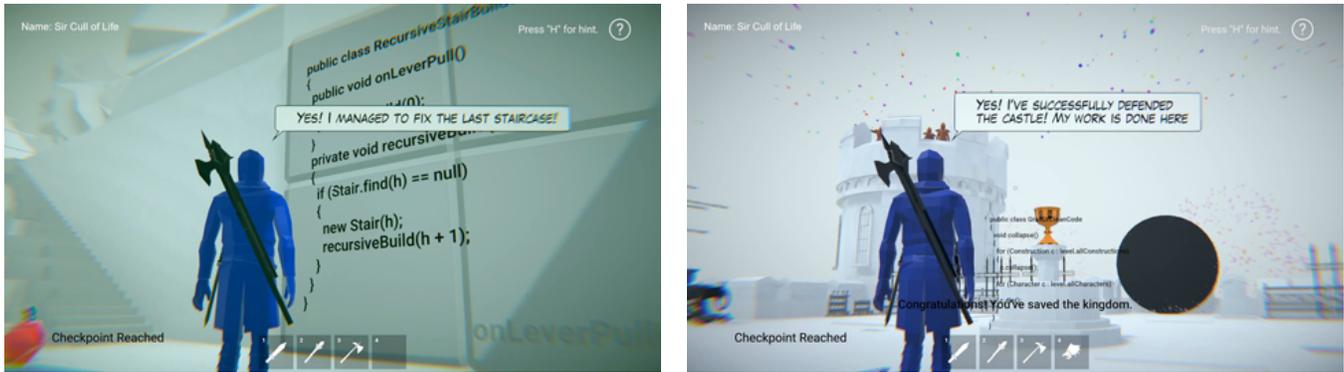


Figure 2: Voice audio occurs in conjunction with speech bubbles that appear on top of the avatar.³



Figure 3: Front view (L) and back view (R) of the four models.

to work together to create two “matching” voice pairs as described in Section 4.2.1. Our goals for the four voices, including the scale of vocal dimensions [82], were clearly articulated to the voice actors. Additionally, both voice actors familiarized themselves with the game by watching video gameplay of CodeBreakers. Both voice actors were also shown the four models that they were voicing. All voices were recorded in the same professional audio recording studio with both voice actors physically copresent. Identical recording equipment and software was used for recording each voice clip: Sennheiser MK-416 (microphone), Universal Audio Arrow (audio interface), and Ableton Live 10 (digital audio workstation). Completed voice clips were reviewed by the project team, and several iterations were made on the voice clips to ensure that our criteria in Section 4.2.1 appeared to be satisfied. A total of 120 voice clips (30 per voice) were recorded and finalized. Sample audio clips can be found at <https://osf.io/mnpsd/>. M1 is male voice one, M2 is male voice two, F1 is female voice one, and F2 is female voice two.

4.2.3 Voice Loudness Normalization. While the same identical recording studio and recording equipment was used for recording each voice, it is possible that relative amplitude (i.e., loudness) could differ between voices, especially between the two different voice actors. To normalize loudness across all voices and voice clips, we adopted the *EBU R 128* (issued by the European Broadcasting Union) standard’s recommendation for loudness normalization [71]. It recommends normalization of audio to -23 ± 0.5 Loudness Units

Full Scale (LUFS), and a max peak of -1 decibel True Peak (dBTP). A professional audio engineer with 15+ years of experience performed this normalization using Nuendo 11 Pro and verified that the loudness normalization recommendation was satisfied.

4.3 Voice Validation

4.3.1 Expert Voice Validation. To ensure that we had created two distinct matching pairs of voices (similarity *within* each pair but variance *between* them), we hired three expert speech pathologists to evaluate each voice. Each speech pathologist was given instructions to listen to a set of voices then asked to rate each voice on a scale. Each speech pathologist was compensated \$25. Speech pathologists all had at least 10 years of professional speech pathology experience ($M=20.0$, $SD=8.19$), with an average age of $M=47.67$ ($SD=4.93$). Before rating the voices, each speech pathologist was instructed to familiarize themselves with the validated scale on perceptual attributes of voice [82].⁶ This scale consists of 17 items, and all items are rated on a Likert scale from 1 to 9. Anchor points for each item are listed in Table 1. Each speech pathologist was provided the 30 voice clips associated with each voice, and each was asked to listen to the entire set of clips belonging to a single voice before rating that voice. Speech pathologists performed the ratings

⁶The scale has been used with speech pathologists revealing modest within-group agreement despite absence of any training in interpretation of the scale descriptors [82].

LOWER ANCHOR—UPPER ANCHOR	M1 (SD)	M2 (SD)	F1 (SD)	F2 (SD)
High Pitch—Low Pitch	6.33 (0.58)	8.00 (0.00)	3.33 (0.71)	4.33 (0.58)
Loud—Soft	4.67 (1.53)	4.67 (2.31)	4.67 (1.41)	4.00 (1.73)
Strong—Weak	2.00 (0.00)	2.33 (1.53)	3.33 (0.71)	3.00 (1.00)
Smooth—Rough	2.33 (0.58)	4.00 (1.73)	2.00 (0.00)	3.67 (1.15)
Pleasant—Unpleasant	1.67 (0.58)	2.33 (0.58)	1.67 (0.71)	3.00 (0.00)
Resonant—Shrill	2.67 (0.58)	1.67 (0.58)	3.67 (2.83)	3.33 (1.15)
Clear—Hoarse	2.33 (0.58)	3.67 (2.89)	2.33 (0.71)	3.67 (2.89)
Unforced—Strained	3.00 (1.00)	4.33 (2.52)	3.00 (0.71)	3.67 (1.53)
Soothing—Harsh	3.33 (0.58)	2.67 (0.58)	2.67 (0.71)	3.33 (1.53)
Melodious—Raspy	3.33 (0.58)	4.33 (2.08)	2.33 (0.00)	4.67 (0.58)
Breathy Voice—Full Voice	7.00 (1.73)	8.33 (0.58)	5.00 (2.83)	7.00 (1.00)
Excessively Nasal—Insufficiently Nasal	5.00 (0.00)	5.00 (0.00)	5.00 (0.00)	4.00 (1.00)
Animated—Monotonous	1.67 (0.58)	4.67 (1.53)	1.67 (0.00)	4.00 (1.73)
Steady—Shaky	2.00 (0.00)	2.33 (0.58)	2.33 (0.00)	2.33 (0.58)
Young—Old	4.33 (0.58)	5.67 (0.58)	3.33 (0.71)	4.33 (1.15)
Slow Rate—Rapid Rate	4.67 (0.58)	5.33 (0.58)	5.33 (0.71)	5.33 (0.58)
I Like This Voice—I Do Not Like This Voice	1.67 (1.15)	2.00 (1.00)	1.67 (1.41)	3.33 (1.53)

Table 1: Mean expert speech pathologist ratings for each voice. All items are rated on a 9-pt Likert scale from 1:Lower Anchor to 9:Upper Anchor.

using their own computers, and they were asked to use the most professional audio equipment available to them to perform the evaluation. Across the three speech pathologists’ ratings, we calculated the intraclass correlation to be $ICC=0.83$, 95% $CI[0.75, 0.89]$ (two-way mixed, average measures [203]), indicating high agreement. Mean ratings for each voice can be seen in Table 1. As a measure of similarity between voices, we then calculated an *absolute mean difference* across the scale between every possible pair of voices. As expected, this difference was lower in the two matched pairs (M1/F1: $M=0.67$; M2/F2: $M=0.88$) when compared to mismatched pairs (M1/F2: $M=2.33$; M2/F1: $M=1.41$) or to same-gender pairs (M1/M2: $M=1.08$; F1/F2: $M=0.98$). Although the same-gender pairs have an absolute mean difference close to the two matched pairs, we attribute some of this due to voice attributes that are often-times known to vary naturally between genders (e.g., pitch [30]). Nevertheless, one potential concern arising from these results is that the same-gender voices may not be perceived as distinct from one another. Therefore, we performed an additional crowdsourced validation.

4.3.2 Crowdsourced Voice Validation. To ensure that we had created two distinct matching pairs of voices, that all voices would be perceived as as being high quality, that voices would be perceived as the stereotypical intended gender, and that voices across the same gender would be perceived as unique and distinct voices, we ran a crowdsourced validation study. This was to reinforce and extend the prior expert validation. We recruited 91 participants (39% self-identified as female) on MTurk to rate voices based on sets of audio clips. Each participant was compensated \$1.00 (USD). Participants had a mean age of 40.62 ($SD=13.82$). All participants were from the U.S. After filling out a consent form, each participant was first presented with, randomly, either a stereotypical male or female voice clip of an English word, which they needed to type correctly. This was to ensure that the participant’s audio was turned on and working. Each of the following questions was equipped with analytics that tracked the amount of time that each participant spent listening to audio clips. These analytics were used to validate

that participants had actually listened to the audio clips before answering the questions. ~10% of participants were removed for not having listened to all audio clips in the study in their entirety.

Participants were then asked to “Please listen to ALL of the following audio clips before answering the question below comparing the first (left-side) and second (right-side) voices.” And to rate: “Besides gender-related voice characteristics, I consider these two voices as similar,” on a scale of 1:Strongly Disagree to 7:Strongly Agree. This question was asked four times comparing the following pairs of voices in a randomized order: M1/F1, M2/F2, M2/F1, and M1/F2. For each comparison, 5 voice clips were selected at random (from the total 30), and those same 5 voice clips were shown for both of the two voices being compared (i.e., the same speech dialog).⁷ Results indicated that matched pairs (M1/F1: $M=5.51$, $SD=1.50$; M2/F2: $M=4.92$, $SD=1.68$) were rated to be more highly similar to one another than unmatched pairs (M1/F2: $M=4.01$, $SD=1.64$; M2/F1: $M=3.13$, $SD=1.71$).

Participants were then asked to “Please listen to ALL of the following audio clips. All clips belong to one voice. After listening to all of the clips, you will be asked a question regarding the voice.” And to rate: “Based on the voice you just listened to, please rate the following: The voice is high-quality,” “The speaker sounds (stereotypically) male,” and “The speaker sounds (stereotypically) female” on a scale of 1:Strongly Disagree to 7:Strongly Agree. This question was asked for each of the four voices in randomized order. For each voice, 5 voice clips were selected at random (from the total 30). Results indicated that all voices were perceived to be relatively high quality (M1: $M=6.02$, $SD=0.80$; F1: $M=6.06$, $SD=0.98$; M2: $M=5.80$, $SD=1.12$; F2: $M=5.60$, $SD=1.08$) and that voices sounded stereotypically male (M1: $M=6.74$, $SD=0.51$; F1: $M=1.20$, $SD=0.56$; M2: $M=6.85$, $SD=0.39$; F2: $M=1.34$, $SD=0.89$) or female (M1: $M=1.32$, $SD=0.77$; F1: $M=6.79$, $SD=0.44$; M2: $M=1.15$, $SD=0.52$; F2: $M=6.70$, $SD=0.55$) as intended.

Participants were then asked to “Please listen to ALL of the following audio clips before answering the question below comparing the first (left-side) and second (right-side) voices.” And to rate: “In comparing the two voices above (left audio clips vs. right audio clips),

⁷Note that randomization is done per participant and per question, so the 5 voice clips selected vary both across questions and across participants.

please rate the following: These two voices are distinct and different from one another,” on a scale of 1:*Strongly Disagree* to 7:*Strongly Agree*. This question was asked twice for voices in each gender (M1/M2 and F1/F2) in a random order. For each comparison, 5 voice clips were selected at random. Results indicated that same-gender voice pairs were perceived to be relatively distinct (M1/M2: $M=5.78$, $SD=1.07$; F1/F2: $M=5.73$, $SD=1.30$). Participants then entered demographic information.

4.4 Model and Voice Integration

4.4.1 WebGL Conversion and Technical Testing. Over 4 months, the original CodeBreakers game, which is playable on machines running either Microsoft Windows or macOS [114], was converted to WebGL to allow for a more convenient play experience. The WebGL version is playable on any PC inside of the browser (e.g., Chrome, Firefox, Safari). This conversion was performed by a professional game development team with expertise in game optimization. During the conversion process, we iterated on the game internally every few days and externally every few weeks. Our main goal during these iterations was to ensure that performance (e.g., frames per second) was adequate and that there were no technical issues (e.g., crashing). Internal iterations were performed by the development and research team where feedback was fed into the next iteration. Performance profiling tools were used extensively to diagnose areas of the game (e.g., code loops, rendering of certain geometry) responsible for increased CPU and memory usage. External iterations were performed when we wanted the game to be tested more widely. We performed iterations with batches of 10-20 participants at a time on MTurk. Participants were asked to play the *entire* game and were provided a walkthrough video in case they were unable to progress. This ensured that each participant would cover the breadth of the entire game. Data, including gameplay metrics, performance, crash logs, and PC details, was automatically logged on the server for further analysis. Participants could report any issues, problems, or concerns they experienced during playtesting. A total of 121 participants, all from the U.S., took part in external playtesting. Each participant was compensated \$10 (USD). Our testing ended when no new technical issues arose in the most recent internal and external iterations, all known technical issues were fixed, and the game performed adequately (e.g., frames per second, load times) under a wide variety of PCs. Additionally, the development and research team agreed that, for all intents and purposes, the WebGL game played and felt identical to the original.

4.4.2 Character Customization UI. A professional game UI designer created four different character customization screens that we requested. These also correspond to our experimental conditions. (See Figure 4.) We made the explicit design decision never to allow mismatched model-voice gender pairings (i.e., male model and female voice or vice versa), since this may be unnatural for players, lacks general ecological validity with existing games, and may be an experimental confound (e.g., in conditions where one or both features are assigned at random). Therefore, avatar customization is, in all cases, a two-step process that involves first choosing or being assigned a model (one of four), then choosing or being assigned a voice (one of two since the model has already

been selected, and there are only two voices corresponding to the designed stereotypical gender).

In *Choice-None*, the player does not have any choice over the model or voice. Both model and voice are randomly assigned. In *Choice-Audio*, a model is randomly preselected, and a player is able to choose the voice. In *Choice-Visual*, the player chooses a model, after which the voice is randomly assigned. In *Choice-All*, the player chooses both model and voice. Note that the two voices corresponding to “Voice 1” and “Voice 2” will differ depending on the model selected. In *Choice-All*, both voice options are grayed out and unavailable until a model has been selected. If a different model is selected after a voice has been selected, the voice is automatically deselected. In all conditions, players must enter a name for their character. For conditions that allow for a model choice (*Choice-Visual* and *Choice-All*), the UI initially shows an empty box where the selected model would normally appear (i.e., no model is selected by default). For conditions that allow for a voice choice (*Choice-Audio* and *Choice-All*), no voice is selected by default (i.e., one of the two voices must be selected manually by the player). When a voice is selected, a single audio clip is played from that voice so that players can compare voices. In all conditions, players must complete all customization options available (e.g., name, model, voice) before the “Start Game” button becomes available. Character customization conditions were designed in this manner to minimize differences between conditions, while still varying the manipulations (visual choice and aural choice).

4.4.3 Expert UI Validation. To assess the appropriateness of our character customization UIs, we performed a validation study with three professional game UI designers. Game UI designers were recruited from the online freelancing platform Upwork, and were each paid \$20 (USD). The job posting was *Assess Character Customization Interface in Educational Game*, and the job description stated that we were looking for expert game UI designers to evaluate a set of character customization interfaces in an educational game. The three UI designers had an average of 9.00 ($SD=4.36$) years of UI design experience and an average of 7.67 ($SD=5.69$) years of game development experience. UI designers all had work experience and portfolios that reflected recent UI design and game development experience (all within one year). UI designers were instructed to give their honest opinions and were told their responses would be anonymous, and proceeded to our survey. Each UI designer was first asked to watch 30 minutes of gameplay footage from CodeBreakers to familiarize themselves with the game. Afterwards, each designer loaded CodeBreakers WebGL on their own machine and interacted with every version of the UI in a randomized order. After interacting with a specific version of the UI, the UI designer was asked to rate “*The character customization interface is appropriate for the game,*” on a scale of 1:*Strongly Disagree* to 7:*Strongly Agree*. UI designers were asked to rate each interface individually, *not* in comparison to the other interfaces they had already seen. UI designers were also able to report open-ended feedback. The survey took approximately 1.5 hours to complete. Responses showed that UI designers generally agreed that the character customization interface was appropriate (*Choice-None*: $M=6.67$, $SD=0.58$; *Choice-Audial*: $M=6.33$, $SD=1.16$; *Choice-Visual*: $M=6.33$, $SD=1.16$; *Choice-All*: $M=7.00$, $SD=0.00$). One UI designer did note as open-ended feedback that they had not

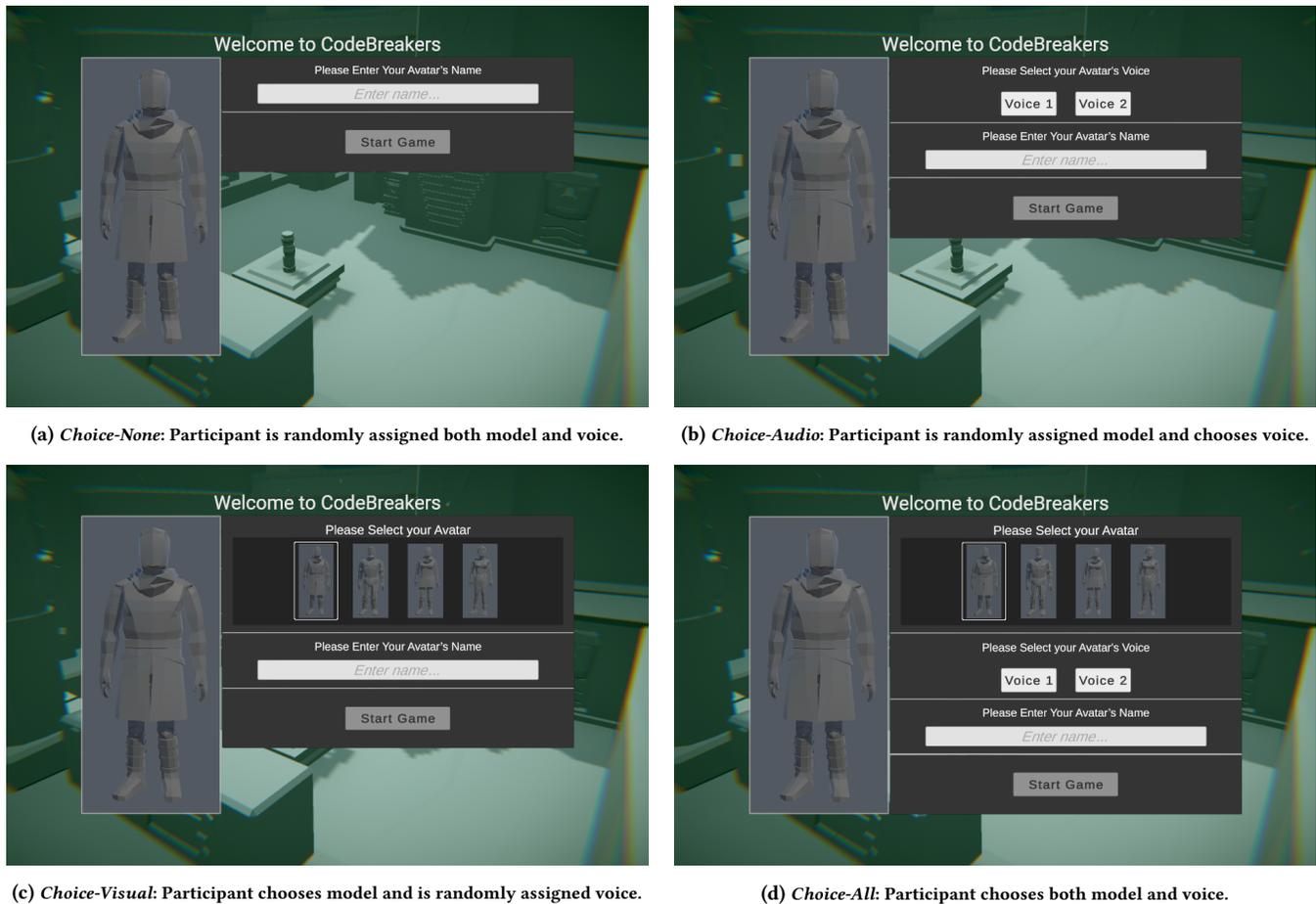


Figure 4: Avatar customization screens.

expected to be able to choose a voice for their character since this is not a commonly available feature in games, but it was stated that this did not play a role in the designer's ratings.

4.4.4 Model and Voice Integration Validation. To assess whether the models and voices that we had developed would be perceived as appropriate for the game, we recruited 120 participants (43% female) on MTurk. All 120 participants played CodeBreakers using the *Choice-None* condition (i.e., randomly assigned model and voice). Participants played the game for a minimum of 5 minutes, but they were allowed to play as long as they liked beyond the 5-minute mark. Random assignments were roughly even across models (24.2%/24.2%/32.5%/19.2%) and voices (24.2%/27.5%/26.7%/21.7%). For the remainder of this section, ratings described for models follow the left-to-right order of models shown in Figure 3. See Figure 5 for graphs summarizing the validation results.⁸

To assess whether models overall visually fit the game, we asked, "How appropriate were your avatar's visual characteristics for the game?" on a scale from 1:*Inappropriate* to 5:*Appropriate*. Scores tended between neutral and appropriate for each model ($M=4.24$,

⁸All validation questions are found in the graphs except for "How appropriate was the avatar design overall?" for which summary statistics are provided in the text.

$SD=0.83$; $M=3.86$, $SD=0.79$; $M=4.18$, $SD=0.76$; $M=4.04$, $SD=0.77$). To assess whether voices overall audially fit the game, we asked "How appropriate was your avatar's voice for the game?" on a scale from 1:*Inappropriate* to 5:*Appropriate*. Scores again tended between neutral and appropriate for each voice ($M1: M=3.91$, $SD=0.82$; $M2: M=4.46$, $SD=0.65$; $F1: M=4.55$, $SD=0.57$; $F2: M=3.97$, $SD=0.98$). To assess whether models and voices in combination fit the game, we asked, "How appropriate were both the visual and audial characteristics combined of your avatar for the game?" on a scale from 1:*Inappropriate* to 5:*Appropriate*. Scores again tended between neutral and appropriate for each model ($M=4.21$, $SD=0.77$; $M=4.00$, $SD=0.80$; $M=4.31$, $SD=0.69$; $M=4.04$, $SD=0.93$) and for each voice ($M1: M=3.88$, $SD=0.79$; $M2: M=4.39$, $SD=0.70$; $F1: M=4.35$, $SD=0.67$; $F2: M=4.09$, $SD=0.88$). To assess whether models' individual visual features (color and clothing) were appropriate for the avatar, and for the game, we asked, "How appropriate was the avatar color for the game?", "How appropriate was the avatar color for the avatar?", "How appropriate was the avatar clothing for the game?", "How appropriate was the avatar clothing for the avatar?", and "How appropriate was the avatar design overall?", on a scale from 1:*Inappropriate* to 5:*Appropriate*. Overall scores were between neutral and appropriate

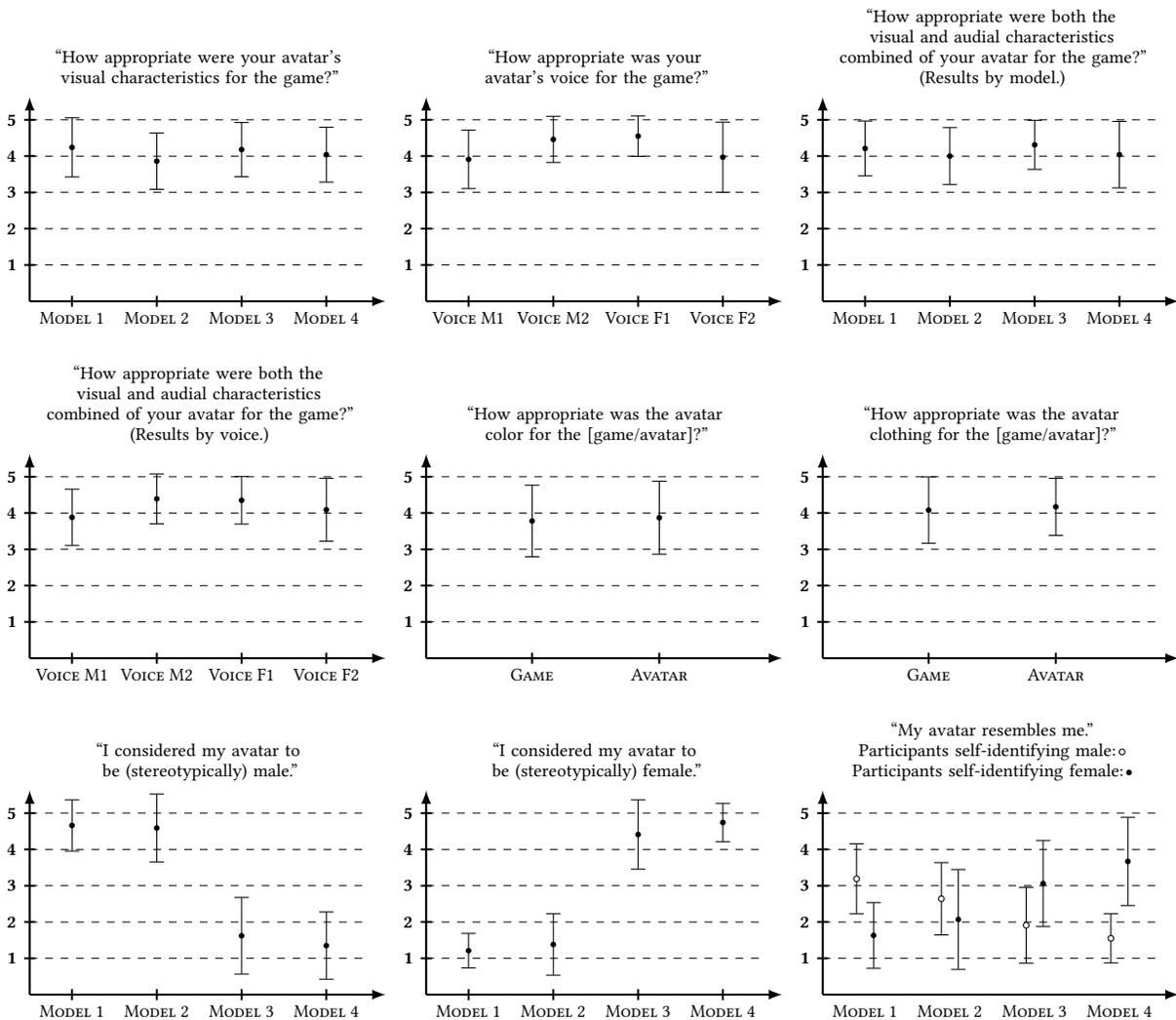


Figure 5: Model and voice validation summary graphs. Error bars show \pm SD.

for appropriateness of avatar color (Game: $M=3.78$, $SD=1.00$; Avatar: $M=3.87$, $SD=1.02$), avatar clothing (Game: $M=4.08$, $SD=0.93$; Avatar: $M=4.17$, $SD=0.80$), and avatar design overall ($M=4.06$, $SD=0.87$). To assess whether models were perceived as the stereotypical gender we had designed them to be, we asked, “I considered my avatar to be (stereotypically) male,” and “I considered my avatar to be (stereotypically) female,” on a scale from 1:Strongly Disagree to 5:Strongly Agree. Participants rated the models designed to be stereotypically male as male ($M=4.66$, $SD=0.72$; $M=4.59$, $SD=0.95$; $M=1.62$, $SD=1.07$; $M=1.35$, $SD=0.94$) and models designed to be stereotypically female as female ($M=1.21$, $SD=0.49$; $M=1.38$, $SD=0.86$; $M=4.41$, $SD=0.97$; $M=4.74$, $SD=0.54$). To assess whether avatars bore a visual similarity with players, we asked participants to rate “My avatar resembles me,” on a scale from 1:Strongly Disagree to 5:Strongly Agree. Participants who self-identified as male had scores tending towards neutral for male models ($M=3.19$, $SD=0.98$; $M=2.64$, $SD=1.01$; $M=1.91$, $SD=1.06$; $M=1.55$, $SD=0.69$) while participants who self-identified as female

had scores tending towards neutral for female models ($M=1.63$, $SD=0.92$; $M=2.07$, $SD=1.39$; $M=3.06$, $SD=1.20$; $M=3.67$, $SD=1.23$). As expected, participants in general did not find close visual similarity with their avatars (likely in part due to their abstract design), with some natural variation across avatars and gender.

4.5 Study Preregistration

Our study was preregistered on the Open Science Framework (OSF). Hypotheses, exploratory analyses, experiment design, data collection, sample size, and measures are contained in our preregistration.⁹

⁹Preregistration: <https://osf.io/dbvvp/>.
Raw Data: <https://osf.io/mnpsd/>.

4.6 Conditions

The study uses a 2 x 2 factorial design. We manipulate visual choice (choice vs. assignment) and audial choice (choice vs. assignment). The manipulations are as follows:

- **Choice-None:** Participant is randomly assigned both model and voice.
- **Choice-Audio:** Participant is randomly assigned model and chooses voice.
- **Choice-Visual:** Participant chooses a model and is randomly assigned voice.
- **Choice-All:** Participant chooses both model and voice.

The only difference between each of these conditions is the character customization interface that appeared at the beginning of the game, which manipulated choice vs. assignment for model and voice. See Figure 4 and Section 4.4.2 for details on how the character customization interface was implemented in these different conditions. All other aspects of the experiment were identical across conditions.

4.7 Measures

In line with best practices on measurement reporting, we report *what* we are measuring, *how* we are measuring, and *why* are we measuring in this way [4].

4.7.1 Avatar Identification (Player Identification Scale). Avatar identification is a “temporary alteration of media users’ self-concept through adoption of perceived characteristics of a media person” [44]. For measuring avatar identification, we use the player identification scale (PIS) [224]. The PIS measures three dimensions of avatar identification on a 5-pt Likert scale (1:*Strongly Disagree* to 5:*Strongly Agree*): similarity identification (e.g., “My character is similar to me”), embodied identification (e.g., “In the game, it is as if I become one with my character”), and wishful identification (e.g., “I would like to be more like my character”). We use the PIS as it has been validated [224] and is used extensively in the HCI literature on avatars—e.g., [25].

4.7.2 Autonomy (Player Experience of Need Satisfaction). Autonomy is the sense that one has volition and is doing activities for interest and personal value [198]. We use the PENS scale [198] to measure autonomy on a 7-pt Likert scale (1:*Do Not Agree* to 7:*Strongly Agree*)—e.g., “The game provides me with interesting options and choices.” We use the PENS autonomy subscale as it has been empirically validated on multiple occasions—e.g., [104].

4.7.3 Intrinsic Motivation (Intrinsic Motivation Inventory). Intrinsic motivation is one’s willingness to engage in an activity because the activity is satisfying in and of itself [196]. The subscale of the IMI, interest/enjoyment, is the primary measure of intrinsic motivation used in the research literature [196]. This is due to both interest and enjoyment being strong contributors to intrinsic motivation [188]. Items are rated on a 7-pt Likert scale (1:*Not At All True* to 7:*Very True*)—e.g., “I enjoyed doing this activity very much.” We chose to use the IMI to measure intrinsic motivation since it is well validated [157].

4.7.4 Immersion (Player Experience Inventory). Immersion is a *sense of immersion and cognitive absorption, experienced by the player* [2]. We use the Player Experience Inventory (PXI) to measure immersion, which uses three items to measure immersion on a 7-pt Likert scale, from -3:*Strongly Disagree* to +3:*Strongly Agree*—e.g., “I was no longer aware of my surroundings while I was playing.” We use the PXI immersion subscale, since it has been extensively validated and was designed specifically for games user research [2].

4.7.5 Motivated Behavior (Time Played). We operationalize motivated behavior as the time spent playing the game. Time on task is a behavioral measure that has been linked to motivation [197, 200] and is an objective measure of motivation in this study. Note that in the current study, participants are *required* to play at least 10 minutes, after which playing longer is optional.

4.7.6 Motivation For Future Play and Likelihood of Game Recommendation. Both motivation for future play and likelihood of game recommendation are measured using questions identical to a previous study [180]. Specifically, motivation for future play was measured using three items based on Ryan, Rigby, and Przybylski [198]: “Given the chance I would play this game in my free time,” “I would like to spend more time playing this game,” and “I would like to continue playing this game” [180]. Participants rated the three items on a 7-pt Likert scale from 1:*Strongly Disagree* to 7:*Strongly Agree*. Likelihood of game recommendation was assessed using the question “How likely would you be to recommend this game to others?” on a 7-pt Likert scale from 1:*Extremely Unlikely* to 7:*Extremely Likely* [180]. These measures allow us to understand how willing a player is to come back to a game, and how willing a player is to recommend the game to others. Both of these measures have been frequently used in the literature—e.g., PENS autonomy has been shown to positively predict both motivation for future play and likelihood of game recommendation in prior studies [180, 184, 198]. Motivation for future play showed good reliability, $\alpha=0.98$.

4.8 Sample Size Determination

To calculate a priori sample size, we perform two separate sample size determination calculations (both of these are specified in our preregistration at <https://osf.io/dbvvp/>). The first calculation is based on a 2 x 2 ANOVA for testing H1 and H2. G*Power 3.1 was used to perform this calculation using an effect size of small (0.1), $\alpha=0.05$, and 95% power. G*Power 3.1 found that a sample size of $N=1302$ would be required [129].

For H3, H4, H5, H6, and H7, our sample size calculation is based on moderated mediation analyses. We performed Monte Carlo simulations in R. We first specify the complete model (i.e., containing X, M1, M2, M3, M4, W, and Y with the appropriate relationships) using the *lavaan* package, with parameter estimates of 0.1 (e.g., correlations between variables). We then use the *simsem* package to create Monte Carlo simulations using 1000 bootstraps.¹⁰ These simulations provided estimations of statistical power for each path, from which we use the lowest power value from all paths as the cutoff. We modified sample size iteratively (± 10) until the necessary power was reached. We performed 10 simulations to confirm that

¹⁰This is considered well above the number of iterations needed: <https://kb.palisade.com/index.php?pg=kb.page&id=125>.

a specified sample size would reach the desired minimum power. The random number generator's seed was re-randomized for every simulation. These Monte Carlo simulations determined that, for a power of 95% and a confidence level of 95%, a sample size of 1500 would be necessary.¹¹ Therefore, to ensure the necessary power across both sample size determinations ($N=1302$ and $N=1500$), we use $N=1500$.

4.9 Participants

We recruited 1527 (47.6% female, 1.2% gender variant, 0.4% transgender) participants with an average age of $M=37.26$ ($SD=11.14$) from MTurk.¹² Workers on MTurk complete Human Intelligence Tasks (HITs), including research experiments. Studies show that MTurk provides data of similar quality [33], diversity [18, 39, 96], and reliability [33, 149] as typical samples (e.g., college students). Participants were each paid \$5.00 (USD). The HIT was available to workers in the U.S. over the age of 18 who had a computer with working audio. For quality control, workers were required to have a HIT approval rate >95%. The Purdue University Institutional Review Board (IRB) approved the study. All participants were asked to provide informed consent.

4.9.1 Data Screening. We screened all participants' responses. Specifically, we carefully screened participants' who had at least three survey measures with zero variance (excluding likelihood of game recommendation, since this was only a single question) or with $\pm 3SD$. A fairly large number of respondents met the criteria of at least three survey measures with zero variance (~40%),¹³ and these responses were scrutinized further (e.g., reverse-coded items and open-ended questions). All responses were deemed legitimate, except for one respondent who responded to all questions (including reverse-coded items) with the same answer. This respondent was removed from further analysis ($N=1526$ remaining participants).

4.9.2 Experience With Video Games and Programming. Participants reported playing an average of $M=8.5$ ($SD=10.5$) hours of video games per week, approximately matching the global average of $M=8.45$ [139]. On a scale from 1: *Minimal* to 7: *Extensive*, participants rated their prior experience playing video games ("How would you rate your prior experience playing video games?") as $M=4.72$ ($SD=1.81$) and their prior programming experience ("How would you rate your prior programming experience?") as $M=2.34$ ($SD=1.64$). Next, we adapted several questions on programming experience from [204]. On a scale from 1: *Very Inexperienced* to 5: *Very Experienced*, participants rated their programming experience compared to experts ("How do you estimate your programming experience compared to experts with 20 years of practical experience?") as $M=1.34$ ($SD=0.85$), their programming experience compared to beginners ("How do you estimate your programming experience compared to beginner programmers?") as $M=2.05$ ($SD=1.18$), their

programming experience in Java specifically ("How experienced are you with the Java programming language?") as $M=1.56$ ($SD=0.93$), and their experience with an object-oriented paradigm ("How experienced are you with the object-oriented programming paradigm?") as $M=1.72$ ($SD=1.13$). Therefore, our sample contains participants who are regularly exposed to video games and have low prior programming experience. ANOVAs found that there were no significant differences between conditions on prior gaming experience ($F[3, 1522]=0.422$, $p=0.737$, $\eta_p^2=0.001$), programming experience ($F[3, 1522]=0.264$, $p=0.851$, $\eta_p^2=0.001$), and Java programming experience ($F[3, 1522]=0.263$, $p=0.852$, $\eta_p^2=0.001$).

4.10 Design

A between-subjects factorial design was used. Each participant was randomly assigned to one of four possible conditions. Participant counts in each condition were approximately equal ($M=381.5$, $SD=5.8$).

4.11 Procedure

Participants first filled out an IRB-approved consent form. Participants were informed that they could exit the game at any time after playing 10 minutes. Participants then began playing CodeBreakers. At the beginning of the game, participants underwent an audio check during which they were required to type a spoken English word. Participants then used the avatar customization interface corresponding to their condition. A robotic agent then engaged in a short conversation with the player. The robot was animated with audio dialogue generated through an automatic voice generator [143]. After a brief introduction, the participant was provided instructions on how to play the game. See Figure 6a. Participants were told they could exit the game at any time after playing 10 minutes by pressing ESC on their keyboard, then clicking quit game. The participant then began playing the game. During gameplay, the text "*Time Remaining for Survey*" appeared at the top of the screen, with a countdown timer starting from 10 minutes. Once the 10 minutes had elapsed, participants were automatically presented an in-game survey which contained the PIS, PENS autonomy, IMI interest/enjoyment, PXI immersion, motivation for future play, and likelihood of game recommendation questions. See Figure 6b. All participant game data was automatically logged (e.g., time played, avatar customization choices). After the survey was completed, a message box appeared, reminding participants that they could now quit at any time, and that they could continue playing for as long as they liked. The message at the top of the game screen which had shown the time remaining was replaced by the message "*You may play for as long as you like and quit at any time by pressing ESC and clicking Quit Game.*" Once participants quit the game (or completed all 6 levels), participants were then asked to describe in their own words any problems encountered. Participants then filled out a set of questions about prior video game experience, programming experience, and demographics.

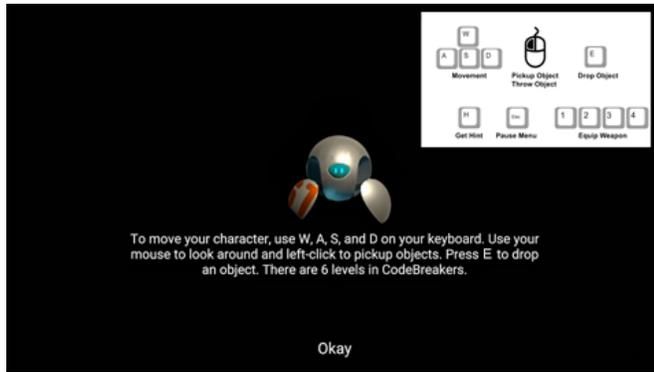
4.12 Analysis

Data was analyzed using SPSS 23 and the PROCESS macro for SPSS [89]. Factorial 2 x 2 ANOVAs were used to study the effects of

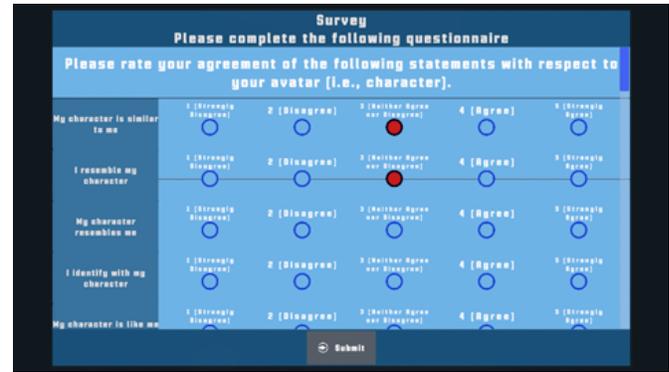
¹¹The sample size calculation takes a similar approach to Schoemann, Boulton, and Short [202].

¹²Note that we explicitly recruited a slightly larger number than we had calculated ($N=1500$) in case of loss of data during data screening.

¹³This large number is not unexpected given that most survey measures are measuring a single concept—e.g., immersion, autonomy, and low variance is expected within these individual survey measures. Nevertheless, it is important to manually inspect these responses for data quality—e.g., "straight-liners" that always pick the same answer option [31].



(a) The robotic agent introduces the game.



(b) Survey that players complete after 10 minutes of gameplay.

Figure 6: Screenshots from the experiment.

visual choice and audial choice on the PIS (H1), and PENS autonomy (H2).¹⁴ We then performed a parallel mediation analysis with visual choice (X), similarity identification ($M1$), embodied identification ($M2$), wishful identification ($M3$), autonomy ($M4$), and IMI interest/enjoyment (Y) (H3). We used PROCESS model 4 [89]. The parallel mediation was repeated using the different outcomes of interest (Y): PXI immersion (H4); time spent playing (H5); motivation for future play (H6); and likelihood of game recommendation (H7). In order to perform exploratory analyses on whether audial choice (W) moderates paths (direct and indirect) between X and Y , we used PROCESS model 59. We used an α of 0.05. These analyses were all preregistered at <https://osf.io/dbvkv/>.

5 RESULTS

5.1 Checking For Model- and Voice-Specific Effects

To ensure that there were no effects of a specific model, or a specific voice, on collected measures (see Section 4.7 for our measures), we used one-way MANOVA. First, we grouped all participants who were assigned a model *randomly*—i.e., participants in the *Choice-None* and *Choice-Audio* conditions. Second, we created another group of participants who were assigned a voice *randomly*—i.e., participants in the *Choice-None* and *Choice-Visual* conditions. We only chose participants who were assigned an avatar or voice *randomly* (and not through choice), since this gives the best approximation of how an avatar or voice may influence a player while avoiding the confound of a self-selection effect. Using the two groups, we then ran two MANOVAs with the IVs of either avatar (group 1) or voice (group 2) and the DVs of our collected measures. Prior to running our MANOVAs, we checked both assumption of homogeneity of variance and homogeneity of covariance by the test of Levene’s Test of Equality of Error Variances and Box’s Test of Equality of Covariance Matrices; and both assumptions were met by the data ($p > 0.05$ for Levene’s, and $p > 0.001$ for Box’s). However, Levene’s test was violated for the measure of *time played* in the MANOVA for group 2 ($p < 0.05$). To deal with this violation, we used the more conservative

¹⁴ANOVAs are considered robust to non-normality, especially at larger sample sizes [28].

Pillai’s Trace [212]. We also set the more conservative significance criterion of $p < 0.01$ (two-tailed) for univariate testing as suggested in the literature [212]. There was no statistically significant difference in our measures based on model, $F[27, 2194]=0.89$, $p=0.630$, Wilk’s $\Lambda=0.969$, $\eta_p^2=0.011$. There was no statistically significant difference in our measures based on voice, $F[27, 2229]=1.241$, $p=0.183$, Pillai’s Trace=0.044, $\eta_p^2=0.015$. Therefore, when assigned randomly, neither a specific model nor a specific voice had a significant effect on our measures.

5.2 H1: Effect of Manipulation on Avatar Identification

From Table 2, factorial 2×2 ANOVAs (choice visual x choice audial) found main effects of choice visual on similarity identification, embodied identification, and wishful identification (H1.1 supported). In contrast, there were no main effects of choice audial (H1.2 not supported). However, a significant interaction effect was found between choice visual and choice audial on similarity identification, embodied identification, and wishful identification (H1.3 not supported). Significant interaction effects were further probed through a simple effects analysis. As this involved two additional tests, the significance threshold was Bonferroni-adjusted to $p=0.025$. Simple effects analysis found that in all cases, the effect of choice audial when there was *no* visual choice was not significant. However, in all cases, the effect of choice audial when there *was* visual choice was significant and positive. Therefore, in the absence of a visual avatar choice, choice of avatar voice has no effect, but in the presence of a visual avatar choice, choice of avatar voice has a significantly positive effect on similarity identification, embodied identification, and wishful identification. Effect sizes (η_p^2) are in the small-to-medium (0.01 to 0.09) range for main effects of choice visual, and small (0.01) for interaction effects.¹⁵

5.3 H2: Effect of Manipulation on Autonomy

From Table 2, a factorial 2×2 ANOVA (choice visual x choice audial) found a main effect of choice visual on autonomy (H2.1 supported).

¹⁵Small effect sizes are not uncommon in games user research due to the complexity of player-game interactions [25, 27, 210, 241].

	PIS Similarity		PIS Embodied		PIS Wishful		PENS Autonomy	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
VISUAL NO CHOICE								
AUDIAL NO CHOICE	2.59	1.08	2.89	1.15	2.50	1.04	4.04	1.70
AUDIAL CHOICE	2.46	0.98	2.75	1.14	2.34	1.00	3.80	1.75
VISUAL CHOICE								
AUDIAL NO CHOICE	2.91	1.08	3.04	1.20	2.67	1.10	4.19	1.73
AUDIAL CHOICE	3.21	1.04	3.35	1.10	2.96	1.07	4.48	1.60
MAIN EFFECT CHOICE VISUAL								
<i>F</i>	98.719		40.104		52.538		23.017	
<i>p</i>	<0.001		<0.001		<0.001		<0.001	
η_p^2	0.061		0.026		0.033		0.015	
MAIN EFFECT CHOICE AUDIAL								
<i>F</i>	2.449		2.041		1.354		0.089	
<i>p</i>	0.118		0.153		0.245		0.765	
η_p^2	0.002		0.001		0.001		0.000	
INTERACTION EFFECT								
<i>F</i>	15.561		14.564		16.998		9.356	
<i>p</i>	<0.001		<0.001		<0.001		0.002	
η_p^2	0.010		0.009		0.011		0.006	
SIMPLE EFFECT CHOICE AUDIAL (VISUAL NO CHOICE)								
<i>F</i>	2.832		2.850		4.378		3.808	
<i>p</i>	0.093		0.092		0.037†		0.051	
η_p^2	0.002		0.002		0.003		0.002	
SIMPLE EFFECT CHOICE AUDIAL (VISUAL CHOICE)								
<i>F</i>	15.178		13.756		13.974		5.638	
<i>p</i>	<0.001		<0.001		<0.001		0.018	
η_p^2	0.010		0.009		0.009		0.004	

Visual Choice $df=1$, Audial Choice $df=1$, Interaction $df=1$, Error $df=1522$

†Not significant due to Bonferroni-adjusted $\alpha=0.025$ for simple effect

Table 2: Results for effects of visual choice and audial choice on PIS (H1) and PENS autonomy (H2). Significant results are bold.

In contrast, there was no main effect of choice audial (H2.2 not supported). However, a significant interaction effect was found between choice visual and choice audial on autonomy (H1.3 not supported). The significant interaction effect was further probed through a simple effect analysis. As this involved two additional tests, the significance threshold was Bonferroni-adjusted to $p=0.025$. The simple effect analysis found that the effect of choice audial when there was *no* visual choice was not significant. However, the effect of choice audial when there *was* visual choice was significant and positive. Therefore, in the absence of a visual avatar choice, choice of avatar voice has no effect, but in the presence of a visual avatar choice, choice of avatar voice has a significantly positive effect on autonomy. The effect size (η_p^2) is considered small.

5.4 H3–H7: Mediation and Moderation Analyses

5.4.1 Assumption Checks. Mediation analyses require several important assumptions to be met [21]: (1) linearity, (2) normality, (3) homoscedasticity, (4) absence of strong multicollinearity, and (5) absence of extreme outliers. (1) To ensure linearity, we plotted scatterplots between each predictor variable and dependent variable. All such permutations were plotted and manually checked to ensure the linearity assumption was satisfied; bivariate correlations were also tested [21]. Linearity was found to be satisfied in all cases. (2) We used PROCESS' bootstrapping option, which makes no assumptions about the distribution of the underlying data [89].

Therefore, the normality assumption is automatically satisfied. (3) We used robust standard errors (HC4 [52]) in all of our analyses, automatically satisfying the assumption of homoscedasticity [89]. (4) To ensure absence of strong multicollinearity, we verified the VIF (Variance Inflation Factor) between all predictor variables and dependent variables. A VIF > 5 is generally a cause for concern, while a VIF > 10 indicates a serious collinearity problem [160]. All VIF scores were below 5, satisfying the assumption of absence of strong multicollinearity. (5) To ensure absence of extreme outliers, we performed outlier testing. The only variable that is at risk of outliers is *time played* (our independent variables are binary and cannot contain outliers by design; similarly, Likert-scale data do not contain outliers). However, outlier testing requires a normal distribution.¹⁶ A Kolmogorov-Smirnov test ($p<0.05$), a Shapiro-Wilk test ($p<0.05$), and a Q-Q plot all indicated that the variable *time played* does not meet the assumption of normality.¹⁷ Therefore, we first perform the data transformation described by Templeton [214]. This is a two-step process: (i) transformation into a percentile rank; and (ii) an inverse-normal transformation. After this process, a Kolmogorov-Smirnov test ($p=0.200$), a Shapiro-Wilk test ($p=0.997$), and a Q-Q plot all indicated that the transformed variable was

¹⁶At a more theoretical level, this is because a distribution must be assumed in order to be able to classify a data point as lying outside the expected range.

¹⁷This was an expected result by design. Because our experimental design was to have a minimum playtime of 10 minutes, we expected a right-skewed distribution with a peak at the 10 minute mark (and no participants below 10 minutes), making the data non-normal.

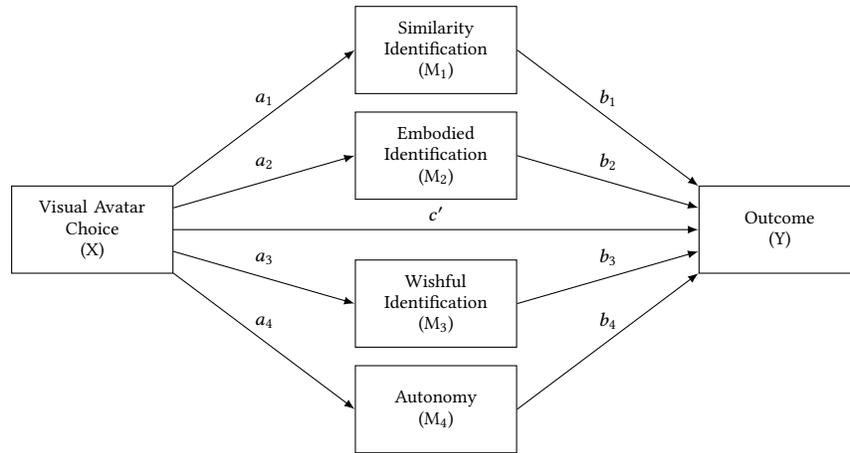


Figure 7: Mediation model being tested for H3 through H7.

SIMILARITY IDENTIFICATION			EMBODIED IDENTIFICATION			WISHFUL IDENTIFICATION		
a_1	b_1	a_1b_1	a_2	b_2	a_2b_2	a_3	b_3	a_3b_3
INTRINSIC MOTIVATION								
0.535***	0.033	0.018; CI[-0.027, 0.065]	0.375***	0.264***	0.099; CI[0.057, 0.148]	0.393***	-0.085*	-0.033; CI[-0.068, -0.001]
IMMERSION								
0.535***	-0.034	-0.018; CI[-0.067, 0.030]	0.375***	0.573***	0.215; CI[0.144, 0.293]	0.393***	-0.059	-0.023; CI[-0.061, 0.010]
TIME SPENT PLAYING								
0.535***	31.65*	16.94; CI[2.091, 33.02]	0.375***	-1.693	-0.635; CI[-10.24, 9.037]	0.393***	33.10*	13.02; CI[2.376, 25.48]
MOTIVATION FOR FUTURE PLAY								
0.535***	0.025	0.013; CI[-0.050, 0.075]	0.375***	0.296***	0.111; CI[0.062, 0.169]	0.393***	0.158**	0.062; CI[0.016, 0.117]
LIKELIHOOD OF GAME RECOMMENDATION								
0.535***	0.084	0.045; CI[-0.013, 0.105]	0.375***	0.187***	0.070; CI[0.028, 0.120]	0.393***	0.174***	0.069; CI[0.023, 0.122]

AUTONOMY			DIRECT EFFECT	TOTAL EFFECT
a_4	b_4	a_4b_4	c'	c
INTRINSIC MOTIVATION				
0.420***	0.686***	0.288; CI[0.171, 0.409]	0.015	0.387***
IMMERSION				
0.420***	0.298***	0.125; CI[0.073, 0.182]	0.048	0.347***
TIME SPENT PLAYING				
0.420***	-0.913	-0.384; CI[-5.683, 4.548]	41.66*	70.60***
MOTIVATION FOR FUTURE PLAY				
0.420***	0.670***	0.281; CI[0.165, 0.400]	0.053	0.521***
LIKELIHOOD OF GAME RECOMMENDATION				
0.420***	0.706***	0.297; CI[0.176, 0.422]	0.030	0.510***

* significant at $p < 0.05$; ** significant at $p < 0.01$; *** significant at $p < 0.005$; significant $a_x b_x$ based on 95% CI

Table 3: Mediation results with visual avatar choice (X), similarity identification (M1), embodied identification (M2), wishful identification (M3), autonomy (M4), and each outcome variable (Y). Regression coefficients a_x ($X \rightarrow M_x$), b_x ($M_x \rightarrow Y$), c' (direct $X \rightarrow Y$), c (total $X \rightarrow Y$), and $a_x b_x$. All presented effects are unstandardized. Significant results are bold.

normally distributed. We then used an Interquartile Range (IQR) multiplier of 2.2 for outlier detection [94], and we found no outliers. Therefore, our mediation analysis assumptions are met.

5.4.2 Hypothesis Tests. The mediation model being tested can be seen in Figure 7. From Table 3, we can see that visual choice has a direct effect (c') on time spent playing only (H5.1 supported; H3.1,

H4.1, H6.1, and H7.1 not supported). A 95% bias-corrected confidence interval based on 10,000 bootstrap samples indicates several significant indirect effects on intrinsic motivation ($a_2 b_2$ and $a_3 b_3$)¹⁸

¹⁸Note the negative coefficient, meaning that higher wishful identification is related to lower intrinsic motivation, which was unexpected. All other significant effects in our model were positive coefficients.

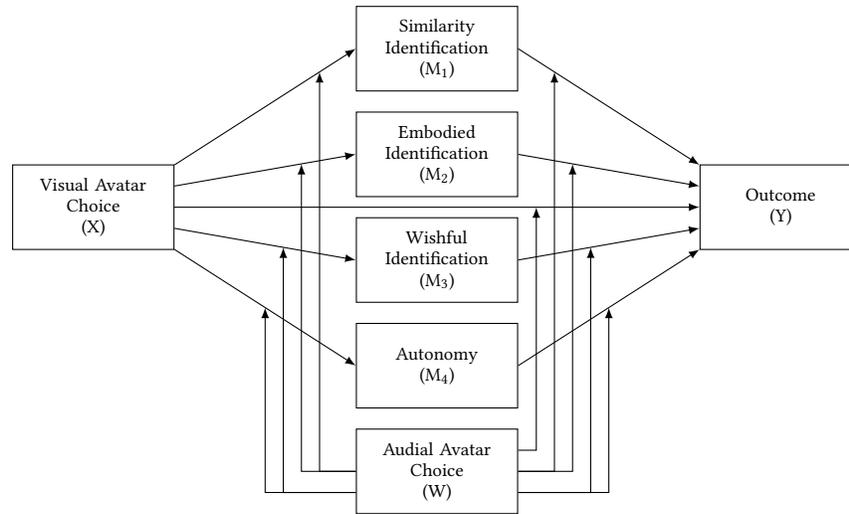


Figure 8: Moderated mediation model being tested in our exploratory analyses.

VARIABLE	VISUAL NO CHOICE				VISUAL CHOICE			
	AUDIAL NO CHOICE		AUDIAL CHOICE		AUDIAL NO CHOICE		AUDIAL CHOICE	
	M	SD	M	SD	M	SD	M	SD
Intrinsic Motivation	4.57	1.70	4.35	1.76	4.72	1.70	4.97	1.47
Immersion	0.77	1.44	0.58	1.59	0.89	1.44	1.15	1.39
Time Spent Playing	864.57	319.14	844.52	292.78	905.35	370.15	944.18	412.18
Motivation for Future Play	3.87	1.95	3.65	2.01	4.10	2.02	4.46	1.96
Likelihood of Game Recommendation	3.94	1.94	3.67	2.00	4.12	1.97	4.50	1.96

Table 4: Descriptives for outcomes in H3 through H7. Immersion was on a Likert scale from -3 to +3. Intrinsic Motivation, Motivation for Future Play, and Likelihood of Game Recommendation were on Likert scales from 1 to 7.

supporting H3.2, a_4b_4 supporting H3.3), immersion (a_2b_2 supporting H4.2, a_4b_4 supporting H4.3), time spent playing (a_1b_1 and a_3b_3 supporting H5.2, H5.3 not supported), motivation for future play (a_2b_2 and a_3b_3 supporting H6.2, a_4b_4 supporting H6.3), and likelihood of game recommendation (a_2b_2 and a_3b_3 supporting H7.2, a_4b_4 supporting H7.3). Therefore, we conclude that visual choice *directly* affects time spent playing, and *indirectly* affects intrinsic motivation (via embodied identification, wishful identification, and autonomy), immersion (via embodied identification and autonomy), time spent playing (via similarity identification and wishful identification), motivation for future play (via embodied identification, wishful identification, and autonomy), and likelihood of game recommendation (via embodied identification, wishful identification, and autonomy). Descriptives for each variable can be seen in Table 4.

5.4.3 Exploratory Analyses. For the exploratory analyses with no a priori hypotheses, we test the model seen in Figure 8. Results of the moderated mediation are found in Table 5. We find evidence of significant moderated mediation through the moderator of audial choice for intrinsic motivation (moderating $X \rightarrow M_1 \rightarrow Y$ and $X \rightarrow M_4 \rightarrow Y$), immersion (moderating $X \rightarrow M_2 \rightarrow Y$ and $X \rightarrow M_4 \rightarrow Y$), time spent playing (moderating $X \rightarrow Y$), motivation for future play (moderating $X \rightarrow M_3 \rightarrow Y$ and $X \rightarrow M_4 \rightarrow Y$), and likelihood of game recommendation (moderating $X \rightarrow M_4 \rightarrow Y$). These effects were then

probed while fixing the value of audial choice to 0 or 1 (see Table 5). When these effects were probed while fixing audial choice to 0, the mediations in *all cases* were non-significant. On the other hand, when fixing audial choice to 1, the mediations in *all cases* were positive and significant. Therefore, audial choice positively moderates different paths across all outcome variables.^{19,20}

¹⁹It is worth noting that despite using a different model with the inclusion of the moderator W , there are overlaps with the results from hypothesis testing (Section 5.4.2). For example, in all 7 cases of significant moderated mediation in this second model, indirect effects in the first model along those same paths are significant (see Table 3). Similarly, in all 5 cases where the index of moderated mediation was *not* significant (or in the case of direct effects, non-existent), but the effect at $AC=1$ was significant, we have a significant effect in the first model along the same path. A slight divergence (2 cases) from the first model appears for the indirect effect $X \rightarrow M_3 \rightarrow Y$. The indirect effect was significant in the first model for intrinsic motivation and time spent playing, but the effect is not significant at either value of AC (0 or 1) in the second model. Therefore, inclusion of the moderator W does slightly affect the results from the model we chose for hypothesis testing. If we were to consider only the results from the second model, our overall hypothesis testing results would remain the same, but with slightly weaker support for H3.2 and H5.2.

²⁰Another point to address is the possibility that the results arise because of participants who are randomly assigned a differently-gendered model. For example, perhaps audial choice alone is not effective at engendering outcomes specifically because gender is then randomly assigned, and will not match the player's gender ~50% of the time. To check if this was the case, we re-ran *all* analyses with *only* participants who self-identified as male and used a male model and participants who self-identified as female and used a female model regardless of condition ($N=808$). We found identical results with respect to each of our hypotheses, and no evidence that random assignment of a differently-gendered avatar affected any of the results.

VARIABLE		$X \rightarrow M_1 \rightarrow Y$	$X \rightarrow M_2 \rightarrow Y$	$X \rightarrow M_3 \rightarrow Y$	$X \rightarrow M_4 \rightarrow Y$	$X \rightarrow Y$
Intrinsic Mot.	Index of MM	-0.080; CI[-0.179, 0.018]	0.132; CI[0.038, 0.233]	0.003; CI[-0.071, 0.077]	0.364; CI[0.128, 0.600]	—
	Effect at AC= 0	0.036; CI[-0.003, 0.084]	0.036; CI[-0.006, 0.089]	-0.022; CI[-0.054, 0.001]	0.104; CI[-0.063, 0.272]	-0.006; CI[-0.153, 0.142]
	Effect at AC=1	-0.044; CI[-0.132, 0.047]	0.168; CI[0.092, 0.259]	-0.019; CI[-0.087, 0.049]	0.468; CI[0.302, 0.637]	0.045; CI[-0.099, 0.190]
Immersion	Index of MM	0.012; CI[-0.098, 0.130]	0.255; CI[0.104, 0.418]	-0.013; CI[-0.098, 0.069]	0.172; CI[0.062, 0.290]	—
	Effect at AC= 0	-0.019; CI[-0.063, 0.018]	0.086; CI[-0.012, 0.188]	-0.013; CI[-0.044, 0.006]	0.042; CI[-0.026, 0.113]	0.026; CI[-0.132, 0.183]
	Effect at AC= 1	-0.006; CI[-0.112, 0.102]	0.341; CI[0.228, 0.471]	-0.026; CI[-0.107, 0.052]	0.214; CI[0.131, 0.309]	0.044; CI[-0.122, 0.210]
Time Spent	Index of MM	24.36; CI[-9.628, 58.76]	-5.954; CI[-27.68, 15.16]	13.28; CI[-12.89, 42.17]	2.453; CI[-9.172, 14.36]	—
	Effect at AC= 0	6.904; CI[-5.576, 21.55]	0.798; CI[-5.627, 7.986]	5.710; CI[-0.823, 16.28]	-0.723; CI[-5.318, 2.348]	28.10; CI[-20.99, 77.18]
	Effect at AC= 1	31.26; CI[0.383, 62.91]	-5.157; CI[-26.42, 15.00]	18.99; CI[-5.25, 46.89]	1.730; CI[-9.468, 12.84]	52.83; CI[3.981, 101.7]
Mot. Fut. Play	Index of MM	-0.111; CI[-0.250, 0.030]	0.096; CI[-0.025, 0.226]	0.174; CI[0.068, 0.293]	0.388; CI[0.157, 0.627]	—
	Effect at AC= 0	0.039; CI[-0.011, 0.096]	0.051; CI[-0.006, 0.122]	0.006; CI[-0.024, 0.044]	0.096; CI[-0.055, 0.256]	0.034; CI[-0.155, 0.223]
	Effect at AC= 1	-0.072; CI[-0.202, 0.057]	0.147; CI[0.046, 0.260]	0.180; CI[0.081, 0.295]	0.484; CI[0.313, 0.669]	0.070; CI[-0.128, 0.267]
Game Rec.	Index of MM	-0.020; CI[-0.147, 0.111]	0.099; CI[-0.003, 0.209]	0.098; CI[-0.001, 0.209]	0.396; CI[0.149, 0.645]	—
	Effect at AC= 0	0.042; CI[-0.006, 0.097]	0.025; CI[-0.004, 0.071]	0.026; CI[-0.003, 0.073]	0.103; CI[-0.065, 0.275]	-0.014; CI[-0.193, 0.165]
	Effect at AC= 1	0.022; CI[-0.098, 0.143]	0.124; CI[0.032, 0.229]	0.124; CI[0.033, 0.229]	0.499; CI[0.323, 0.685]	0.063; CI[-0.130, 0.255]

Table 5: Moderated mediation results for each path with the inclusion of the moderator W (audial choice). For each variable and path, the index of moderated mediation (MM) and the conditional effects when Audial Choice (AC) is set to 0 and 1 are shown. Direct effects ($X \rightarrow Y$) do not have an index of moderated mediation. Significant results are bold. Significant results are based on 95% CI.

6 DISCUSSION

Existing work on avatar customization has focused almost exclusively on visual aspects of customization. While there are many benefits to avatar customization, it is unknown whether audial avatar customization confers similar benefits.

We conducted a 2×2 (visual choice x audial choice) experiment. Visual customization directly increases avatar identification and autonomy. Visual customization directly increases time spent playing and indirectly (through avatar identification and autonomy as mediators) increases intrinsic motivation,²¹ immersion, time spent playing, motivation for future play, and likelihood of game recommendation. Audial customization did not lead to a direct increase in avatar identification and autonomy. A significant interaction effect showed that audial customization directly increases avatar identification and autonomy, but only when visual customization was also available. Audial customization significantly moderated eight paths between visual customization and the outcome variables intrinsic motivation, immersion, time spent playing, motivation for future play, and likelihood of game recommendation. The moderation was such that when audial customization was unavailable, the path had a non-significant effect on the outcome, but when audial customization was available, the path had a significant effect on the outcome.

²¹ An exception here is for the indirect effect of visual choice on intrinsic motivation through wishful identification, which was the *only* significant result across all analyses with a negative effect. The reason for this is not immediately apparent, since the other indirect effects through wishful identification on time spent playing, motivation for future play, and likelihood of game recommendation were all significant and positive. One potential explanation is that when we identify wishfully with an avatar, we may view the avatar as a more competent version of ourselves (e.g., a better programmer). Even if we view the game as being interesting, this could detract from the enjoyability of the game (e.g., we feel “less than” our avatar), but not from intention to play or recommend. This is not the first time that such a discrepancy has been noted in the literature for wishful identification. Wishful identification has been found to be positively associated with PX outcomes but negatively associated with quality of created artifacts in an educational play and making context [113]. In other studies which measure wishful identification—e.g., in an entertainment-oriented context [25], no such negative associations have been found. It is possible that wishful identification (which is known to be correlated with lower psychological well-being [22, 92, 165]) has a more two-sided nature in educational contexts, potentially because they may feel more achievement-oriented rather than for “fun.” Additional controlled studies which manipulate game type and/or framing are needed to make more conclusive claims.

Based on these results, we conclude that audial customization plays an important role in affecting outcomes.

However, we make the argument that although audial customization is important, it appears to have a weaker effect in comparison to visual customization. This argument is based on two facets of the results: (1) visual customization *alone* has a significant effect on avatar identification and autonomy, whereas audial customization has a significant effect only within the group of participants who also have visual customization available;²² and (2) audial customization’s effects on avatar identification and autonomy have lower effect sizes (small) when compared to visual customization (small-to-medium) [47, 162]. The first point suggests that audial customization plays an enhancing role for visual customization (i.e., when visual customization was present, audial customization further increased avatar identification and autonomy compared to no audial customization). Both points together suggest that audial customization, although important, is somewhat weaker than visual customization.

6.1 Visual Customization Has a Stronger Effect Than Audial Customization

Many possibilities exist for why visual customization had a stronger effect than audial customization. One possibility is that players are simply more familiar with visual customization. People are known to prefer things due to familiarity alone. The familiarity principle (also called the mere-exposure effect) describes the phenomenon of preference for things merely due to familiarity [240]. Therefore, the effects of visual customization could have been enhanced through familiarity.

Additionally, the total exposure time to the audial customization aspects of the avatar (i.e., voice) was only a fraction of the exposure time to the visual aspects of the avatar (i.e., model). While the audial aspect of the avatar is infrequent and typically only occurs before and after each puzzle, the visual aspect of the avatar is always

²²Note that this is still the case even when only considering participants whose self-reported gender matched the avatar’s as discussed in footnote 20.

present on screen. Moreover, the audial aspect of the avatar was interleaved with other sounds (game audio and background noise). Such factors could have all served to reduce the impactfulness of audial customization. Studying games with frequent voice lines (e.g., a narrative adventure such as *The Walking Dead* [213]) would help to balance the exposure between visual and audial aspects of the avatar. Such studies would help to understand if the reason for the discrepancy between visual and audial customization effects stems from exposure.

Visual aspects of an avatar might also inherently (at a fundamental level) be more important than audial aspects. Humans have been shown to have better visual memory than auditory memory and that there appear to be fundamental differences between visual and auditory processing [48]. The *picture superiority effect* describes the phenomenon whereby pictures and images are more often remembered compared to words [43]. Reasons for why the picture superiority effect happens are still being debated. However, this fundamental asymmetry between visual and auditory stimuli would give credibility to the argument that visual aspects of an avatar are inherently more important than audial aspects of the avatar.

It may also be possible to explain the audial-visual discrepancy through investment of effort. If participants view the visual aspect of their avatar as more important, then they may invest more effort into visual customization than audial customization. According to Cialdini's commitment and consistency principle, people tend to behave in ways consistent with how they have acted in the past [45] (i.e., future behavior often resembles past behavior). To maintain consistency with the effort in customizing the avatar visually, players would also invest more effort into the game. This would increase outcomes (e.g., avatar identification). Future studies could study the customization process itself more closely—e.g., time spent on customizing visual vs. audial aspects, measuring cognitive load in customizing visual vs. audial aspects.

6.2 Audial Customization Is Effective When Paired With Visual Customization, But Not Alone

Interestingly, audial customization was only effective at increasing avatar identification and autonomy when visual customization was also present. This was true even when we re-performed all analyses with only participants with a matching avatar gender (see footnote 20). The reason for this is not immediately apparent. Although the character customization conditions were designed carefully and validated with expert UI designers, it is possible that the ability to customize voice (and especially in the absence of model selection) did not match players' expectations. A more in-depth investigation into the avatar customization process itself may help shed light on this phenomenon. Based on our results, we recommend pairing audial customization options with visual customization options to enhance outcomes.

6.3 Implications for Research on Avatar Customization

This research has examined both the effects of avatar customization (e.g., [25, 221]) and avatar customization interfaces (e.g., [152, 154,

156, 177]). Our contribution is a large-scale preregistered study showing that audial avatar customization, when paired with visual avatar customization, engenders important outcomes. Audial avatar customization was effective in increasing *all* types of avatar identification (similarity, embodied, wishful) beyond the degree of avatar identification induced by visual avatar customization alone. Although prior studies (and the current study) show that *visual* customization is effective at increasing avatar identification, we show that audial customization (in the form of a minimal set of voices) can also influence all aspects of avatar identification. This result suggests that even simple audial avatar customization—the selection of one voice from two options—is sufficient to increase perceived similarity with the avatar, the sense of being embodied within the avatar, and the idealization of the avatar. These three elements of identification are sometimes but not always consistently influenced by facets of avatar use [224], so this finding is particularly notable. Further, additional audial avatar customization options (e.g., pitch, loudness, pace, resonance, intonation) might facilitate even higher levels of avatar identification. Future studies could also investigate audial avatar customization in additional domains (e.g., exercise applications [9], social media and VR [125, 236]), using additional methodological techniques (e.g., player interviews [15]), and the social inclusivity of audial customization interfaces (e.g., gender and race [156, 235]).

The finding that audial choice significantly moderates the effect of visual choice on game outcomes (as mediated by identification and autonomy) provides further evidence for the importance of audial avatar customization. For example, visual customization was associated with greater intrinsic motivation (finding the game satisfying), sense of immersion, motivation for future play, and game recommendation, but only when there was audial customization, and all of these associations were fully mediated by embodied identification. In other words, visual customization alone did not sufficiently induce an association between embodied identification and these game outcomes, but visual together with audial customization did. Similarly, visual and audial customization together induced greater time spent playing the game and this effect was partially mediated by similarity identification. Together, these findings suggest that audial customization is a notable contributor not only to the subjective experience of identification with the avatar, but also to the *outcomes* of identification with the avatar within the game.

This work is also relevant to the Proteus Effect, a phenomenon whereby users tend to conform to the expected behaviors of their avatars [239]. This has been studied extensively with respect to visual characteristics [186], but not audial characteristics. For example, physically healthy-looking avatars can promote physical activity [135], and avatars perceived as creative can promote creative brainstorming [87]. However, allowing users to create audial avatar identities could also be a powerful avenue for inducing the Proteus Effect. In the present context of learning games, this research suggests that using an avatar with a voice that sounds more capable of success in a computer-science context (e.g., intelligent, persistent) might empower players to perform better in the game and thus learn the educational content more effectively. Further research could be designed to confirm this expectation first by pretesting the

perceived intelligence/persistence of different voices and then assigning exemplary voices as customization options within a similar game.

6.4 Potential Applications for Audial Customization

The amount of dialogue in CodeBreakers can be considered minimal compared to most games that contain voiced dialogue—e.g., *Mass Effect* [161]. Nevertheless, audial avatar customization promoted all outcomes studied (e.g., autonomy, intrinsic motivation, immersion). Games for learning, health, and entertainment would all benefit from increases in the outcomes studied. Audial customization could have even broader implications in real-world devices. Examples include in-home devices incorporating voice interaction [79–81] and conversational agents more generally (e.g., Alexa [10]) [16, 17, 93]. For example, it is not well understood what the effects of changing the voices of these devices are (e.g., to be more similar to the user). This includes other companion devices, such as robotic learning companions [145, 216, 243] and other dialogue-capable digital agents [32, 118, 130].

Audial customization could enhance video instruction [40, 166] (e.g., lecture videos [120, 121, 164]), massive open online courses (MOOCs) [77], intelligent tutoring [42, 173], e-books [50, 194], and collaborative platforms [119, 128]. This could involve different modalities such as tangibles [72], tabletop displays [116, 147], interactive installations [144, 191, 195], augmented reality [29, 35] and virtual reality [13, 74, 146], and digital streaming [41, 179]. Additional investigation into different domains and modalities would elucidate whether audial customization can be applied more generally to increase user engagement. It is also important to investigate how the design choices behind audial customization can influence user identities—e.g., underrepresented minorities in STEM [5]—and how those design choices can be either exclusionary or inclusionary [106, 189] and influence phenomena including stereotype threat [187, 190] and user anxiety [181]. Further research is needed on audial customization to understand more generally the potential use cases.

7 LIMITATIONS

Despite the robust design of this controlled experiment, there are some limitations to the study’s external and internal validity that should be considered in future research. First, participants were given only two visual and audial customization choices for each gender. Many games provide a greater number of choices during avatar customization, suggesting that avatar identification in such games is generally higher than it was in our study. Further, participants were likely more familiar with visual avatar customization than audial customization given that the former is more prevalent in current games and social media. Hence, the choice of avatar appearance—even based on just two options—was more likely to remind participants of previous avatar customization experiences that involved choices over many visual aspects of an avatar. In contrast, audial customization could potentially include a wide range of avatar characteristics that were not included in the present study (e.g., footsteps, whistling, grunting noises, pitch modification), but the participants’ choice of avatar voice was less likely to remind

them of these possibilities. Moreover, avatar identification may have been limited for players who do not conform to stereotypical representations of “male” and “female” voices. For these reasons, future research on this topic should include a larger set of customization options, especially for audial avatar characteristics.

The study also included a potential confound relating to the attention paid to audial and visual cues. Namely, in order to proceed in the game, participants were required to solve visual puzzles that did not include audial elements. This prioritization of visual stimuli may have led to a greater focus on the avatar’s appearance compared to avatar’s speech, partially explaining why visual avatar customization was more consequential in the study outcomes. Another related but minor issue is that the quality of the sound hardware may have varied between players’ computers causing noise in the data (i.e., less attention to audial cues), but this was likely not confounded with experiment condition given random assignment. Further, all participants performed an audio check, so a threshold of audial attention can be inferred.

The study relied on participants being paid to play the game, like most research in this field, which potentially limits ecological validity. Further, generalizability was not established beyond the single, education-oriented game designed for this research. Relatedly, the study cannot determine how specific facets of this particular game design (e.g., pacing) influenced the study outcomes. For one, the game was designed to highlight the avatar’s voice for a single user, so the study findings do not directly speak to multi-user games which offer voice-based communication [231, 232]. However, algorithmic voice modification (e.g., pitch modulation to mask gender) is an increasingly popular multi-user technology for games (e.g., [151, 229]) that could potentially help mitigate toxic behavior between players [225, 233]. The present findings indirectly suggest that customizing such voice modification might also be beneficial to the user’s experience in other ways.

The study required participants to play the game for a minimum of 10 minutes, which is significantly less time than many people tend to play video games [70]. However, this length of exposure is sufficient to induce avatar identification [61], as other studies have found [6], and the present study was not intended to examine changes in identification over time. We should also note that 10-minute exposures are common in video-game experiments, perhaps due to operational constraints, but these studies tend to find sufficient effects on their outcomes of interest with such durations.

8 CONCLUSION

Avatar customization is known to positively affect crucial outcomes in numerous domains, including health, entertainment, and education. However, studies on avatar customization have focused almost exclusively on visual aspects of customization. It is unknown whether audial customization can confer the same benefits as visual customization. We presented one of the first studies to date on audial avatar customization. Participants with visual choice experienced higher avatar identification and autonomy. Participants with audial choice experienced higher avatar identification and autonomy, but only within the group of participants who had visual choice available. Visual choice led to an increase in time spent and indirectly led to increases in intrinsic motivation, immersion,

time spent, future play motivation, and likelihood of game recommendation. Audial choice moderated the majority of these effects. Our results suggest that audial customization, although having a moderately weaker effect compared to visual customization, plays an important role in enhancing all outcomes compared to visual customization alone. We discussed the implications for research and potential applications of audial avatar customization. This work takes an important first step in developing a baseline understanding of audial avatar customization.

REFERENCES

- [1] 2021. UMA 2 - Unity Multipurpose Avatar: 3D Characters: Unity asset store. <https://assetstore.unity.com/packages/3d/characters/uma-2-unity-multipurpose-avatar-35611>
- [2] Vero Vanden Abeele, Katta Spiel, Lennart Nacke, Daniel Johnson, and Kathrin Gerling. 2020. Development and validation of the player experience inventory: A scale to measure player experiences at the level of functional and psychosocial consequences. *International Journal of Human Computer Studies* 135, January 2019 (2020), 102370. <https://doi.org/10.1016/j.ijhcs.2019.102370>
- [3] Alexandra Adkins, Kristopher Kohm, Rui Zhang, and Nicholas Gustafson. 2020. Lost in Space: An Audio Maze Game for the Visually Impaired. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–6. <https://doi.org/10.1145/3334480.3381660>
- [4] Lena Fanya Aeschbach, Sebastian Andrea Caesar Perrig, Lorena Weder, Klaus Opwis, and Florian Brühlmann. 2021. A Systematic Literature Review of Transparency in Measurement Reporting at CHI PLAY. (2021).
- [5] June Ahn, Mega Subramaniam, Elizabeth Bonsignore, Anthony Pellicone, Amanda Waugh, and Jason Yip. 2014. I Want to be a game designer or scientist: Connected learning and developing identities with urban, African-American youth. Boulder, CO: International Society of the Learning Sciences.
- [6] Johnie J. Allen and Craig A. Anderson. 2021. Does avatar identification make unjustified video game violence more morally consequential? *Media Psychology* 24, 2 (2021), 236–258. <https://doi.org/10.1080/15213269.2019.1683030>
- [7] Fraser Allison, Marcus Carter, and Martin Gibbs. 2020. Word Play: A History of Voice Interaction in Digital Games. *Games and Culture* 15, 2 (March 2020), 91–113. <https://doi.org/10.1177/1555412017746305>
- [8] Fraser Allison, Marcus Carter, Martin Gibbs, and Wally Smith. 2018. Design Patterns for Voice Interaction in Games. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play*. ACM, Melbourne VIC Australia, 5–17. <https://doi.org/10.1145/3242671.3242712>
- [9] Aishat Aloba, Gianne Flores, Jaida Langham, Zari McFadden, John Bell, Nikita Dagar, Shaghayegh Esmaili, and Lisa Anthony. 2020. Toward Exploratory Design with Stakeholders for Understanding Exergame Design. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–8.
- [10] Amazon. 2020. Alexa. <https://developer.amazon.com/en-US/alexa>
- [11] Autodesk.com. 2021. Welcome to Autodesk® character generator. <https://charactergenerator.autodesk.com/>
- [12] Laura Aymerich-Franch, Cody Karutz, and Jeremy N Bailenson. 2012. Effects of Facial and Voice Similarity on Presence in a Public Speaking Virtual Environment. In *Proceedings of the International Society for Presence Research Annual Conference*. 24–26.
- [13] Laura Aymerich-Franch, René F Kizilcec, and Jeremy N Bailenson. 2014. The relationship between virtual self similarity and social anxiety. *Frontiers in human neuroscience* 8 (2014), 944.
- [14] Rachel Bailey, Kevin Wise, and Paul Bolls. 2009. How Avatar Customizability Affects Children's Arousal and Subjective Presence During Junk Food-Sponsored Online Video Games. *CyberPsychology & Behavior* 12, 3 (June 2009), 277–283. <https://doi.org/10.1089/cpb.2008.0292>
- [15] Jaime Banks and Nicholas David Bowman. 2016. Avatars are (sometimes) people too: Linguistic indicators of parasocial and social ties in player-avatar relationships. *New Media & Society* 18, 7 (2016), 1257–1276.
- [16] Erin Beneteau, Ashley Boone, Yuxing Wu, Julie A Kientz, Jason Yip, and Alexis Hiniker. 2020. Parenting with Alexa: exploring the introduction of smart speakers on family dynamics. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–13.
- [17] Erin Beneteau, Olivia K Richards, Mingrui Zhang, Julie A Kientz, Jason Yip, and Alexis Hiniker. 2019. Communication breakdowns between families and Alexa. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.
- [18] Adam J Berinsky, Gregory A Huber, and Gabriel S Lenz. 2012. Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. *Political Analysis* 20, 3 (2012), 351–368.
- [19] Axel Berndt. 2011. Diegetic Music: New Interactive Experiences. In *Game Sound Technology and Player Interaction: Concepts and Developments*. IGI Global, 60–77.
- [20] Axel Berndt and Knut Hartmann. 2008. The functions of music in interactive media. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 5334 LNCS (2008), 126–131. https://doi.org/10.1007/978-3-540-89454-4_19
- [21] William D Berry. 1993. *Understanding regression assumptions*. Vol. 92. Sage.
- [22] K Bessière, AF Seay, and S Kiesler. 2007. The ideal elf: Identity exploration in World of Warcraft. *CyberPsychology & Behavior* (2007). <http://online.liebertpub.com/doi/abs/10.1089/cpb.2007.9994>
- [23] Bethesda Game Studios. 2011. *The Elder Scrolls V: Skyrim*. Game [Multiple Platforms]. Bethesda Softworks, Maryland, USA.
- [24] Bethesda Game Studios. 2015. *Fallout 4*. Game [Multiple Platforms]. Bethesda Softworks, Maryland, USA.
- [25] Max V. Birk, Cheralyn Atkins, Jason T. Bowey, and Regan L. Mandryk. 2016. Fostering Intrinsic Motivation through Avatar Identification in Digital Games. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, San Jose California USA, 2982–2995. <https://doi.org/10.1145/2858036.2858062>
- [26] Max V. Birk and Regan L. Mandryk. 2018. Combating Attrition in Digital Self-Improvement Programs Using Avatar Customization. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal QC Canada, 1–15. <https://doi.org/10.1145/3173574.3174234>
- [27] Max V. Birk, Regan L. Mandryk, Matthew K. Miller, and Kathrin M. Gerling. 2015. How self-esteem shapes our interactions with play technologies. In *CHI PLAY 2015 - Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play*. <https://doi.org/10.1145/2793107.2793111>
- [28] Maria J. Blanca, Rafael Alarcón, Jaume Arnau, Roser Bono, and Rebecca Bendayan. 2017. Non-normal data: Is ANOVA still a valid option? *Psychothema* (2017). <https://doi.org/10.7334/psicothema2016.383>
- [29] Elizabeth Bonsignore, Derek Hansen, Anthony Pellicone, June Ahn, Kari Kraus, Steven Shumway, Kathryn Kaczmarek, Jeff Parkin, Jared Cardon, Jeff Sheets, et al. 2016. Traversing transmedia together: Co-designing an educational alternate reality game for teens, with teens. In *Proceedings of the 15th International Conference on Interaction Design and Children*. 11–24.
- [30] Barbara Borkowska and Boguslaw Pawlowski. 2011. Female voice frequency in the context of dominance and attractiveness perception. *Animal Behaviour* 82, 1 (2011), 55–59. <https://doi.org/10.1016/j.anbehav.2011.03.024>
- [31] Florian Brühlmann and Elisa D Mekler. 2018. Surveys in games user research. *Games User Research* (2018), 141–162.
- [32] Amanda Buddemeyer, Leshell Hatley, Angela Stewart, Jaemarie Solyst, Amy Ogan, and Erin Walker. 2021. Agentic Engagement with a Programmable Dialog System. In *Proceedings of the 17th ACM Conference on International Computing Education Research*. 423–424.
- [33] Michael Buhrmester, Tracy Kwang, and Samuel D Gosling. 2011. Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on psychological science* 6, 1 (2011), 3–5.
- [34] Stéphanie Buisine, Jérôme Guegan, Jessy Barré, Frédéric Segonds, and Améziane Aoussat. 2016. Using Avatars to Tailor Ideation Process to Innovation Strategy. *Cognition, Technology & Work* 18, 3 (Aug. 2016), 583–594. <https://doi.org/10.1007/s10111-016-0378-y>
- [35] Su Cai, Xu Wang, and Feng-Kuang Chiang. 2014. A case study of Augmented Reality simulation system application in a chemistry course. *Computers in human behavior* 37 (2014), 31–40.
- [36] Capcom. 2018. *Monster Hunter: World*. Game [Multiple Platforms].
- [37] Marcus Carter, Fraser Allison, John Downs, and Martin Gibbs. 2015. Player Identity Dissonance and Voice Interaction in Games. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play*. ACM, London United Kingdom, 265–269. <https://doi.org/10.1145/2793107.2793144>
- [38] CD Projekt Red. 2020. *Cyberpunk 2077*. Game [Multiple Platforms]. CD Projekt, Höfen, Austria.
- [39] Jesse Chandler and Danielle Shapiro. 2016. Conducting clinical research using crowdsourced convenience samples. *Annual Review of Clinical Psychology* 12 (2016).
- [40] Minsuk Chang, Anh Truong, Oliver Wang, Maneesh Agrawala, and Juho Kim. 2019. How to design voice based navigation for how-to videos. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–11.
- [41] Xinyue Chen, Si Chen, Xu Wang, and Yun Huang. 2021. "I was afraid, but now I enjoy being a streamer!" Understanding the Challenges and Prospects of Using Live Streaming for Online Education. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (2021), 1–32.
- [42] Min Chi, Pamela Jordan, and Kurt VanLehn. 2014. When is tutorial dialogue more effective than step-based tutoring?. In *International Conference on Intelligent Tutoring Systems*. Springer, 210–219.
- [43] Terry L Childers and Michael J Houston. 1984. Conditions for a picture-superiority effect on consumer memory. *Journal of consumer research* 11, 2 (1984), 643–654.

- [44] Klimmt Christoph, Hefner Dorothee, and Vorderer Peter. 2009. The Video Game Experience as "True" Identification: A Theory of Enjoyable Alterations of Players' Self-Perception. *Communication Theory* 19, 4 (Nov. 2009), 351–373. <https://doi.org/10.1111/j.1468-2885.2009.01347.x>
- [45] Robert B. Cialdini, Wilhelmina Wosinska, Daniel W. Barrett, Jonathan Butner, and Malgorzata Gornik-Durose. 1999. Compliance with a request in two cultures: The differential influence of social proof and commitment/consistency on collectivists and individualists. *Personality and Social Psychology Bulletin* 25, 10 (1999), 1242–1253. <https://doi.org/10.1177/0146167299258006>
- [46] Jonathan Cohen. 2001. Defining Identification: A Theoretical Look at the Identification of Audiences With Media Characters. *Mass Communication and Society* 4, 3 (Aug. 2001), 245–264. https://doi.org/10.1207/S15327825MCS0403_01
- [47] Jacob Cohen. 2013. *Statistical power analysis for the behavioral sciences*. Routledge.
- [48] Michael A Cohen, Todd S Horowitz, and Jeremy M Wolfe. 2009. Auditory recognition memory is inferior to visual recognition memory. *Proceedings of the National Academy of Sciences* 106, 14 (2009), 6008–6010.
- [49] Karen Collins. 2008. *Game sound: an introduction to the history, theory, and practice of video game music and sound design*. Mit Press.
- [50] Luca Colombo, Monica Landoni, and Elisa Rubegni. 2014. Design guidelines for more engaging electronic books: insights from a cooperative inquiry study. In *Proceedings of the 2014 conference on Interaction design and children*. 281–284.
- [51] Mia Consalvo. 2003. *It's a Queer World after All: Studying The Sims and Sexuality*. Gload.
- [52] Francisco Cribari-Neto. 2004. Asymptotic inference under heteroskedasticity of unknown form. *Computational Statistics and Data Analysis* 45, 2 (2004), 215–233. [https://doi.org/10.1016/S0167-9473\(02\)00366-3](https://doi.org/10.1016/S0167-9473(02)00366-3)
- [53] Crystal Dynamics. 2004. *World of Warcraft*. Game [Multiple Platforms]. Blizzard Entertainment, California, USA.
- [54] Crystal Dynamics. 2013. *Tomb Raider*. Game [Multiple Platforms]. Eidos Interactive (Square Enix), Tokyo, Japan.
- [55] James J. Cummings and Jeremy N. Bailenson. 2016. How Immersive Is Enough? A Meta-Analysis of the Effect of Immersive Technology on User Presence. *Media Psychology* 19, 2 (2016), 272–309. <https://doi.org/10.1080/15213269.2015.1015740>
- [56] Alwin de Rooij, Sarah van der Land, and Shelly van Erp. 2017. The Creative Proteus Effect: How Self-Similarity, Embodiment, and Priming of Creative Stereotypes with Avatars Influences Creative Ideation. In *Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition*. ACM, Singapore Singapore, 232–236. <https://doi.org/10.1145/3059454.3078856>
- [57] Martin J Dechant, Max V Birk, Youssef Shiban, Knut Schnell, and Regan L Mandryk. 2021. How Avatar Customization Affects Fear in a Game-based Digital Exposure Task for Social Anxiety. *CHI PLAY '21* (2021).
- [58] Mats Deutschmann, Anders Steinvall, and Anna Lagerström. 2011. Gender-Bending in Virtual Space: Using Voice-Morphing in Second Life to Raise Sociolinguistic Gender Awareness. In *V-Lang International Conference, Warsaw, 17th November 2011*. Warsaw Academy of Computer Science, Management and Administration, 54–61.
- [59] Igor Dolgov, William J Graves, Matthew R Nereants, Jeremy D Schwark, and C Brooks Volkman. 2014. Effects of cooperative gaming and avatar customization on subsequent spontaneous helping behavior. *Computers in human behavior* 33 (2014), 49–55.
- [60] Sebastian Domsch. 2017. Dialogue in video games. In *Dialogue across Media*. John Benjamins, 251–270.
- [61] Edward Downs, Nicholas D. Bowman, and Jaime Banks. 2019. A Polythetic Model of Player-Avatar Identification: Synthesizing Multiple Mechanisms. *Psychology of Popular Media Culture* 8, 3 (July 2019), 269–279. <https://doi.org/10.1037/ppm0000170>
- [62] Nicolas Ducheneaut, Ming-Hui Wen, Nicholas Yee, and Greg Wadley. 2009. Body and Mind: A Study of Avatar Personalization in Three Virtual Worlds. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Boston MA USA, 1151–1160. <https://doi.org/10.1145/1518701.1518877>
- [63] Nicolas Ducheneaut, Nicholas Yee, Eric Nickell, and Robert J. Moore. 2006. "Alone Together?": Exploring the Social Dynamics of Massively Multiplayer Online Games. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, Montréal Québec Canada, 407–416. <https://doi.org/10.1145/1124772.1124834>
- [64] EA Black Box. 2007. *Need for Speed: ProStreet*. Game [Multiple Platforms]. Electronic Arts, California, USA.
- [65] EA Canada. 2013. *FIFA 14*. Game [Multiple Platforms]. EA Sports, California, USA.
- [66] Inger Ekman. 2008. Psychologically Motivated Techniques for Emotional Sound in Computer Games. *Proc. AudioMostly 2008* January 2008 (2008), 20–26. <https://meaningfulnoise.wordpress.com/psychologically-motivated-techniques-for-emotional-sound-in-computer-games/>
- [67] Inger Ekman. 2013. On the desire to not kill your players: Rethinking sound in pervasive and mixed reality games. *FDG* (2013), 142–149.
- [68] Electronic Arts. 2014. *The Sims 4*. Game [Multiple Platforms].
- [69] Andrew J. Elliot, Vincent Payen, Jeanick Brisswalter, Francois Cury, and Julian F. Thayer. 2011. A subtle threat cue, heart rate variability, and cognitive performance. *Psychophysiology* (2011). <https://doi.org/10.1111/j.1469-8986.2011.01216.x>
- [70] ESA. 2021. 2021 Essential Facts About the Video Game Industry. *ESA Report 2021* (2021). <https://www.theesa.com/resource/2021-essential-facts-about-the-video-game-industry/>
- [71] European Broadcasting Union. 2011. Loudness normalisation and permitted maximum level of audio signals. (2011).
- [72] Min Fan, Uddipana Baishya, Elgin-Skye McLaren, Alissa N Antle, Shubhra Sarker, and Amal Vincent. 2018. Block talks: a tangible and augmented reality toolkit for children to learn sentence construction. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–6.
- [73] Mary Flanagan. 1999. Mobile Identities, Digital Stars, and Post-Cinematic Selves. *Wide Angle* 21, 1 (1999), 77–93. <https://doi.org/10.1353/wan.1999.0002>
- [74] Guo Freeman and Divine Maloney. 2021. Body, avatar, and me: The presentation and perception of self in social virtual reality. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (2021), 1–27.
- [75] Johnny Friberg and Dan Gärdenfors. 2004. Audio Games: New Perspectives on Game Audio. In *Proceedings of the 2004 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology - ACE '04*. ACM Press, Singapore, 148–154. <https://doi.org/10.1145/1067343.1067361>
- [76] FromSoftware. 2015. *Bloodborne*. [PlayStation 4]. Electronic Arts, California, USA.
- [77] Dilrukshi Gamage, Shantha Fernando, and Indika Perera. 2015. Quality of MOOCs: A review of literature on effectiveness and quality aspects. In *2015 8th International Conference on Ubi-Media Computing (UMEDIA)*. IEEE, 224–229.
- [78] Franco Eusébio Garcia and Vânia Paula de Almeida Neris. 2013. Design Guidelines for Audio Games. In *Human-Computer Interaction. Applications and Services*, David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Doug Tygar, Moshe Y. Vardi, Gerhard Weikum, and Masaaki Kurosu (Eds.). Vol. 8005. Springer Berlin Heidelberg, Berlin, Heidelberg, 229–238. https://doi.org/10.1007/978-3-642-39262-7_26
- [79] Radhika Garg, Hua Cui, and Yash Kapadia. 2021. "Learn, Use, and (Intermittently) Abandon": Exploring the Practices of Early Smart Speaker Adopters in Urban India. (2021).
- [80] Radhika Garg and Subhasree Sengupta. 2020. Conversational Technologies in In-home Learning: Using Co-Design to Understand Children's and Parents' Perspectives. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–13.
- [81] Radhika Garg and Subhasree Sengupta. 2020. He is just like me: a study of the long-term use of smart speakers by parents and children. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4, 1 (2020), 1–24.
- [82] Marylou Pausewang Gelfer. 1988. Perceptual attributes of voice: Development and use of rating scales. *Journal of Voice* 2, 4 (1988), 320–326. [https://doi.org/10.1016/S0892-1997\(88\)80024-9](https://doi.org/10.1016/S0892-1997(88)80024-9)
- [83] Giovanni Ribeiro, Katja Rogers, Maximilian Altmeyer, Thomas Terkildsen, and Lennart E. Nacke. 2020. Game Atmosphere Effects of Audiovisual Thematic Cohesion on Player Experience and Psychophysiology. (2020). <https://doi.org/10.1145/3410404.3414245>
- [84] Timo Gnamb, Markus Appel, and Bernard Batinic. 2010. Color red in web-based knowledge testing. *Computers in Human Behavior* 26, 6 (2010), 1625–1631. <https://doi.org/10.1016/j.chb.2010.06.010>
- [85] Mark Grimshaw. 2007. Sound and Immersion in the First-Person Shooter. (2007).
- [86] Mark Grimshaw, Craig Lindley, and Lennart Nacke. 2008. Sound and immersion in the first-person shooter: Mixed measurement of the player's sonic experience. In *Audio Mostly-a conference on interaction with sound*. www.audiomostly.com.
- [87] Jérôme Guegan, Stéphanie Buisine, Fabrice Mantelet, Nicolas Maranzana, and Frédéric Segonds. 2016. Avatar-Mediated Creativity: When Embodying Inventors Makes Engineers More Creative. *Computers in Human Behavior* 61 (Aug. 2016), 165–175. <https://doi.org/10.1016/j.chb.2016.03.024>
- [88] Perttu Hämäläinen, Teemu Mäki-Patola, Ville Pulkki, and Matti Airas. 2004. Musical Computer Games Played by Singing. In *Proc. 7th Int. Conf. on Digital Audio Effects (DAFx'04), Naples*.
- [89] Andrew F Hayes. 2017. *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford publications.
- [90] Sylvie Hébert, Renée Béland, Odrée Dionne-Fournelle, Martine Crête, and Sonia J. Lupien. 2005. Physiological stress response to video-game playing: The contribution of built-in music. *Life Sciences* 76, 20 (2005), 2371–2380. <https://doi.org/10.1016/j.lfs.2004.11.011>
- [91] Dorothee Hefner, Christoph Klimmt, and Peter Vorderer. 2007. Identification with the Player Character as Determinant of Video Game Enjoyment. In *Entertainment Computing - ICEC 2007*. 39–48. https://doi.org/10.1007/978-3-540-74873-1_6
- [92] E Tory Higgins. 1987. Self-discrepancy: a theory relating self and affect. *Psychological review* 94, 3 (1987), 319.

- [93] Alexis Hiniker, Amelia Wang, Jonathan Tran, Mingrui Ray Zhang, Jenny Radesky, Kiley Sobel, and Sungsoo Ray Hong. 2021. Can Conversational Agents Change the Way Children Talk to People? (2021).
- [94] David C. Hoaglin and Boris Iglewicz. 1987. Fine-Tuning Some Resistant Rules for Outlier Labeling. *J. Amer. Statist. Assoc.* 82, 400 (1987), 1147–1149. <https://doi.org/10.1080/01621459.1987.10478551>
- [95] Thomas Holmes. 2021. Defining Voice Design in Video Games. (2021).
- [96] John J Horton, David G Rand, and Richard J Zeckhauser. 2011. The online laboratory: Conducting experiments in a real labor market. *Experimental Economics* 14, 3 (2011), 399–425.
- [97] Rex Hsieh and Hisashi Sato. 2021. Evaluation of Avatar and Voice Transform in Programming E-Learning Lectures. *Journal on Multimodal User Interfaces* 15, 2 (June 2021), 121–129. <https://doi.org/10.1007/s12193-020-00349-5>
- [98] Bart Hulshof. 2013. The influence of colour and scent on people's mood and cognitive performance in meeting rooms. *Master Thesis* May (2013), 1–97.
- [99] Takeo Igarashi and John F Hughes. 2001. Voice as sound: using non-verbal voice input for interactive control. In *Proceedings of the 14th annual ACM symposium on User interface software and technology*. 155–156.
- [100] Katherine Isbister. 2006. *Better game characters by design: A psychological approach*. Elsevier/Morgan Kaufmann.
- [101] Jamie Banks, Nicholas David Bowman, and Joseph Wasserman. 2017. A Bard in the Hand: The Role of Materiality in Player-Character Relationships. *Imagination, Cognition and Personality* (2017). <https://doi.org/10.1177/0276236617748130>
- [102] Josh Jarrett. 2021. Gaming the Gift: The Affective Economy of League of Legends 'Fair' Free-to-Play Model. *Journal of Consumer Culture* 21, 1 (Feb. 2021), 102–119. <https://doi.org/10.1177/1469540521993932>
- [103] Colby Johanson and Regan L. Mandryk. 2016. Scaffolding Player Location Awareness through Audio Cues in First-Person Shooters. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems - CHI '16* (2016), 3450–3461. <https://doi.org/10.1145/2858036.2858172>
- [104] Daniel Johnson, M John Gardner, and Ryan Perry. 2018. Validation of two game experience scales: the Player Experience of Need Satisfaction (PENS) and Game Experience Questionnaire (GEQ). *International Journal of Human-Computer Studies* (2018). <https://doi.org/10.1016/j.ijhcs.2018.05.003>
- [105] Yasmin B. Kafai, Deborah A. Fields, and Melissa S. Cook. 2010. Your Second Selves: Player-Designed Avatars. *Games and Culture* 5, 1 (Jan. 2010), 23–42. <https://doi.org/10.1177/1555412009351260>
- [106] Yasmin B Kafai, Gabriela T Richard, and Brendesha M Tynes. 2017. *Diversifying Barbie and Mortal Kombat: Intersectional perspectives and inclusive designs in gaming*. Lulu. com.
- [107] Dominic Kao. 2019. The effects of anthropomorphic avatars vs. non-anthropomorphic avatars in a jumping game. In *Proceedings of the 14th International Conference on the Foundations of Digital Games*. 1–5.
- [108] Dominic Kao. 2019. Infinite Loot Box: A platform for simulating video game loot boxes. *IEEE Transactions on Games* 12, 2 (2019), 219–224.
- [109] Dominic Kao. 2019. JavaStrike: A Java Programming Engine Embedded in Virtual Worlds. In *Proceedings of The Fourteenth International Conference on the Foundations of Digital Games*.
- [110] Dominic Kao. 2020. The effects of juiciness in an action RPG. *Entertainment Computing* 34 (2020), 100359.
- [111] Dominic Kao and D. Fox Harrell. 2016. Exploring the Impact of Avatar Color on Game Experience in Educational Games. *Proceedings of the 34th Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI 2016)* (2016).
- [112] Dominic Kao and D. Fox Harrell. 2017. MazeStar: A Platform for Studying Virtual Identity and Computer Science Education. In *Foundations of Digital Games*.
- [113] Dominic Kao and D. Fox Harrell. 2018. The Effects of Badges and Avatar Identification on Play and Making in Educational Games. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal QC Canada, 1–19. <https://doi.org/10.1145/3173574.3174174>
- [114] Dominic Kao, Rabindra Ratan, Christos Mousas, and Alejandra Magana. 2021. The Effects of a Self-Similar Avatar Voice in Educational Games. *CHI PLAY* (2021).
- [115] Oleksandra Keehl and Edward Melcer. 2019. Radical tunes: exploring the impact of music on memorization of stroke order in logographic writing systems. In *Proceedings of the 14th International Conference on the Foundations of Digital Games*. 1–6.
- [116] Ahmed Kharrufa, David Leat, and Patrick Olivier. 2010. Digital mysteries: designing for learning at the tabletop. In *ACM International Conference on Interactive Tabletops and Surfaces*. 197–206.
- [117] Changsoo Kim, Sang-Gun Lee, and Minchoel Kang. 2012. I Became an Attractive Person in the Virtual World: Users' Identification with Virtual Communities and Avatars. *Computers in Human Behavior* 28, 5 (Sept. 2012), 1663–1669. <https://doi.org/10.1016/j.chb.2012.04.004>
- [118] Soomin Kim, Joonhwan Lee, and Gahgene Gweon. 2019. Comparing data from chatbot and web surveys: Effects of platform and conversational style on survey response quality. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–12.
- [119] Tae Soo Kim, Seungsu Kim, Yoonseo Choi, and Juho Kim. 2021. Linking Speech and Visual Objects to Support Communication in Asynchronous Collaboration. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–17.
- [120] René F Kizilcec, Jeremy N Bailenson, and Charles J Gomez. 2015. The instructor's face in video instruction: Evidence from two large-scale field studies. *Journal of Educational Psychology* 107, 3 (2015), 724.
- [121] René F Kizilcec, Kathryn Papadopoulos, and Lalida Sritanyaratana. 2014. Showing face in video instruction: effects on information retention, visual attention, and affect. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 2095–2102.
- [122] Christoph Klimmt. 2003. Dimensions and determinants of the enjoyment of playing digital games: A three-level model. In *Level up: Digital games research conference*. 246–257.
- [123] Christoph Klimmt, Dorothee Hefner, Peter Vorderer, Christian Roth, and Christopher Blake. 2010. Identification with video game characters as automatic shift of self-perceptions. *Media Psychology* 13, 4 (2010), 323–338.
- [124] Christoph Klimmt, Daniel Possler, Nicolas May, Hendrik Auge, Louisa Wanjek, and Anna-Lena Wolf. 2019. Effects of Soundtrack Music on the Video Game Experience. *Media Psychology* 22, 5 (Sept. 2019), 689–713. <https://doi.org/10.1080/15213269.2018.1507827>
- [125] Anya Kolesnichenko, Joshua McVeigh-Schultz, and Katherine Isbister. 2019. Understanding emerging design practices for avatar systems in the commercial social vr ecology. In *Proceedings of the 2019 on Designing Interactive Systems Conference*. 241–252.
- [126] Jordan Koulouris, Zoe Jeffery, James Best, Eamonn O'Neill, and Christof Luteroth. 2020. Me vs. Super(Wo)Man: Effects of Customization and Identification in a VR Exergame. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, Honolulu HI USA, 1–17. <https://doi.org/10.1145/3313831.3376661>
- [127] Christof Kuhbandner and Reinhard Pekrun. 2013. Joint effects of emotion and color on memory. *Emotion (Washington, D.C.)* 13, 3 (2013), 375–9. <https://doi.org/10.1037/a0031821>
- [128] Rohit Kumar, Gahgene Gweon, Mahesh Joshi, Yue Cui, and Carolyn Penstein Rosé. 2007. Supporting students working together on math with social dialogue. In *Workshop on Speech and Language Technology in Education*. Citeseer.
- [129] Daniel Lakens. 2013. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in psychology* 4 (2013), 863.
- [130] Monica Landoni, Emiliana Murgia, Theo Huibers, and Maria Soledad Pera. 2019. My Name is Sonny, How May I help You Searching for Information?. In *18th ACM International Conference on Interaction Design and Children, IDC 2019*. Association for Computing Machinery (ACM).
- [131] Pontus Larsson, Aleksander Våljamäe, Daniel Västfjäll, Ana Tajadura-Jiménez, and Mendel Kleiner. 2010. Auditory-Induced Presence in Mixed Reality Environments and Related Technology. (2010), 143–163. https://doi.org/10.1007/978-1-84882-733-2_8 arXiv:arXiv:1011.1669v3
- [132] Eun Ju Lee, Clifford Nass, and Scott Brave. 2000. Can computer-generated speech have gender? An experimental test of gender stereotype. In *CHI'00 extended abstracts on Human factors in computing systems*. 289–290.
- [133] Kwan Min Lee, Katharine Liao, and Seoungho Ryu. 2007. Children's responses to computer-synthesized speech in educational media: gender consistency and gender similarity effects. *Human communication research* 33, 3 (2007), 310–329.
- [134] Benjamin J. Li and May O. Lwin. 2016. Player See, Player Do: Testing an Exergame Motivation Model Based on the Influence of the Self Avatar. *Computers in Human Behavior* 59 (June 2016), 350–357. <https://doi.org/10.1016/j.chb.2016.02.034>
- [135] Benjamin J Li, May O Lwin, and Younbo Jung. 2014. Wii, Myself, and Size: The Influence of Proteus Effect and Stereotype Threat on Overweight Children's Exercise Motivation and Behavior in Exergames. *Games for health: Research, Development, and Clinical Applications* 3, 1 (2014), 40–48.
- [136] Gen-Yih Liao, TCE Cheng, and Ching-I Teng. 2019. How do avatar attractiveness and customization impact online gamers' flow and loyalty? *Internet Research* (2019).
- [137] Mats Liljedahl. 2011. Sound for Fantasy and Freedom. In *Game Sound Technology and Player Interaction: Concepts and Developments*. IGI Global, 22–43.
- [138] Sohye Lim and Byron Reeves. 2009. Being in the Game: Effects of Avatar Choice and Point of View on Psychophysiological Responses During Play. *Media Psychology* 12, 4 (Nov. 2009), 348–370. <https://doi.org/10.1080/15213260903287242>
- [139] Lighthouse Labs. 2021. State of Online Gaming 2021. (2021). <https://www.lighthouse.com/lp/state-of-online-gaming-2021/>
- [140] Lorraine Lin, Dhaval Parmar, Sabarish V. Babu, Alison E. Leonard, Shaundra B. Daily, and Sophie Jörg. 2017. How Character Customization Affects Learning in Computational Thinking. In *Proceedings of the ACM Symposium on Applied Perception*. ACM, Cottbus Germany, 1–8. <https://doi.org/10.1145/3119881.3119884>
- [141] Linden Lab. 2003. *Second Life*. Game [Multiple Platforms]. Linden Lab, San Francisco, USA.

- [142] Conor Linehan, George Bellord, Ben Kirman, Zachary H. Morford, and Bryan Roche. 2014. Learning curves: Analysing pace and challenge in four successful puzzle games. *CHI PLAY 2014 - Proceedings of the 2014 Annual Symposium on Computer-Human Interaction in Play* (2014), 181–190. <https://doi.org/10.1145/2658537.2658695>
- [143] LingoJam. 2020. Robot Voice Generator. <https://lingojam.com/RobotVoiceGenerator>
- [144] Duri Long, Hannah Guthrie, and Brian Magerko. 2018. Don't steal my balloons: designing for musical adult-child ludic engagement. In *Proceedings of the 17th ACM Conference on Interaction Design and Children*. 657–662.
- [145] Nichola Lubold, Erin Walker, and Heather Pon-Barry. 2021. Effects of adapting to user pitch on rapport perception, behavior, and state with a social robotic learning companion. *User Modeling and User-Adapted Interaction* 31, 1 (2021), 35–73.
- [146] Michelle Lui, Rhonda McEwen, and Martha Mullally. 2020. Immersive virtual reality for supporting complex scientific knowledge: Augmenting our understanding with physiological monitoring. *British Journal of Educational Technology* 51, 6 (2020), 2181–2199.
- [147] Roberto Martinez Maldonado, Judy Kay, Kalina Yacef, and Beat Schwendimann. 2012. An interactive teacher's dashboard for monitoring groups in a multi-tabletop learning environment. In *International Conference on Intelligent Tutoring Systems*. Springer, 482–492.
- [148] Hazel Markus and Paula Nurius. 1986. Possible selves. *American psychologist* 41, 9 (1986), 954.
- [149] Winter Mason and Siddharth Suri. 2012. Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods* 44, 1 (2012), 1–23. <https://doi.org/10.3758/s13428-011-0124-6>
- [150] Richard E Mayer, Kristina Sobko, and Patricia D Mautone. 2003. Social cues in multimedia learning: Role of speaker's voice. *Journal of educational Psychology* 95, 2 (2003), 419.
- [151] Oscar Mayor, Jordi Bonada, and Jordi Janer. 2009. Kaleivoicope: Voice transformation from interactive installations to video-games. In *Proceedings of the AES International Conference*.
- [152] Victoria McArthur. 2017. The UX of Avatar Customization. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, Denver Colorado USA, 5029–5033. <https://doi.org/10.1145/3025453.3026020>
- [153] Victoria McArthur. 2018. Challenging the User-Avatar Dichotomy in Avatar Customization Research. 9, 1 (2018), 21.
- [154] Victoria McArthur. 2019. Making Mi: Studying the Effects of Methodological Approaches and Gaming Contexts on Avatar Customization. *Behaviour & Information Technology* 38, 3 (March 2019), 230–243. <https://doi.org/10.1080/0144929X.2018.1526969>
- [155] Victoria McArthur and Jennifer Jenson. 2014. E Is for Everyone? Best Practices for the Socially Inclusive Design of Avatar Creation Interfaces. In *Proceedings of the 2014 Conference on Interactive Entertainment*. ACM, Newcastle NSW Australia, 1–8. <https://doi.org/10.1145/2677758.2677783>
- [156] Victoria McArthur, Robert John Teather, and Jennifer Jenson. 2015. The Avatar Affordances Framework: Mapping Affordances and Design Trends in Character Creation Interfaces. In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play*. ACM, London United Kingdom, 231–240. <https://doi.org/10.1145/2793107.2793121>
- [157] Edward McAuley, Terry Duncan, and Vance V Tammen. 1989. Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: A confirmatory factor analysis. *Research Quarterly for Exercise and Sport* 60, 1 (1989), 48–58.
- [158] Ravi Mehta and Rui(Juliet) Zhu. 2008. Blue or Red? Exploring the Effect of Color on Cognitive Task Performances. *Science* 323, February (2008), 1226–1229. <https://doi.org/10.1126/science.1169144>
- [159] M.A. Meier, Russell A. Hill, Andrew J. Elliot, and R.A. Barton. 2015. Color in Achievement Contexts in Humans. *Handbook of Color Psychology* 44, February (2015), 0–103. <https://doi.org/10.1063/1.2756072>
- [160] Scott Menard. 2002. *Applied logistic regression analysis*. Vol. 106. Sage.
- [161] Microsoft Game Studios and Electronic Arts. 2007. Mass Effect. Game [Multiple Platforms].
- [162] Jeremy Miles and Mark Shevlin. 2001. *Applying regression and correlation: A guide for students and researchers*. Sage.
- [163] Modulate. 2021. VoiceWear. <https://www.modulate.ai/voice-wear>
- [164] Toni-Jan Keith Palma Monserrat, Shengdong Zhao, Kevin McGee, and Anshul Vikram Pandey. 2013. Netvideo: Facilitating navigation of blackboard-style lecture videos. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 1139–1148.
- [165] Marlene M Moretti and E Tory Higgins. 1990. Relating self-discrepancy to self-esteem: The contribution of discrepancy beyond actual-self ratings. *Journal of Experimental Social Psychology* 26, 2 (1990), 108–123.
- [166] Briana B Morrison and Betsy DiSalvo. 2014. Khan academy gamifies computer science. In *Proceedings of the 45th ACM technical symposium on Computer science education*. 39–44.
- [167] Sheila C Murphy. 2004. 'Live in your world, play in ours': The spaces of video game identity. *Journal of visual culture* 3, 2 (2004), 223–238.
- [168] Lennart E. Nacke and Mark Grimshaw. 2011. Player-Game Interaction Through Affective Sound. *Game Sound Technology and Player Interaction* (2011), 264–285. <https://doi.org/10.4018/978-1-61692-828-5.ch013>
- [169] Lennart E. Nacke, Mark N. Grimshaw, and Craig A. Lindley. 2010. More than a Feeling: Measurement of Sonic User Experience and Psychophysiology in a First-Person Shooter Game. *Interacting with Computers* 22, 5 (Sept. 2010), 336–343. <https://doi.org/10.1016/j.intcom.2010.04.005>
- [170] Lisa Nakamura. 2013. *Cybertypes: Race, Ethnicity, and Identity on the Internet*. Routledge.
- [171] Clifford Nass and Kwan Min Lee. 2000. Does Computer-Generated Speech Manifest Personality? An Experimental Test of Similarity-Attraction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems - CHI '00*. ACM Press, The Hague, The Netherlands, 329–336. <https://doi.org/10.1145/332040.332452>
- [172] Clifford Nass and Kwan Min Lee. 2001. Does Computer-Synthesized Speech Manifest Personality? Experimental Tests of Recognition, Similarity-Attraction, and Consistency-Attraction. *Journal of Experimental Psychology: Applied* 7, 3 (2001), 171–181. <https://doi.org/10.1037/1076-898X.7.3.171>
- [173] Greg L Nelson, Benjamin Xie, and Andrew J Ko. 2017. Comprehension first: evaluating a novel pedagogy and tutoring system for program tracing in CS1. In *Proceedings of the 2017 ACM Conference on International Computing Education Research*. 2–11.
- [174] Tobias Neuhold. [n.d.]. The Role of Audio for Spatial and Affective Involvement in Survival Horror Games. (n.d.), 16.
- [175] Raymond Ng and Robb Lindgren. 2013. Examining the Effects of Avatar Customization and Narrative on Engagement and Learning in Video Games. In *Proceedings of CGAMES'2013 USA*. IEEE, Louisville, KY, 87–90. <https://doi.org/10.1109/CGames.2013.6632611>
- [176] Nintendo EAD. 2006. *The Legend of Zelda: Twilight Princess*. Game [Multiple Platforms]. Nintendo, Kyoto, Japan.
- [177] Tyler Pace, Aaron Houssian, and Victoria McArthur. 2009. Are Socially Exclusive Values Embedded in the Avatar Creation Interfaces of MMORPGs? *Journal of Information, Communication and Ethics in Society* (2009).
- [178] Pearl Abyss. 2015. Black Desert Online. Game [Multiple Platforms].
- [179] Anthony J Pellicone and June Ahn. 2017. The Game of Performing Play: Understanding streaming as cultural production. In *Proceedings of the 2017 CHI conference on human factors in computing systems*. 4863–4874.
- [180] Wei Peng, Jih Hsuan Lin, Karin A. Pfeiffer, and Brian Winn. 2012. Need Satisfaction Supportive Game Features as Motivational Determinants: An Experimental Study of a Self-Determination Theory Guided Exergame. *Media Psychology* 15, 2 (2012), 175–196. <https://doi.org/10.1080/15213269.2012.673850>
- [181] Daniel Pimentel and Sri Kalyanaraman. 2020. Your Own Worst Enemy: Implications of the Customization, and Destruction, of Non-Player Characters. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. ACM, Virtual Event Canada, 93–106. <https://doi.org/10.1145/3410404.3414269>
- [182] Cale Plut and Philippe Pasquier. 2019. Music Matters: An Empirical Study on the Effects of Adaptive Music on Experienced and Perceived Player Affect. In *2019 IEEE Conference on Games (CoG)*. IEEE, London, United Kingdom, 1–8. <https://doi.org/10.1109/CIG.2019.8847951>
- [183] Andrew K. Przybylski, C. Scott Rigby, and Richard M. Ryan. 2010. A Motivational Model of Video Game Engagement. *Review of General Psychology* 14, 2 (2010), 154–166. <https://doi.org/10.1037/a0019440>
- [184] Andrew K Przybylski, Richard M Ryan, and C Scott Rigby. 2009. The motivating role of violence in video games. *Personality and social psychology bulletin* 35, 2 (2009), 243–259.
- [185] Rabindra Ratan. 2017. Companions & Vehicles. In *Avatars, Assembled: The Sociotechnical Anatomy of Digital Bodies*, Jaime Banks (Ed.). Peter Lang, Digital Formations Series.
- [186] Rabindra Ratan, David Beyea, Benjamin J. Li, and Luis Graciano. 2020. Avatar Characteristics Induce Users' Behavioral Conformity with Small-to-Medium Effect Sizes: A Meta-Analysis of the Proteus Effect. *Media Psychology* 23, 5 (Sept. 2020), 651–675. <https://doi.org/10.1080/15213269.2019.1623698>
- [187] Rabindra Ratan and Young June Sah. 2015. Leveling up on stereotype threat: The role of avatar customization and avatar embodiment. *Computers in Human Behavior* 50 (2015), 367–374.
- [188] Johnmarshall Reeve. 1989. The interest-enjoyment distinction in intrinsic motivation. *Motivation and emotion* 13, 2 (1989), 83–103.
- [189] Gabriela T Richard and Kishonna L Gray. 2018. Gendered play, racialized reality: Black cyberfeminism, inclusive communities of practice, and the intersections of learning, socialization, and resilience in online gaming. *Frontiers: A Journal of Women Studies* 39, 1 (2018), 112–148.
- [190] Gabriela T Richard and Christopher M Hoadley. 2013. Investigating a supportive online gaming community as a means of reducing stereotype threat vulnerability across gender. *Proceedings of Games, Learning & Society* 9 (2013), 261–266.
- [191] Jessica Roberts, Amartya Banerjee, Annette Hong, Steven McGee, Michael Horn, and Matt Matcuk. 2018. Digital exhibit labels in museums: promoting visitor engagement with cultural artifacts. In *Proceedings of the 2018 CHI Conference on*

- Human Factors in Computing Systems*. 1–12.
- [192] Rockstar Games. 2018. Red Dead Redemption 2. Game [Multiple Platforms].
- [193] Katja Rogers. 2017. Exploring the Role of Audio in Games. In *Extended Abstracts Publication of the Annual Symposium on Computer-Human Interaction in Play*. ACM, Amsterdam The Netherlands, 727–731. <https://doi.org/10.1145/3130859.3133227>
- [194] Elisa Rubegni, Rebecca Dore, Monica Landoni, and Ling Kan. 2021. “The girl who wants to fly”: Exploring the role of digital technology in enhancing dialogic reading. *International Journal of Child-Computer Interaction* 30 (2021), 100239.
- [195] Elisa Rubegni, Vito Gentile, Alessio Malizia, Salvatore Sorce, and Niko Kargas. 2020. Child–display interaction: Lessons learned on touchless avatar-based large display interfaces. *Personal and Ubiquitous Computing* (2020), 1–14.
- [196] Richard M Ryan and Edward L Deci. 2000. Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary educational psychology* 25, 1 (2000), 54–67.
- [197] Richard M Ryan, Richard Koestner, and Edward L Deci. 1991. Ego-involved persistence: When free-choice behavior is not intrinsically motivated. *Motivation and emotion* 15, 3 (1991), 185–205.
- [198] Richard M. Ryan, C. Scott Rigby, and Andrew Przybylski. 2006. The Motivational Pull of Video Games: A Self-Determination Theory Approach. *Motivation and Emotion* 30, 4 (2006), 344–360. <https://doi.org/10.1007/s11031-006-9051-8>
- [199] Timothy Sanders and Paul Cairns. 2010. Time perception, immersion and music in videogames. *Proceedings of HCI 2010 24* (2010), 160–167.
- [200] Carol Sansone, Charlene Weir, Lora Harpster, and Carolyn Morgan. 1992. Once a Boring Task Always a Boring Task?: Interest as a Self-Regulatory Mechanism. *Journal of Personality and Social Psychology* (1992). <https://doi.org/10.1037/0022-3514.63.3.379>
- [201] Mike Schmierbach, Anthony M. Limperos, and Julia K. Woolley. 2012. Feeling the Need for (Personalized) Speed: How Natural Controls and Customization Contribute to Enjoyment of a Racing Game Through Enhanced Immersion. *Cyberpsychology, Behavior, and Social Networking* 15, 7 (July 2012), 364–369. <https://doi.org/10.1089/cyber.2011.0025>
- [202] Alexander M. Schoemann, Aaron J. Boulton, and Stephen D. Short. 2017. Determining Power and Sample Size for Simple and Complex Mediation Models. *Social Psychological and Personality Science* 8, 4 (2017), 379–386. <https://doi.org/10.1177/1948550617715068>
- [203] P E Shrout and J L Fleiss. 1979. Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin* 86, 2 (1979), 420–428. <https://doi.org/10.1037/0033-2909.86.2.420>
- [204] Janet Siegmund, Christian Kästner, Jörg Liebig, Sven Apel, and Stefan Hanenberg. 2014. Measuring and modeling programming experience. *Empirical Software Engineering* 19, 5 (2014), 1299–1334. <https://doi.org/10.1007/s10664-013-9286-4>
- [205] Siu-Lan Tan, John Baxa, and Matthew P. Spackman. 2010. Effects of Built-in Audio versus Unrelated Background Music on Performance in an Adventure Role-Playing Game. *International Journal of Gaming and Computer-mediated Simulations* (2010). <https://doi.org/10.4018/jgcms.2010070101>
- [206] Peter Smucker. 2018. Gaming Sober, Playing Drunk: Sound Effects of Alcohol in Video Games. *The Computer Games Journal* 7, 4 (2018), 291–311.
- [207] Alistair Raymond Bryce Soutter and Michael Hitchens. 2016. The Relationship between Character Identification and Flow State within Video Games. *Computers in Human Behavior* 55 (Feb. 2016), 1030–1038. <https://doi.org/10.1016/j.chb.2015.11.012>
- [208] Square Enix. 2013. Final Fantasy XIV. Game [Multiple Platforms].
- [209] Standing Stone Games. 2007. *The Lord of the Rings Online*. Game [Multiple Platforms]. Warner Bros. Interactive Entertainment, California, USA.
- [210] Sharon T. Steinemann, Elisa D. Mekler, and Klaus Opwis. 2015. Increasing Donating Behavior Through a Game for Change. <https://doi.org/10.1145/2793107.2793125>
- [211] Axel Stockburger. 2010. The play of the voice: The role of the voice in contemporary video and computer games. *Voice: Vocal aesthetics in digital arts and media*. Cambridge: MIT Press. van Leeuwen (2010).
- [212] Barbara G Tabachnick and Linda S Fidell. 2007. *Using multivariate statistics*. Allyn & Bacon/Pearson Education.
- [213] Telltale Games. 2012. The Walking Dead. Game [Multiple Platforms].
- [214] Gary F Templeton. 2011. A two-step approach for transforming continuous variables to normal: implications and recommendations for IS research. *Communications of the Association for Information Systems* 28, 1 (2011), 4.
- [215] Ching-I Teng. 2021. How can avatar’s item customizability impact gamer loyalty? *Telematics and Informatics* 62 (2021), 101626.
- [216] Xiaoyi Tian, Nichola Lubold, Leah Friedman, and Erin Walker. 2020. Understanding Rapport over Multiple Sessions with a Social, Teachable Robot. In *International Conference on Artificial Intelligence in Education*. Springer, 318–323.
- [217] Sabine Trepte and Leonard Reinecke. 2010. Avatar Creation and Video Game Enjoyment: Effects of Life-Satisfaction, Game Competitiveness, and Identification with the Avatar. *Journal of Media Psychology* 22, 4 (Jan. 2010), 171–184. <https://doi.org/10.1027/1864-1105/a000022>
- [218] Stefano Triberti, Ilaria Durosini, Filippo Aschieri, Daniela Villani, and Giuseppe Riva. 2017. Changing Avatars, Changing Selves? The Influence of Social and Contextual Expectations on Digital Rendition of Identity. *Cyberpsychology, Behavior, and Social Networking* 20, 8 (Aug. 2017), 501–507. <https://doi.org/10.1089/cyber.2016.0424>
- [219] Selen Turkey and Sonam Adinolf. 2010. Free to Be Me: A Survey Study on Customization with World of Warcraft and City Of Heroes/Villains Players. *Procedia - Social and Behavioral Sciences* 2, 2 (2010), 1840–1845. <https://doi.org/10.1016/j.sbspro.2010.03.995>
- [220] Selen Turkey and Charles K. Kinzer. 2014. The Effects of Avatar-Based Customization on Player Identification. *International Journal of Gaming and Computer-Mediated Simulations* 6, 1 (Jan. 2014), 1–25. <https://doi.org/10.4018/jgcms.2014010101>
- [221] Selen Turkey and Charles K Kinzer. 2015. The effects of avatar-based customization on player identification. In *Gamification: Concepts, methodologies, tools, and applications*. IGI Global, 247–272.
- [222] Ubisoft Kyiv Ubisoft Montreal, Ubisoft Milan. 2013. *Assassin’s Creed IV: Black Flag*. Game [Multiple Platforms]. Ubisoft, Paris, France.
- [223] Ubisoft Toronto. 2013. *Tom Clancy’s Splinter Cell: Blacklist*. Game [Multiple Platforms]. Ubisoft, Paris, France.
- [224] Jan Van Looy, Cédric Courtois, Melanie De Vocht, and Lieven De Marez. 2012. Player Identification in Online Games: Validation of a Scale for Measuring Identification in MMOGs. *Media Psychology* 15, 2 (May 2012), 197–221. <https://doi.org/10.1080/15213269.2012.674917>
- [225] Kellie Vella, Madison Klarkowski, Selen Turkey, and Daniel Johnson. 2020. Making friends in online games: gender differences and designing for greater social connectedness. *Behaviour and Information Technology* (2020). <https://doi.org/10.1080/0144929X.2019.1625442>
- [226] VentureBeat. 2020. Newzoo: U.S. gamers are in love with skins and in-game cosmetics. <https://venturebeat.com/2020/12/18/newzoo-u-s-gamers-are-in-love-with-skins-and-in-game-cosmetics/>
- [227] Daniela Villani, Elena Gatti, Stefano Triberti, Emanuela Confalonieri, and Giuseppe Riva. 2016. Exploration of Virtual Body-Representation in Adolescence: The Role of Age and Sex in Avatar Customization. *SpringerPlus* 5, 1 (Dec. 2016), 740. <https://doi.org/10.1186/s40064-016-2520-y>
- [228] Jellyvision Vivarium. 1999. *Seaman*. Game [Multiple Platforms]. Sega, Tokyo, Japan.
- [229] Voicemod. 2020. Voicemod. <https://www.voicemod.net/>
- [230] Volition and Deep Silver. 2013. Saints Row IV. Game [Multiple Platforms].
- [231] Greg Wadley, Marcus Carter, and Martin Gibbs. 2015. Voice in virtual worlds: The design, use, and influence of voice chat in online play. *Human-Computer Interaction* 30, 3-4 (2015), 336–365.
- [232] Greg Wadley, Martin Gibbs, and Peter Benda. 2007. Speaking in character: using voice-over-IP to communicate within MMORPGs. In *Proceedings of the 4th Australasian conference on Interactive entertainment*. 1–8.
- [233] Greg Wadley, Martin R. Gibbs, and Nicolas Ducheneaut. 2009. You can be too rich: Mediated communication in a virtual world. In *Proceedings of the 21st Annual Conference of the Australian Computer-Human Interaction Special Interest Group - Design: Open 24/7, OZCHI '09*. <https://doi.org/10.1145/1738826.1738835>
- [234] Zachary Charles Waggoner. 2007. *Passage to morrowind: (dis)locating virtual and real identities in video role-playing games*. Arizona State University.
- [235] Helen Wauck, Gale Lucas, Ari Shapiro, Andrew Feng, Jill Boberg, and Jonathan Gratch. 2018. Analyzing the effect of avatar self-similarity on men and women in a search and rescue game. *Conference on Human Factors in Computing Systems - Proceedings 2018-April* (2018). <https://doi.org/10.1145/3173574.3174059>
- [236] David Westerman, Ron Tamborini, and Nicholas David Bowman. 2015. The effects of static avatars on impression formation across different contexts on social networking sites. *Computers in Human Behavior* 53 (2015), 111–117.
- [237] Hanna Elina Wirman and Rhys Jones. 2017. *Voice and Sound: Player Contributions to Speech*. Peter Lang, Digital Formations Series.
- [238] Wizet. 2003. *MapleStory*. [Microsoft Windows]. Nexon, Seoul, South Korea.
- [239] Nick Yee and J Bailenson. 2007. The Proteus Effect: The Effect of Transformed Self-Representation on Behavior. *Human communication research* (2007), 1–38. <http://onlinelibrary.wiley.com/doi/10.1111/j.1468-2958.2007.00299.x/full>
- [240] Robert B Zajonc. 2001. Mere exposure: A gateway to the subliminal. *Current directions in psychological science* 10, 6 (2001), 224–228.
- [241] David Zentle, Paul Cairns, and Daniel Kudenko. 2015. Higher graphical fidelity decreases players’ access to aggressive concepts in violent video games. In *CHI PLAY 2015 - Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play*. <https://doi.org/10.1145/2793107.2793113>
- [242] Lotus Zhang, Lucy Jiang, Nicole Washington, Augustina Ao Liu, Jingyao Shao, Adam Fournery, Meredith Ringel Morris, and Leah Findlater. 2021. Social Media through Voice: Synthesized Voice Qualities and Self-Presentation. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (April 2021), 1–21. <https://doi.org/10.1145/3449235>
- [243] Oren Zuckerman, Dina Walker, Andrey Grishko, Tal Moran, Chen Levy, Barak Lisak, Iddo Yehoshua Wald, and Hadas Erel. 2020. Companionship Is Not a Function: The Effect of a Novel Robotic Object on Healthy Older Adults’ Feelings of “Being-Seen”. In *Proceedings of the 2020 CHI conference on human factors in*

computing systems. 1–14.