# AVOCUS: A Voice Customization System for Online Personas

Hyeon Jeong Byeon
cat@ewhain.net
Ewha Womans University
Seoul, Korea, Republic of

Seungjin Ha
hastar2027@ewhain.net
Ewha Womans University
Seoul, Korea, Republic of

Uran Oh*
uran.oh@ewha.ac.kr
Ewha Womans University
Seoul, Korea, Republic of

## ABSTRACT

Many digital applications offer avatar customization options, positively affecting user experience. However, the adoption of auditory aspects in avatar customization has often been neglected and may have been understudied for its potential. Inspired by prior research that uncovers end-user's demands for voice customization, we seek to apply the identified implications into practice and discover end-user's voice preferences and behavior towards voice customization systems. To this end, we designed and deployed AVOCUS, a web application that enables users to search for specific voices or manipulate voice-related parameters to generate a voice similar to a target voice. Our findings suggest that (1) searching for specific voice using hashtags were perceived to be easy, (2) customized voices generated from voice reflection and voice parameter control functions had high satisfaction, and (3) participants tend to reflect the features of their desired voices when customizing their own voice.

## CCS CONCEPTS

• **Human-centered computing → Empirical studies in interaction design**.

## KEYWORDS

voice customization, voice perception, voice quality, online communication, user evaluation

**ACM Reference Format:**
Hyeon Jeong Byeon, Seungjin Ha, and Uran Oh. 2023. AVOCUS: A Voice Customization System for Online Personas. *J. ACM* 37, 4, Article 111 (August 2023), 6 pages. https://doi.org/10.1145/3544549.3585892

## 1 INTRODUCTION

As the new era of the metaverse rapidly approaches, the boundaries between real life and virtual reality are becoming thinner. Virtual avatars are gaining popularity on the internet, expanding their influence from social media to live-streaming platforms. Many digital applications offer users to choose avatars to represent their owners. Such applications provide avatar customization experiences, allowing the users to modify the visual features of the avatar such as physical (*e.g.,* body shape), demographic (*e.g.,* gender, age, race), and transient (*e.g.,* clothes, ornaments) aspects [13]. As avatar customization is ubiquitous among systems, a large corpus of literature has focused on the positive effects of avatar customization on users' online experiences. Research has shown that avatar customization can higher avatar similarity with the user, resulting in higher satisfaction and self-presence in a virtual world [10]. Also, customized avatars are effective in offering greater game enjoyment [3] and physiological arousal [17]. However, the adoption of auditory aspects in avatar customization has often been neglected due to substantial overhead (*e.g.,* multiple voice actors, region localization) and indifference among researchers [13, 27]. In addition, researchers discuss the difficulties of establishing a consensus of recognizable terminologies that describe voice [21], burdening the vagueness of voice customization.

Yet, audio forms a significant part of people's individuality both online and offline. Prior research has proven that voice influences the evaluation of other people's impressions [8]. Also, users participating in online games with vocal interactions experience increased physiological responses [9], emotional authenticity [4, 6], performance [12], and immersion [7, 14, 16, 19, 20]. Given prior work indicating the importance of modifying the visual aspects of avatars and online vocal communications separately, voice customization for online personas may be understudied for its potential. For instance, voice customization can be used to set different voices for different contexts, as prior research has proven that there is demand for preferable voices that represent the user's profile for each social media platform [29]. Also, voice customization can keep the anonymity for users even in vocal communications, where there is a prominent tendency among female users in online communication to hinder their identity (*e.g.,* gender) for particular situations [23, 28]. Lastly, voice customization can diversify the choices of artificial voices generated by Text-to-speech engines. Although synthetic speech has approached human-level naturalness, there are limited options to personalize the results to more closely match the vocal identity of the users [25]. Customized artificial voices may support social communication and identity display for people with speech or hearing disorders.

Motivated by prior research that uncovers how individuals wish to change their voice in what circumstances [5], we seek to apply the identified implications from prior work into practice and discover end-user's voice preferences and behavior toward voice customization systems. Research questions are below.

- *RQ1. How do individuals find ease in use in voice search engines?*
- *RQ2. How could we build efficient tools with intuitive prototypes for voice customization?*
- *RQ3. How do individuals wish to change their voice using what interfaces?*

To address these questions, we first designed and deployed AVOCUS, a voice customization system, as a web application that

*The corresponding author.


0004-5411/2023/8-ART111
https://doi.org/10.1145/3544549.3585892

enables users to search for specific voices for customization or manipulate voice-related parameters to generate a voice similar to a target voice. Then, we conducted a user study where participants were asked to complete a set of tasks using our system. Our findings identified that (1) searching for a specific voice using hashtags as search keywords was perceived to be easier than similarity and voice attributes, (2) customized voices generated from both voice reflection and voice parameter control functions had high satisfaction compared to other conditions that utilized either one and (3) participants tend to reflect the features of their desired voices when customizing their own voice. As future work, we plan to add more intuitive voice attributes (e.g., accent, intonation, and volume) to enhance user experience with voice customization systems.

## 2 RELATED WORK

### 2.1 Voice Filters

Several companies offer software programs that change the users' voices by applying simple filters. A mobile application called *My Talking Tom*[1] had gained popularity among users by distorting voices for entertainment. While *My Talking Tom* was limited to one option of voice filter, *Voice Changer*[2] expanded options by providing more than 50 voice filters. Filters include entertaining voices such as ghosts, zombies, and aliens. Users can experience these filters on the website by uploading their recorded voices. Meanwhile, Discord users experience real-time voice modification by downloading external applications such as *Voicemod*[3]. It offers more than 50 filters and allows users to communicate using different voices while playing games. While many of these systems successfully intrigued curiosity among users, voice filters are somewhat limited to only a few options. It restrains users from controlling specific values of a voice. Also, as many such systems were developed for entertainment, filtered voices are often exaggerated. Therefore, it is difficult to expect these systems to be used for other purposes, such as in business meetings and phone calls.

### 2.2 Text-to-Speech Voice Synthesizers

The endmost goal of text-to-speech synthesis is to convert plain text into an indistinguishable acoustic signal from human speech. Text-to-speech synthesizers consist of two parts: the front-end that interprets high-level linguistic features of human speech and the back-end that handles functions related to phonetics, acoustics, and signal processing [18]. State-of-the-art models have shown high performance in implementing human-like speeches. Wang *et al.* presented Tacotron and end-to-end text-to-speech model that generates speech from characters. The Tacotron model enhances the sequence-to-sequence (seq2seq) [22] by applying attention paradigm [2] to all decoder steps. As text-to-speech models reach a high performance of naturalness, there have been approaches to diversify synthesized voices by training models with various voice datasets. *Voicemaker*[4] presents over 60 voice personas that speak diverse languages. Users can select a voice persona by language, gender, and age and control the volume to generate text into an audio file. While inspiring, such systems present voice personas trained in advance, neglecting the user's voice identity. As prior work has revealed [5], individuals show concern about voice customization systems modifying users' entire voice identity when communicating with acquaintances. Therefore, text-to-speech synthesizers with several options may not be the cue for customized voices. Our system embraces the user's voice identity by generating customized voices from the user's voice input and target voices that the user desires to sound alike.

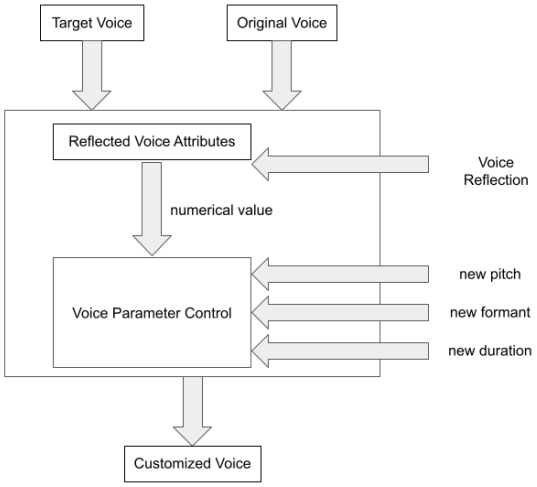


**Figure 1: The system overview of AVOCUS**

## 3 SYSTEM

AVOCUS is a web application for people who wish to customize their voice used in online vocal communication. In this section, we describe the instructions and the implementation of functions. Please refer to our system overview shown in Figure 1.

The voice features defined in AVOCUS are pitch, formant, and duration. Pitch determines the degree of highness or lowness of a tone. Formant refers to the lowest frequency that a voice resonates, which consists of f1, f2, f3, and f4. F1 ranges between 0Hz and 1000Hz, f2 ranges between 1000Hz and 2000Hz, f3 ranges between 2000Hz and 3000Hz, and f4 ranges between 3000Hz and 4000Hz. A shrill voice results in formants with high frequency, and a gravelly voice results in formants with low frequency. Lastly, the duration controls the rate of speech. Rapid speech leads to a high value in duration.

### 3.1 Implementations

As for the implementation, we used Python and Django, a Python-based open-source web platform for our framework, and S3 buckets in Amazon Web Services (AWS) to deploy the website. As shown in Figure 2, AVOCUS consists of two functions: target voice reflection and voice parameter control. To implement the functions, we referred to a Python open-source library Praat-Parselmouth [11]. We chose Praat-Parselmouth, as it offers manipulation of voice by calculating the exact value of voice features.

[1]https://apps.apple.com/us/app/my-talking-tom/id657500465
[2]https://voicechanger.io/
[3]https://www.voicemod.net/discord-voice-changer/
[4]https://voicemaker.in/

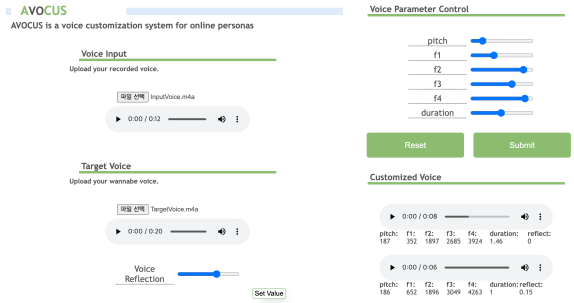


**Figure 2: A screenshot of AVOCUS**

*3.1.1 Target Voice Reflection.* The reflection value determines the ratio between the input voice and target voice. Larger reflection value leads to an output voice that resembles the voice features of the target voice.

In our formula, the new value for the output voice uses the reflection value. Pitch and formant features are used to generate a new voice, setting the new feature value of the output voice derived from the following equation:

$$F_{new} = (F_{source} - F_{input}) \times r + F_{input} \quad (1)$$

In this equation, $F_{new}$ indicates the new feature value, $F_{source}$ indicates the target voice's feature, $F_{input}$ indicates the input voice's feature, and $r$ indicates the reflection value. Feature value refers to the frequency of pitch and each component of formant (f1, f2, f3, and f4).

*3.1.2 Voice Parameter Control.* Once the user uploads the input voice, the system automatically sets sliders that indicates the voice's feature values. Each end of the slider represents each feature's maximum and minimum values. The reset function sets all sliders to return to the original values of an input voice. When submitted without any adjustments on the slider, it gives the guide to set the new feature value as it shows the original feature values of an input voice.

# 4 METHOD

## 4.1 Participants

We recruited 24 participants (16 female, 8 male) via word-of-mouth, the university board, and online communities for college students. Anyone aged between 18 and 65 were allowed to participate in the experiment. Participants' age was 23.88 on average ($SD$ = 6.73; range 18-53).

## 4.2 Apparatus

All participants were interviewed through Zoom. While the researcher was sharing her screen, participants remotely controlled the system to complete the tasks. After a brief explanation on how to use the system at the beginning of each task, participants started to solve the task without any guidance of the researcher. The interview was recorded under participants' consensus.

For the experiment, participants were asked to submit a preliminary questionnaire. The preliminary questionnaire collected participants' voice recording under participants' consensus.

One of the question asked participants to record their voice reading an excerpt from the speech accent archive [26]. Also, when participants mentioned a celebrity with *'good voices'*, researchers prepared the celebrities' voice recordings from interviews or Youtube videos. Lastly, to prevent the bias of task completion time, we used Latin square of order 3 for each task.

## 4.3 Procedure

*4.3.1 Preliminary Questionnaire.* Before the interview, participants were asked to submit a preliminary questionnaire. The survey contained questions about participants' demographic information, prior experience with voice modification systems, desired voices, and the future interest in using voice customisation systems.

*4.3.2 Task 1. Finding the most similar voice.* Participants were asked to search for the closest voice with the answer voice provided on the top of the website. $Cond_{sim.}$ was to find the closest voice among the list of voice database which is sorted by statistical similarity. The similarity was calculated based on the numeric values of voice parameters. The formula to calculate similarity is as follows:

$$s = 100 - |F_{orig} - F_{comp}|/F_{orig} \times 100 \quad (2)$$

In this equation, $s$ refers to similarity, $F_{orig}$ refers to the feature values of original voice, and $F_{comp}$ refers to the feature values of the voice in comparison, where feature values mean pitch and formants (f1, f2, f3, and f4). $Cond_{att.}$ was to find the closest voice using voice attributes as search keywords, derived from phonetics papers [5]. The attributes consists of breathiness, hoarseness, pitch, smoothness, speed, variation, and volume. $Cond_{hash.}$ asked participants to find the closest voice using hashtag keywords. Hashtags were manually annotated by two researchers. The answer voice was annotated as slow paced, low-pitched, monotonous, clear, and good pronunciation.

*4.3.3 Task 2. Customizing a voice to match the target voice.* The participants were asked to customize a voice to match the target voice. $Cond_{ref.}$ was to use a voice reflection. The original voice is the input voice, and the participants were to utilize the target voice, making the original voice to match the target voice. $Cond_{par.}$ was to manipulate a numerical value of voice parameters of the original voice to match the target voice. $Cond_{ref.\&par.}$ was to utilize both voice reflection and voice parameter control.

*4.3.4 Task 3. Customizing one's own voice.* Task 3 asked participants to customize their own voice using voice recordings of desired voices. For Task 3, participants were able to choose the most preferred prototype from Task 2. Using their favorite prototype, participants freely customized their voice that they wish to use for online communication. Desired voice recordings were mostly renowned celebrities' voices, and they were provided in advance for reference.

*4.3.5 Post Survey.* After completing each task, we conducted a semi-structured interview using a post-survey questionnaire asking about the overall experiences about the prototype. The survey asked

participants to answer satisfaction of result voice, satisfaction of system, and the difficulty of the task, in a 5-point Likert score.

# 5 FINDINGS

## 5.1 Interview: Prior Experience and Preference

*5.1.1 Most tried voice modification for fun.* Although we did not specify to recruit individuals who are familiar with voice customization, the majority of the participants (83.3%; 20 out of 24) answered that they had tried voice modification systems. Participants were familiar with Voice Changer for Discord, *My Talking Tom*, and voice filters for voice call through *KakaoTalk*[5]. Then, we asked what intrigued them to have experience with voice modification systems: all of the participants answered that they had tried it for entertainment. To understand the strengths and weaknesses of current systems, we asked about participants' experiences using such systems. All of the participants who had prior experience ($N$ = 20) found voice modification systems useful for amusement. However, participants found inconvenience in using such systems for practical use ($N$ = 13), found the modified voice exaggerated ($N$ = 9), awkward ($N$ = 4), and limited to only a few options ($N$ = 1).

*5.1.2 Most preferred low-pitched, calm and comforting voice with good context delivery.* In our next section of the survey, we asked the participants to mention celebrities who are considered to have *'good voices'*. Participants mentioned renowned singers, movie stars, announcers, and voice actors. The majority of the participants ($N$ = 21) mentioned celebrities of the same gender.

During our in-depth interview, we asked our participants to describe the celebrity's voice they mentioned in the preliminary questionnaire as *'good voices'* using three keywords. The following keywords were popular among participants: good content delivery ($N$ = 9), calm and comforting ($N$ = 8), low-pitched ($N$ = 7). While 29.2% of the participants described their desired voices as low-pitched (N = 7), only one participant described his/her wished voice as high-pitched.

## 5.2 Task 1. Finding the most similar voice

Participants were asked to find the closest voice that matches the answer among the voice database described in terms of: statistical similarity ($Cond_{sim.}$) vs. phonetic voice attributes ($Cond_{att.}$) vs. hashtags ($Cond_{hash.}$)

Participants chose the most similar voice under $Cond_{hash.}$ ($m$ = 96.213%), and the least similar voice under $Cond_{sim.}$ ($m$ = 93.170%). Please refer to Figure 3 to compare similarity among three conditions.

To find out whether each condition was easy to search for the most similar voice, we asked our participants to rate the difficulty of Task 1 under three conditions. Participants found the task less demanding under $Cond_{hash.}$ ($m$ = 2.92, $SD$ = 1.14).

Participants commented that they preferred $Cond_{hash.}$ over $Cond_{att.}$ when searching voices because they were unfamiliar with the terms used in $Cond_{att.}$. They found the task most difficult under $Cond_{sim.}$, due to low confidence in the calculation of similarity.

[5]www.kakao.com

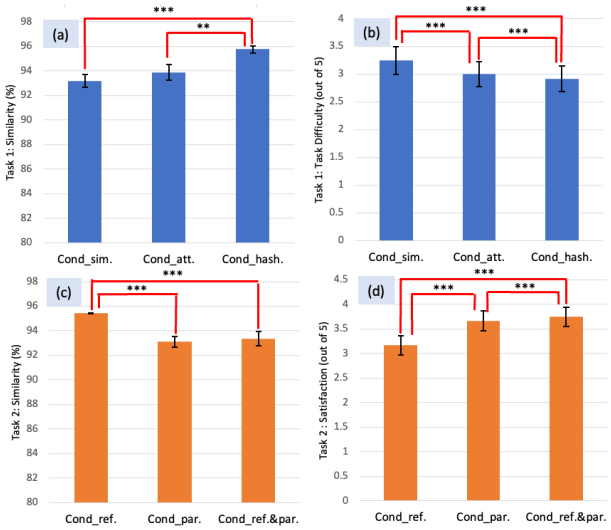


**Figure 3: (a) Average similarity for Task 1 (b) average task difficulty for Task 1 (c) average similarity for Task 2 (d) average voice satisfaction for Task 2. Error bars indicate standard errors (N=24). '*', '**', '***' indicate p < .05, p < .01, and p < .001 respectively.**

## 5.3 Task 2. Customizing a voice to match the target voice

For Task 2, participants customized a voice to match the target voice using three interfaces: voice reflection ($Cond_{ref.}$) vs. numerical value of voice parameters ($Cond_{par.}$) vs. voice reflection & numerical value of voice parameters (AVOCUS, $Cond_{ref.\&par.}$)

Participants customized the closest voice to the answer when they were under $Cond_{ref.}$ ($m$ = 95.425%, $SD$ = 0.172). $Cond_{par.}$ and $Cond_{ref.\&par.}$ resulted an average of 93.116% ($SD$ = 2.206), 93.382% ($SD$ = 2.803) of similarity.

Participants showed the highest satisfaction ($m$ = 3.75, $SD$ = 0.94) for our system, $Cond_{ref.\&par.}$, where it allows the users to customize voices using both of the reflection functions and the voice parameter control. Participants commented that they were satisfied with the wide range of options for voice customization, leading to a high degree of freedom using the system. On the contrary, participants found $Cond_{ref.}$ difficult to familiarize themselves with the reflection function, as the effects seemed insignificant when the target voice was the same gender as the original voice. However, participants suggested that the voice parameter control redeems such shortages by finely manipulating the voice attributes.

## 5.4 Task 3. Customizing one's own voice

For Task 3, participants were asked to choose the most preferred interface from Task 2, and customize their own voice using their desired voices. The majority of participants chose $Cond_{ref.\&par.}$ ($N$ = 14). Seven participants chose $Cond_{par.}$ and three participants chose $Cond_{ref.}$ for Task 3.

After performing Task 3, we asked participants to rate the satisfaction of customized voices and how customized voices differ from participant's original voices. Participants who chose our system,

$Cond_{ref.\&par.}$, showed the highest satisfaction level on customized voices ($m$ = 4.357, $SD$ = 0.841). Participants answered that they used the system to make their voices with better content delivery ($N$ = 10) and calming voice ($N$ = 5). Some participants tried to mimic their desired voices' distinct characteristics ($N$ = 3). We found out that participants reflected the desired attributes of good voices when customizing their own voices.

## 6 DISCUSSION

Here we summarize findings from the study, discuss design recommendations for voice customization systems, and how it can be made in use for practical purposes.

### 6.1 Preferred Voices

Prior research proves that human voice pitch are perceived to deliver attractiveness, strength, and social dominance [15, 24]. Also, researchers determined that low-pitched (*i.e.,* masculine) voices are generally preferred by both male and female participants and are often considered to have suitable voices for leadership [1]. Our findings analyzed the features of preferred voices, confirming the findings of prior research about individuals' desire for low-pitched voices. The majority of participants wished to have a low-pitched, calm and comforting voice with good context delivery. We concluded that participants desired to reflect the attributes of desired voices. Thus, we recommend voice customization tools to provide guidelines of how to customize a preferable voice.

### 6.2 Describing Voices

From our findings, we concluded that participants preferred contextual information over numeric information when describing voice. Hence, numeric information (i.e., similarity) was easily neglected while completing the task and gained low reliance among participants.

Participants found hashtags the most intuitive and understandable and wished to gain descriptions of voices used in a common language. Some participants commented that they were confused about voice attributes from phonetics studies, as some attributes contain complex terminologies. We highly recommend using hashtags that are verified by the audience when explaining voices.

### 6.3 Functional Recommendations for Voice Customization Systems

Contrary to concerns that too many features in voice customization systems would be tiresome, participants wished the system to contain more specified voice attributes. AVOCUS consists of seven sliders, which control pitch, formants (f1, f2, f3, f4), and duration. We plan to implement additional features to customize voices. For instance, participants wished to change the accent, intonation, and volume of their voices. We expect specified functions would lead to higher satisfaction from end-users.

### 6.4 Limitations and Future Work

Our system has several limitations to be improved in future work. First, we plan to track down additional voice parameters that can be quantified. During the in-depth interview, our participants commented that formants were difficult to understand for users without any knowledge about phonetics. Second, voice customization systems should offer customization guidance for first-time users. For example, if a participant wishes to transform his voice to a calm and comforting voice, the system may recommend the average numeric values of calm voices. In this way, first-time users can easily follow the guidance and refer to such standards. Last but not least, we plan to implement our system to be experienced in various platforms. Until now, voice modification functions were mostly experienced in entertaining platforms. We believe that voice customization systems can be broadly used in platforms for work, everyday communication, and multimedia content creation.

## 7 CONCLUSION

Motivated by prior research that discovered the demand for customized voices for specific contexts, we designed a voice customization system called AVOCUS. Here, we seek to put preliminary foundations into practice and determine user behaviors toward voice customization systems by conducting a user study with 24 participants. Our findings confirmed our system's usability and end-user behavior toward voice customization. As future work, we plan to add more intuitive voice attributes (e.g., accent, intonation, and volume) to enhance user experience with voice customization systems.

## ACKNOWLEDGMENTS

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2023-2020-0-01460) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation).

## REFERENCES

[1] Rindy C Anderson and Casey A Klofstad. 2012. Preference for leaders with masculine voices holds in the case of feminine leadership roles. *PloS one* 7, 12 (2012), e51216.
[2] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473* (2014).
[3] 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 (2009), 277–283.
[4] Axel Berndt and Knut Hartmann. 2008. The functions of music in interactive media. In *Joint International Conference on Interactive Digital Storytelling*. Springer, 126–131.
[5] Hyeon Jeong Byeon, Chaerin Lee, Jeemin Lee, and Uran Oh. 2022. "A Voice that Suits the Situation": Understanding the Needs and Challenges for Supporting End-User Voice Customization. In *CHI Conference on Human Factors in Computing Systems*. 1–10.
[6] Inger Ekman. 2008. Psychologically motivated techniques for emotional sound in computer games. *Proc. AudioMostly* (2008), 20–26.
[7] Inger Ekman. 2013. On the desire to not kill your players: Rethinking sound in pervasive and mixed reality games.. In *FDG*. 142–149.
[8] Jerry Bryan Fuller, Tim Barnett, Kim Hester, Clint Relyea, and Len Frey. 2007. An exploratory examination of voice behavior from an impression management perspective. *Journal of Managerial Issues* (2007), 134–151.
[9] 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.
[10] Rosalie Hooi and Hichang Cho. 2014. Avatar-driven self-disclosure: The virtual me is the actual me. *Computers in Human Behavior* 39 (2014), 20–28.
[11] Yannick Jadoul, Bill Thompson, and Bart De Boer. 2018. Introducing parselmouth: A python interface to praat. *Journal of Phonetics* 71 (2018), 1–15.

[12] Colby Johanson and Regan L Mandryk. 2016. Scaffolding player location awareness through audio cues in first-person shooters. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 3450–3461.

[13] 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*. 1–27.

[14] 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.

[15] Casey A Klofstad, Rindy C Anderson, and Susan Peters. 2012. Sounds like a winner: voice pitch influences perception of leadership capacity in both men and women. *Proceedings of the Royal Society B: Biological Sciences* 279, 1738 (2012), 2698–2704.

[16] Pontus Larsson, Aleksander Väljamäe, Daniel Västfjäll, Ana Tajadura-Jiménez, and Mendel Kleiner. 2010. Auditory-induced presence in mixed reality environments and related technology. In *The engineering of mixed reality systems*. Springer, 143–163.

[17] 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 (2009), 348–370.

[18] Marian Macchi. 1998. Issues in text-to-speech synthesis. In *Proceedings. IEEE International Joint Symposia on Intelligence and Systems (Cat. No. 98EX174)*. IEEE, 318–325.

[19] Lennart E Nacke and Mark Grimshaw. 2011. Player-game interaction through affective sound. In *Game sound technology and player interaction: Concepts and developments*. IGI global, 264–285.

[20] Timothy Sanders and Paul Cairns. 2010. Time perception, immersion and music in videogames. (2010).

[21] Aatto Sonninen and Pertti Hurme. 1992. On the terminology of voice research. *Journal of Voice* 6, 2 (1992), 188–193.

[22] Ilya Sutskever, Oriol Vinyals, and Quoc V Le. 2014. Sequence to sequence learning with neural networks. *Advances in neural information processing systems* 27 (2014).

[23] Stefano Taddei and Bastianina Contena. 2013. Privacy, trust and control: Which relationships with online self-disclosure? *Computers in human behavior* 29, 3 (2013), 821–826.

[24] Cara C Tigue, Diana J Borak, Jillian JM O'Connor, Charles Schandl, and David R Feinberg. 2012. Voice pitch influences voting behavior. *Evolution and Human Behavior* 33, 3 (2012), 210–216.

[25] Christophe Veaux, Junichi Yamagishi, and Simon King. 2012. Using HMM-based speech synthesis to reconstruct the voice of individuals with degenerative speech disorders. In *Thirteenth Annual Conference of the International Speech Communication Association*.

[26] Steven Weinberger. 2015. Speech accent archive. george mason university. *Online:< http://accent. gmu. edu* (2015).

[27] Hanna Elina Wirman and Rhys Jones. 2017. Voice and Sound: Player Contributions to Speech. In *Avatar Assembled: The Social and Technical Anatomy of Digital Bodies*. Polity Press.

[28] Kuang-Wen Wu, Shaio Yan Huang, David C Yen, and Irina Popova. 2012. The effect of online privacy policy on consumer privacy concern and trust. *Computers in human behavior* 28, 3 (2012), 889–897.

[29] Lotus Zhang, Lucy Jiang, Nicole Washington, Augustina Ao Liu, Jingyao Shao, Adam Fourney, 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 (2021), 1–21.