THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF MARKETING

Voice Assistant Realism: How Perceived Realism of a Device's Voice Affects Consumer Satisfaction and Affinity

OLIVIA HIRT
SPRING 2024

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in Marketing
with honors in Marketing

Reviewed and approved* by the following:

Brett Christenson
Assistant Clinical Professor of Marketing
Thesis Supervisor

Jennifer Coupland
Clinical Professor of Marketing
Honors Adviser

* Electronic approvals are on file.

ABSTRACT

Voice assistants are becoming increasingly prominent in our world today, providing firms with an opportunity to capitalize on optimizing voice assistant design in order to better satisfy consumers, thus promoting success of the product and the firm. This paper is intended to provide insight into how firms can adapt the voice of the assistant by utilizing the concept of anthropomorphism, or the attribution of human characteristics to non-human agents (Epley, Waytz, & Cacioppo, 2007), to create positive connect with consumers. In my research, I measure two dependent variables: 1) consumer satisfaction with the assistant's recommendation and 2), consumer affinity for the device itself. I perform two studies, in which I manipulate the level of human likeness or "realism" of the voice to create a range from robotic to human, then measure participants' responses. Study 1 looks at how the varying levels of voice realism impacts consumer recommendation satisfaction and affinity. Study 2 considers whether the context of the task the assistant is performing is an influencing factor, measuring whether matching the level of voice realism to the seriousness of the task context has an influence on consumer response. The results suggest that consumers prefer humanlike or "realistic" voice assistant voice, with respondents' recommendation satisfaction and affinity (specifically in terms of their perception of friendliness of the device) generally increasing as realism levels increased. They also indicate that task context does not impact this effect, showing a preference for realism no matter the task the assistant was performing. These results inform my ultimate recommendation for firms to design their voice assistants' voices to be as realistic as possible to maximize positive consumer response overall.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank Dr. Brett Christenson for all the time and effort he put into guiding me through this process from beginning to end. Dr. C took me on over two years ago when I was writing my proposal to get into Schreyer Honors College and has seen me through until now, as I turn in my senior thesis. Having no prior experience with conducting research or thesis writing, I greatly appreciate his patience and kindness as he helped me take on this project. His support and encouragement not only made the process easier and less intimidating, but also fun and rewarding. I'm honored to have been his first thesis student.

I also want to thank Dr. Jennifer Coupland for her continuous support throughout my time in Schreyer. Dr. Coupland is a wealth of knowledge and resources, and her warm and encouraging attitude always gives me inspiration to be the best I can be. I often tell my marketing friends to take her classes so they can experience her magic.

I also want to thank Dr. Meg Meloy for helping me jumpstart my thesis. Her consumer behavior course inspired my idea for my thesis topic, and she helped guide me through the core beginnings of my project. I'm very grateful for her wisdom and support.

Finally, I want to thank my family for their love and support throughout my college journey. I'm inspired by and grateful for them every day.

Chapter 1

Introduction

"Alexa, how old is Elton John?" "Alexa, what's the Liverpool score?" "Alexa, give me a recipe for banana bread." "Alexa, what's the Barbie movie about?" "Alexa, sing Happy Birthday." "Alexa, thank you." These are just a few of the most-asked questions, requests, and utterances of the millions that UK customers posed to Amazon's Alexa voice assistant in 2023 ("Alexa's Most Asked Questions," 2023). Questions vary from the meaningful, the mundane, the hilarious, and the heartfelt.

Voice assistants such as the Amazon Alexa, Google Home, and Apple's Siri have become widely used by consumers to answer questions, provide recommendations, and perform tasks in an effort to make their everyday lives easier. These devices meet consumer demand for reliable, convenient content that is delivered in real time, relevant, and personalized to cater to one's own individual wants and needs (Brill, Munoz, & Miller, 2019). Predictions show that there will be approximately 8.4 billion voice assistants in use worldwide by the end of 2024, the total doubling from that of 2020 (Shewale, 2024). Given this statistic, voice assistants are clearly becoming widely used in society. Brands such as Apple, Amazon, and Google can use these devices to get consumers to buy into their digital ecosystems, utilizing them as strategic tools for gaining market share. Therefore, it is wise that firms pay attention to this influential product and perfect its design to capitalize on an opportunity to profit and better satisfy customers.

One defining aspect of most voice assistants is the way they tend to talk and act like real human beings rather than machines, a component strategically added to the devices' design by firms. This concept is called “anthropomorphism,” or the attribution of human characteristics and emotions to non-human agents (Epley, Waytz, & Cacioppo, 2007). Whether consumers view products as human depends on the presence of features that signify a sense of humanity, such as overall shape, facial features, sounds and voices, and non-physical components like intentionality, imitation, and communication ability (Aggarwal & McGill, 2007). Thus, voice, as one of these key features, plays a significant role in how consumers interact with anthropomorphized products like voice assistants (Niculescu et al., 2013). In fact, the voice, or linguistics, can influence consumer thoughts, feelings, and behavior toward the anthropomorphized agents (Kim, Chen, & Zhang, 2016).

The growing presence of voice assistants demonstrates that they are becoming a prominent and successful product in our modern world. As more consumers embrace them, it is increasingly important to know how to effectively utilize sensory elements such as voice when it comes to voice assistant design. As explained earlier, a better design could generate higher consumer satisfaction and affinity toward the product, and subsequently more adoption of the recommendations of the voice assistant, higher sales, more brand equity, etc. Thus, to approach this opportunity, firms have started to adapt their assistants’ voices to make them more appealing to consumers. For example, Apple has added a variety of accents and genders to Siri’s voice options to promote greater diversity and inclusion, asking users which voice they prefer during set-up instead of defaulting to the female one (“Apple adds two Siri voices,” 2022), and Amazon allows customers to change Alexa’s voice to a chosen celebrity or character voice such as those of Shaquille O’Neal, Samuel L. Jackson, or Deadpool (Wells, 2022). However, because these devices are still relatively new, firms still do not fully understand how to manipulate voice properties to maximize consumer satisfaction of their recommendations and affinity toward the device. For instance, there is limited research on voice assistants regarding which properties of the voice will lead to positive responses from the consumer and to what degree those properties should be adjusted. Again, these unexplored elements are important for firms to consider, as they can contribute to the success of voice assistants and, consequently, the success of the firms. Thus, these gaps in research leave room for students like me to explore this topic. In this paper, I will study consumer responses to a voice assistant device's voice by measuring two dependent variables: recommendation satisfaction (respondent satisfaction with the recommendation the assistant makes) and affinity (respondent affinity toward the device itself). With current technology, there are ways for researchers to manipulate the individual properties of speech to closely investigate the intricacies of the voice. Indeed, one under-researched area of voice that I will focus on for my project is "timbre" (Christenson et al., 2020), a property which I expand upon in the following Theoretical Framework section. The goal of my research is to provide insight on how firms can refine the design of voice assistants to maximize their value and better meet consumer preferences. Specifically, I will focus on the degree to which anthropomorphism, also known as the level of "realism," of the assistant's voice affects the perceived consumer value of voice assistants. I will measure the perceived value by analyzing the dependent variables of consumer recommendation satisfaction and affinity for the device.

Chapter 2

Theoretical Framework

Marketing is a critical aspect of the business process that focuses on identifying and catering to consumer wants and needs to ultimately create value for customers (“Definitions of Marketing,” n.d.). Of the numerous approaches to marketing, one growing method I have found intriguing is sensory marketing, or “marketing that engages the consumer’s senses and affects their behaviors” (Krishna, 2012). It involves the utilization of the five senses: haptics, smell, taste, vision, and audition. Haptics, or touch, can involve humans touching products, humans touching humans, and products touching products (Krishna, 2012). Smell is connected to memory and emotion (Cahill et al., 1995), and some scents can affect brand and product recall and recognition (Morrin & Ratneshwar, 2003). Taste perception depends on the presence of other senses (Krishna, 2012) and external influences regarding a product (Hoch & Ha, 1986). Vision is the dominant sense, and there is significant research on product spatial configuration (Meyers-Levy & Zhu, 2007) and aesthetic aspects (Hagtvedt & Patrick, 2008). Research on audition, or sound, the sense I’m focusing on in my studies, has mostly focused on influences on product and advertisement evaluation in radio and television advertising and the perceived attractiveness of jingles and songs. This area of research covers how ambient music in a store, signature product sounds, or even reading or hearing a product or brand name can influence consumers’ attitudes (Krishna, 2012). To relate current audition literature to my own studies, the way a consumer interprets stimuli such as music or voice can affect how they think or feel about a product or brand (Meyers-Levy, 2010; Dahl, 2011), which constitutes the foundation of my experiments. In particular, I assess how respondents interpret the stimulus of a voice assistant’s voice, as measured by the dependent variables of their satisfaction with the assistant's recommendation and their affinity toward the device itself.

2.1 The voice

The voice will always be relevant to consumers, as sensitivity to voice and language cues has been important in human social groups throughout evolution. When humans communicate face-to-face, the voice provides cues that inform the listener about the speaker's gender, age, personality, emotional state, or place of origin. Socially intelligent individuals then use this information to decide who to like and trust (Niculescu et al., 2013). Thus, according to this explanation, in a consumer interaction with a voice assistant, the assistant's voice provides cues that consumers take in and use to decide how they feel about the assistant. They can determine whether they like, trust, and respect the device and its recommendations. So, it is wise for firms to assess their assistants' voices to understand how consumers are likely to feel about the product. Indeed, given that how a consumer "feels" about the voice assistant is a very general concept and can be intercepted in many ways, to assess this phenomenon, it is important to define specific and measurable variables. This brings me to the two dependent variables I measure in my studies: recommendation satisfaction and affinity toward the device.

2.2 Recommendation satisfaction and affinity

While there are many valuable potential variables to measure in this situation, I narrowed my scope to two dependent variables that I believe encompass vital traits and functions for a voice assistant to possess and provide to maximize its value and best meet consumer preferences.

The first dependent variable is recommendation satisfaction. This term represents the consumer's overall satisfaction with the voice assistant's recommendations (Yang 2021). It is important to measure due to the fact that voice assistants function through the method of voice command. The fact that voice assistants provide information vocally creates added complication. Humans have more difficulty retaining vocal information in their short-term memory because voice information subsists on a single temporal dimension (Bjork, 1970). Thus, when a consumer asks an assistant for a recommendation, if the first one the assistant provides doesn't satisfy them, they are increasingly less likely to ask for a second, third, or fourth one due to their inability to remember and compare all the different options. So, it is common for people to simply accept or reject the first recommendation the assistant provides (Klaus & Zaichkowsky, 2020). Because of this limitation, it is important that the information provided is high quality, so consumers are satisfied from the beginning of their usage experience and do not give up on the assistant's functionality. Therefore, it is necessary to measure recommendation satisfaction to ensure that this key function is properly addressed.

The second dependent variable is affinity. This term represents the consumer's natural liking or empathy toward the voice assistant itself (Mori, et al., 2012). I believe this is another important variable to measure because it represents the anthropomorphism factor that transforms the assistant into more than just a mindless, task-performing machine. This feature is one of the main draws of the voice assistant, as its ability to convey human characteristics and emotions allows consumers to accept and enjoy the assistant as a coequal social partner (Becker, et al., 2007). Today, research shows that people's views toward their relationship with artificial, autonomous humanoid agents has shifted from a simple "user-tool" relationship to more of a partnership (Negrotti, 2005). Thus, because these agents can interact more socially with humans, a major goal should be to integrate emotions and personality into their design (Becker, et al., 2007). Therefore, affinity toward the device is important to measure because it addresses the assistant's more meaningful role for consumers.

Depending on consumer responses to the device's voice as measured by recommendation satisfaction and affinity, firms should adapt the voice accordingly. This calls forth the question of *how* exactly firms can and should adapt it, discussed next.

2.3 Timbre

The voice can be broken up into four properties: volume, pitch, speech rate, and timbre. Volume is how loud or soft a sound is, tempo is the speed of the sound, pitch is how high or low the sound is, and timbre is vocal quality (Christenson et al., 2020). Adjusting any of the four properties can affect consumer response to a voice. For my experiment, I decided to manipulate timbre, a texture-related component also known as voice quality or harmonics (Bruner, 1990). Timbre is difficult to define and probably the least researched property of voice because it is a complex construct characterized by many acoustic parameters. Scientifically, timbre perception is based on the physical characteristics of sound waves (Bolden, 2009). However, a more relatable explanation for the property of timbre is that it is based on texture (Bruner, 1990), in other words, the roughness or smoothness of a sound. Another reason timbre is still quite uninvestigated is because it is difficult to define clear measurement parameters with which to study it. Properties like volume and speech rate are easily quantifiable by using a simple numerical scale, but there is no easy way to quantify abstract characteristics of timbre such as "scratchiness" (Oakes & North, 2006). Yet, despite this ambiguity, current technology and proper study design has allowed researchers to begin scratching the surface of timbre manipulation.

For my study, one aspect of timbre that guides my project is that, when manipulated, it can presumably affect how consumers perceive the degree of realism of a voice. Based on its definition, as a representation of texture (Bruner, 1990), timbre can seemingly be assessed on a scale of rough to smooth. Additionally, research equates timbre to masculinity, suggesting that people with different vocal timbres can indicate different physicalities; for instance, certain levels of timbre tend to be associated with a physically larger body size, and such perceptions of physicality lead to inferences that the voice belongs to a male (Efthymiou et al., 2023). So, research suggests a connection between roughness and smoothness.

Thus, I believe that this dimension of roughness versus smoothness can be applied to the dimension of realism, as more robotic voices sound rougher in comparison to smoother humanlike voices. Therefore, if audio engineering involving a manipulation of the sound wave can create variations of a voice to sound more rough or smooth, then it can be used to depict different degrees of realism. Rougher voices would represent lower degrees of realism (robotic) and smoother voices would represent higher degrees of realism (humanlike). Again, there is little research that directly addresses the effect of timbre on perceived realism, or how consumers respond to different levels of timbre, and my thought process here is based on logic from the limited information that currently exists. Thus, it is the goal of my experiment to provide some more insight on these topics. In order to inform my hypotheses, I turned to the theoretical frameworks provided by literature on anthropomorphized technology to help explain my proposed effects.

2.4 Technology and Anthropomorphism

As touched on earlier, many firms strategically design their products with an anthropomorphism component in mind. This happens frequently with products that act as a sort of helper or assistant, such as a chatbot on a website, a digital helper in a video game, or the product I am studying: a voice assistant like Amazon Alexa or Google Home. For these types of products, “humanness” is often achieved with visual or linguistic representations of anthropomorphism (Kim, Chen, and Zhang, 2016).

In general, even without strategic anthropomorphic product design, humans tend to attribute familiar humanlike intentions and emotions to non-human objects because it can make the object feel more familiar, explainable, or predictable (Fink 2012). So, when designing socially interactive robots, firms use anthropomorphism to increase acceptance of the product and facilitate interaction (Fink 2012). This research is applicable to my own project, as recommendation satisfaction and affinity, the two variables I am measuring, can effectively lead to product acceptance and facilitate interaction with the assistant.

Additionally, besides the fact that humans have a natural preference for anthropomorphism, another important point to note is that these devices are meant to mimic the role of a real-life assistant. Thus, if the voice assistant has more humanlike features, a consumer-assistant interaction will be more characteristic of a natural human interaction and likely resonate more with the consumer. Thus, firms should incorporate certain elements of personality that real assistants often possess, such as the ability to effectively communicate and display a friendly and accommodating attitude toward customers (Niculescu et al., 2013). Of course, the ability for the assistant to successfully perform concrete duties like providing recommendations and completing tasks is essential and foremost. However, once those abilities are implemented, I believe that adding an element of smooth timbre and thus, realism, to the assistant’s repertoire would elevate its design and better satisfy consumers.

However, in order for firms to apply this knowledge to the design of their products, it is important to determine the degree to which they should anthropomorphize. While literature suggests firms should humanize their products greatly to appeal to consumers' proposed preferences, the specific degree of humanization recommended is yet to be determined. Therefore, I hope to shed some light on this gap in knowledge through my own research. It should be noted that my project does not address the physical appearance of the robot, and as discussed earlier, visual features can play a role in perceived anthropomorphism. However, since digital assistants are not currently designed to look like human beings but rather speak like human beings, I focus only on the voice within my research.

To summarize, current literature suggests that consumers will respond more positively to a more realistic, or humanlike, voice assistant voice. However, research has not yet determined exactly how humanlike the voice should be. Should it be as human as possible, almost indistinguishable from a real human voice? Or will that make consumers feel uneasy? Perhaps there is some sort of middle ground where the voice is sufficiently humanlike, but not too much so? To answer these questions, I plan to investigate how respondents react to an increasingly humanlike voice assistant voice. In terms of marketing implications, this research is intended to gain insight on the role of anthropomorphism in products, especially those driven by AI. As the presence of these devices continues to grow in the technology industry, this could help firms maximize their design to meet customer preferences.

To evaluate this effect, with the help of a professional audio technician, I manipulated the timbre of a voice to create different voices that depict varying degrees of realism from unrealistic (robotic) to realistic (completely human). Then, I studied how participants responded to the different voices, specifically measuring their satisfaction with the assistant's recommendation and their affinity toward the assistant. The resulting data would help determine where within the scale from unrealistic to realistic a consumer's highest level of recommendation satisfaction and affinity occurs. Based on the literature, it seems probable that participants would respond increasingly positively to the more realistic voices, with the most positive responses occurring at Voice D, which is completely human. Accordingly, my first two hypotheses are as follows:

H_1: When voice realism increases from unrealistic to realistic, recommendation satisfaction will also increase.

H_2: When voice realism increases from unrealistic to realistic, affinity for the voice assistant will also increase.

2.5 Matching theory

Another relevant theory that adds more complexity to my research is called the matching theory. The matching theory suggests that the appearance and social behavior of a robot should match the seriousness of a task or situation (Goetz, et al., 2002). For example, if a service robot is delivering bedside medications to patients in a hospital, a cheerier and more enthusiastic robot might work well for a patient in less serious condition. However, for more grave situations, a more authoritative demeanor would be more appropriate. Hence, according to the matching theory, the robot should match its behavior to the user's situation for the best response. In one study, a robot that was more humanlike, attractive, or playful in multiple contextual situations was not perceived positively by participants across the board. Instead, the robot that matched how participants expected it to look and act given the task context was better perceived, as it increased their sense of the robot's compatibility with its job and their compliance with the robot (Goetz et al., 2002).

My first two hypotheses, which are indeed based on extensive research, predict a steady increase in satisfaction with the assistant's recommendations and affinity for the device as realism increases. However, the matching theory suggests that the effect is not that simple, and that the task the assistant is performing also plays a role. Thus, adding in a task context factor in addition to the realism factor qualifies my experiment. According to the matching theory, if the voice assistant's voice appropriately aligns with the context of the task the assistant is performing, the respondents will respond more positively (their recommendation satisfaction and affinity will increase). I decided to measure this effect by manipulating the task context, testing a playful task condition and a serious task condition (more specifics on this will be detailed in the Study 2 chapter). My prediction, according to the matching theory, is that the respondent would prefer a more humanlike voice for the playful condition and a more robotic voice for the serious task condition. Therefore, my second two hypotheses are as follows:

H_3: When the realistic (human) voice is paired with the playful task condition, recommendation satisfaction will increase; when the less realistic (robotic) voice is paired with the serious task condition, recommendation satisfaction will increase.

H_4: When the realistic (human) voice is paired with the playful task condition, affinity for the assistant will increase; when the less realistic (robotic) voice is paired with the serious task condition, affinity for the assistant will increase.

Chapter 3

Overview of Studies

I test my prediction in a series of two studies. Study 1 establishes the impact of voice realism alone on consumer recommendation satisfaction and affinity for the device (H_1 and H_2). Following this, Study 2 manipulates the context of the task the assistant performs along with the voice realism to measure the interaction of the two factors (H_3 and H_4). Further details on research method and design and results and analysis for each study are discussed in the following designated chapters.

Chapter 4

Study 1

4.1 Research Method and Design

To test H_1 and H_2, I conducted a Qualtrics questionnaire that involved participants listening to each of the voices of varying realism and measuring their satisfaction of the voice's recommendation and affinity toward the voice.

All stimuli were created with the help of a professionally trained audio technician, manipulating only timbre and keeping volume, pitch, and speech rate constant. The voice creation involved recording a human male and then manipulating the sound file using digital audio manipulation software to create recordings for each level of timbre. The four voices ranged from robotic to human. Voice A was 60% robot and 40% human, Voice B was 40% robot and 60% human, Voice C was 20% robot and 80% human, and Voice D was 100% human. These distributions were chosen based on the rationale that current voice assistants do not use completely robotic voices; in fact, they all sound somewhat lifelike, so the experiment was focused on voices that were at least 40% human. See Appendix A for the sound stimuli files.

254 students at Penn State University participated in the questionnaire for partial class credit. They were each randomly assigned one of the voices to listen to, so roughly 63 to 64 students heard each voice. See Section 4.2 for a description of the sample demographics. The questionnaire began with a prompt telling participants that they were going to be listening to a recording of a new voice for a digital assistant. They were instructed to imagine that they have used the assistant to search for a local restaurant to eat at, and that they are going to hear the assistant's response. They were also instructed to pay attention to the voice and its information. On the next page, the participants were randomly assigned to listen to one audio file (out of the four possible files). Participants were instructed to play the audio file and listen carefully. The audio file provided the name and description of the restaurant, including details about the atmosphere of the restaurant, type of food served, dine-in and carry-out options, and location, then asked if the user would be interested in going. Participants could listen to the recording as many times as they wanted, but once they clicked to the next page, they could not go back to relisten. After interaction with the voice assistant, participants responded to measures for the variables of interest, manipulation checks, and demographic information. Twenty-nine participants either responded with half or less of the manipulation check questions incorrectly or did not answer them at all, indicating they may have skipped or not paid attention to the file content, and thus were removed, leaving 225 participants. See Appendix B for the audio file script and Appendix C for the full questionnaire.

The dependent variables of recommendation satisfaction and affinity toward the assistant were measured using 7-point Likert scales. The first main portion of the questionnaire was designed to measure satisfaction of the recommendation. Participants rated the recommendation provided by the assistant on many dimensions, including whether needs and expectations were met, whether the amount of information was appropriate, the effectiveness and efficiency of the response, the potential for repeated use, etc. The second main portion was designed to measure affinity toward the assistant. Again, participants rated the voice on many dimensions, including likability, speed, friendliness, roughness, authenticity, trustworthiness, competence, etc. Participants then completed manipulation check questions testing the recall of information from the audio file. Finally, they filled out demographic information regarding their personal voice assistant usage habits, attitude about usefulness of digital assistants, preference for the realism level of digital assistants, age, gender, political affiliation, race, ethnicity, and socioeconomic status.

4.2 Results and Analysis

In terms of demographics, the experiment utilized a sample of 253 Penn State students, where they were asked to take an online questionnaire in return for partial course credit. The sample ranged from 18 to over 22 years old, with diversity of demographic information and varying habits and attitudes regarding voice assistants. See Appendix C for the questionnaire.

With respect to the recommendations section of the questionnaire, after a factor analysis of the 10 recommendation scale questions from a Varimax rotation, two factors emerged explaining 65.71% of the variance of rating responses on the basis of Eigenvalues greater than 1.

I called these factors “RecSat” (“Recommendation Satisfaction”) and “KeepSearch” (“Keep Searching”) and performed two one-way ANOVA tests to find statistical differences between groups for the factors. “RecSat” characterized all the questions except the last one, Question 10 (Rest_10: “I will gather additional information about the recommended restaurant.”), based on loadings greater than .635. “KeepSearch” characterized Question 10, which was loaded independently.

As determined by one-way ANOVA, the RecSat factor showed statistically significant differences between groups of the varying conditions ($F(3,247) = 3.668$, $p = .013$), showing a meaningful relationship between satisfaction of the recommendation and realism of the voice. The KeepSearch factor, however, showed no significant difference between groups ($F(3,250) = 1.729$, $p = .162$), so there is no meaningful relationship between the willingness to keep searching for information about the recommended restaurant and realism of the voice, thus I did not complete further testing on this factor. See Appendix D and E for the RecSat and KeepSearch One-way ANOVA tests.

Since the ANOVA tests show that there are differences occurring between groups for RecSat, I used a Dunnett test to reveal which individual groups are displaying the differences. Based on the logic of my hypothesis and the literature review, where the 100% human voice is presumably the level of realism that would maximize consumer recommendation satisfaction, the Dunnett test was set to compare Voice A, B, and C within the RecSat factor against Voice D, the completely human voice. The differences detected by the ANOVA tests occur when comparing Voice A to Voice D (Mean diff. = -.4532, $p = .041$, 95% CI (-.891 to -.015)) and when comparing Voice B to Voice D (Mean diff = -.5413, $p = .011$, 95% CI (-.979 to -.015)). However, there are no differences between Voice C and D for RecSat (Mean diff. = -.1656, p =

.706, 95% CI (-.607 to .276)). Therefore, there are significant differences between the two least realistic voices and the completely realistic voice when looking at the factor of recommendation satisfaction. This effect suggests that the degree of voice realism affects recommendation satisfaction, and that participants found that the human voice provided significantly better recommendations than the two most robotic voices. See Appendix F for the RecSat Dunnett test.

These results are represented by the mean scores in RecSat for Voice A (M = 4.884, SD = 1.086), Voice B (M = 4.795, SD = 1.042), Voice C (M = 5.171, SD = 1.021), and Voice D (M = 5.337, SD = 1.024). Looking at all four mean scores, there is a general increase as the voice becomes more realistic, but it is not a perfect continuous rise as predicted in my hypothesis; there is a decrease from Voice A to Voice B. See Figure 1 for the RecSat means.

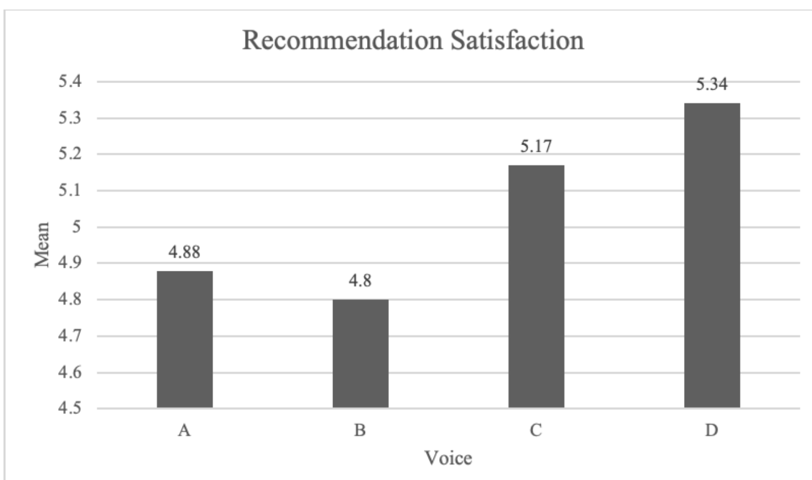


Figure 1. Study 1 Recommendation Satisfaction Across Voice Realism

Finally, a Tukey's test was performed to test all pairwise comparisons to determine whether the differences between the four means were significant. The Tukey's test showed no significant differences between any of the consecutive pairs (for Voice A and B (Mean diff. = .0882, p = .965, 95% CI (-.3928 to .5692), for Voice B and C (Mean diff. = -3758, p = .189, 95% CI (-.8607 to .1091), and for Voice C and D (Mean diff. = -.1656, p = .812, 95% CI (-.6486 to

.3174)). See Appendix G for the RecSat Tukey's test. Nonetheless, even though these differences were not significant and there was a brief decrease in the beginning, they did display a general increase across the four voices.

Thus, the RecSat ANOVA results do not fully support my hypothesis, which predicted that there would be a steady increase in recommendation satisfaction as the assistant's voice became more realistic. Even though there was a general increase across the voices, there was still a slight decrease between Voice A and B. The Tukey's test did reveal that this drop was not significant, however the results still show a clear decrease. While a definitive explanation behind this drop is unknown, it should be noted that it occurred earlier in the scale, when the voice was more robotic. Perhaps the two most robotic voices are not very differentiable and do not have enough of an effect on how people respond to them. Nonetheless, overall, according to the ANOVA for RecSat, voice realism affects recommendation satisfaction, with participants rating the recommendation higher as the voice becomes more humanlike. Thus, the data suggests that the realism of a voice assistant's voice is relevant when recommendation satisfaction is important to consumers, and that more humanlike voices increase satisfaction.

With respect to the voice section of the questionnaire, after a factor analysis of the 18 voice scale questions from a Varimax rotation, four factors emerged explaining 64.33% of the variance of rating responses on the basis of Eigenvalues greater than 1. I called these factors "Friend," "Expert," "Unpleasant," and "Slow," and performed a one-way ANOVA test, shown in the table below, to find statistical differences between groups for the factors. "Friend" characterized Questions 1, 2, 5 10, and 12 (Fun, Likable, Friendly, Engaging, Authentic), based on loadings greater than .635, "Expert" characterized Questions 14, 16, 17, and 18 (Professional, Knowledgeable, Informative, Competent) based on loadings greater than .715, "Unpleasant" characterized Questions 6, 9, 11, and 15 (Rough, Unrelatable, Annoying, Creepy) based on loadings greater than .685, and “Slow” (representing perceptions of the voice’s speed) characterized Question 4 (Slow), which was loaded independently. See Appendix C for the specific questionnaire questions.

As determined by one-way ANOVA, three out of the four factors showed statistically significant differences between groups of the varying conditions: Friend ($F(3,246) = 4.410$, $p = .005$), Unpleasant ($F(3,247) = 9.290$, $p < .001$), and Slow ($F(3,250) = 2.781$, $p = .042$). These results show that there are meaningful relationships between the perceived friendliness, unpleasantness, and slowness of the voice and the realism of the voice. The factor that showed no significant difference between groups from the ANOVA was Expert ($F(3,246) = 1.729$, $p = .162$), and therefore there is no meaningful relationship between the perceived expertise of the voice and the realism of the voice, thus I did not complete further testing on this factor. See Appendix H for the Friend, Expert, Unpleasant, and Slow One-way ANOVA test.

Again, a Dunnett test was performed to reveal the groups displaying the differences shown by the ANOVA test, comparing Voices A, B, and C within each factor against Voice D, the completely human voice. For Friend, differences occur when comparing Voice A to Voice D (Mean diff. = -.7659, $p = .001$, 95% CI (-1.28 to -.254)), but not when comparing Voices B or C to Voice D (Mean diff. = -.3159, $p = .331$, 95% CI (-.829 to .197); Mean diff. = -.5082, $p = .057$, 95% CI (-1.027 to .011). The same pattern occurred for the Unpleasant factor, where there was a difference between Voice A and D (Mean diff. = 1.1352, $p < .001$, 95% CI (.599 to 1.672), but no differences when comparing B and C to D (Mean diff. = -.3304, $p = .329$, 95% CI (-.204 to .865); Mean diff. = -.2562, $p = .536$, 95% CI (-.281 to .793). Therefore, when looking at the factors of perceived friendliness and unpleasantness, there is a significant difference between the least realistic voice and the completely realistic voice. For Slow, while the ANOVA test did suggest that there was a difference between groups in that factor, the Dunnett test did not reveal any differences when comparing Voices A, B, and C to Voice D within Slow (however the mean pattern for Slow can still be considered). Therefore, the results suggest that the degree of voice realism affects perception of friendliness and unpleasantness of the voice, and participants found the human voice to be significantly more friendly and less unpleasant than the most robotic voice. See Appendix I for the Dunnett test.

Disregarding the Expert factor because it is not relevant, the results for the factors Friend, Unpleasant, and Slow are represented by mean scores for Voices A, B, C, and D (for Friend, Voice A ($M = 3.060$, $SD = 1.223$), Voice B ($M = 3.756$, $SD = 1.209$), Voice C ($M = 3.564$, $SD = 1.123$), and Voice D ($M = 4.072$, $SD = 1.294$); for Unpleasant (Voice A ($M = 3.452$, $SD = 1.492$, $SE = .153$), Voice B ($M = 2.647$, $SD = 1.285$), Voice C ($M = 2.573$, $SD = .1.252$), and Voice D ($M = 2.316$, $SD = 1.029$, $SE = .189$); and for Slow (Voice A ($M = 3.719$, $SD = 1.548$), Voice B ($M = 4.141$, $SD = 1.582$), Voice C ($M = 3.355$, $SD = 1.621$), and Voice D ($M = 3.563$, $SD = 1.602$).) See Appendix J for the Friend, Expert, Unpleasant, and Slow descriptives.

The only factor that shows a general increase is Friend. There was not a perfect steady increase however, as demonstrated by a slight decrease between Voices B and C. Unpleasant and Slow show prominent decreases between voices that do not support my hypothesis. See Appendix K and L for the Friend and Unpleasant means plots.

Again, a Tukey's test was performed to test all pairwise comparisons for Friend in order to determine whether the differences between the four condition means were significant. Like the RecSat factor, the Tukey's test showed no significant differences between any of the consecutive pairs (for Voice A to B (Mean diff. = -.4500, $p = .157$, 95% CI (-1.0051 to .1051), for Voice B to

C (Mean diff. = .1923, p = .812, 95% CI (-.3695to .7542), and for Voice C to D (Mean diff. = -.5082, p = .098, 95% CI (-1.0767 to .0603)). See Appendix M for the Friend Tukey's test. Again, even though these differences were not significant and there was one slight decrease, the general upward trend in affinity is noted.

Thus, similar to the RecSat ANOVA results, the Friend ANOVA results do not fully support my hypothesis, which predicted that there would be a steady increase in affinity for the assistant as the voice became more realistic. Even though there was a general increase across the voices, there was still a slight decrease between Voice B and C. The Tukey's test did reveal that this drop was not significant, however the results still show a clear decrease. Nonetheless, according to the ANOVA for Friend, voice realism affects affinity in terms of the factor of perceived friendliness. For the Friend factor, participants have a greater affinity for the assistant as the voice becomes more humanlike. See Figure 2 for the Friend means.

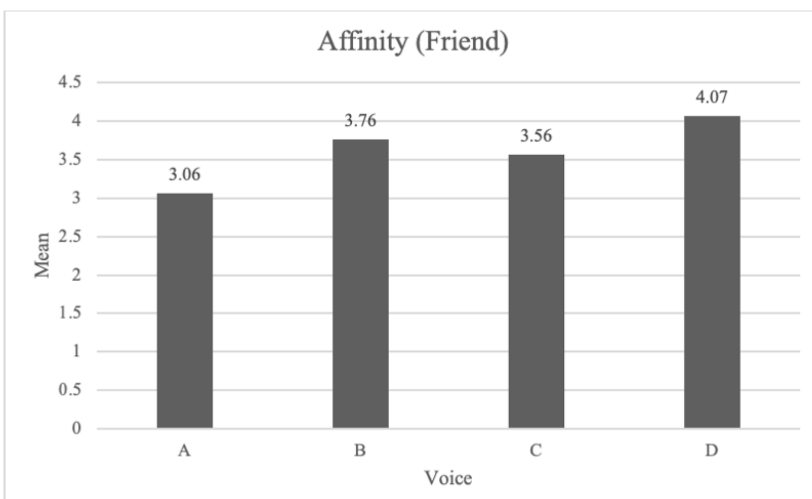


Figure 2. Study 1 Affinity (Friend) Across Voice Realism

However, in terms of the other three factors Expert, Unpleasant, and Slow, there is no relevant effect. Additionally, in terms of affinity as a whole (the means for each of the four factors averaged together), there is no relevant effect. See Figure 3 for the overall Affinity means.

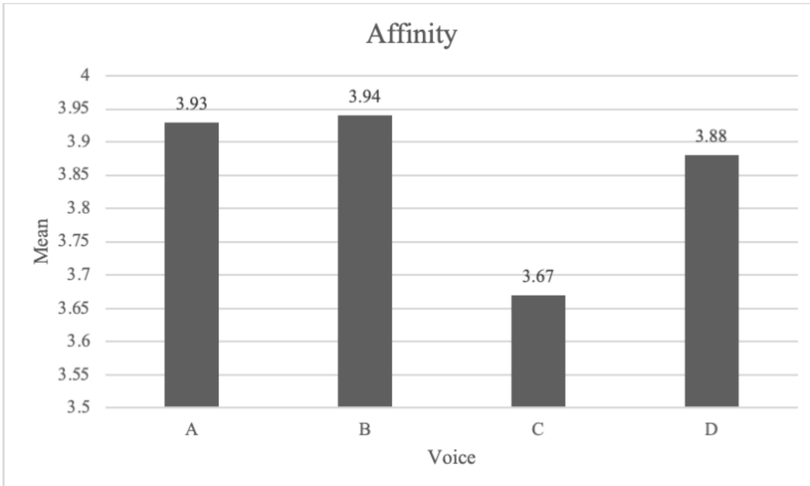


Figure 3. Study 1 Affinity Across Voice Realism

Therefore, the data suggests that the realism of a voice assistant's voice is relevant when friendliness of the assistant is important to consumers, and that more humanlike voices increase affinity in terms of the perceived friendliness factor.

Chapter 5

Study 2

5.1 Research Method and Design

To test H_3 and H_4, I conducted a Qualtrics questionnaire that involved participants listening to each of the voices of varying realism and contextual situation and measuring their satisfaction of the voice's recommendation and affinity toward the voice.

All stimuli were created with the help of a professionally trained audio technician, manipulating only timbre and keeping volume, pitch, and speech rate constant. The voice creation involved recording a human male and then manipulating the sound file using digital audio manipulation software to create recordings for each level of realism and contextual situation. The four voices varied in realism (robot vs. human) and context (serious vs. playful). Voice A was robot and serious, Voice B was human and serious, Voice C was robot and playful, and Voice D was human and playful. See Appendix N for the sound stimuli files.

A total of 291 students at Penn State University participated in the questionnaire for partial class credit. They were each randomly assigned one of the voices to listen to, so roughly 72 to 73 students heard each voice. See Section 5.2 for a description of the sample demographics. The questionnaire began with a randomized prompt telling participants that they were going to be listening to a recording of a digital assistant. They were instructed to imagine that they have asked the assistant to tell them either a fun or important fact (depending on the context they were randomly assigned), and that they are going to hear the assistant's response.

They were also instructed to pay attention to the voice and its information. On the next page, the participants were randomly assigned to listen to one audio file (out of the four possible files). Participants were instructed to play the audio file and listen carefully. The audio file provided the fact. Participants could listen to the recording as many times as they wanted, but once they clicked to the next page, they could not go back to relisten. See Appendix O for the audio file script and Appendix P for the full questionnaire.

The dependent variables of recommendation satisfaction and affinity toward the assistant were measured using 7-point Likert scales which matched those from Study 1. Again, the first main portion of the questionnaire was designed to measure satisfaction of the recommendation. Participants rated the recommendation provided by the assistant on many dimensions, including whether needs and expectations were met, whether the amount of information was appropriate, the effectiveness and efficiency of the response, the potential for repeated use, etc. The second main portion was designed to measure affinity toward the assistant. Participants rated the voice on many dimensions, including likability, speed, friendliness, roughness, authenticity, trustworthiness, competence, etc. Again, participants then completed manipulation check questions testing the recall of information from the audio file. Finally, they filled out demographic information regarding their personal voice assistant usage habits, attitude about usefulness of digital assistants, preference for the realism level of digital assistants, age, gender, political affiliation, race, ethnicity, and socioeconomic status.

5.2 Results and Analysis

In terms of demographics, the experiment utilized a sample of 291 Penn State students, where they were asked to take an online questionnaire in return for partial course credit. The sample ranged from 18 to over 22 years old, with diversity of demographic information and varying habits and attitudes regarding voice assistants. See Appendix P for the audio file scripts and Appendix Q for the full questionnaire.

With respect to the recommendations section of the questionnaire, in an initial test of the results, a 2x2 Between-Subjects ANOVA was conducted on recommendation satisfaction. Results revealed a significant main effect of voice realism ($F(1, 261) = 17.442$, $p < .001$), indicating that there is a significant difference in respondents' recommendation satisfaction between the robotic voice and the human voice. Additionally, there was a significant main effect of command context ($F(1, 261) = 15.937$, $p < .001$), suggesting that there is a significant difference in respondents' recommendation satisfaction between the playful context and serious context. See Appendix R and S for the main effects for RecSat and command context.

However, the interaction effect between voice realism and command context was non-significant ($F(1, 161) = .769$, $p < .381$, suggesting that the combined influence of both factors did not significantly impact recommendation satisfaction. Post hoc tests were conducted to further explore the nature of this interaction, but the results did not reveal any significant differences. See Appendix T for the interaction effect between Recsat and command context. Therefore, for the recommendation satisfaction variable, I will only be discussing the main effects going forward.

For the voice realism factor, the mean for the human voice ($F(4.306, 4.718) = 4.512$, $p < .001$) was significantly higher than the mean for the robotic voice ($F(3.671, 4.097) = 3.884$, $p < .001$), indicating that respondents preferred the more human voice. For the command context factor, the mean for the playful context ($F(3.690, 4.105) =3.898$, $p < .001$) was significantly lower than the mean for the serious context ($F(4.287, 4.710) =4.498$, $p < .001$), indicating that respondents preferred the more serious context. See Tables 4 and 5 for the RecSat means.

Table 1. Study 2 RecSat Means (Voice Realism)

Voice Realism	Mean
Unrealistic (Robotic)	3.884
Realistic (Human)	4.512

Table 2. Study 2 RecSat Means (Command Context)

Command Context	Mean
Serious	4.498
Playful	3.898

The results of Study 2 provide an expansion and verification of my findings in Study 1. Not only did Study 2 replicate and confirm earlier results about the realism factor, but it also expanded my scope to consider whether a contextual condition played a meaningful role in the recommendation satisfaction and affinity factors.

Regarding recommendation satisfaction, there are significant main effects for both voice realism and command context yet no significance for the interaction between the two. These results do not support H_3, as respondents' preferences for the realism level of the voice assistant's voice (measured by recommendation satisfaction) is not impacted by the context of the task that they have assigned it to perform. Therefore, it does not matter whether the assistant is performing a more fun task or a more serious task, either way, the respondent prefers a humanlike voice. But even though my findings do not support H_3, they do reveal clear conclusions. That is, when recommendation satisfaction is important to customers, realism of the voice is relevant to consider in firms' voice assistant design. More specifically, to maximize recommendation satisfaction, firms should make the assistant's voice as realistic, or humanlike, as possible. It is also notable to mention that these findings do support H_1, as I initially predicted that recommendation satisfaction would increase when voice realism increases.

With respect to the voice section of the questionnaire, a 2x2 Between-Subjects ANOVA was conducted on recommendation satisfaction. Results revealed a significant main effect of voice realism ($F(1, 260) = 24.603$, $p < .001$), indicating that there is a significant difference in respondents' affinity for the voice between the robotic voice and the human voice. However, there was not a significant main effect of command context ($F(1, 260) = .008$, $p < .930$), suggesting that there is no meaningful relationship between respondents' affinity for the voice and the context of the command being given, thus I did not complete further testing on this factor.

Since the command context factor was not significant for the affinity variable, the interaction effect between voice realism and command context was non-significant ($F(1, 260) = .079$, $p < .779$), therefore the combined influence of both factors did not significantly impact recommendation satisfaction. Post hoc tests were conducted to further explore the nature of this interaction, but the results did not reveal any significant differences. Therefore, for the affinity variable, I will only be discussing the voice realism main effect going forward.

For the voice realism factor, the mean for the human voice ($F(3.976, 4.161) = 4.068$, $p < .001$) was significantly higher than the mean for the robotic voice ($F(3.637, 3.829) = 3.733$, $p < .001$), indicating that respondents preferred the more human voice. As mentioned before, the command context factor was disregarded as it did not show a significant main effect. Therefore, in terms of affinity, there is meaningful difference between the command contexts. See Table 4 for the Affinity means.

Table 3. Study 2 Affinity Means (Voice Realism)

Affinity	Mean
Unrealistic (Robotic)	3.733
Realistic (Human)	4.068

Regarding affinity for the voice, there is a significant main effect for voice realism yet no significant main effect for command context, and thus no significance for the interaction between the two. Similar to before, these results do not support H_4, as respondents' preferences for the realism level of the voice assistant's voice (measured by affinity for the device) is not impacted by the context of the task that they have assigned it to perform. Again, either way, the respondent prefers a humanlike voice. Instead, my findings show that when affinity is important to consumers, realism of the voice is relevant to consider in firms' voice assistant design. Firms should make the assistant's voice as realistic, or humanlike, as possible. Once again, these findings support H_2, as I initially predicted that affinity for the voice assistant would increase when voice realism increases.

Chapter 6

Conclusions, Limitations, and Future Research

By adjusting the voice's realism and measuring its effect on consumer recommendation satisfaction and affinity toward the device, I was able to gain insight for firms to consider for optimization of their voice assistant design.

Study 1 measured respondents' recommendation satisfaction and affinity for the assistant across four different voices on a scale from robotic to human. I predicted that recommendation satisfaction (H_1) and affinity (H_2) would increase as realism increases. Study 2 added the additional variable of task context, addressing the matching theory by measuring respondents' recommendation satisfaction and affinity for the assistant across two realism levels (robotic and human) and two task contexts (playful and serious). I predicted that recommendation satisfaction (H_3) and affinity (H_4) would increase when the realistic voice was paired with the playful task and when the less realistic voice was paired with the serious task.

Regarding recommendation satisfaction, my results showed that voice realism does indeed affect this variable, with respondents rating recommendation satisfaction higher as the assistant's voice becomes more humanlike, supporting H_1. Additionally, results showed that this effect is not impacted by task context and thus did not support H_3. They suggest that no matter the context of the task the consumer assigns the assistant, the consumer is better satisfied by the recommendation provided if the assistant has a more humanlike voice. Therefore, according to my research, when recommendation satisfaction is important to the consumer, firms should make the assistant's voice as humanlike as possible. Practical examples of when recommendation satisfaction might be important are when critical tasks are being performed by the assistant, such as giving directions, calculating a math problem, or sending a professional message to someone. When people use the assistant for such tasks, they want it to perform exactly to their specifications, as mistakes can be especially harmful.

Regarding affinity for the assistant, my results showed that voice realism affects this variable as the assistant's voice becomes more humanlike in terms of friendliness as a factor of affinity, roughly supporting H_1. The Friend factor was the only one that showed a general increase, indicating that consumer perceptions of friendliness specifically increase when voice realism increases. As for affinity overall, however, when I averaged all four affinity factors to determine how the realism level affected affinity overall, there was no pattern of increase. The increase pattern of Friend and lack of increase pattern for affinity overall is interesting and a bit unusual; further testing could help determine greater insight here and verify the effect. Additionally, results showed that this effect is not impacted by task context and thus did not support H_4. Again, they suggest that no matter the context of the task the consumer assigns the assistant, the consumer perceives the assistant as friendlier if it has a more humanlike voice. Therefore, according to my research, when friendliness is important to the consumer, firms should make the assistant's voice as humanlike as possible. A practical example of when friendliness might be important is when children are using the assistant for educational or play purposes, as friendlier tones and words are important in such interactions. Again, this example also demonstrates that a humanlike voice best suits both types of tasks contexts, serious (educational purposes) and playful (play purposes).

My ultimate recommendation, based on the data from this study, is for firms to make their voice assistants' voices as humanlike as possible. From my studies, recommendation satisfaction and friendliness as a factor of affinity increase along with increased realism. While other factors within these variables did not show a clear increase, they also did not show a decrease detrimental enough to advise against making the voice humanlike. Thus, to maximize positive consumer response overall, firms should design their assistants' voices to be as humanlike as possible.

Like any research, there are some limitations in my studies. However, those limitations invite potential opportunities for future research and greater insight.

First, as discussed in the demographics sections, one limitation is that my samples only included Penn State students and therefore mainly 18- to 22-year-olds. The study is intended to look at consumers of voice assistants, so only studying college student consumers certainly eliminates a significant group of the target market. Future research could use samples of many ages and occupations to better represent the true market of voice assistant consumers.

Another limitation is the contextual scope of my studies, as I used only a male voice for my stimuli and only measured two types of tasks, the restaurant recommendation and request for either a fun or important fact. The choice to use a male voice was intentional, as I believed that familiarity with the female voice (from well-known assistants like Siri, Alexa, etc.) might be a confound in the study. Also, I chose to measure the restaurant recommendation and request for a fact for simplification's sake. However, both could be factors that impact participants' responses. Future research could assess more stimuli contexts, such as using a female or gender-neutral voice, or evaluating other types of tasks, like telling the user the weather, helping the user create a grocery list, or answering more complex user questions to see if similar responses occur.

Finally, another limitation is the range of realism of voice in my studies. Study 1 had four voices, and Study 2 only had two. For both studies, especially Study 2, measuring on a scale of only a few different voices limits the results. Additionally, a wider range of human-robot voices might improve calibration to demonstrate clearer results. Future research could add more individual voices within the realism range to more directly pinpoint where effects occur.

Evidently, I have many ideas for where my research could go. As voice assistants become increasingly more relevant in our world today, they also present firms with opportunities to better satisfy consumers and ultimately drive success and profit to continue innovation. If firms can realize this potential and take advantage of it, they can make consumers' lives more convenient, knowledge-driven, meaningful, and fun. The essence of a voice assistant is indeed its voice, which can be harnessed to facilitate greater connection between user and assistant. The questions, requests, and utterances of billions of users are powerful; they just need a voice to speak back to them.

Appendix

Appendix A: Study 1 Sound Stimuli

Voice A: 40%Human-60%Robot Voice B: 60%Human-40%Robot

Voice C: 80%Human-20%Robot Voice D: 100%Human

Appendix B: Study 1 Script

"Based upon your search, I've identified Roberto's Italian Bistro as a good option. It's an intimate Italian BYOB serving large portions of authentic cuisine, plus vegetarian and health-minded options. They offer dine-in and carry-out and are located downtown. Would you be interested in going?"

Appendix C: Study 1 Questionnaire

Start of Block: Intro

Thank you Thank you for participating in this portion of the study today. It is being conducted by Olivia Hirt for her Honors option in MKTG 330. Please follow the directions. We will start by asking you to listen to a recording.

Please take the headset at your desk, place it on your head, and get comfortable.

Page Break

Q2 You are going to be listening to a recording of a digital assistant. **Imagine that you have used this assistant to search for a local restaurant to eat at. What you will hear will be the assistant's response.** Please pay particular attention to the voice and its information.

Is your headset on? Remember to pay careful attention. When you are ready, proceed to the next page.

Q41 Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: Intro

Start of Block: 40-60 Human Robot

Q56 Please play the voice file below and listen carefully.

Q42 Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: 40-60 Human Robot

Start of Block: 60-40 Human Robot

Q57 Please play the voice file below and listen carefully.

Q43 Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: 60-40 Human Robot

Start of Block: 80-20 Human Robot

Q58 Please play the voice file below and listen carefully.

Q45 Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: 80-20 Human Robot

Start of Block: Human

Q59 Please play the voice file below and listen carefully.

Q47 Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: Human

Start of Block: Recommendations

Rest Please indicate the degree to which you agree or disagree with each statement.

	Strongly disagree (1)	Disagree (2)	Somewhat disagree (3)	Neither agree nor disagree (4)	Somewhat agree (5)	Agree (6)	Strongly agree (7)
I am likely to go to this restaurant. (1)	○	○	○	○	○	○	○
This recommendation met my needs. (2)	○	○	○	○	○	○	○
This recommendation met my expectations. (3)	○	○	○	○	○	○	○
The voice assistant clearly answered my question. (4)	○	○	○	○	○	○	○
The voice assistant efficiently answered my question. (5)	○	○	○	○	○	○	○
The voice assistant gave me the right amount of information in its response. (6)	○	○	○	○	○	○	○
I am confident in the voice assistant's ability to effectively help me with tasks like this. (7)	○	○	○	○	○	○	○

I am likely to use the voice assistant to help me with other decisions. (8)	○	○	○	○	○	○	○
I am likely to recommend this voice assistant to a peer. (9)	○	○	○	○	○	○	○
I will gather additional information about the recommended restaurant. (10)	○	○	○	○	○	○	○

Page Break

Voices Please rate the voice assistant's voice on the following dimensions:

	Not at all (1)	2 (2)	3 (3)	4 (4)	5 (5)	6 (6)	Extremely (7)
Fun (1)	○	○	○	○	○	○	○
Likable (2)	○	○	○	○	○	○	○
Fast (3)	○	○	○	○	○	○	○
Slow (4)	○	○	○	○	○	○	○
Friendly (5)	○	○	○	○	○	○	○
Rough (6)	○	○	○	○	○	○	○
Smooth (7)	○	○	○	○	○	○	○
Boring (8)	○	○	○	○	○	○	○
Unrelatable (9)	○	○	○	○	○	○	○
Engaging (10)	○	○	○	○	○	○	○
Annoying (11)	○	○	○	○	○	○	○
Authentic (12)	○	○	○	○	○	○	○
Trustworthy (13)	○	○	○	○	○	○	○
Professional (14)	○	○	○	○	○	○	○
Creepy (15)	○	○	○	○	○	○	○
Knowledgeable (16)	○	○	○	○	○	○	○

Informative (17)	○	○	○	○	○	○	○
Competent (18)	○	○	○	○	○	○	○

Page Break

Q8 How much did you like the voice assistant overall?

- ◯ Dislike a great deal (1)
- ◯ Dislike a moderate amount (2)
- ◯ Dislike a little (3)
- ◯ Neither like nor dislike (4)
- ◯ Like a little (5)
- ◯ Like a moderate amount (6)
- ◯ Like a great deal (7)

End of Block: Recommendations

Start of Block: Manipulation Checks

Q9 Which restaurant did the voice assistant recommend?

- ◯ Mateo's Mexican Cuisine (1)
- ◯ Mieko's Sushi House (2)
- ◯ Roberto's Italian Bistro (3)
- ◯ Samuel's Steak and Seafood (4)

Q10 What kind of food does the restaurant serve?

- ◯ Mexican (1)
- ◯ Japanese (2)
- ◯ Italian (3)
- ◯ American (4)

Q11 Where is the restaurant located?

- Downtown State College (1)
- Toftrees (2)
- Bellefonte (3)
- Boalsburg (4)

Q12 Is the restaurant dine-in, carry-out, or both?

- Dine-in (1)
- Carry-out (2)
- Both (3)

Q13 Does the restaurant serve alcohol, or is it BYOB?

- Serves alcohol (1)
- BYOB (2)

Q14 What is your expectation of the atmosphere of the restaurant?

- Lively (1)
- Lavish (2)
- Intimate (3)
- Modern (4)

End of Block: Manipulation Checks

Start of Block: Demographics

Q16 How often do you use voice assistants?

- Never (1)
- Rarely (2)
- Sometimes (3)
- Often (4)
- Always (5)

Q17 Do you own a voice assistant (e.g., Amazon Alexa, Google Home)?

- Yes (1)
- No (2)

Q18 How much do you like voice assistants in general?

- Dislike a great deal (1)
- Dislike somewhat (2)
- Neither like nor dislike (3)
- Like somewhat (4)
- Like a great deal (5)

Q19 How useful do you think voice assistants are?

- Not at all useful (1)
- Slightly useful (2)
- Moderately useful (3)
- Very useful (4)
- Extremely useful (5)

Q20 How would you rate the quality of recommendations that voice assistants typically provide?

- Poor (1)
- 2 (2)
- 3 (3)
- 4 (4)
- 5 (5)
- 6 (6)
- Excellent (7)

Q7 Would you prefer if the voice assistant sounded more human-like or more robotic?

- Definitely more robotic (1)
- 2 (2)
- 3 (3)
- 4 (4)
- 5 (5)
- 6 (6)
- Definitely more human-like (7)

Page Break

Q22 What is your age?

◯ 18 (1)

◯ 19 (2)

◯ 20 (3)

◯ 21 (4)

◯ 22+ (5)

Q23 What is the gender that you identify with?

◯ Male (1)

◯ Female (2)

◯ Non-binary / third gender (3)

◯ Prefer not to say (4)

End of Block: Demographics

Start of Block: Demographics Political

Q2 Generally speaking, do you usually think of yourself as Republican, Democrat, Independent, or something else?

◯ Republican (1)

◯ Democrat (2)

◯ Independent (3)

◯ Other (4) ____

◯ No preference (5)

End of Block: Demographics Political

Start of Block: Demographics Base/Universal

Q5 Choose one or more races that you consider yourself to be:

- ☐ White (1)
- ☐ Black or African American (2)
- ☐ American Indian or Alaska Native (3)
- ☐ Asian (4)
- ☐ Native Hawaiian or Pacific Islander (5)
- ☐ Other (6) ____

Q3 Are you Spanish, Hispanic, or Latino or none of these?

- ○ Yes (1)
- ○ None of these (2)

Q7 Based upon your best guess, please indicate the answer that includes your entire household income (in the previous year) before taxes.

- Less than $10,000 (1)
- $10,000 to $19,999 (2)
- $20,000 to $29,999 (3)
- $30,000 to $39,999 (4)
- $40,000 to $49,999 (5)
- $50,000 to $59,999 (6)
- $60,000 to $69,999 (7)
- $70,000 to $79,999 (8)
- $80,000 to $89,999 (9)
- $90,000 to $99,999 (10)
- $100,000 to $149,999 (11)
- $150,000 or more (12)

Page Break

Q39 Thank you very much for participating in today's study. If you have any questions about it, please email Olivia Hirt (ovh5082@psu.edu).

Have a great day!

End of Block: Demographics Base/Universal

Appendix D: RecSat One-way ANOVA

ANOVA

RecSat

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	11.986	3	3.995	3.668	.013
Within Groups	269.017	247	1.089		
Total	281.003	250			

Appendix E: KeepSearch One-way ANOVA

ANOVA

KeepSearch

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	9.696	3	3.232	1.729	.162
Within Groups	467.331	250	1.869		
Total	477.028	253			

Appendix F: RecSat Dunnett Test

Multiple Comparisons

Dependent Variable: RecSat
Dunnett t (2-sided)[a]

					95% Confidence Interval	
(I) CondNumeric	(J) CondNumeric	Mean Difference (I-J)	Std. Error	Sig.	Lower Bound	Upper Bound
1	4	-.45321*	.18522	.041	-.8912	-.0152
2	4	-.54139*	.18522	.011	-.9794	-.1034
3	4	-.16559	.18674	.706	-.6072	.2760

*. The mean difference is significant at the 0.05 level.

a. Dunnett t-tests treat one group as a control, and compare all other groups against it.

Appendix G: RecSat Tukey's Test

Multiple Comparisons

Dependent Variable: RecSat

	(I) CondNumeric	(J) CondNumeric	Mean Difference (I–J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Tukey HSD	1	2	.08818	.18595	.965	-.3928	.5692
		3	-.28762	.18746	.419	-.7725	.1973
		4	-.45321	.18522	.071	-.9323	.0259
	2	1	-.08818	.18595	.965	-.5692	.3928
		3	-.37581	.18746	.189	-.8607	.1091
		4	-.54139*	.18522	.020	-1.0205	-.0623
	3	1	.28762	.18746	.419	-.1973	.7725
		2	.37581	.18746	.189	-.1091	.8607
		4	-.16559	.18674	.812	-.6486	.3174
	4	1	.45321	.18522	.071	-.0259	.9323
		2	.54139*	.18522	.020	.0623	1.0205
		3	.16559	.18674	.812	-.3174	.6486

*. The mean difference is significant at the 0.05 level.

Appendix H: Friend, Expert, Unpleasant, and Slow One-way ANOVA test

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Friend	Between Groups	19.494	3	6.498	4.410	.005
	Within Groups	362.478	246	1.473		
	Total	381.973	249			
Expert	Between Groups	5.841	3	1.947	1.729	.162
	Within Groups	276.965	246	1.126		
	Total	282.806	249			
Unpleasant	Between Groups	45.162	3	15.054	9.290	<.001
	Within Groups	400.263	247	1.620		
	Total	445.425	250			
Slow	Between Groups	21.042	3	7.014	2.781	.042
	Within Groups	630.615	250	2.522		
	Total	651.657	253			

Appendix I: Friend, Expert, Unpleasant, and Slow Dunnett Test

Multiple Comparisons

Dunnett t (2-sided)[a]

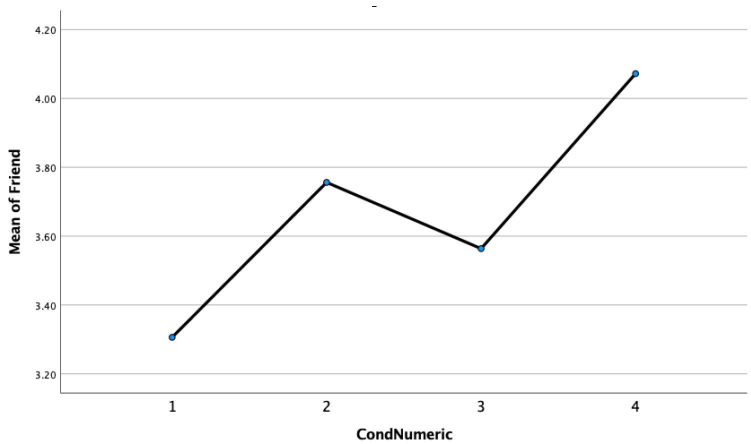
Dependent Variable	(I) CondNumeric	(J) CondNumeric	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Friend	1	4	-.76588*	.21721	.001	-1.2789	-.2528
	2	4	-.31588	.21721	.331	-.8289	.1972
	3	4	-.50820	.21980	.057	-1.0274	.0110
Expert	1	4	-.31855	.19057	.227	-.7689	.1318
	2	4	-.35370	.18908	.154	-.8005	.0931
	3	4	-.37903	.19057	.120	-.8294	.0713
Unpleasant	1	4	1.13521*	.22684	<.001	.5988	1.6716
	2	4	.33042	.22593	.329	-.2038	.8647
	3	4	.25617	.22684	.536	-.2802	.7926
Slow	1	4	.15625	.28076	.901	-.5074	.8199
	2	4	.57813	.28076	.103	-.0856	1.2418
	3	4	-.20766	.28302	.805	-.8767	.4613

Appendix J: Friend, Expert, Unpleasant, and Slow Descriptives

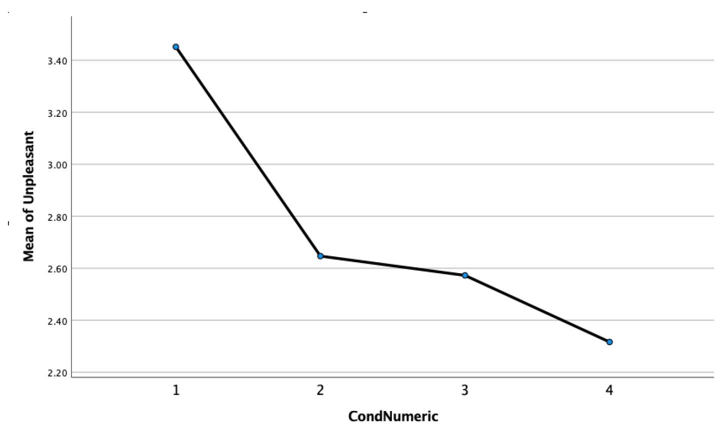
Descriptives

		N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean Lower Bound	95% Confidence Interval for Mean Upper Bound	Minimum	Maximum
Friend	1	64	3.3062	1.22304	.15288	3.0007	3.6118	1.00	6.60
	2	64	3.7562	1.20868	.15109	3.4543	4.0582	1.40	6.20
	3	61	3.5639	1.12339	.14384	3.2762	3.8516	1.20	6.20
	4	61	4.0721	1.29411	.16569	3.7407	4.4036	1.40	7.00
	Total	250	3.6712	1.23856	.07833	3.5169	3.8255	1.00	7.00
Expert	1	62	5.2500	1.12807	.14326	4.9635	5.5365	2.00	7.00
	2	64	5.2148	1.20504	.15063	4.9138	5.5159	1.00	7.00
	3	62	5.1895	.97687	.12406	4.9414	5.4376	2.75	7.00
	4	62	5.5685	.90215	.11457	5.3394	5.7977	3.25	7.00
	Total	250	5.3050	1.06572	.06740	5.1722	5.4378	1.00	7.00
Unpleasant	1	62	3.4516	1.49167	.18944	3.0728	3.8304	1.00	6.50
	2	63	2.6468	1.28471	.16186	2.3233	2.9704	1.00	5.25
	3	62	2.5726	1.25155	.15895	2.2547	2.8904	1.00	5.25
	4	64	2.3164	1.02860	.12858	2.0595	2.5733	1.00	5.25
	Total	251	2.7430	1.33480	.08425	2.5771	2.9090	1.00	6.50
Slow	1	64	3.7188	1.54785	.19348	3.3321	4.1054	1.00	7.00
	2	64	4.1406	1.58231	.19779	3.7454	4.5359	1.00	7.00
	3	62	3.3548	1.62053	.20581	2.9433	3.7664	1.00	7.00
	4	64	3.5625	1.60233	.20029	3.1622	3.9628	1.00	7.00
	Total	254	3.6969	1.60491	.10070	3.4985	3.8952	1.00	7.00

Appendix K: Friend Means Plots



Appendix L: Unpleasant Means Plots



Appendix M: Friend Tukey's Test

Multiple Comparisons

Dependent Variable: Friend
Tukey HSD

(I) CondNumeric	(J) CondNumeric	Mean Difference (I–J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	-.45000	.21458	.157	-1.0051	.1051
	3	-.25768	.21721	.636	-.8195	.3042
	4	-.76588*	.21721	.003	-1.3277	-.2040
2	1	.45000	.21458	.157	-.1051	1.0051
	3	.19232	.21721	.812	-.3695	.7542
	4	-.31588	.21721	.467	-.8777	.2460
3	1	.25768	.21721	.636	-.3042	.8195
	2	-.19232	.21721	.812	-.7542	.3695
	4	-.50820	.21980	.098	-1.0767	.0603
4	1	.76588*	.21721	.003	.2040	1.3277
	2	.31588	.21721	.467	-.2460	.8777
	3	.50820	.21980	.098	-.0603	1.0767

*. The mean difference is significant at the 0.05 level.

Appendix N: Study 2 Sound Stimuli

Voice A: SeriousRobot

Voice B: FunRobot

Voice C: SeriousHuman

Voice D: FunHuman

Appendix O: Study 2 Script

Playful: *"The unicorn is the national animal of Scotland."*

Serious: *"Fire kills more Americans per year than all natural disasters combined."*

Appendix P: Study 2 Questionnaire

Start of Block: S: Intro

Q81 Thank you for participating in this portion of the study today. It is being conducted by Olivia Hirt for her Honors thesis. Please follow the directions. We will start by asking you to listen to a recording.

Please take the headset at your desk, place it on your head, and get comfortable.

You are going to be listening to a recording of a digital assistant. Imagine that you have asked this assistant to tell you an important fact. What you will hear will be the assistant's response. Please pay particular attention to the voice and its information.

Is your headset on? Remember to pay careful attention. When you are ready, proceed to the next page.

Page Break

Q69 Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: S: Intro

Start of Block: P: Intro

Q100 Thank you for participating in this portion of the study today. It is being conducted by Olivia Hirt for her Honors thesis. Please follow the directions. We will start by asking you to listen to a recording.

Please take the headset at your desk, place it on your head, and get comfortable.

You are going to be listening to a recording of a digital assistant. Imagine that you have asked this assistant to tell you a fun fact. What you will hear will be the assistant's response. Please pay particular attention to the voice and its information.

Is your headset on? Remember to pay careful attention. When you are ready, proceed to the next page.

Page Break

Q41 Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: P: Intro

Start of Block: S: Robot

Q56 Please play the voice file below and listen carefully.

Q42 Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: S: Robot

Start of Block: S: Human

Q57 Please play the voice file below and listen carefully.

Q43 Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: S: Human

Start of Block: P: Robot

Q58 Please play the voice file below and listen carefully.

Q45 Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: P: Robot

Start of Block: P: Human

Q59 Please play the voice file below and listen carefully.

Q47 Timing
First Click (1)
Last Click (2)
Page Submit (3)
Click Count (4)

End of Block: P: Human

Start of Block: Recommendations

Rest Please indicate the degree to which you agree or disagree with each statement.

	Strongly disagree (1)	Disagree (2)	Somewhat disagree (3)	Neither agree nor disagree (4)	Somewhat agree (5)	Agree (6)	Strongly agree (7)
I appreciate this fact. (1)	○	○	○	○	○	○	○
The fact provided met my needs. (2)	○	○	○	○	○	○	○
The fact provided met my expectations. (3)	○	○	○	○	○	○	○
The voice assistant clearly achieved my request. (4)	○	○	○	○	○	○	○
The voice assistant efficiently achieved my request. (5)	○	○	○	○	○	○	○
The voice assistant gave me the right amount of information in its response. (6)	○	○	○	○	○	○	○
I am confident in the voice assistant's ability to effectively perform tasks like this. (7)	○	○	○	○	○	○	○

I am likely to use the voice assistant to help me with other tasks. (8)	○	○	○	○	○	○	○
I am likely to recommend this voice assistant to a peer. (9)	○	○	○	○	○	○	○
I will gather additional information like this in the future. (10)	○	○	○	○	○	○	○

Page Break

Voices Please rate the voice assistant's voice on the following dimensions:

	Not at all (1)	2 (2)	3 (3)	4 (4)	5 (5)	6 (6)	Extremely (7)
Fun (1)	○	○	○	○	○	○	○
Likable (2)	○	○	○	○	○	○	○
Fast (3)	○	○	○	○	○	○	○
Slow (4)	○	○	○	○	○	○	○
Friendly (5)	○	○	○	○	○	○	○
Rough (6)	○	○	○	○	○	○	○
Smooth (7)	○	○	○	○	○	○	○
Boring (8)	○	○	○	○	○	○	○
Unrelatable (9)	○	○	○	○	○	○	○
Engaging (10)	○	○	○	○	○	○	○
Annoying (11)	○	○	○	○	○	○	○
Authentic (12)	○	○	○	○	○	○	○
Trustworthy (13)	○	○	○	○	○	○	○
Professional (14)	○	○	○	○	○	○	○
Creepy (15)	○	○	○	○	○	○	○
Knowledgeable (16)	○	○	○	○	○	○	○

Informative (17)	○	○	○	○	○	○	○
Competent (18)	○	○	○	○	○	○	○

Page Break

Q8 How much did you like the voice assistant overall?

○ Dislike a great deal (1)

○ Dislike a moderate amount (2)

○ Dislike a little (3)

○ Neither like nor dislike (4)

○ Like a little (5)

○ Like a moderate amount (6)

○ Like a great deal (7)

End of Block: Recommendations

Start of Block: Manipulation Check S

Q80 Which fact did the assistant provide?

○ If you're not wearing a seat belt, you are 30 times more likely to be thrown from a vehicle during a collision. (1)

○ On average, smokers die 10 years earlier than nonsmokers. (2)

○ Fire kills more Americans every year than all natural disasters combined. (3)

○ Over 25,000 children under the age of five are admitted to the emergency room every year because of accidental poisoning. (4)

End of Block: Manipulation Check S

Start of Block: Manipulation Check P

Q83 Which fact did the assistant provide?

○ British military tanks are equipped to make tea. (1)

○ A blue whale's tongue can weigh as much as a young elephant. (2)

○ The unicorn is the national animal of Scotland. (3)

○ A chef's hat has 100 pleats, representing the 100 ways you can cook an egg. (4)

End of Block: Manipulation Check P

Start of Block: Demographics

Q16 How often do you use voice assistants?

○ Never (1)

○ Rarely (2)

○ Sometimes (3)

○ Often (4)

○ Always (5)

Q17 Do you own a voice assistant (e.g., Amazon Alexa, Google Home)?

○ Yes (1)

○ No (2)

Q18 How much do you like voice assistants in general?

○ Dislike a great deal (1)

○ Dislike somewhat (2)

○ Neither like nor dislike (3)

○ Like somewhat (4)

○ Like a great deal (5)

Q19 How useful do you think voice assistants are?

- Not at all useful (1)
- Slightly useful (2)
- Moderately useful (3)
- Very useful (4)
- Extremely useful (5)

Q20 How would you rate the quality of the tasks that voice assistants typically complete?

- Poor (1)
- 2 (2)
- 3 (3)
- 4 (4)
- 5 (5)
- 6 (6)
- Excellent (7)

Q7 Would you prefer if the voice assistant sounded more human-like or more robotic?

- Definitely more robotic (1)
- 2 (2)
- 3 (3)
- 4 (4)
- 5 (5)
- 6 (6)
- Definitely more human-like (7)

Q22 What is your age?

- 18 (1)
- 19 (2)
- 20 (3)
- 21 (4)
- 22+ (5)

Q23 What is the gender that you identify with?

- Male (1)
- Female (2)
- Non-binary / third gender (3)
- Prefer not to say (4)

End of Block: Demographics

Start of Block: Demographics Political

Q2 Generally speaking, do you usually think of yourself as Republican, Democrat, Independent, or something else?

- Republican (1)
- Democrat (2)
- Independent (3)
- Other (4) ____
- No preference (5)

End of Block: Demographics Political

Start of Block: Demographics Base/Universal

Q5 Choose one or more races that you consider yourself to be:

- ☐ White (1)
- ☐ Black or African American (2)
- ☐ American Indian or Alaska Native (3)
- ☐ Asian (4)
- ☐ Native Hawaiian or Pacific Islander (5)
- ☐ Other (6) ____

Q3 Are you Spanish, Hispanic, or Latino or none of these?

- ○ Yes (1)
- ○ None of these (2)

Q7 Based upon your best guess, please indicate the answer that includes your entire household income (in the previous year) before taxes.

- ○ Less than $10,000 (1)
- ○ $10,000 to $19,999 (2)
- ○ $20,000 to $29,999 (3)
- ○ $30,000 to $39,999 (4)
- ○ $40,000 to $49,999 (5)
- ○ $50,000 to $59,999 (6)
- ○ $60,000 to $69,999 (7)
- ○ $70,000 to $79,999 (8)
- ○ $80,000 to $89,999 (9)
- ○ $90,000 to $99,999 (10)
- ○ $100,000 to $149,999 (11)
- ○ $150,000 or more (12)

Q98 Thank you very much for participating in today's study. If you have any questions about it, please email Olivia Hirt (ovh5082@psu.edu).

Have a great day!

End of Block: Gap Project

BIBLIOGRAPHY

Aggarwal, P., McGill, A. L., & article., J. D. served as editor and J. A. served as associate editor for this. (2007). Is That Car Smiling at Me? Schema Congruity as a Basis for Evaluating Anthropomorphized Products. *Journal of Consumer Research*, *34*(4), 468–479. https://doi.org/10.1086/518544

Alexa's most asked questions and requests from UK customers in 2023. (2023, December 15). UK About Amazon. https://www.aboutamazon.co.uk/news/devices/alexa-most-asked-questions-2023

Apple adds two Siri Voices, will no longer default to female voice. HT Tech. (2021, April 1). https://tech.hindustantimes.com/tech/news/apple-adds-two-siri-voices-will-no-longer-default-to-female-voice-71617247927168.html

Becker, C., Kopp, S., & Wachsmuth, I. (2007). Why emotions should be integrated into conversational agents. *Conversational informatics: an engineering approach*, 49-68.

Bjork, R. A. (1970). Positive forgetting: The noninterference of items intentionally forgotten. *Journal of Verbal Learning and Verbal Behavior*, *9*(3), 255-268.

Bolden, B. (2009). Timbre. *The Canadian Music Educator*, *51*(1), 2–3.

Brill, T. M., Munoz, L., & Miller, R. J. (2019, November 1). *Siri, Alexa, and other digital assistants: A study of customer satisfaction...: EBSCOhost*. https://web-p-ebscohost-com.ezaccess.libraries.psu.edu/ehost/pdfviewer/pdfviewer?vid=0&sid=7cd1f901-7ebb-4df5-a32d-3619ce21a68a%40redis

Bruner, G. C. (1990). Music, Mood, and Marketing. *Journal of Marketing*, *54*(4), 94–104. https://doi.org/10.1177/002224299005400408

Cahill, L., Babinsky, R., Markowitsch, H. J., & McGaugh, J. L. (1995). The amygdala and emotional memory. *Nature*, *377*(6547), Article 6547. https://doi.org/10.1038/377295a0

Christenson, B. A., Ringler, C., Sirianni, N. J., Co-Chair, C., Allaway, A. W., Reynolds, K. E., & Harms, P. D. (n.d.). *SPEAKING FAST AND SLOW: HOW SPEECH RATE OF DIGITAL ASSISTANTS AFFECTS LIKELIHOOD TO USE*. 90.

Dahl, D. W. (2011). Understanding the role of spokesperson voice in broadcast advertising. *Sensory Marketing*, 169-182.

Definitions of Marketing. American Marketing Association. (n.d.). https://www.ama.org/the-definition-of-marketing-what-is-marketing/

Efthymiou, F., Hildebrand, C., de Bellis, E., & Hampton, W. H. (2023). The Power of AI-Generated Voices: How Digital Vocal Tract Length Shapes Product Congruency and Ad Performance. *Journal of Interactive Marketing*. https://doi.org/10.1177/10949968231194905

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, *114*(4), 864–886. https://doi.org/10.1037/0033-295X.114.4.864

Fink, J. (2012). Anthropomorphism and Human Likeness in the Design of Robots and Human-Robot Interaction. In S. S. Ge, O. Khatib, J.-J. Cabibihan, R. Simmons, & M.-A. Williams (Eds.), *Social Robotics* (Vol. 7621, pp. 199–208). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-34103-8_20

Goetz, J., Kiesler, S., & Powers, A. (2002, December 31). Matching Robot Appearance and Behavior to Tasks to Improve Human-Robot Cooperation.

https://www.cs.cmu.edu/~kiesler/anthropomorphism.org/pdf/AppearanceAndBehavior.pdf

Hagtvedt, H., & Patrick, V. M. (2008). Art Infusion: The Influence of Visual Art on the Perception and Evaluation of Consumer Products. *Journal of Marketing Research*, *45*(3), 379–389.

Hoch, S. J., & Ha, Y.-W. (1986). Consumer Learning: Advertising and the Ambiguity of Product Experience. *Journal of Consumer Research*, *13*(2), 221–233.

Kim, S., Chen, R. P., & Zhang, K. (2016). Anthropomorphized Helpers Undermine Autonomy and Enjoyment in Computer Games. *Journal of Consumer Research*, *43*(2), 282–302.

Klaus, P., & Zaichkowsky, J. (2020). AI voice bots: a services marketing research agenda. *Journal of Services Marketing*, *34*(3), 389-398.

Krishna, A. (2012). An integrative review of sensory marketing: Engaging the senses to affect perception, judgment and behavior. *Journal of Consumer Psychology*, *22*(3), 332–351. https://doi.org/10.1016/j.jcps.2011.08.003

Meyers-Levy, J. (2010). The sounds of the marketplace: The role of audition in marketing. *Sensory Marketing: Research on the Sensuality of Products*.

Meyers-Levy, J., Zhu, R. (Juliet). (2007). The Influence of Ceiling Height: The Effect of Priming on the Type of Processing That People Use. *Journal of Consumer Research*, *34*(2), 174–186. https://doi.org/10.1086/519146

Mori, M., MacDorman, K., & Kageki, N. (2012). The Uncanny Valley [From the Field]. *IEEE Robotics & Automation Magazine*, *19*, 98–100. https://doi.org/10.1109/MRA.2012.2192811

Morrin, M., & Ratneshwar, S. (2003). Does It Make Sense to Use Scents to Enhance Brand Memory? *Journal of Marketing Research*, *40*(1), 10–25. http://www.jstor.org/stable/30038832

Negrotti, M. (2005). Humans and naturoids: From use to partnerships. In *Yearbook of the Artificial. Nature, Culture & Technology. Volume 3. Cultural Dimension of the User* (pp. 9-15). Peter Lang.

Niculescu, A., van Dijk, B., Nijholt, A., Li, H., & See, S. L. (2013). Making Social Robots More Attractive: The Effects of Voice Pitch, Humor and Empathy. *International Journal of Social Robotics*, *5*(2), 171–191. https://doi.org/10.1007/s12369-012-0171-x

Oakes, S., & North, A. C. (2006). The impact of background musical tempo and timbre congruity upon ad content recall and affective response. *Applied Cognitive Psychology*, *20*(4), 505–520. https://doi.org/10.1002/acp.1199

Shewale, R. (2024, January 11). *67 Voice Search Statistics for 2024 (Updated Data)*. DemandSage. https://www.demandsage.com/voice-search-statistics/

Wells, Robert Earl, III. (2022, April 25). *How to get Alexa Celebrity Voices for Your Amazon Echo Device*. Lifewire. https://www.lifewire.com/get-celebrity-voices-for-alexa-5196719

Yang, Xue. (2021, February). *Determinants of consumers' continuance intention to use social recommender systems: A self-regulation perspective*. https://doi.org/10.1016/j.techsoc.2020.101464