

The Impact of Perceived Tone, Age, and Gender on Voice Assistant Persuasiveness in the Context of Product Recommendations

Sabid Bin Habib Pias
Indiana University
Bloomington, USA

Ran Huang
Indiana University
Bloomington, USA

Donald Williamson
Ohio State University
Columbus, USA

Minjeong Kim
Indiana University
Bloomington, USA

Apu Kapadia
Indiana University
Bloomington, USA

ABSTRACT

Voice Assistants (VAs) can assist users in various everyday tasks, but many users are reluctant to rely on VAs for intricate tasks like online shopping. This study aims to examine whether the vocal characteristics of VAs can serve as an effective tool to persuade users and increase user engagement with VAs in online shopping. Prior studies have demonstrated that the perceived tone, age, and gender of a voice influence the perceived persuasiveness of the speaker in interpersonal interactions. Furthermore, persuasion in product communication has been shown to affect purchase decisions in online shopping. We investigate whether variations in a VA voice's perceived tone, age, and gender characteristics can persuade users and ultimately affect their purchase decisions. Our experimental study showed that participants were more persuaded to make purchase decisions by VA voices having positive or neutral tones as well as middle-aged male or younger female voices. Our results suggest that VA designers should offer users the ability to easily customize VA voices with a range of tones, ages, and genders. This customization can enhance user comfort and enjoyment, potentially leading to higher engagement with VAs. Additionally, we discuss the boundaries of ethical persuasion, emphasizing the importance of safeguarding users' interests against unwarranted manipulation.

CCS CONCEPTS

• Human Centered Computing → Empirical studies in HCI.

KEYWORDS

voice assistants, persuasiveness, paralinguistic traits, trust, survey

ACM Reference Format:

Sabid Bin Habib Pias, Ran Huang, Donald Williamson, Minjeong Kim, and Apu Kapadia. 2024. The Impact of Perceived Tone, Age, and Gender on Voice Assistant Persuasiveness in the Context of Product Recommendations. In *ACM Conversational User Interfaces 2024 (CUI '24)*, July

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions.acm.org.

CUI '24, July 8–10, 2024, Luxembourg, Luxembourg

© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 979-8-4007-0511-3/24/07...\$15.00

<https://doi.org/10.1145/3640794.3665545>

8–10, 2024, Luxembourg, Luxembourg. ACM, New York, NY, USA, 15 pages.
<https://doi.org/10.1145/3640794.3665545>

1 INTRODUCTION

Voice assistants (VA) like Amazon's Alexa, Google Assistant, and Apple's Siri [19, 118, 119] have become incredibly popular due to their hands-free convenience and wide range of applications [49, 112]. Popular functionalities of voice assistants encompass information seeking, including weather checks, recent news, or cooking instruction [51, 72, 73], entertainment, such as music or jokes, and controlling external devices [72]. Because of recent advancements in natural language processing (NLP) [100] through the use of large language models (LLMs), VAs have vastly expanded their capabilities to perform intricate tasks, such as engaging in a conversation with users or offering online shopping recommendations [24, 81]. Furthermore, VA recommendations have demonstrated greater efficacy compared to text-based recommendations [37], making them a suitable alternative for online shopping guidance. However, people are less comfortable with AI assistance in risky tasks that may have financial consequences compared to tasks with low risks [48, 89]. Evidence suggests that the perceived humanlike characteristics of a machine increase users' perceived comfort [56]. Moreover, Li and Sung posited that anthropomorphism reduces the psychological distance in human-AI assistant interactions [70]. Therefore, improving the social appearance of a VA can be a stepping stone toward increasing the adoption of intricate tasks in VAs.

To increase anthropomorphism in VAs, researchers have explored the effect of human-like vocal characteristics in VAs [30, 99, 132, 133]. People perceive news stories as more interesting if the synthesized voice expresses them in a happier voice, as opposed to a sad tone [85]. Additionally, Belanche et al. demonstrated that integrating perceived warmth into a service robot's voice increased user expectations regarding the quality of the robot's service [7] and studies on interaction with synthesized voices found that users generally prefer female and extroverted VA voices [35]. Furthermore, users generally prefer VAs to have a friendly and engaging tone rather than sounding robotic or monotone [43]. However, the implications of differing tones, ages, and genders of VA voices on the perceived persuasiveness of VAs and users' purchase decisions remain largely unexplored.

Current voice assistants let users personalize how they sound to some extent—Amazon's Alexa [118] offers one male and one female voice each for a limited number of accents, such as US or UK accents. In addition, Alexa lets users adjust the speaking rate or switch to a

whisper mode [102]. Google Assistant provides 12 different voices, including male and female voices in varied accents [134]. Siri also offers a few different male or female voices in various accents [16]. However, these voice assistants offer limited voice options, with minimal diversity in vocal age and tone for user preference.

Besides linguistic variation, paralinguistic traits, such as the tone and the perceived age and gender of the voice, are effective in building trust in interpersonal relationships [41]. Evidence suggests that pitch variance, speaker's age, and fluency persuade listeners [8, 109]. Furthermore, being perceived as entertaining and interesting positively influences the adoption and retention of voice assistants among users [144]. Like interpersonal relationships, users likely have individual preferences in conversational tones, and thus, respond to various voices differently, whether consciously or unconsciously [21, 22, 62]. Therefore, variations in tone, age, or gender of the VA voice have the potential to influence users' perceptions of voice assistants in critical decision-making contexts. If making VAs sound different in tone, age, or gender helps users feel more at ease when they make critical decisions such as online shopping, then VA designers can utilize different voice attributes to broaden the adoption of VAs in critical contexts. Moreover, such flexibility will facilitate research exploring variations in VA tasks within analogous contexts such as banking and bill payments. Enhancing trust and comfort in using VAs for such intricate tasks can encourage more users to adopt VAs for intricate daily activities, paving the way to make individuals' daily lives easier through hands-free alternatives, particularly benefiting elderly individuals or those with disabilities or visual impairments. However, ensuring healthy persuasion is vital to prevent unwanted manipulation and to protect the users' interests. VA designs should follow ethical guidelines that prioritize user welfare and consent. In addition, clear disclosure of persuasive intent and flexibility to control the vocal characteristics are also critically important in VA design.

The interplay between perceived age, tone, and gender of VAs in shaping persuasiveness and purchase decisions is yet to be explored. To address the gap in current literature, we are particularly interested in evaluating how the perceived tone, age, and gender of a VA's voice influence users' perceptions of persuasiveness and their purchase decisions. In particular, we sought answers to the following questions:

RQ1: "How does the perceived tone of a voice assistant's voice persuade participants and subsequently affect participants' purchase decisions?" We aimed to explore how positive, negative, and neutral tones used by VAs persuade participants, and consequently, whether the participants follow VA suggestions for purchase decisions. Our objective was to examine the effect of changing tones regardless of the age or gender of the voice.

RQ2: "How do the perceived gender and age of a voice assistant's voice persuade participants and subsequently affect participants' purchase decisions?" We investigated whether the persuasiveness of middle-aged and younger adult male/female voices in voice assistants are different and whether the persuasiveness affects participants' purchase decisions.

To investigate these questions, we developed an online experimental study with synthesized voices to examine how the perceived tone, age group, and gender of a VA's voice persuade participants

and how the persuasiveness affects participants' purchase intentions. We also collected open-ended responses from the participants to further investigate their preferences toward particular voice types.

Our results showed that positive and neutral tones significantly persuaded the participants. As a result of the persuasion, the participants were more likely to follow VA recommendations in purchase decisions for positive and neutral tones. Interestingly, middle-aged male and younger female voices significantly persuaded the users and showed a similar influence on participants' purchase decisions.

With the advancement of voice assistants that require increased user engagement, the findings of this research can help subsequent research endeavors. Diversifying the tones, age, and gender of VA voices can effectively increase user involvement in intricate tasks with financial risks through healthy persuasion. However, an ethical framework to govern the use of persuasion and effective disclosure of potential persuasion is also important to safeguard users from unexpected manipulation. Through responsible VA design, future VAs can enhance user involvement in intricate tasks, fostering a closer connection between users and VAs.

2 RELATED WORKS

In this section, we discuss the literature on a) user behavior in online shopping with voice assistants, and b) approaches to understanding the interplay between VA voice persuasiveness and the tone, age, and gender of the VA voice in interpersonal and human-computer interaction (HCI) contexts.

2.1 Decision Making for Online Shopping with Voice Assistants

VA platforms offer decision-making assistance in online shopping through conversations [95, 124]. The conversations provide options including looking up suitable products and browsing customer reviews [74, 75]. However, the perceived risk of negative consequences [13, 69] and the lack of controllability of VAs challenge the wide adoption of online shopping with VAs. Rzepka et al. demonstrated that users find VAs enjoyable and convenient for general tasks, but are perceived as risky because of the perceived lack of reliability, opacity, and controllability of VAs in online shopping [107]. Muthukumaran and Vani posited that consumers are hesitant to use VAs for shopping because of the possible risk of payment malfunction [84]. Hong et al. showed that uncertainty with the technology causes user anxiety when online shopping with VAs [46]. Perceived risk in adoption of a new technology has been observed in internet banking [66] and similar high involvement contexts [128, 141].

However, increasingly human-like behavior by the VAs can ameliorate users' involvement in online shopping with VAs. Lee et al. showed that social presence of the VA helps increase the perceived usefulness of VAs in an online shopping context. As a result, users ultimately show more intention to use VAs for online shopping [67]. Dellaert suggested that human-like communication by VAs increases users' online shopping activity with VAs [26]. Clausen et al. found humorous responses by VAs effective in recovering mistakes [17]. These findings support the implication of the 'computers are social actors' (CASA) paradigm [86] in an online shopping context with VAs. The CASA paradigm posits that social cues in HCI

can potentially trigger similar social behaviors involving real-life scenarios. In this context, users react to human-like activities of VAs as they would react to a human being. For instance, Rhee and Choi demonstrated a positive effect of a VA with friendly language in a voice shopping context [103]. Zhu et al. showed that users liked voice chatbots when the chatbots displayed emotional expressiveness [145]. Similarly, McLean et al. demonstrated the importance of social presence, perceived intelligence, and social attraction in building user engagement in online shopping [80]. Lee et al. showed that perceived emotion toward VAs can strengthen users' mental involvement with the VA [65]. In contrast, there are opposing views on integrating human-like attributes into conversational agents in recommendation tasks. Starke and Lee argued that the quality of conversational aspects of a conversational recommender system might be slightly lowered if the recommender does an appropriate job in recommending a personalized item [120]. Moreover, Roesler et al. demonstrated that people prefer lower anthropomorphism in robots for specific contexts, such as industrial domains [105].

There are various scholarly perspectives on whether designing VAs with human-like characteristics may enhance user comfort and consequently increase the likelihood of users adhering to the VAs' recommendations. This diversity of viewpoints has sparked our interest in investigating the impact of vocal attributes on the persuasiveness of VAs, with a specific focus on determining whether such characteristics passively influence users' inclination to follow VA suggestions in the context of online shopping.

2.2 Voice Characteristics and Persuasiveness

Numerous studies have explored persuasiveness as a way to influence users' shopping outcomes in online shopping contexts [1, 78]. Persuasion is the process of influencing a person's attitude or behavior directly or indirectly by some form(s) of action [33, 90]. Peripheral cues including trustworthiness, attractiveness, and tone [38, 70, 76, 101, 104] can influence the outcome of persuasion in interpersonal interaction. For instance, Wirz et al. found that perceived emotion in a political speech persuades the audience [138]. Moreover, confidence expressed by a human voice was shown to effectively enhance persuasion among audience members [130].

A similar effect of vocal characteristics on user perception has also been observed in the human-computer interaction domain. Kessens et al. designed a robot to educate and motivate children and showed that emotional vocal and visual cues effectively persuaded the children [58]. Similarly, people perceive news stories as more interesting if the synthesized voice expresses them in a happier voice, as opposed to a sad tone [85]. Additionally, Belanche et al. demonstrated that integrating perceived warmth into a service robot's voice increased user expectations regarding the quality of the robot's service [7] and studies on interaction with synthesized voices found that users generally prefer female and extroverted VA voices [35]. Kim et al. found that urgency in a voice agent's tone increased user trust in emergency situations [59]. Dubiel et al. showed that people consider synthetic voices more truthful when they have a debating tone compared to a story-telling tone [32]. Torre et al. observed that including the sound of a smile in the recommendation of a voice agent elicited higher trust in an investment gaming context [129]. Papenmeier and Topp demonstrated

that meaningful acknowledgment from conversational agents increases the perceived competence of the agent [93].

For VAs, studies have largely explored how the persuasiveness and impression of voice agents can be improved with linguistic or visual expressions [31, 45, 50, 68, 71]. Linguistic improvisations including shorter and longer responses [42, 52], and informal and formal responses [54, 143] have demonstrated persuasiveness in diverse conversational agent contexts. Similarly, there is an extensive body of research on user perceptions of synthesized voice attributes including tone, age, and gender [4, 5, 108, 115]. However, there is a dearth of research specifically examining the interplay of tone, age, and gender in VAs' voices concerning their perceived persuasiveness. Our research addresses this gap, probing the associations between the persuasiveness of VAs and the tone, age, and gender of their voices, and subsequently how the persuasiveness leads users to consider VAs' suggestions. Previous studies have explored the impact of positive and negative words uttered by the user to increase the effectiveness of conversational interfaces' recommendations for online fashion choice [139]. We have taken a similar approach, but for the tone of the VA voice, selecting a positive and negative tone of the VA voice.

Overall, the vocal attributes of tone, age, and gender collectively contribute to how a speaker is perceived [110, 142]. Hence, these attributes can play a pivotal role in shaping users' perceptions of the persuasiveness of VAs. Therefore, our study emphasizes select vocal characteristics—tone, age, and gender—to examine the perceived persuasiveness of voice assistants.

3 METHOD

First, we generated voice stimuli with varying tones, age groups, and genders. Next, we validated the voice stimuli and measured the user behavior for varying voices of the VA. In the following subsections, we discuss the components of the study.

3.1 Stimuli Generation

We used Microsoft speech studio¹ to generate voices that varied by tone, age group, and gender following the age group and gender criteria outlined by Waller et al. [115]. We selected female and male voices and age groups including younger adults (20–30 years old), middle-aged adults (40–50 years old), and older adults (60–70 years old). We applied positive, neutral, and negative tones to each voice category. In Microsoft Speech Studio, we used a few template voices with emotional tones. To avoid complexity, we categorized these into three tones—positive tones including happy, excited, and cheerful; neutral, reflecting the default or normal and flat tone; and negative tones encompassing sad and frustrated tones. In addition, we varied the pitch and speech rate (Table 5) of the voices to create different age group impressions, based on previous research showing that listeners perceive higher-pitched voices as younger and lower-pitched voices as older [137]. We edited the recordings in Audacity² to normalize for intensity [115].

We used positive and negative product reviews from the popular online shopping platform Amazon³ and chose products from

¹<https://speech.microsoft.com/>

²<https://www.audacityteam.org/>

³<https://www.amazon.com/>

the most popular product categories purchased in 2023: Home & Kitchen, Office Supplies, and Bags & Luggage [18, 55, 63]. We selected gender-neutral products to avoid gender bias toward a product, including categories such as tumblers, desks, chairs, etc. Positive product reviews were generated with either positive or neutral tones and negative product reviews were generated with either negative or neutral tones. We excluded negative tone for positive reviews and positive tone for negative reviews, as the cross-valence effect was beyond the scope of our research. Overall, we generated 24 combinations of voices (two genders x three age groups x two review valences x two tone valences for each review valence). The reviews contained 26 to 33 words, aligned to the text length of the speech materials by Waller et al. [115]. Thereby, the corresponding audio clips were between 12 and 16 seconds.

Following several iterations of voice synthesis, we preselected a group of voices that appeared to be appropriate for the corresponding target age groups and tones. Subsequently, we conducted a stimulus validation study to ensure that the voices were perceived as intended in terms of age group and tone.

3.2 Stimuli Validation

A preliminary study was conducted to identify the perceived tone and age group of the voices ($N = 78$). We adapted the voice perception test for assessing the perceived age and gender of synthesized voices developed by Baird et al. [4, 5].

We conducted several pilot studies and found that the participants faced difficulty in detecting the age of the older adult voices. Therefore, we removed the older adult voices from the stimulus set. In the final stimuli validation study, we used 48 audio clips containing 16 different combinations of the voices (two genders x two age groups x two review valences x two tones for each review valence) and each combination had three different voices.

Each participant was presented with 14 audio clips including two attention check audio clips. We paid all the participants regardless of their attention check scores, but we used only the responses of the participants with zero attention check failure. We randomly presented only male or female voices to a participant so that the participants did not have a direct bias toward a particular gender, with 38 listening to male voices and 41 to female voices. For all audio clips, the participants responded to questions about their perception of the tone of the voice (seven-point Likert scale: strongly negative = -3 to strongly positive = 3) [3] and age-group of the voice (younger adult 20–30, middle-aged adult 40–50). The participants were able to play the audio multiple times and could stop, pause, and resume the audio. We included a test audio to let the participants adjust the volume of the audio on their devices. We calculated the average ratings across the participants for each voice. Each voice was then categorized into negative, neutral, or positive tones based on the tone ratings, and younger adults or middle-aged adults based on the frequency of age-group selection. We paid \$3.50 per participant according to the minimum wage guidance in the study location [114]. Participant demographics are detailed in Appendix Table 6.

3.3 User Behavior Study

3.3.1 Stimulus and Procedure. This experiment investigated user perceptions of varying vocal characteristics of a voice assistant in the context of online shopping and was conducted using the online survey platform Prolific⁴ in line with similar studies [47, 126]. Participants ($N = 335$) were given a scenario to imagine themselves searching for a certain product and checking product reviews using a voice assistant. Each participant was presented with eight audio clips including two attention check clips. To avoid an idiosyncratic effect of a specific product and enhance the generalizability of our findings beyond a single product [39], we used stimulus sampling of the reviews from six different household products such as dishwasher soap or hand soap, for example.

We conducted the study with a 2 (review valence: positive vs. negative) x 2 (voice assistant gender: female vs. male) x 2 (voice assistant age: younger adult vs. middle-aged adult) x 3 (voice assistant tone: positive/negative vs. neutral) between-subjects factorial design. Hence, one participant encountered only one voice type. The between-factor design for all variables was done to avoid any carry-over effect. In total, there were 96 audio clips for six products, with 16 clips each (eight positive and eight negative reviews for each product). After adjusting the device volume in an audio test, participants listened to the audio reviews one at a time. These reviews contained 26 to 35 words and the audio clips ranged from 12 seconds to 17 seconds. For each audio clip, we measured participants' purchase likelihood of the corresponding product (*I am likely to purchase this product* [135]).

The following questions focused on participants' perceptions of the simulated VA and the VA voice. We measured participants' perceived persuasiveness of the voice (Table 1) alongside the perceived trustworthiness, usefulness, and enjoyment of the VA. The participants responded to these questions only once during the study.

Factor	Item	Chron.
Persuasiveness [98]	The voice is compelling	0.89
	The voice is convincing	

Table 1: Question Items about Voice Persuasiveness. Chron. = Chronbach's alpha

Next, we presented the participants with two qualitative open-ended questions:

- What do you like/dislike about the voice you heard during the survey? Please elaborate.
- If you could create a voice for the voice assistant, what would that voice sound like? Please elaborate.

The remaining questions focused on VA usage, disposition of trust, and demographics (Appendix Tables 4).

3.3.2 Participants. We adopted the a priori G*Power analysis [36] by Ischen et al. [50] for a between-subjects one-way analysis of variance (ANOVA) with four groups and found a required sample size of 320. We invited a total of 350 participants in Prolific and the responses of 335 without attention check failure were used for

⁴<https://app.prolific.co/>

further analyses. More than 94% participants had experience using a voice assistant (Appendix Table 6). In this study, 75% of the participants completed the survey in 17 minutes or less and were paid \$5 to conform with the minimum wage recommendation in the study location [114]. Of the 335 eligible participants, 168 were presented with female voices and 167 listened to male voices. Furthermore, 159 participants listened to middle-aged adult voices and 176 heard younger adult voices. In terms of tones, voices considered positive, negative, and neutral voices were distributed to 80, 75, and 180 participants respectively.

3.4 Quantitative Analysis

We used the causal mediation analysis framework [6, 125] in R using the package ‘lavaan’ [106] to estimate the direct effect of the perceived tone, age, and gender of the VA voice on the purchase decision of the participants and the indirect effect through the perceived persuasiveness of the VA voice. We chose the best mediation model based on the comparative fit index (CFI) [9] and significant model test statistics. In addition, we conducted linear regression to evaluate the relationship between VA voice persuasiveness and VA vocal characteristics, i.e., tone, age, and gender of the VA voice. The data we collected met the assumptions of homoscedasticity and normality of the residuals.

The purchase likelihood score was reversed into ‘purchase decision’ for negative reviews to represent the influence of the VA in the purchase decision variable. We treated VA voice persuasiveness and purchase decision as dependent variables. We had four between-subject independent variables (gender, age group, tone of voice, and review valence); the independent variables were encoded as categorical variables. In all models, the participant ID was encoded as a random categorical variable. For the dependent variables with multiple items in the dataset, we used Chronbach’s alpha [20] to measure the internal consistency among the items in a group and we took the mean of the items to represent the corresponding dependent variable. In addition, we used Chronbach’s alpha to measure the internal consistency among the responses for the six products (Cronbach’s alpha = 0.96). We used estimated marginal means to compute the pairwise comparisons following the regression analysis to help make sense of the significant effects across the interaction effects. We adjusted the p-values using the Tukey method [136] as it considers multiple comparisons and adjusts the p-value to minimize the risk of Type I errors.

3.5 Qualitative Analysis

We used an inductive (thematic) analysis process for analyzing the open-ended questions about liking or disliking the voices and participants’ preferences for the VA voice. We chose to perform thematic analysis because we did not derive the codes from any theory or prior research and our purpose was not to build new theories in the work. We used Delve⁵ for coding the responses and identifying the themes. Initially, the researchers developed a code book for codes in an iterative bottom-up approach by weekly meetings. Then, based on the code book, we coded the rest of the responses. From relevant codes, we came up with the main themes. We did not compute inter-rater reliability as we conducted thematic

analysis on multiple iterations and the themes were developed through refinement of the codes [79].

3.6 Hypotheses

Significant research efforts have presented that vocal characteristics of human [104, 138] or robot [58, 130] influence listeners’ perceived persuasiveness. Moreover, studies have found perceived persuasiveness to be a significant factor in consumers’ purchase decisions in online shopping [113]. Vocal characteristics refer to the vocal elements beyond just the words we say, including the tone, personality, gender, and age of the speaker [76, 101]. Therefore, we examined the effect of perceived tone, gender, and age of VA voice on the perceived persuasiveness of a VA and how the persuasiveness ultimately impacts participants’ purchase decisions considering the VA’s recommendation. We present the following hypotheses based on our research questions.

H1: The positive and negative tones of a voice assistant’s voice persuade participants more compared to the neutral tone and subsequently, the persuasiveness affects participants’ purchase decisions.

H2: Varying interactions between the perceived age and gender of a voice assistant’s voice persuade participants differently and subsequently, the persuasiveness affects participants’ purchase decisions.

4 FINDINGS

We used mediation analysis to estimate how perceived tone, age, and gender of the VA voice persuade participants, and subsequently, how the persuasiveness affects participants’ purchase decisions. In the following sections, we discuss our findings based on *H1* and *H2*.

4.1 Effect of Perceived VA tone on VA Voice Persuasiveness and Purchase Decision

Table 2 shows that the indirect effect of VA tone on purchase decision was significant, indicating a mediating role of VA voice persuasiveness. This result means that VA tone significantly affected (large effect size) the VA voice persuasiveness, and VA voice persuasiveness in turn significantly affected participants’ purchase decisions (Figure 1). A linear regression between VA vocal tone and VA voice persuasiveness revealed that positive (large effect) and neutral tones (medium effect) persuaded participants much more than VA voices with negative tones (Fig. 2a). These results provided partial support for *H1*, where positive and neutral tones had significant effects on persuasiveness and persuasiveness affected users’ purchase decisions.

In summary, our findings suggest that perceived positive and neutral VA tones were significantly more persuasive than negative VA tones, and VA voice persuasiveness significantly affected participants’ purchase decisions. In simple terms, neutral VA tones were significantly more persuasive than negative VA tones for negative reviews and ultimately reduced users’ purchase likelihood. Concurrently, the positive and neutral VA tones had similar persuasiveness for positive reviews.

⁵<https://delvetool.com/>

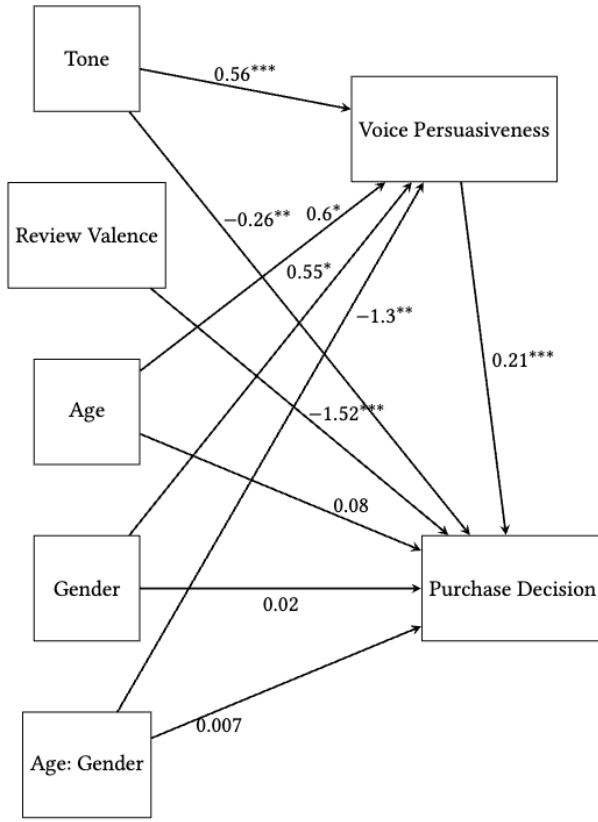


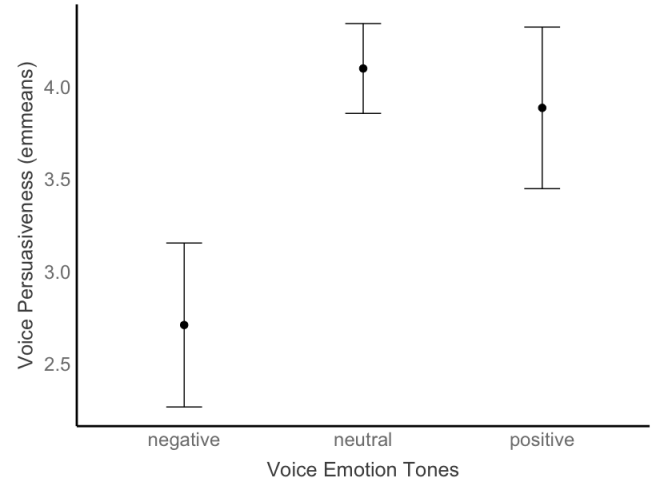
Figure 1: Path model diagram for the mediation of voice persuasiveness in the relationship between purchase decision and the tone, review valence, age, and gender of the VA voice. (* = $p < 0.05$, ** = $p < 0.01$, * = $p < 0.001$). Baselines for age = younger adult, gender = male**

	Estimate	Std. Error
Tone	0.115**	0.034
Gender	0.11*	0.05
Age	0.12*	0.05
Age: Gender	-0.27**	0.09

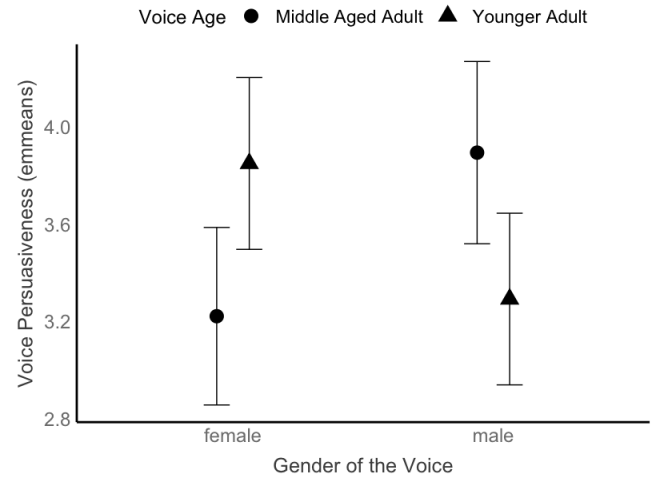
Table 2: Indirect effect of VA tone and the interaction between VA age and VA gender in the proposed mediation model (* = $p < 0.05$, ** = $p < 0.01$, * = $p < 0.001$)**

4.2 Effect of Perceived Gender and Age of VA Voice on VA Voice Persuasiveness and Purchase Decision

We found a significant effect in the interaction of VA voice's perceived age and gender on VA voice persuasiveness (Figure 1). However, we did not find a significant direct effect of the age: gender interaction on participants' purchase decisions. It indicates that there was a full mediation of the voice's persuasiveness between



(a) Effects of VA tone



(b) Interaction effects between voice age and gender of the voice

Figure 2: Differences in the perceived persuasiveness of voice (estimated marginal means)

Voice Age	Voice Gender	Est.	Effect Size
Middle Aged	Female < Male	0.67*	0.4 (small)
Middle Aged < Younger	Female	0.63*	0.38 (small)

Table 3: Pairwise comparison of Voice Persuasiveness for the interaction between the gender and age of the voice. Significance after Tukey adjustment is indicated with p-values. Effect size indicates Cohen's d . For voice age group, 'Middle Aged' is the shorter form for 'Middle Aged Adult' and 'Younger' is the shorter form of 'Younger Adult'.

the age: gender interaction and purchase decision of the participants, meaning that age: gender interaction indirectly affected purchase decision through VA voice's persuasiveness. Furthermore, we performed a linear regression between the perceived gender and age of the VA voice and perceived VA voice persuasiveness to

determine which combinations of age and gender influenced the VA voice's persuasiveness. We also conducted a post hoc pairwise comparison with Tukey adjustments (Table 3). Our findings suggest that participants perceived middle-aged male and younger female voices as more persuasive than younger male voices or middle-aged female voices (small effect size). The effects of the age and gender of the voice are visualized in Figure 2. These results support H2 that varying age: gender interactions have various effects on VA persuasiveness, and consequently, the persuasiveness affected participants' purchase decisions.

Overall analysis suggests that participants found middle-aged male and younger female voices to be more persuasive compared to the other voice types, and the perceived persuasiveness led them to accept the VA recommendation in their purchase decision.

4.3 Reasons for Voice Characteristics Preference

To better understand why our participants preferred certain vocal characteristics, we coded the open-ended responses for additional insights. This helped us comprehend why the participants liked certain voice tones and combinations of ages and genders and were likely to accept suggestions from corresponding VAs. Moreover, this analysis yielded crucial design implications for increased user satisfaction and a user-friendly online shopping interface.

4.3.1 Positive tone conveyed authenticity. A group of participants' reflections about the voice revealed that a positive VA voice created a sense of authenticity.

"The voice felt and sounded authentic and spoke pretty naturally" (P36)

The genuine feeling came from the comfortable atmosphere caused by the voice assistant's friendly tone. The positive tones of the voices made the listener feel at ease and welcomed.

"The voice assistant had a friendly and inviting quality that made me feel comfortable." (P157)

In a follow-up question, we asked the participants about their personal VA voice preferences. Some participants stated that enthusiastic voices portrayed sincerity in helping out with decision-making. Sometimes participants thought of VAs as sincere salespersons. When the VA sounded keen to help, participants felt that the VA was sincere. Furthermore, multiple participants expressed that flat voices were boring and took away the positive experience from the conversation.

"I want my voice assistant to be high energy and sound like they were happy to help me instead of (displaying) disinterest." (P290)

"The (neutral) voice was very flat and monotone. It was boring to listen to at times." (P89)

4.3.2 Neutral tone was reassuring and safe. In some cases, participants showed a fondness for neutral voices because of the calmness portrayed by the neutrality of the voice. The flat nature of the tone assured the participants that the VA was not biased in stating the recommendation. Furthermore, a neutral tone created a sense of objectivity which helped participants rely on the recommendation.

"I like that the voice was neutral sounding; it did not give off any type of extreme emotion, which helped it seem unbiased while it was giving the review of the products." (P87)

"I liked that (neutral voice) because I felt like it presented information in an objective manner." (P148)

"I liked the (neutral) voice it wasn't too much or too loud, I liked how calm the voice was and it was reassuring." (P191)

4.3.3 Positive or negative tone can be overwhelming. Some participants expressed their dislike for both positive and negative tones, mentioning that the emotional intensity in these tones can be overwhelming. They preferred calmness because an overenthusiastic voice sounded suspicious (P90). Moreover, some participants suggested that they did not like the negative tone as the tone made it too distracting to focus on the reviews (P146).

"The (enthusiastic) voice was suspicious that it's selling something" (P90)

"The voice with a negative tone sounded uninterested or almost sad. It made it hard to concentrate on the recommendations." (P146)

"The voice with negative tone seemed pessimistic towards the majority of things." (P19)

Overall, a large number of participants indicated their dislike of positive and negative tones as they wanted to focus more on the content of the reviews and did not want to be tricked. These participants preferred the content and objectivity of the reviews over entertainment.

4.3.4 Middle-aged male voices sounded knowledgeable and experienced. A group of participants seemed to trust middle-aged male voices because they felt the voices portrayed confidence, knowledgeability, and reassurance. The confidence portrayed by the low steady pitch of the middle-aged adult voices created a sense of experience as well. As a result, participants perceived the reviews as reliable.

"The voice seemed confident and as if they knew what they were talking about." (P59)

"It sounded like he had been doing this for a while." (P111)

4.3.5 Younger female voices were soothing. A few participants favored younger female voices due to their perceived soothing and appealing quality. The softness in the younger female voice created a soothing ambiance for the participants that triggered a favorable response toward younger female VA voices.

"I liked how the voice was soft and very easy to understand everything that was being said." (P24)

"I feel that overall a female voice is more trustworthy and approachable. Her (VA) tone was pleasant and calm" (P278)

4.3.6 Preference toward familiar or celebrity voice. Some participants expressed an interesting intention to use celebrity or familiar voices as VA voices. They thought the familiarity of the voice would create a comfortable and safe ambiance and they would

enjoy shopping with a VA having a familiar voice. A few suggestions included ‘radio DJ from 1989’ and ‘gaming voices’.

“I would create a voice like Ellen DeGeneres, I love everything about her but most of all I love her voice.” (P126)

“I would use actor voices like Ellen McLain, Morgan Freeman, people that have smooth voices I suppose I like older voices they feel warmer and trustworthy.” (P10)

4.3.7 Customizable and flexible voice control. While articulating their preferences and attitudes toward the voice characteristics of VAs, several participants pointed out that they would have liked different VA voices in varying contexts. Some participants suggested the incorporation of multiple voice characteristics and the flexibility to switch between these voices.

“I would make a voice pack with different emotions and voices that a user could change depending on their mood.” (P20)

“I think she (the preferred VA) would change her tone depending on the product she was discussing.” (P126)

5 DISCUSSION

We now discuss the implications of our findings and the limitations of the study.

5.1 Positive and Neutral Tones are More Persuasive and Influence Purchase Decisions

Our research indicates that both positive and neutral tones in VA voices were perceived as more persuasive than negative tones. Furthermore, this perceived persuasiveness substantially increased the likelihood of participants following the VA recommendation. For instance, participants were persuaded by the VA and were more likely to purchase a product when the VA provided a positive recommendation with positive tones. On the other hand, participants were more persuaded by a neutral tone for negative reviews, and as a result, they were more likely to refrain from purchasing the corresponding product. The positive effect of positive and neutral tone on voice persuasiveness is also evident from the open-ended responses by the participants. Several participants reported that they preferred a positive tone due to the feelings of ‘comfort’ it generated, creating a sense of ‘authenticity and sincerity.’ Likewise, participants found the neutral tone objective and calm. The sense of calmness was reportedly helpful as the participants did not feel forced into making a specific decision, which in turn may have helped the voice appear more persuasive. This outcome can be explained by prior research supporting calm language to be persuasive at a linguistic level [122]. Furthermore, participants indicated an aversion toward the negative tones of VAs, indicating that the negative tones induced negative feelings such as pessimism. This can be explained by prior studies where negative facial expressions and negative activities in a dialogue provoked anxiety in speakers and a neutral or positive attitude from an audience created a sense of comfort [96].

The elaboration likelihood model (ELM) [97] of persuasion states superficial cues such as vocal tone or emotional appeal influence persuasion when people are not motivated to process a message

from a communicator deeply. Our model presents evidence that VA tone influenced persuasion, which may indicate participants’ lack of motivation or ability to comprehend the content of the reviews. The effect of vocal tone overshadowing the review content might be explained by Delin’s argument that the conversational linguistic tone of reviews can make a brand feel ‘more socially close’ to the consumer and ultimately create an attachment with ‘warm’ and ‘approachable’ texts [25]. It is evident that, in the era of VAs’ humanlike communication capability, vocal cues may effectively influence users’ decision-making.

5.2 Middle-Aged Male and Younger Female Voices Are More Persuasive and Influence Purchase Decisions

Our findings indicated that participants attributed higher levels of persuasiveness to the voices of middle-aged males and younger females when compared to other voice types. Because of this increased persuasiveness, they were more likely to follow the VA recommendation for or against product purchase. This result is consistent with participants’ open-ended responses on their preference for the VA voice. Participants perceived middle-aged male voices as confident, knowledgeable, and experienced because of their low pitch and calmness. Our findings are consistent with previous research showing that people with lower-pitched voices are more persuasive in various contexts. For example, studies have found that older male voices have a lower pitch than the voices of other age groups or genders. Subsequently, people with lower-pitched voices are more likely to be elected to leadership positions [61], receive donations [77], and receive higher ratings of persuasiveness [83]. In contrast, some participants indicated female voices as being more friendly, happier, and soothing. This result is also consistent with the findings of prior studies in which female voices in synthesized form were perceived as more attractive and soothing to participants [35, 53].

5.3 Designing Acceptable and Useful Voice Assistants

Our findings indicate that participants had varying preferences for vocal characteristics of VAs. This suggests that VA design cannot follow a ‘one size fits all’ approach. Some participants were persuaded by a more upbeat voice that conveyed a sense of fun and excitement. One group of participants was persuaded by middle-aged male voices which they perceived as experienced and knowledgeable, while another group was persuaded by a younger female voice which they perceived as soothing and youthful. The efficacy of humanlike voice is aligned with the ‘computers are social actors’ (CASA) paradigm [86]. The CASA paradigm states that people often apply social heuristics in their interactions with machines and humanlike activities of a machine may lead to a higher level of engagement. In our study, we also found that some participants find VAs with humanlike voices more engaging and authentic. In contrast, some participants were persuaded by a neutral and flat tone that emphasized factual information over emotional appeals. One underlying reason for this preference might align with the research findings that users sometimes find personalized VA attitudes intrusive and a threat to their privacy [28]. Furthermore,

a few participants flagged varying tones as a ‘distraction’ while listening to the reviews. This preference for a neutral tone can be explained by the central route to persuasion by the elaboration likelihood model (ELM) [97]. The central route involves a higher level of cognitive processing: keeping the focus on the evidence and logical arguments.

This grouping in preferences might result from participants’ inherent motivation. Prior studies have suggested that users motivated by social interaction are likely to perceive a VA as socially attractive and as a friend, whereas users motivated by efficiency are likely to perceive it as an assistant [15]. Differences in user motivation underscore the need to customize VA characteristics. VA designs can align the personality of the VA with that of the user to increase motivation. Several studies in both human-computer interaction (HCI) and psychological fields revealed a significant increase in likability and trustworthiness when a voice matches the user’s personality [11, 21, 22, 116]. This aligns with the *similarity-attraction* theory: people are more likely to be attracted to those who are like them [12]. Another avenue for VA personalization lies in enhancing accessibility for marginalized and vulnerable populations. For example, older adults may favor the voices of middle-aged adults because they are generally easier to understand and they speak at a slower rate than younger adults [137].

In addition to personalized voices, many participants favored VAs with the capability to change vocal expression based on a situation or a function. Desai and Twidale demonstrated that users might perceive virtual assistants (VAs) as different personas depending on the context, using various metaphors to interact with VAs accordingly [27]. Adapting VA tones to the situation can enhance the user’s impression. The preference for situation-aware voices is aligned with previous findings that users prefer personalized conversation by chatbots based on varying functions [131]. Such adaptability could be beneficial in specific scenarios. For example, the use of a warm vocal tone can be an effective way to mitigate user disappointment when a VA fails to communicate with the user as expected [47]. Furthermore, the ability of a VA to reflect multiple personalities can evoke the social presence heuristic [64] which suggests that users are more likely to believe and trust a VA [60]. In addition, the ability to easily change VA tone based on the mood might be a delightful experience for the users.

The current versions of voice assistants such as Alexa, Google Nest, or Siri provide limited variation in the voices, encompassing only gender or accent-based variations [102, 134]. Our study indicates that introducing variations in tone or age for VA voices could yield a more personalized and engaging user experience in VA usage, particularly enhancing trust and comfort during complex tasks such as online shopping. Furthermore, as technology advances, there is a potential for VAs to create completely realistic and credible speech, incorporating natural intonations and emotions according to appropriate context [88, 92]. These enhancements of VA voices can facilitate more VA engagement, ultimately increasing VA adoption and providing users with a more seamless and intuitive experience, particularly beneficial for disabled and older adult individuals seeking hands-free alternatives. However, alongside the advancement of VA voices, it is imperative to consider the ethical ramifications of creating AI voices indistinguishable from human voices to prevent their potential misuse for deception or manipulation of users.

Therefore, future studies should not only focus on advancing the technology to make VA voices humanlike and context-aware but also ethically evaluate the consequences and potential misuse of such advancements and formulate intelligent policies to mitigate the risk of manipulative VA voices.

5.4 Influence of Cultural and Linguistic Backgrounds

Currently, voice assistants speak multiple languages for users from different linguistic backgrounds [117]. Studies have shown the effect of culturally specific principles in voice emotion recognition—affecting the perception of the valence of vocal tones in different languages [94, 140]. Our study evaluated American participants’ perceived tones of the VAs. Hence, the findings of this study should be interpreted carefully considering the cultural background of the participants, as the perception of tone or age in a VA voice may vary in different cultural and linguistic contexts. For example, research has shown that Japanese people pay more attention to the emotional tones in voices compared to Dutch people [123]. Also, there are differences in how native Mandarin and native American speakers understand emotions in voice tones [14]. Furthermore, users from Japanese, Mandarin, and Brazilian Portuguese backgrounds were found to have different perceptions of the positivity and negativity of the vocal tones in the same linguistic contents [34]. Therefore, future research should consider the cultural backgrounds of VA users and compare VAs with different languages to understand how people from different backgrounds comprehend the tone or age of the voice.

5.5 Ethical Consideration and Healthy Persuasion

Although our findings suggest the effectiveness of VA vocal characteristics to influence users’ purchase decisions, safeguarding user interest is an important aspect of VA designs. Ethical considerations, including the permissible degree of persuasive tactics and mechanisms to ensure informed consumer decisions, are critical. Moreover, it is crucial to distinguish between helping users make good decisions (healthy persuasion) and deceiving them (manipulation) [44, 87]. This is especially true for special interest groups susceptible to manipulative practices, including older adults or individuals with visual impairments [57]. Tailoring personalized VA voices fosters inclusivity and empowers people with visual impairment or older adults to utilize hands-free technologies with trust and confidence. However, these vulnerable demographics have limited resources to verify the source of new information, often relying on other people for assistance [2]. In addition, potential cognitive or physical limitations may limit their capability to discern manipulative tactics, making them a potential target for exploitation through manipulation [111]. One strategy to mitigate the risk of manipulation is to provide comprehensible disclaimers and offer adequate training. Moreover, utilizing automated mechanisms to identify voice-based manipulation could be valuable in preventing deceptive practices via VAs. Furthermore, research found that a subset of older adults prefers robots to exhibit robotic characteristics rather than human-like qualities [10], aligning with our study’s findings that some participants favor robotic voices in VAs

as a reminder of the machine nature of the technology. Therefore, VA designs should prioritize user preferences, particularly those of vulnerable populations, before implementing alterations to VA voices [91].

We also need to be careful how we interpret the findings related to preference for a particular gendered voice. Previous research has observed a tendency toward gender stereotypes in preferences for synthesized voices [29]. However, the result of this study may not be applicable in different contexts or with different voice stimuli. Therefore, it is crucial to ensure that voice assistant design does not inadvertently perpetuate gender stereotypes. Rather, the VA design should focus on enhancing user comfort and provide gendered or gender-ambiguous voice options [127] as a personalization alternative. Moreover, letting users easily change the VA voice can protect their interests. This way, users can choose voices that make them feel comfortable and enjoy shopping more. By integrating brief disclaimers and easy customization, VA designs respect users' choices and make sure they have a good experience without feeling pressured.

5.6 Limitation and Future Work

There are a few limitations in this study. First, the voice dataset we used consisted of only male and female voices, as there are limited resources on generating gender-ambiguous [121] voices with the desired tones and age groups. Significant efforts have been made toward making VAs more inclusive, with gender-ambiguous text-to-speech (TTS) systems that can produce non-binary voices [23]. However, a considerable proportion of users are yet to adopt the concept of a gender-ambiguous voice [82]. Future work should incorporate variations in the vocal tone and age of gender-ambiguous voices to yield more precise results free of gender stereotypical perceptions.

Furthermore, we conducted the study in an online platform with synthesized voices to minimize variability in the simulated scenarios. However, this study design did not allow researchers to control environmental factors such as the volume of the voices or the participants' surroundings. Moreover, this approach did not provide participants with an authentic VA interaction experience involving a physical VA and the inherent motivation of users to engage in product shopping. Although we curated the most popular gender-neutral products suitable for online shopping with voice-assisted reviews, ensuring minimal visual inspection, the limited selection of predefined product categories may not fully capture responses across diverse product types. Therefore, there is a possibility that participants' responses may differ in a real-world VA interaction setting. Future work should evaluate the feasibility of conducting VA experiments in a controlled laboratory setting with a physical VA and a broader array of product categories.

Moreover, we did not explore the distinction between persuasion and manipulation, despite their correlation and overlap in scope. Similar research on persuasiveness should also consider the effects of manipulation.

6 CONCLUSION

As voice assistants (VAs) become more capable of assisting people in complex tasks, it is essential to identify the factors that enhance

their persuasiveness and the strategies through which VAs can engender user trust, thereby enhancing reliance on VA-provided recommendations. Our findings suggest that voice characteristics play an important role in shaping user decisions in online shopping with VAs. Participants found neutral and positive-toned simulated voices more persuasive than negative-toned voices and tended to follow positive and neutral-toned recommendations more often. Additionally, middle-aged male and younger female voices were perceived as more persuasive, prompting users to more frequently follow recommendations delivered in these vocal tones. Our findings suggest that varying voice characteristics can create a sense of comfort and authenticity among users, improving user adoption of VAs in tasks with financial consequences such as online shopping. However, ethical considerations, such as implementing disclaimers regarding the VA's vocal tone or measures to prevent the misuse of vocal characteristics for deceptive purposes, need to be reinforced in VA design to differentiate legitimate persuasion from manipulation. Moreover, it is imperative to exercise caution before generalizing findings related to gender stereotypical perceptions to avoid oversimplification and further amplification of gender biases.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation under grant CNS-2207019 and the Social Science Research Funding Program (SSRFP) at Indiana University. We would also like to thank Hannah Bolte and Elizabeth Ray at the Indiana Statistical Council Center (ISCC) for their guidance on statistics analysis.

REFERENCES

- [1] Ifeoma Adaji, Kiemute Oyibo, and Julita Vassileva. 2020. E-Commerce Shopping Motivation and the Influence of Persuasive Strategies. *Frontiers in Artificial Intelligence* 3 (2020), 1–13. <https://www.frontiersin.org/articles/10.3389/frai.2020.00067>
- [2] Taslima Akter, Bryan Dosono, Tousif Ahmed, Apu Kapadia, and Bryan Semaan. 2020. "I am uncomfortable sharing what I can't see": Privacy Concerns of the Visually Impaired with Camera Based Assistive Applications. In *29th USENIX Security Symposium (USENIX Security 20)*. 29th USENIX Security Symposium (USENIX Security 20), Santa Clara, CA, 1929–1948. <https://www.usenix.org/conference/usenixsecurity20/presentation/akter>
- [3] Mary Jean Amon, Rakibul Hasan, Kurt Hugenberg, Bennett I Bertenthal, and Apu Kapadia. 2020. Influencing photo sharing decisions on social media: A case of paradoxical findings. In *2020 IEEE Symposium on Security and Privacy (SP)*. IEEE, IEEE, Virtual, 1350–1366.
- [4] Alice Baird, Stina Hasse Jørgensen, Emilia Parada-Cabaleiro, Nicholas Cummins, Simone Hantke, and Björn Schuller. 2018. The perception of vocal traits in synthesized voices: Age, gender, and human likeness. *Journal of the Audio Engineering Society* 66, 4 (2018), 277–285.
- [5] Alice Baird, Stina Hasse Jørgensen, Emilia Parada-Cabaleiro, Simone Hantke, Nicholas Cummins, and Björn Schuller. 2017. Perception of paralinguistic traits in synthesized voices. In *Proceedings of the 12th International Audio Mostly Conference on Augmented and Participatory Sound and Music Experiences*. ACM, London United Kingdom, 1–5.
- [6] Reuben M. Baron and David A. Kenny. 1986. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology* 51, 6 (1986), 1173–1182. <https://doi.org/10.1037/0022-3514.51.6.1173> Place: US Publisher: American Psychological Association.
- [7] Daniel Belanche, Luis V Casaló, Jeroen Schepers, and Carlos Flavián. 2021. Examining the effects of robots' physical appearance, warmth, and competence in frontline services: The Humanness-Value-Loyalty model. *Psychology & Marketing* 38, 12 (2021), 2357–2376.
- [8] Pascal Belin, Bibi Boehme, and Phil McAleer. 2019. The sound of trustworthiness: Acoustic-based modulation of perceived voice personality (vol 12, e0185651, 2017). *PLOS ONE* 14, 1 (2019).

- [9] P. M. Bentler. 1990. Comparative fit indexes in structural models. *Psychological Bulletin* 107, 2 (1990), 238–246. <https://doi.org/10.1037/0033-2909.107.2.238> Place: US Publisher: American Psychological Association.
- [10] Clara Berridge, Yuanjin Zhou, Julie M. Robillard, and Jeffrey Kaye. 2023. Companion robots to mitigate loneliness among older adults: Perceptions of benefit and possible deception. *Frontiers in Psychology* 14 (Feb. 2023). <https://doi.org/10.3389/fpsyg.2023.1106633> Publisher: Frontiers.
- [11] Michael Braun, Anja Mainz, Ronée Chadowitz, Bastian Pfleging, and Florian Alt. 2019. At your service: Designing voice assistant personalities to improve automotive user interfaces. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–11.
- [12] Donn Byrne, William Griffitt, and Daniel Stefaniak. 1967. Attraction and similarity of personality characteristics. *Journal of Personality and Social Psychology* 5, 1 (1967), 82.
- [13] Margaret C. Campbell and Ronald C. Goodstein. 2001. The Moderating Effect of Perceived Risk on Consumers' Evaluations of Product Incongruity: Preference for the Norm. *Journal of Consumer Research* 28, 3 (Dec. 2001), 439–449. <https://doi.org/10.1086/323731>
- [14] Peiyao Chen, Ashley Chung-Fat-Yim, Taomei Guo, and Viorica Marian. 2023. Cultural background and input familiarity influence multisensory emotion perception. *Cultural Diversity and Ethnic Minority Psychology* -, - (2023). <https://doi.org/10.1037/cdp0000577> Place: US Publisher: Educational Publishing Foundation.
- [15] Tae Rang Choi and Minette E Drumwright. 2021. “OK, Google, why do I use you?” Motivations, post-consumption evaluations, and perceptions of voice AI assistants. *Telematics and Informatics* 62 (2021), 101628.
- [16] Jason Cipriani. 2023. Siri's 4 new voices: Here's where to find them and how to make the switch. CNET. <https://www.cnet.com/tech/mobile/siri-4-new-voices-heres-where-to-find-them-and-how-to-make-the-switch/>
- [17] Mikkel Clausen, Mikkel Peter Kyhn, Eleftherios Papachristos, and Timothy Merritt. 2023. Exploring Humor as a Repair Strategy During Communication Breakdowns with Voice Assistants. In *Proceedings of the 5th International Conference on Conversational User Interfaces (CUI '23)*. Association for Computing Machinery, New York, NY, USA, 1–9. <https://doi.org/10.1145/3571884.3597127>
- [18] Brian Connolly. 2023. Top Amazon Product Categories in 2023. Jungle Scout. Retrieved June 28, 2023 from <https://www.junglescout.com/blog/amazon-product-categories/>
- [19] Apple Correspondent. 2022. Apple's Siri. <https://www.apple.com/siri/>. Accessed: 2022-02-10.
- [20] Lee J Cronbach. 1951. Coefficient alpha and the internal structure of tests. *psychometrika* 16, 3 (1951), 297–334.
- [21] Nils Dahlbäck, Seema Swamy, Clifford Nass, Fredrik Arvidsson, and Jörgen Skågeby. 2001. Spoken interaction with computers in a native or non-native language—same or different. In *Proceedings of INTERACT*. Citeseer, Springer Berlin, Heidelberg, Berlin, Heidelberg, 294–301.
- [22] Nils Dahlbäck, QianYing Wang, Clifford Nass, and Jenny Alwin. 2007. Similarity is more important than expertise: Accent effects in speech interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 1553–1556.
- [23] Andreea Danielescu, Sharon A Horowitz-Hendler, Alexandria Pabst, Kenneth Michael Stewart, Eric M Gallo, and Matthew Peter Aylett. 2023. Creating inclusive voices for the 21st century: A non-binary text-to-speech for conversational assistants. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–17.
- [24] Celso M De Melo, Kangsoo Kim, Nahal Norouzi, Gerd Bruder, and Gregory Welch. 2020. Reducing cognitive load and improving warfighter problem solving with intelligent virtual assistants. *Frontiers in psychology* 11 (2020), 554706.
- [25] Judy Delin. 2005. Brand Tone of Voice: a linguistic analysis of brand positions. *Journal of Applied Linguistics* 2, 1 (2005).
- [26] Benedict G C Dellaert. 2020. Consumer decisions with artificially intelligent voice assistants. *Marketing Letters* 31 (2020), 335–347.
- [27] Smit Desai and Michael Twidale. 2023. Metaphors in Voice User Interfaces: A Slippery Fish. *ACM Trans. Comput.-Hum. Interact.* 30, 6, Article 89 (sep 2023), 37 pages. <https://doi.org/10.1145/3609326>
- [28] Jayati Dev. 2022. *Privacy-Preserving Conversational Interfaces*. Ph.D. Indiana University, United States – Indiana. <https://www.proquest.com/docview/2703500357/abstract/F24B9345758B401CPQ/1> ISBN: 9798841768746.
- [29] Leyla Dogruel and Sven Joeckel. 2023. Gender stereotypes and voice assistants: do users' gender and conversation topic matter? *Behaviour & Information Technology* 0, 0 (2023), 1–11. <https://doi.org/10.1080/0144929X.2023.2235021> Publisher: Taylor & Francis_eprint: <https://doi.org/10.1080/0144929X.2023.2235021>
- [30] Philip R Doyle, Justin Edwards, Odile Dumbleton, Leigh Clark, and Benjamin R Cowan. 2019. Mapping perceptions of humanness in intelligent personal assistant interaction. In *Proceedings of the 21st international conference on human-computer interaction with mobile devices and services*. Association for Computing Machinery, New York, NY, USA, 1–12.
- [31] Mateusz Dubiel, Sylvain Daronnat, and Luis A Leiva. 2022. Conversational Agents Trust Calibration: A User-Centred Perspective to Design. In *Proceedings of the 4th Conference on Conversational User Interfaces*. ACM, Glasgow, UK, 1–6.
- [32] Mateusz Dubiel, Martin Halvey, Pilar Oplustil Gallegos, and Simon King. 2020. Persuasive synthetic speech: Voice perception and user behaviour. In *Proceedings of the 2nd Conference on Conversational User Interfaces*. ACM, New York City, New York, 1–9.
- [33] Alice H Eagly and Shelly Chaiken. 1984. Cognitive theories of persuasion. In *Advances in experimental social psychology*. Vol. 17. Elsevier, Amsterdam, Netherlands, 267–359.
- [34] Donna Erickson, Shigeto Kawahara, Albert Rilliard, Ryoko Hayashi, Toshiyuki Sadanobu, Yongwei Li, Hayato Daikuhara, João Moraes, and Kerrie Obert. 2020. *Cross cultural differences in arousal and valence perceptions of voice quality*. Proceedings of the International Conference on Speech Prosody, Virtual. <https://doi.org/10.21437/SpeechProsody.2020-147> Pages: 724.
- [35] Claus-Peter H Ernst and Nils Herm-Stapelberg. 2020. The Impact of Gender Stereotyping on the Perceived Likability of Virtual Assistants.. In *AMCIS*. Association for Information Systems (AIS), Atlanta, GA.
- [36] Franz Faul, Edgar Erdfelder, Albert-Georg Lang, and Axel Buchner. 2007. G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods* 39, 2 (2007), 175–191.
- [37] Carlos Flavián, Khaoula Akdim, and Luis V Casaló. 2023. Effects of voice assistant recommendations on consumer behavior. *Psychology & Marketing* 40, 2 (2023), 328–346.
- [38] Brian J Fogg. 2002. Persuasive technology: using computers to change what we think and do. *Ubiquity* 2002, December (2002), 2.
- [39] Gail A Fontenelle, Amanda P Phillips, and David M Lane. 1985. Generalizing across stimuli as well as subjects: A neglected aspect of external validity. *Journal of Applied Psychology* 70, 1 (1985), 101.
- [40] David Gefen. 2000. E-commerce: the role of familiarity and trust. *Omega* 28, 6 (2000), 725–737.
- [41] Judith E Glaser. 2016. *Conversational intelligence: How great leaders build trust and get extraordinary results*. Routledge, London, UK.
- [42] Anais Gretry, Csilla Horváth, Nina Belei, and Allard CR van Riel. 2017. “Don't pretend to be my friend!” When an informal brand communication style backfires on social media. *Journal of Business Research* 74 (2017), 77–89.
- [43] Abhijit Guha, Timna Bressgott, Dhruv Grewal, Dominik Mahr, Martin Wetzels, and Elisa Schweiger. 2023. How artificiality and intelligence affect voice assistant evaluations. *Journal of the Academy of Marketing Science* 51, 4 (2023), 843–866.
- [44] Rom Harré. 1985. Persuasion and manipulation. *Discourse and communication. New approaches to the analysis of mass media discourse and communication* -, - (1985), 126–42.
- [45] Eelco Herder and Sven Herden. 2023. Context-Dependent Use of Authority and Empathy in Lifestyle Advices Given By Persuasive Voice Assistants. In *Adjunct Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization*. Association for Computing Machinery, New York, NY, USA, 132–139.
- [46] Areum Hong, Changi Nam, and Seongcheol Kim. 2020. What will be the possible barriers to consumers' adoption of smart home services? *Telecommunications Policy* 44, 2 (March 2020), 101867. <https://doi.org/10.1016/j.telpol.2019.101867>
- [47] Bo Huang and Sylvain Sénécal. 2023. How should voice assistants be heard? The mitigating effect of verbal and vocal warmth in voice assistant service failure. *The Service Industries Journal* -, - (2023), 806–826.
- [48] Tatum Hunter. 2022. *Needy, overconfident voice assistants are wearing on their owners' last nerves*. Washington Post. <https://www.washingtonpost.com/technology/2022/03/07/voice-assistants-wrong-answers/>
- [49] Hanna Ingber. 2023. *Alexa, Why Do We Keep Buying You?* New York Times. <https://www.nytimes.com/2023/01/07/technology/personaltech/alexa-amazon-uses.html>
- [50] Carolin Ischen, Theo B Araujo, Hilde AM Voorveld, Guda Van Noort, and Edith G Smit. 2022. Is voice really persuasive? The influence of modality in virtual assistant interactions and two alternative explanations. *Internet Research* 32, 7 (2022), 402–425.
- [51] Razan Jaber, Sabrina Zhong, Sanna Kuoppamäki, Aida Hosseini, Iona Gessinger, Duncan P Brumby, Benjamin R. Cowan, and Donald Mcmillan. 2024. Cooking With Agents: Designing Context-aware Voice Interaction. In *Proceedings of the CHI Conference on Human Factors in Computing Systems* (<conf-loc>, <city>Honolulu</city>, <state>HI</state>, <country>USA</country>, <conf-loc>) (CHI '24). Association for Computing Machinery, New York, NY, USA, Article 551, 13 pages. <https://doi.org/10.1145/3613904.3642183>
- [52] Ana Javornik, Raffaele Flieri, and Ralph Gumann. 2020. “Don't forget that others are watching, too!” The effect of conversational human voice and reply length on observers' perceptions of complaint handling in social media. *Journal of Interactive Marketing* 50, 1 (2020), 100–119.
- [53] Iris Jestin, Joel Fischer, Maria Jose Galvez Trigo, David Large, and Gary Burnett. 2022. Effects of Wordings and Gendered Voices on Acceptability of Voice Assistants in Future Autonomous Vehicles. In *Proceedings of the 4th Conference on Conversational User Interfaces (CUI '22)*. Association for Computing Machinery,

- New York, NY, USA, 1–11. <https://doi.org/10.1145/3543829.3543836>
- [54] Marius Johnen and Oliver Schnittka. 2019. When pushing back is good: The effectiveness of brand responses to social media complaints. *Journal of the Academy of Marketing Science* 47 (2019), 858–878.
- [55] Bradley Johnson. 2023. *The 23 Best High Demand Products to Sell In 2023*. Blue Cart. <https://www.bluecart.com/blog/high-demand-products>
- [56] Hiroko Kamide and Tatsuo Arai. 2017. Perceived Comfortableness of Anthropomorphized Robots in U.S. and Japan. *International Journal of Social Robotics* 9, 4 (Sept. 2017), 537–543. <https://doi.org/10.1007/s12369-017-0409-8>
- [57] Steven Kemp and Nieves Erades Pérez. 2023. Consumer Fraud against Older Adults in Digital Society: Examining Victimization and Its Impact. *International Journal of Environmental Research and Public Health* 20, 7 (Jan. 2023), 5404. <https://doi.org/10.3390/ijerph20075404> Number: 7 Publisher: Multidisciplinary Digital Publishing Institute.
- [58] Judith M Kessens, Mark A Neerinx, Rosemarijn Looije, Melanie Kroes, and Gertrud Bloothoof. 2009. Facial and vocal emotion expression of a personal computer assistant to engage, educate and motivate children. In *2009 3rd international conference on affective computing and intelligent interaction and workshops*. IEEE, International Conference on Affective Computing and Intelligent Interaction and Workshops, Amsterdam, Netherlands, 1–7.
- [59] Jieun Kim, Gonzalo Gonzalez-Pumariaga, Soyee Park, and Susan R. Fussell. 2023. Urgency Builds Trust: A Voice Agent's Emotional Expression in an Emergency. In *Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing (CSCW '23 Companion)*. Association for Computing Machinery, New York, NY, USA, 343–347. <https://doi.org/10.1145/3584931.3606979>
- [60] Ki Joon Kim. 2016. Interacting socially with the Internet of Things (IoT): effects of source attribution and specialization in human–IoT interaction. *Journal of Computer-Mediated Communication* 21, 6 (2016), 420–435.
- [61] Casey A Klostad, 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.
- [62] Brigitte Krenn, Stephanie Schreitter, and Friedrich Neubarth. 2017. Speak to me and I tell you who you are! A language-attitude study in a cultural-heritage application. *AI & society* 32 (2017), 65–77.
- [63] Alexander Kunst. 2023. *Most popular categories for online purchases in the U.S. 2023*. Statista. Retrieved August 25, 2023 from <https://www.statista.com/forecasts/997093/most-popular-categories-for-online-purchases-in-the-us>
- [64] Kwan Min Lee and Clifford Nass. 2004. The multiple source effect and synthesized speech: Doubly-disembodied language as a conceptual framework. *Human communication research* 30, 2 (2004), 182–207.
- [65] Minha Lee, Lily Frank, Yvonne De Kort, and Wijnand IJsselstein. 2022. Where is Vincent? Expanding our emotional selves with AI. In *Proceedings of the 4th Conference on Conversational User Interfaces (CUI '22)*. Association for Computing Machinery, New York, NY, USA, 1–11. <https://doi.org/10.1145/3543829.3543835>
- [66] Ming-Chi Lee. 2009. Factors influencing the adoption of internet banking: An integration of TAM and TPB with perceived risk and perceived benefit. *Electronic Commerce Research and Applications* 8, 3 (May 2009), 130–141. <https://doi.org/10.1016/j.elerap.2008.11.006>
- [67] Sangwook Lee, Jeeyun Oh, and Won-Ki Moon. 2023. Adopting Voice Assistants in Online Shopping: Examining the Role of Social Presence, Performance Risk, and Machine Heuristic. *International Journal of Human-Computer Interaction* 39, 14 (Aug. 2023), 2978–2992. <https://doi.org/10.1080/10447318.2022.2089813> Publisher: Taylor & Francis _eprint: <https://doi.org/10.1080/10447318.2022.2089813>
- [68] Ge Li and Katie Seaborn. 2024. No Joke: An Embodied Conversational Agent Greeting Older Adults with Humour or a Smile Unrelated to Initial Acceptance. In *Extended Abstracts of the 2024 CHI Conference on Human Factors in Computing Systems* (<conf-loc> <city>Honolulu</city> <state>HI</state> <country>USA</country> </conf-loc>) (CHI EA '24). Association for Computing Machinery, New York, NY, USA, Article 252, 7 pages. <https://doi.org/10.1145/3613905.3650918>
- [69] Jui-Min Li, Jen-Shou Yang, and Hsin-Hsi Wu. 2009. Analysis of competency differences among frontline employees from various service typologies: integrating the perspectives of the organisation and customers. *The Service Industries Journal* 29, 12 (Dec. 2009), 1763–1778. <https://doi.org/10.1080/02642060902793664> Publisher: Routledge _eprint: <https://doi.org/10.1080/02642060902793664>
- [70] Xinge Li and Yongjun Sung. 2021. Anthropomorphism brings us closer: The mediating role of psychological distance in User–AI assistant interactions. *Computers in Human Behavior* 118 (2021), 106680.
- [71] Christine Liebrecht, Christina Tsaousi, and Charlotte van Hooijdonk. 2021. Linguistic elements of conversational human voice in online brand communication: Manipulations and perceptions. *Journal of Business Research* 132 (2021), 124–135.
- [72] Irene Lopatovska, Katrina Rink, Ian Knight, Kieran Raines, Kevin Cosenza, Harriet Williams, Perachya Sorsche, David Hirsch, Qi Li, and Adrianna Martinez. 2019. Talk to me: Exploring user interactions with the Amazon Alexa. *Journal of Librarianship and Information Science* 51, 4 (2019), 984–997.
- [73] Ewa Luger and Abigail Sellen. 2016. "Like Having a Really Bad PA" The Gulf between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 5286–5297.
- [74] Xiao Ma and Ariel Liu. 2020. Challenges in Supporting Exploratory Search through Voice Assistants. In *Proceedings of the 2nd Conference on Conversational User Interfaces (CUI '20)*. Association for Computing Machinery, New York, NY, USA, 1–3. <https://doi.org/10.1145/3405755.3406152>
- [75] Alex Mari, Andreina Mandelli, and René Algesheimer. 2020. The Evolution of Marketing in the Context of Voice Commerce: A Managerial Perspective. In *HCI in Business, Government and Organizations (Lecture Notes in Computer Science)*, Fiona Fui-Hoon Nah and Keng Siau (Eds.). Springer International Publishing, Cham, 405–425. https://doi.org/10.1007/978-3-030-50341-3_32
- [76] Zachariah NK Marrero, Samuel D Gosling, James W Pennebaker, and Gabriella M Harari. 2022. Evaluating voice samples as a potential source of information about personality. *Acta Psychologica* 230 (2022), 103740.
- [77] Josefa D Martin-Santana, Clara Muela-Molina, Eva Reinares-Lara, and Miriam Rodriguez-Guerra. 2015. Effectiveness of radio spokesperson's gender, vocal pitch and accent and the use of music in radio advertising. *BRQ Business Research Quarterly* 18, 3 (2015), 143–160.
- [78] Ewa Maslowska, Edward Malthouse, and Stefan Bernritter. 2017. The Effect of Online Customer Reviews' Characteristics on Sales. In *Advances in Advertising Research (Vol. VII)*. Springer, Berlin, Heidelberg, 87–100. https://doi.org/10.1007/978-3-658-15220-8_8
- [79] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. 2019. Reliability and inter-rater reliability in qualitative research: Norms and guidelines for CSCW and HCI practice. *Proceedings of the ACM on human-computer interaction* 3, CSCW (2019), 1–23.
- [80] Graeme McLean, Kofi Osei-Frimpong, and Jennifer Barhorst. 2021. Alexa, do voice assistants influence consumer brand engagement? – Examining the role of AI powered voice assistants in influencing consumer brand engagement. *Journal of Business Research* 124 (Jan. 2021), 312–328. <https://doi.org/10.1016/j.jbusres.2020.11.045>
- [81] Macy Meyer and Alina Bradford. 2022. *Alexa Shopping and Voice Purchasing*. Amazon. <https://www.cnet.com/home/smart-home/alexa-shopping-and-voice-purchasing-a-step-by-step-guide/>
- [82] Sandra Mooshammer and Katrin Etzrodt. 2022. Gender ambiguity in voice-based assistants: Gender perception and influences of context. *Human-Machine Communication* 5 (Aug. 2022), 49–74. <https://doi.org/10.3316/informit.869744164360916> Publisher: Communication and Social Robotics Labs.
- [83] John W Mullennix, Steven E Stern, Stephen J Wilson, and Corrie-lynn Dyson. 2003. Social perception of male and female computer synthesized speech. *Computers in Human Behavior* 19, 4 (2003), 407–424.
- [84] A. P. Kavitha Muthukumaran and H. Vani. 2020. Optimizing the usage of voice assistants for shopping. *Indian Journal of Science and Technology* 13, 43 (Dec. 2020), 4407–4416. <https://doi.org/10.17485/IJST/v13i43.1911> Publisher: The Indian Society of Education and Environment.
- [85] Clifford Nass, Ulla Foehr, Scott Brave, and Michael Somoza. 2001. The effects of emotion of voice in synthesized and recorded speech. In *Proceedings of the AAAI symposium emotional and intelligent II: The tangled knot of social cognition*. AAAI North Falmouth, MA, AAAI, North Falmouth, MA.
- [86] Clifford Nass, Jonathan Steuer, and Ellen R Tauber. 1994. Computers are social actors. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 72–78.
- [87] Robert Noggle. 2018. Manipulation, salience, and nudges. *Bioethics* 32, 3 (2018), 164–170. <https://doi.org/10.1111/bioe.12421> _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/bioe.12421>
- [88] Camille Noufi, Lloyd May, and Jonathan Berger. 2023. The Role of Vocal Persona in Natural and Synthesized Speech. In *2023 IEEE 17th International Conference on Automatic Face and Gesture Recognition (FG)*. IEEE Computer Society, Los Alamitos, CA, USA, 1–4. <https://doi.org/10.1109/FG57933.2023.10042714>
- [89] Ekaterina Novozhilova, Kate Mays, and James E. Katz. 2024. Looking towards an automated future: U.S. attitudes towards future artificial intelligence instantiations and their effect. *Humanities and Social Sciences Communications* 11, 1 (Jan. 2024), 1–11. <https://doi.org/10.1057/s41599-024-02625-1> Number: 1 Publisher: Palgrave.
- [90] Daniel J O'keefe. 2015. *Persuasion: Theory and research*. Sage Publications, Thousand Oaks, California.
- [91] Christina Oumard, Julian Kreimeier, and Timo Götzelmann. 2022. Pardon? An Overview of the Current State and Requirements of Voice User Interfaces for Blind and Visually Impaired Users. In *Computers Helping People with Special Needs*, Klaus Miesenberger, Georgios Kouroupetroglou, Katerina Mavrou, Roberto Manduchi, Mario Covarrubias Rodriguez, and Petr Penáz (Eds.). Springer International Publishing, Cham, 388–398. https://doi.org/10.1007/978-3-031-08648-9_45
- [92] Giridhar Pamisetty and K. Sri Rama Murty. 2023. Prosody-TTS: An End-to-End Speech Synthesis System with Prosody Control. *Circuits, Systems, and Signal Processing* 42, 1 (Jan. 2023), 361–384. <https://doi.org/10.1007/s00034-022-02126->

- z
- [93] Andrea Papenmeier and Elin Anna Topp. 2023. Ah, Alright, Okay! Communicating Understanding in Conversational Product Search. In *Proceedings of the 5th International Conference on Conversational User Interfaces*. ACM, Eindhoven Netherlands, 1–5. <https://doi.org/10.1145/3571884.3604318>
 - [94] Silke Paulmann and Ayse K Uskul. 2014. Cross-cultural emotional prosody recognition: Evidence from Chinese and British listeners. *Cognition & emotion* 28, 2 (2014), 230–244.
 - [95] Sarah Perez. 2018. Voice shopping estimated to hit \$40+ billion across U.S. and U.K. by 2022. <https://techcrunch.com/2018/03/02/voice-shopping-estimated-to-hit-40-billion-across-u-s-and-u-k-by-2022/>
 - [96] David-Paul Pertaub, Mel Slater, and Chris Barker. 2002. An experiment on public speaking anxiety in response to three different types of virtual audience. *Presence* 11, 1 (2002), 68–78.
 - [97] Richard E. Petty and John T. Cacioppo. 1986. The Elaboration Likelihood Model of Persuasion. In *Communication and Persuasion: Central and Peripheral Routes to Attitude Change*, Richard E. Petty and John T. Cacioppo (Eds.). Springer, New York, NY, 1–24. https://doi.org/10.1007/978-1-4612-4964-1_1
 - [98] Michel Tuan Pham and Tamar Avnet. 2004. Ideals and oughts and the reliance on affect versus substance in persuasion. *Journal of consumer research* 30, 4 (2004), 503–518.
 - [99] Rachel Phinnemore, Mohi Reza, Blaine Lewis, Karthik Mahadevan, Bryan Wang, Michelle Annett, and Daniel Wigdor. 2023. Creepy Assistant: Development and Validation of a Scale to Measure the Perceived Creepiness of Voice Assistants. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1–18.
 - [100] Yury Pinsky. 2021. Loud and clear: AI is improving Assistant conversations.
 - [101] Fernando Poyatos. 1993. *Paralanguage: A linguistic and interdisciplinary approach to interactive speech and sounds*. Vol. 92. John Benjamins Publishing, Amsterdam, Netherlands.
 - [102] Erica Rawes. 2023. *How to enable whisper mode on Alexa, and how exactly it works*. Digital Trends. <https://www.digitaltrends.com/home/how-to-enable-whisper-mode-on-alexa/>
 - [103] Chong Eun Rhee and Junho Choi. 2020. Effects of personalization and social role in voice shopping: An experimental study on product recommendation by a conversational voice agent. *Computers in Human Behavior* 109 (2020), 106359.
 - [104] Matthew D Rocklage, Derek D Rucker, and Loran F Nordgren. 2018. Persuasion, emotion, and language: The intent to persuade transforms language via emotionality. *Psychological science* 29, 5 (2018), 749–760.
 - [105] Eileen Roesler, Lara Naendrup-Poell, Dietrich Manzey, and Linda Onnasch. 2022. Why Context Matters: The Influence of Application Domain on Preferred Degree of Anthropomorphism and Gender Attribution in Human-Robot Interaction. *International Journal of Social Robotics* 14, 5 (July 2022), 1155–1166. <https://doi.org/10.1007/s12369-021-00860-z>
 - [106] Yves Rosseel. 2012. lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software* 48 (May 2012), 1–36. <https://doi.org/10.18637/jss.v048.i02>
 - [107] Christine Rzepka, Benedikt Berger, Anton Koslow, and Thomas Hess. 2023. Voice Assistant, Buy Coffee Capsules!: Understanding the Determinants of Consumers' Intention to Use Voice Commerce. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems* 54, 3 (Aug. 2023), 137–159. <https://doi.org/10.1145/3614178.3614185>
 - [108] Robert T Sataloff and Karen M Kost. 2020. The effects of age on the voice, Part 1. *Journal of Singing* 77, 1 (2020), 63–70.
 - [109] Annett Schirmer, Man Hey Chiu, Clive Lo, Yen-Ju Feng, and Trevor B Penney. 2020. Angry, old, male—and trustworthy? How expressive and person voice characteristics shape listener trust. *Plos one* 15, 5 (2020), e0232431.
 - [110] Bjorn Schuller and Anton Batliner. 2013. *Computational Paralinguistics: Emotion, Affect and Personality in Speech and Language Processing* (1st ed.). Wiley Publishing, Hoboken, New Jersey.
 - [111] Yuxi Shang, Zhongxian Wu, Xiaoyu Du, Yanbin Jiang, Beibei Ma, and Meihong Chi. 2022. The psychology of the internet fraud victimization of older adults: A systematic review. *Frontiers in Psychology* 13 (Sept. 2022), -. <https://doi.org/10.3389/fpsyg.2022.912242> Publisher: Frontiers.
 - [112] Rohit Shewale. 2023. *67 Voice Search Statistics For 2023*. Demand Sage. <https://www.demandpage.com/voice-search-statistics/>
 - [113] Megan B. Shreffler. 2014. The Persuasiveness of Electronic Word of Mouth on Attitudes and Behaviors: An Elaboration-Likelihood-Model Perspective. *International Journal of Sport Communication* 7, 3 (Sept. 2014), 377–398. <https://proxyiub.uits.iu.edu/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=s3h&AN=98181060&site=ehost-live&scope=site> Publisher: Human Kinetics Publishers, Inc..
 - [114] M Six Silberman, Bill Tomlinson, Rochelle LaPlante, Joel Ross, Lilly Irani, and Andrew Zaldivar. 2018. Responsible research with crowds: pay crowdworkers at least minimum wage. *Commun. ACM* 61, 3 (2018), 39–41.
 - [115] Sara Skoog Waller, Mårten Eriksson, and Patrik Sörqvist. 2015. Can you hear my age? Influences of speech rate and speech spontaneity on estimation of speaker age. *Frontiers in psychology* 6 (2015), 978.
 - [116] Eugene C Snyder, Sanjana Mendu, S Shyam Sundar, and Saeed Abdullah. 2023. Busting the One-Voice-Fits-All Myth: Effects of Similarity and Customization of Voice-Assistant Personality. *International Journal of Human-Computer Studies* -, (2023), 103126.
 - [117] Amazon Staff. 2020. How Alexa is bringing multilingual households closer together. <https://www.aboutamazon.com/news/devices/how-alexa-is-bringing-multilingual-households-closer-together> Section: Devices.
 - [118] Amazon Staff. 2022. Amazon Echo and Alexa Devices. <https://www.amazon.com/smart-home-devices/b?ie=UTF8&node=9818047011>. Accessed: 2022-02-10.
 - [119] Google Staff. 2024. Google Assistant. <https://assistant.google.com/>. Accessed: 2022-02-10.
 - [120] Alain D. Starke and Minha Lee. 2022. Unifying Recommender Systems and Conversational User Interfaces. In *Proceedings of the 4th Conference on Conversational User Interfaces*. ACM, Glasgow United Kingdom, 1–7. <https://doi.org/10.1145/3543829.3544524>
 - [121] Selina Jeanne Sutton. 2020. Gender Ambiguous, not Genderless: Designing Gender in Voice User Interfaces (VUIs) with Sensitivity. In *Proceedings of the 2nd Conference on Conversational User Interfaces (CUI '20)*. Association for Computing Machinery, New York, NY, USA, 1–8. <https://doi.org/10.1145/3405755.3406123>
 - [122] Chenhao Tan, Vlad Niculae, Cristian Danescu-Niculescu-Mizil, and Lillian Lee. 2016. Winning arguments: Interaction dynamics and persuasion strategies in good-faith online discussions. In *Proceedings of the 25th international conference on world wide web*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 613–624.
 - [123] Akihiro Tanaka, Ai Koizumi, Hisato Imai, Saori Hiramatsu, Eriko Hiramoto, and Beatrice de Gelder. 2010. I Feel Your Voice: Cultural Differences in the Multisensory Perception of Emotion. *Psychological Science* 21, 9 (Sept. 2010), 1259–1262. <https://doi.org/10.1177/0956797610380698> Publisher: SAGE Publications Inc.
 - [124] Vito Tassiello, Jack S. Tillotson, and Alexandra S. Rome. 2021. “Alexa, order me a pizza!”: The mediating role of psychological power in the consumer–voice assistant interaction. *Psychology & Marketing* 38, 7 (2021), 1069–1080. <https://doi.org/10.1002/mar.21488> _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/mar.21488>
 - [125] Dustin Tingley, Teppei Yamamoto, Kentaro Hirose, Luke Keele, and Kosuke Imai. 2014. **mediation**: R Package for Causal Mediation Analysis. *Journal of Statistical Software* 59, 5 (2014), 6–12. <https://doi.org/10.18637/jss.v059.i05>
 - [126] Suzanne Tolmeijer, Naim Zierau, Andreas Janson, Jalil Sebastian Wahdatehagh, Jan Marco Marco Leimeister, and Abraham Bernstein. 2021. Female by default?—exploring the effect of voice assistant gender and pitch on trait and trust attribution. In *Extended abstracts of the 2021 CHI conference on human factors in computing systems*. Association for Computing Machinery, New York, NY, USA, 1–7.
 - [127] Suzanne Tolmeijer, Naim Zierau, Andreas Janson, Jalil Sebastian Wahdatehagh, Jan Marco Marco Leimeister, and Abraham Bernstein. 2021. Female by Default? – Exploring the Effect of Voice Assistant Gender and Pitch on Trait and Trust Attribution. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems (CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, 1–7. <https://doi.org/10.1145/3411763.3451623>
 - [128] Saba Torki Biucky, Neda Abdolvand, and Saeedeh Rajae Harandi. 2017. The Effects of Perceived Risk on Social Commerce Adoption Based on TAM Model. <https://papers.ssrn.com/abstract=3926443>
 - [129] Ilaria Torre, Jeremy Goslin, and Laurence White. 2020. If your device could smile: People trust happy-sounding artificial agents more. *Computers in Human Behavior* 105 (2020), 106215.
 - [130] Alex B Van Zant and Jonah Berger. 2020. How the voice persuades. *Journal of personality and social psychology* 118, 4 (2020), 661.
 - [131] Sarah Theres Völkel, Penelope Kempf, and Heinrich Hussmann. 2020. Personalised Chats with Voice Assistants: The User Perspective. In *Proceedings of the 2nd Conference on Conversational User Interfaces (CUI '20)*. Association for Computing Machinery, New York, NY, USA, 1–4. <https://doi.org/10.1145/3405755.3406156>
 - [132] Katja Wagner, Frederic Nimmermann, and Hanna Schramm-Klein. 2019. Is it human? The role of anthropomorphism as a driver for the successful acceptance of digital voice assistants. -, (2019).
 - [133] Katja Wagner and Hanna Schramm-Klein. 2019. Alexa, Are You Human? Investigating Anthropomorphism of Digital Voice Assistants-A Qualitative Approach.. In *ICIS*. Association for Information Systems (AIS), Atlanta, GA.
 - [134] Ara Wagoner. 2019. *Hear the rainbow: Putting personality to the Google Assistant*. Android Central. <https://www.androidcentral.com/google-assistant-voice-personalities>
 - [135] David Williams, Victoria Crittenden, and Amy Henley. 2022. Third-party procedural justice perceptions: The mediating effect on the relationship between eWOM and likelihood to purchase. *Journal of Marketing Theory and Practice* 30, 1 (2022), 86–107.
 - [136] Valerie SL Williams, Lyle V Jones, and John W Tukey. 1999. Controlling error in multiple comparisons, with examples from state-to-state differences in educational achievement. *Journal of educational and behavioral statistics* 24, 1 (1999), 42–69.

- [137] Ralf Winkler. 2007. Influences of pitch and speech rate on the perception of age from voice. In *Proceedings of the 16th International Congress of Phonetic Sciences, Saarbrücken*. -, -, 1849–1852.
- [138] Dominique Wirz. 2018. Persuasion through emotion? An experimental test of the emotion-eliciting nature of populist communication. *International Journal of Communication* 12 (2018), 1114–1138.
- [139] Yaxiong Wu, Craig Macdonald, and Iadh Ounis. 2022. Multimodal Conversational Fashion Recommendation with Positive and Negative Natural-Language Feedback. In *Proceedings of the 4th Conference on Conversational User Interfaces (CUI '22)*. Association for Computing Machinery, New York, NY, USA, 1–10. <https://doi.org/10.1145/3543829.3543837>
- [140] Yang Yang, Li Wang, and Qi Wang. 2021. Take Your Word or Tone for It? European American and Chinese Children's Attention to Emotional Cues in Speech. *Child Development* 92, 3 (2021), 844–852. <https://doi.org/10.1111/cdev.13576> _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/cdev.13576>.
- [141] Silvia Zaharia and Matthias Würfel. 2021. Voice Commerce - Studying the Acceptance of Smart Speakers. In *Human Interaction, Emerging Technologies and Future Applications III (Advances in Intelligent Systems and Computing)*, Tareq Ahram, Redha Taiar, Karine Langlois, and Arnaud Choplin (Eds.). Springer International Publishing, Cham, 449–454. https://doi.org/10.1007/978-3-030-55307-4_68
- [142] Yue Zhang, Felix Weninger, Boqing Liu, Maximilian Schmitt, Florian Eyben, and Björn Schuller. 2017. A Paralinguistic Approach To Speaker Diarisation: Using Age, Gender, Voice Likability and Personality Traits. In *Proceedings of the 25th ACM International Conference on Multimedia (Mountain View, California, USA) (MM '17)*. Association for Computing Machinery, New York, NY, USA, 387–392. <https://doi.org/10.1145/3123266.3123338>
- [143] Xinyan Zhao and Mengqi Monica Zhan. 2019. Appealing to the Heart: How Social Media Communication Characteristics Affect Users' Liking Behavior During the Manchester Terrorist Attack. *International Journal of Communication* 13 (2019), 22.
- [144] Runtong Zhong, Mengyao Ma, Yutong Zhou, Qingxia Lin, Leiling Li, and Nengjing Zhang. 2022. User acceptance of smart home voice assistant: a comparison among younger, middle-aged, and older adults. *Universal Access in the Information Society* -, - (2022), 1–18.
- [145] Qingxiaoyang Zhu, Austin Chau, Michelle Cohn, Kai-Hui Liang, Hao-Chuan Wang, Georgia Zellou, and Zhou Yu. 2022. Effects of Emotional Expressiveness on Voice Chatbot Interactions. In *Proceedings of the 4th Conference on Conversational User Interfaces (CUI '22)*. Association for Computing Machinery, New York, NY, USA, 1–11. <https://doi.org/10.1145/3543829.3543840>

Categories	Frequency	
	Stimulus Validation	Main Study
Gender		
Female	38 (48.7%)	161 (48.1%)
Male	38 (48.7%)	168 (50.1%)
Non-binary	2 (2.5%)	5 (1.5%)
Others	0 (0.0%)	1 (0.3%)
Age Group		
18-29	36 (46.1%)	135 (40.3%)
30-49	31 (39.7%)	144 (43.0%)
50-64	10 (12.8%)	46 (13.7%)
65+	1 (1.3%)	10 (3.0%)
Highest Education Level		
No high school	3 (3.8%)	2 (0.6%)
High school or equivalent	34 (43.6%)	63 (18.8%)
Undergrad./2 year degree	35 (44.8%)	227 (67.75%)
Graduate degree	6 (7.7%)	43 (12.8%)
Racial or Ethnic Background		
Hispanic or Latino	5 (6.4%)	25 (7.5%)
Black or African American	7 (9.0%)	31 (9.25%)
American Indian	0 (0%)	4 (1.2%)
Asian	7 (9.0%)	26 (7.8%)
White	54 (69.2%)	222 (66.3%)
Mixed	4 (5.1%)	25 (7.5%)
Other	1 (1.3%)	2 (6.0%)
Using Voice Assistant		
Never	8 (10.4%)	19 (5.7%)
Less than once a month	10 (13.0%)	39 (11.6%)
Once a month	5 (6.5%)	24 (7.2%)
Multiple times a month	13 (16.9%)	70 (20.9%)
Once a week	2 (2.6%)	29 (8.7%)
Multiple times a week	14 (18.2%)	75 (22.4%)
Once a day	5 (6.5%)	14 (4.2%)
Multiple times a day	20 (26.0%)	65 (19.4%)
	N= 78	N= 335

Table 6: Demographic and voice assistant usage information of final participants

Factor	Item
VA Usage	How often do you interact with a Voice Assistant?
Disposition of Trust [40]	I generally trust other people
	I tend to count upon other people
	I generally have faith in humanity
	I generally trust other people unless they give me a reason not to

Table 4: Question Items about usage of voice assistant and disposition of trust

	Positive	Negative	Neutral
Younger Male	204.7 (6.7)	120.7 (5.8)	121.5 (6.5)
Younger Female	258.2 (6.5)	203.4 (5.3)	209.7 (6.4)
Middle-Aged Male	155 (6)	125.9 (5.4)	129 (5.4)
Middle-Aged Female	247.8 (5.9)	164.8 (5.1)	182.7 (5.6)

Table 5: Mean pitch (Hz) and speech rate (in parenthesis) of tone, gender, and age combinations for the synthesized speech in the study