INTERSPEECH 2020
October 25–29, 2020, Shanghai, China

Individual variation in language attitudes toward voice-AI: The role of listeners' autistic-like traits

Michelle Cohn[1], Melina Sarian[1], Kristin Predeck[1], Georgia Zellou[1]

[1]Phonetics Laboratory, Department of Linguistics, UC Davis, USA
{mdcohn, msarian, kpredeck, gzellou} @ucdavis.edu

Abstract

More and more, humans are engaging with voice-activated artificially intelligent (voice-AI) systems that have names (e.g., Alexa), apparent genders, and even emotional expression; they are in many ways a growing 'social' presence. But to what extent do people display sociolinguistic attitudes, developed from human-human interaction, toward these disembodied text-to-speech (TTS) voices? And how might they vary based on the cognitive traits of the individual user? The current study addresses these questions, testing native English speakers' judgments for 6 traits (intelligent, likeable, attractive, professional, human-like, and age) for a naturally-produced female human voice and the US-English default Amazon Alexa voice. Following exposure to the voices, participants completed these ratings for each speaker, as well as the Autism Quotient (AQ) survey, to assess individual differences in cognitive processing style. Results show differences in individuals' ratings of the likeability and human-likeness of the human and AI talkers based on AQ score. Results suggest that humans transfer social assessment of human voices to voice-AI, but that the way they do so is mediated by their own cognitive characteristics.

Index Terms: language attitudes, voice-activated artificially intelligent (voice-AI) systems, sociolinguistic competence

1. Introduction

Prior work has shown that people have strong beliefs and opinions of other individuals based solely on speech patterns [1]. These language attitudes have been linked to associations between speech variation and geographic region [2], [3], gender [4], social class [5], and native language [6]. Language attitudes are often a proxy for people's social attitudes toward individuals from a particular regional or social group. Specifically, people attribute a person who speaks with an accent as being inherently 'pleasant' and 'smart', or inversely 'harsh' and 'not intelligent', if they associate people from the social group the accent indexes as having that attribute [7], [8]. In other words, people have intricate folk beliefs about inherent cognitive and social characteristics of speakers based purely on their voice and speech patterns.

An open question is whether people attribute similar language attitudes to humans and to voice-activated artificially intelligent (voice-AI) systems (e.g., Apple's Siri, Google Assistant, and Amazon's Alexa). Given the ubiquity of these systems [9], there is growing interest in the implications of social characteristics of voice-AI, such as gender stereotyping of predominantly female assistants [10]–[12]. But our scientific understanding of how people attribute social variables to voice-AI is still limited.

The current study examines human and voice-AI language attitudes through the lens of the *Computers as Social Actors* (CASA) theoretical framework [13], [14]. CASA posits that people automatically apply social behaviors from human-human interaction to their behaviors toward computers given that a cue of humanity is expressed by the system. In particular, the use of spoken language has been proposed to 'activate' human social norms toward technology [15]. For example, [16] found that participants perceived differences in personality across synthesized text-to-speech (TTS) voices (e.g., labeled some as being 'introverted' or 'extroverted') based on the acoustic parameters of the voices. Relatedly, [17] found that a robot with a higher pitched voice was given higher ratings in overall appearance, voice appeal, behavior, and personality, relative to with a lower pitched voice. For modern voice-AI, TTS synthesis is based on datasets of productions from real human speakers, via concatenative TTS [18] or neural vocoders trained on human pronunciation patterns (e.g., Tacotron or Wavenet [19], [20]). Given their more 'human-like' voices, a CASA account might predict that listeners will ascribe social judgments to voice-AI in the same ways that they would toward a real human, even if they know that they are interacting with a non-human entity.

Related work in human-robot/computer interaction provides some support for the possibility that humans will rate TTS voices using human-based attributes. For example, individuals' attitudes of a robot have been shown to vary based on the dialect it 'speaks' (e.g., Acapela TTS child voices presented via NAO robots in [21]; US, UK, NZ dialect ratings for a health robot in [22]). Similarly, listeners assign age and gender to TTS voices presented (e.g., IBM TTS in [23]), suggesting that social attribution of a voice does not rely on the presence of a physical form. Even fewer studies have made a direct comparison between both TTS and naturally-produced human voices. Such a comparison is essential for evaluating predictions made by the CASA framework: people might apply social rules from human-human interaction to interactions with technology, but it is possible that there is still a distinction between real human versus voice-AI. For example, [24] found that a human voice was rated as nicer, less eerie, less supernatural, less hair-raising, and less shocking than a TTS voice. [25] found that a human voice was rated as having stronger social presence and behavioral intentions, while a TTS voice received lower trust ratings.

1.1. Individual variation in language attitudes

An additional consideration is whether computer personification responses (here, application of language attitudes to voice-AI) might vary *across individuals*. Individual variation in behavioral responses to speech is well attested: there are differences across people within the same age, social group, and region in how speech patterns are perceived [26]. In particular, variation across listeners has

http://dx.doi.org/10.21437/Interspeech.2020-1339

been linked to differences in individuals' cognitive processing style (see [27] for review); that is, how people cognitively vary in how they process sensory information [28].

There is increasing evidence that there is individual variation in personification of technology, as well. For example, in a self-reported diary study of human-Alexa interactions, [29] found that only half of participants reported human social behaviors toward Alexa (e.g., saying "please" and "thank you"). This suggests that the application of politeness norms from human-human interaction to interactions with technology (cf. [30]) varies across users. For example, [31] found that the extent to which participants responded positively to a computer's flattering praise varied as a function of their cognitive style: individuals with less analytical and more intuition-driven traits were more greatly affected by the computer's flattery. Also, [32] found that participants' subconscious vocal entrainment behavior toward human and Siri voices varied based on their cognitive processing styles, measured by the Autism Quotient (AQ).

The AQ [33] is a common non-clinical instrument across studies of speech and language behavior used to assess differences in individuals' cognitive processing style [26], [32]. The AQ has been shown to capture variation within neurotypical populations and is consistent with those formally diagnosed with Autism Spectrum Disorder (ASD), a condition that results in significant atypicality in social, emotional, and communicative behavior (DSM-5 [34]). In a general population of people, without a clinical ASD diagnosis, autistic-like traits manifest to varying degrees and can be quantified [35]. The AQ has also been shown to capture differences in behavior across individuals. For instance, people with higher AQ scores, signaling more autistic-like traits, were less accurate in detecting whether a robot's actions were pre-programmed or human-controlled [36].

Given that one of the primary characteristics for individuals with more autistic-like traits is a deficit in emotion perception [37], [38], presenting participants with emotionally expressive stimuli is one way to further probe the social nature of these interactions and emphasize possible sources of more subtle individual variation. This has been previously demonstrated for emotion expression in robots: individuals with more autistic-like traits display weaker sensitivity to a robot's fear and disgust facial expressions [39], [40]. For voice-AI, emotional expressiveness can be conveyed in some systems. For instance, the Amazon Alexa (US-English) voice has 'Speechcons' [41]: words and phrases that have been recorded in a hyper-expressive manner by the voice actor. Prior work suggests that participants respond positively to these emotionally expressive productions by voice-AI, displaying more vocal entrainment toward emotionally expressive interjections by Alexa [42] and rating interactions with an Alexa social bot more highly if it contained them [43]. We predict that differences based on AQ will manifest in different social judgments of Alexa and human voices who use emotionally expressive speech. Therefore, in the present study, we exposed participants to these same emotionally expressive interjections to further increase possible variation in responses based on AQ.

1.2. Current Study

We designed the current study to explore differences in people's social judgments of voice-AI talkers, compared to a human's utterances. Participants completed a short interactive task, where they heard neutral and hyper-expressive interjections produced by the default Amazon Alexa voice and a real human female speaker. Then, participants rated the speakers across 6 social traits: how intelligent, professional, likeable, attractive, human-like, and old each voice sounded. We selected these ratings based on related work in human-human (e.g., 'intelligence' in [7]; 'professional' in [44]; 'likeable' in [45]; 'attractive' in [46]; 'age' in [47]), and human-computer interaction ('human-like' and 'age' in [23]). Participants also completed the Autism Quotient (AQ) [33]. We then tested the extent the relationship between AQ scores and ratings for the voice-AI and the human talkers.

We have several predictions about the relationship between an individual's autistic-like traits and their social ratings of a voice-AI and a naturally produced human voice. First, we predict that people who display more autistic-like traits will show more variation in their social ratings in general (i.e., for humans and voice-AI); this is in line with prior work showing that individuals with ASD display difficulty with social evaluation [48].

Second, we predict that individuals with higher AQ scores will attribute more positive social judgments to the voice-AI talker (relative to the human talker). This prediction stems from work showing that individuals with ASD often *prefer* interactions with technology over those with humans [49], [50]. One way to understand this relationship is through the lens of the *Uncanny Valley of the Mind* [51], a function proposed to capture the dynamics of increasing human-likeness on likeability: while usually increasing humanness correlates with increasing likeability, at a point nearing the 'human' boundary, humans respond with discomfort or disgust. This 'valley', however, has been proposed to be shifted in individuals with ASD [52], occurring with the most 'human-like' humans. Put another way, increasing autistic-like traits may led to greater uncanniness of the *human*, relative to the voice-AI, talker. In the present study, this could be realized as lower ratings for the human voice.

2. Methods

2.1. Stimuli, Participants, and Procedure

Stimuli for the interactive task consisted of 24 interjections, used in [42], generated in neutral and emotionally-expressive prosodies: *awesome, bravo, bummer, cheers, cool, darn, ditto, dynamite, eureka, great, howdy, hurray, jinx, roger, shucks, splash, super, wow, wowzer, yikes, yuck, yum, zap, zing*. Items were selected from the Alexa 'Speechcons' available [41]. The neutral Alexa productions were generated using the Alexa Skills Kit (ASK). For the human voice, a female native English speaker (age 25) was recorded producing the same set of interjections. These productions were elicited using instructions to speak in an emotionally neutral or expressive manner; the speaker did not imitate the Alexa voice. The recordings were made in a sound attenuated booth while the speaker wore a head-mounted microphone (Shure WH20 XLR) at a 44.1 kHz sampling rate. All stimuli were amplitude normalized to 65 dB in Praat. (Though normalizing might reduce acoustic cues to expressiveness, we did so to maintain perceptual loudness across items).

Participants (n=34), native English speakers recruited from the UC Davis Psychology subjects pool (21 M, 13 F; mean age = 20.12 years, sd = 2.2) were first familiarized with the words by reading them aloud. Then, participants were exposed to both speakers' productions. In this phase, participants were first introduced to each interlocutor, one at a time; either the voice-AI system ('Alexa') or human ('Melissa') first (Speaker order blocked and counterbalanced across participants). On a trial, participants heard an item produced by a talker and were asked to repeat the word. Both neutral and emotionally expressive productions were randomly presented within each block (2 blocks per speaker).

After the exposure phase, participants rated each talker's voice for 6 traits using a sliding scale (ranging from 0-100) (see Table 1). Order of speaker block (human, voice-AI) was counterbalanced across participants. Following the ratings task, participants completed the AQ questionnaire.

Table 1: *Social Attribute Ratings*

Professional	How professional did ___ sound? (0=not professional, 100 =extremely professional)
Likeable	How likeable did ___ sound? (0=not likeable, 100 =extremely likeable)
Attractive	How attractive did ___ sound? (0=not attractive, 100 =extremely attractive)
Intelligent	How intelligent did ___ sound? (0=not intelligent, 100 =extremely intelligent)
Human-like	How much like a real person did ___ sound? (0=not like a real person, 100=extremely realistic)
Age	How old did ___ sound? (0-100 in years)

2.2. Autism Quotient

The AQ questionnaire [33] consists of 50 statements designed to quantify the extent of autistic-like traits in adults of normal intelligence in a non-clinical setting. There are 5 categories of questions, assessing cognitive dimensions specifically associated with ASD: social skills, attention switching, attention to detail, communication, and imagination. For each statement, participants pick one of four answers "definitely agree", "slightly agree", "slightly disagree", and "definitely disagree". We followed the binary coding of responses as 1 or 0, with 1 corresponding to a more autistic-like response; 0 corresponding to a less autistic-like response. The total score is summed such that a higher value indicates more autistic-like traits, ranging from 0 (no autistic-like) to 50 (highly autistic-like).

3. Analysis & Results

3.1. AQ Scores

We observed variation in participants' overall AQ scores (range=8-31, mean = 17.7, sd = 5.7). The standard deviation of all social ratings (collapsed across the six variables) was modeled with a linear regression with a fixed effect of AQ score (continuous 0-50) with the *lme4* R package [53]. The model did not reveal an effect of AQ score on overall ratings variation [β=-0.01, t=-0.03, p=0.97]. Overall, the intercept for all ratings was 60.2.

3.2. Social Ratings Models

We modeled each social rating as a continuous dependent variable (0-100) with separate linear mixed effects models. Each model contained identical fixed and random effects structure. Fixed effects included Talker (2 levels: Alexa, human; contrasts were sum coded), AQ score (continuous: 0-50), and their interaction. Random effects included by-Subject random intercepts. (Models with the added by-Subject random slope for Talker did not converge; note that a separate model, with AQ as a 4-point scale, did not improve model fit).

3.2.1. *Attractiveness, intelligence, professionalness*

The attractiveness, intelligence, and professionalness ratings models all showed no significant effects or interactions. The intercepts were all above 50 (51.9 attractive, 64.9 intelligent, 70.5 professional), indicating that participants rated both voices as similarly attractive, intelligent, and professional.

3.2.2. *Likeability, age, human-likeness*

The likeability ratings model computed a significant main effect of Talker, where participants judged the Alexa speaker to more likeable, overall, than the human speaker [β=-15.2, t=-2.7, p<0.05]. This effect was additionally modulated by an interaction between Talker and AQ score [β=-0.99, t=-3.2, p<0.01]. This interaction is depicted in Figure 1.A; as participants' AQ scores increase, they are more likely to report distinct likeability ratings for the voice-AI and for the human talker: higher for the real person and lower for Alexa.

For participants' estimates of the speakers' ages, the model computed a main effect of Talker: the Alexa voice was rated as sounding older than the human voice [β=6.5, t=2.6, p<0.05]. Figure 1.B displays this main effect. There were no other significant effects or interactions.

The human-likeness ratings model revealed a significant interaction between Talker and AQ score [β=-1.2 t=-3.2, p<0.01]. This interaction can be seen in Figure 1.C: as participants' AQ scores increase, they are more likely to rate the human as more human-like and the Alexa talker as *less* human-like. Participants with lower AQ scores (indicating less autistic-like traits) rate the Alexa speaker and the human speaker as more similar in their human-likeness. No other effects were observed.

4. Discussion

In this study, we examined whether participants attribute language attitudes to a voice-AI interlocutor (here, Amazon's Alexa) and a real human talker in similar ways, and whether there are patterns of individual variation in these ratings. Overall, we found no difference in the ratings for the Alexa and human voice with respect to three dimensions: how intelligent, professional, and attractive they sound. Still, listeners did hear other differences in the voices: they rated the Alexa voice as more likeable and slightly older than the human voice. These patterns show that listeners extract subtle personality and age-related cues for TTS voices. Here, Alexa was rated as being in her 30s, while the human voice was rated as being in her 20s. Together, these overall ratings patterns provide support for the *CASA* personification framework [13], [14]: humans are applying social labels to voice-AI that are, in some cases, similar across the interlocutors (intelligence, professional, attractive).

In addition to these general patterns, we also tested whether there was variation in ratings based on an individual's autistic-like traits. While our first prediction was that individuals with more autistic-like traits would show greater *variation* in scores in general, we did not find evidence for this: there was no relationship between increasing AQ score and overall variation in ratings. This is contra what was observed in [48], where they saw differences in social evaluations in individuals

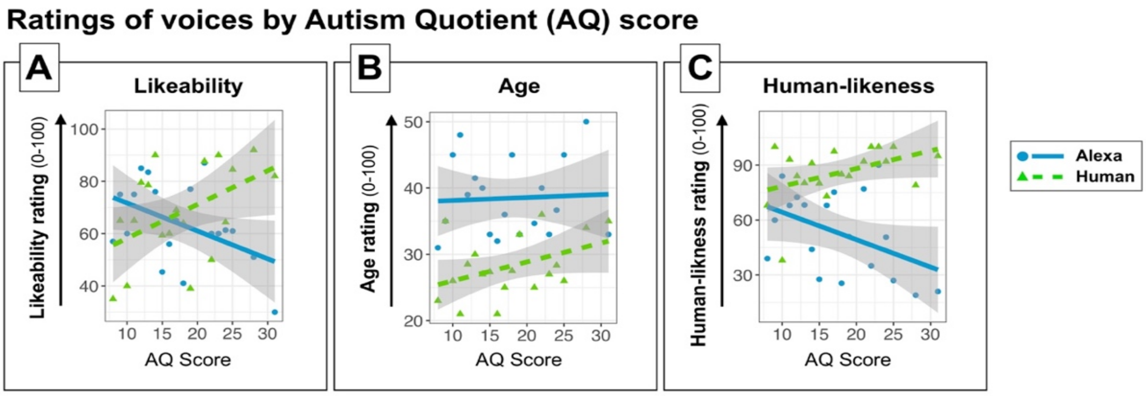

Figure 1: *Mean Ratings by Autism Quotient (AQ) score for (A) Likeability, (B) Age, and (C) Human-likeness for the Human and Alexa voices. Ribbons depict standard errors.*

with ASD. As our participant pool consisted of individuals without a formal ASD diagnosis – in assessing individual differences in the general undergraduate population – this could be one explanation as to why differences in our study were minimal.

At the same time, we did observe patterns of variation with some of the ratings: individuals with more autistic-like traits were more likely to provide *distinct* ratings for the Alexa and human voices. Contrary to our prediction, however, these differences were not in the expected direction; rather, higher AQ score was associated with a *decrease* in likeability and human-likeness of the Alexa voice and an increase in both dimensions for the human voice. One way to interpret this finding is that individuals with more autistic-like traits categorize human and voice-AI interlocutors as distinct social categories, and subsequently rate them more distinctly. These findings can add to prior work outlining *Uncanny Valley* [51]: here, we see that a voice-AI interlocutor who produces speech with neutral and expressive emotion is less likeable and less human-like for individuals with greater autistic-like traits. While we did not see an uncanny 'cliff' for autistic-like traits, as proposed for ASD [52], this is not to say that such cliff does not exist; rather, our findings suggest that the *poles* of human-likeness (machine ← → human) may be more distinct for individuals with greater autistic-like traits.

While this work provides evidence that humans apply language attitudes toward voice-AI, there are many questions that remain. For one, how these attitudes may interact with gender is an open question. In this study, we held gender constant, only including female voices, given the availability of 'Speechcons' [41] for the Alexa female voice. Yet, expanding to other genders – and comparing *multiple* voices – will be critical next steps in this line of research. Furthermore, gender of the participants may also be a relevant factor in these social evaluations. Additionally, a person's experience with voice-AI may be another factor in whether, and to what degree, they might apply language attitudes toward voice-AI in similar ways as they do for human voices.

Finally, another open avenue for future work is whether the top-down label of voice-AI and human may lead to different social ratings. In the current study, these labels always matched (i.e., TTS acoustic productions paired with knowledge that the speaker was a voice-AI system). Future work could test how listeners, varying in AQ, differently weigh these factors (different voice quality or different speaker category).

5. Acknowledgments

This material is based upon work supported by the National Science Foundation SBE Postdoctoral Research Fellowship under Grant No. 1911855 to MC. This work was partially supported by an Amazon Faculty Research Award to GZ.

6. References

[1] E. Anisfeld and W. E. Lambert, "Evaluational reactions of bilingual and monolingual children to spoken languages," *J. Abnorm. Soc. Psychol.*, vol. 69, no. 1, pp. 89–97, 1964.

[2] H. Giles and N. Niedzielski, "German sounds awful, but Italian is beautiful," *Lang. Myths*, pp. 85–93, 1998.

[3] M. Bucholtz, N. Bermudez, V. Fung, L. Edwards, and R. Vargas, "Hella Nor Cal or Totally So Cal?: The Perceptual Dialectology of California," J. Engl. Linguist., vol. 35, no. 4, pp. 325–352, 2007.

[4] L. Bilaniuk, "Gender, language attitudes, and language status in Ukraine," *Lang. Soc.*, vol. 32, no. 1, pp. 47–78, Jan. 2003.

[5] M. A. Stewart, E. B. Ryan, and H. Giles, "Accent and Social Class Effects on Status and Solidarity Evaluations," *Pers. Soc. Psychol. Bull.*, vol. 11, no. 1, pp. 98–105, Mar. 1985.

[6] M. Dragojevic, H. Giles, A.-C. Beck, and N. T. Tatum, "The fluency principle: Why foreign accent strength negatively biases language attitudes," *Commun. Mono.*, vol. 84, no. 3, pp. 385–405, Jul. 2017.

[7] D. R. Preston, "What's Old and What's New in Perceptual Dialectology?," *Bord. Lang. Dialect*, vol. 21, p. 16, 2018.

[8] D. Preston, "Perceptual dialectology," *Handb. Dialectol.*, 2017.

[9] F. Bentley, C. Luvogt, M. Silverman, R. Wirasinghe, B. White, and D. Lottridge, "Understanding the long-term use of smart speaker assistants," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 2, no. 3, pp. 1–24, 2018.

[10] F. Habler, V. Schwind, and N. Henze, "Effects of Smart Virtual Assistants' Gender and Language," in *Proc. of Mensch und Computer 2019*, 2019, pp. 469–473.

[11] G. Hwang, J. Lee, C. Y. Oh, and J. Lee, "It sounds like a woman: exploring gender stereotypes in south Korean voice assistants," in *2019 CHI*, 2019, pp. 1–6.

[12] A. Piper, "Stereotyping Femininity in Disembodied Virtual Assistants," *Grad. Theses Diss.*, Jan. 2016.

[13] C. Nass, J. Steuer, and E. R. Tauber, "Computers are social actors," in *Proc. of the SIGCHI Conf. on Human factors in computing systems*, 1994, pp. 72–78.

[14] C. Nass, Y. Moon, J. Morkes, E.-Y. Kim, and B. J. Fogg, "Computers are social actors: A review of current research," *Hum. Values Des. Comput. Technol.*, vol. 72, pp. 137–162, 1997.

[15] C. I. Nass and S. Brave, *Wired for speech: How voice activates and advances the human-computer relationship*. MIT press Cambridge, MA, 2005.

[16] C. Nass and K. M. Lee, "Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction.," *J. Exp. Psychol. Appl.*, vol. 7, no. 3, p. 171, 2001.

[17] A. Niculescu, B. van Dijk, A. Nijholt, H. Li, and S. L. See, "Making social robots more attractive: the effects of voice pitch, humor and empathy," *Int. J. Soc. Robot.*, vol. 5, no. 2, pp. 171–191, 2013.

[18] T. Merritt *et al.*, "Comprehensive Evaluation of Statistical Speech Waveform Synthesis," in *2018 IEEE Spoken Lang. Tech. (SLT)*, Dec. 2018, pp. 325–331.

[19] A. Van Den Oord *et al.*, "WaveNet: A generative model for raw audio.," in *SSW*, 2016, p. 125.

[20] Y. Wang *et al.*, "Tacotron: Towards end-to-end speech synthesis," *ArXiv Prepr. ArXiv170310135*, 2017.

[21] A. Sandygulova and G. M. P. O'Hare, "Children's Perception of Synthesized Voice: Robot's Gender, Age and Accent," in *Soc. Robotics*, Cham, 2015, pp. 594–602.

[22] R. Tamagawa, C. I. Watson, I. H. Kuo, B. A. MacDonald, and E. Broadbent, "The Effects of Synthesized Voice Accents on User Perceptions of Robots," *Int. J. Soc. Robot.*, vol. 3, no. 3, pp. 253–262, Aug. 2011.

[23] A. Baird, S. H. Jørgensen, E. Parada-Cabaleiro, S. Hantke, N. Cummins, and B. Schuller, "Perception of Paralinguistic Traits in Synthesized Voices," in *Proc. Conf. on Aug. & Particip Sound & Music Exp.*, 2017, p. 17.

[24] A. Abdulrahman, D. Richards, and A. Aysin Bilgin, "A Comparison of Human and Machine-Generated Voice," in *Symp. on VR Soft. and Tech.*, Parramatta, Australia, Nov. 2019, pp. 1–2.

[25] E. Chérif and J.-F. Lemoine, "Anthropomorphic virtual assistants and the reactions of Internet users: An experiment on the assistant's voice," *Rech. Appl. En Mark. Engl. Ed.*, vol. 34, no. 1, pp. 28–47, 2019.

[26] A. C. L. Yu, C. Abrego-Collier, and M. Sonderegger, "Phonetic Imitation from an Individual-Difference Perspective: Subjective Attitude, Personality and 'Autistic' Traits," *PLoS One*, vol. 8, no. 9, p. e74746, Sep. 2013.

[27] A. C. L. Yu and G. Zellou, "Individual Differences in Language Processing: Phonology," *Annu. Rev. Linguist.*, vol. 5, no. 1, pp. 131–150, 2019.

[28] L. J. Ausburn and F. B. Ausburn, "Cognitive styles: Some information and implications for instructional design," *ECTJ*, vol. 26, no. 4, pp. 337–354, Dec. 1978.

[29] I. Lopatovska and H. Williams, "Personification of the Amazon Alexa: BFF or a Mindless Companion," in *Conf. on Human Info. Int. & Retr.l*, New York, NY, USA, 2018, pp. 265–268.

[30] B. J. Fogg and C. Nass, "Silicon sycophants: the effects of computers that flatter," *Int. J. Hum.-Comput. Stud.*, vol. 46, no. 5, pp. 551–561, 1997.

[31] E.-J. Lee, "The more humanlike, the better? How speech type and users' cognitive style affect social responses to computers," *Comput. Hum. Behav.*, vol. 26, no. 4, pp. 665–672, 2010.

[32] C. Snyder, M. Cohn, and G. Zellou, "Individual variation in cognitive processing style predicts differences in phonetic imitation of device and human voices.," in *Proc. of the Annual Conf. of the Int'l Speech Comm. Association*, Graz, Austria, 2019, pp. 116–120.

[33] S. Baron-Cohen, S. Wheelwright, R. Skinner, J. Martin, and E. Clubley, "The Autism-Spectrum Quotient (AQ): Evidence from Asperger Syndrome/High-Functioning Autism, Malesand Females, Scientists and Mathematicians," *J. Autism Dev. Disord.*, vol. 31, no. 1, pp. 5–17, Feb. 2001.

[34] A. P. Association, *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®)*. American Psychiatric Pub, 2013.

[35] S. Fletcher-Watson and F. Happé, *Autism: A New Introduction to Psychological Theory and Current Debate*. Routledge, 2019.

[36] A. Wykowska, J. Kajopoulos, K. Ramirez-Amaro, and G. Cheng, "Autistic traits and sensitivity to human-like features of robot behavior," *Interact. Stud.*, vol. 16, no. 2, pp. 219–248, Jan. 2015.

[37] S. Bölte and F. Poustka, "The recognition of facial affect in autistic and schizophrenic subjects and their first-degree relatives," *Psychol. Med.*, vol. 33, no. 5, pp. 907–915, Jul. 2003.

[38] S. Kuusikko *et al.*, "Emotion Recognition in Children and Adolescents with Autism Spectrum Disorders," *J. Autism Dev. Disord.*, vol. 39, no. 6, pp. 938–945, Jun. 2009.

[39] P. E. McKenna, A. Ghosh, R. Aylett, F. Broz, I. Keller, and G. Rajendran, "Robot Expressive Behaviour and Autistic Traits," in *Proc. Conf. on Autonomous Agents & MultiAgent Systems*, Richland, SC, 2018, pp. 2239–2241.

[40] F. Askari, H. Feng, A. Gutierrez, T. Sweeny, and M. Mahoor, "How Children with Autism Spectrum Disorder Recognize Facial Expressions Displayed by a Rear-Projection Humanoid Robot," *Electr. Comput. Eng. Grad. Stud. Scholarsh.*, Jan. 2018.

[41] Amazon, "Speechcon Reference (Interjections): English (US) | Custom Skills," 2018. https://developer.amazon.com/docs/custom-skills/speechcon-reference-interjections-english-us.html (accessed Dec. 09, 2018).

[42] M. Cohn and G. Zellou, "Expressiveness influences human vocal alignment toward voice-AI," *Interspeech 2019*, pp. 41–45, 2019.

[43] M. Cohn, C.-Y. Chen, and Z. Yu, "A Large-Scale User Study of an Alexa Prize Chatbot: Effect of TTS Dynamism on Perceived Quality of Social Dialog," *SIGdial Meeting*, 2019, pp. 293–306.

[44] R. E. Evans and P. Kortum, "Voice personalities inducing trust and satisfaction in a medical interactive voice response system," in *Proc. of the Human Factors and Ergonomics Society Annual Meeting*, 2009, vol. 53, no. 18, pp. 1456–1460.

[45] B. Weiss and F. Burkhardt, "Voice attributes affecting likability perception," in *Conf. of the Int'l Speech Com. Assoc*, 2010.

[46] B. Borkowska and B. Pawlowski, "Female voice frequency in the context of dominance and attractiveness perception," *Anim. Behav.*, vol. 82, no. 1, pp. 55–59, Jul. 2011.

[47] T. Shipp, Y. Qi, R. Huntley, and H. Hollien, "Acoustic and temporal correlates of perceived age," *J. Voice*, vol. 6, no. 3, pp. 211–216, Jan. 1992.

[48] B. Forgeot d'Arc *et al.*, "Atypical Social Judgment and Sensitivity to Perceptual Cues in Autism Spectrum Disorders," *J. Autism Dev. Disord.*, vol. 46, no. 5, pp. 1574–1581, May 2016.

[49] S. H. Hedges, S. L. Odom, K. Hume, and A. Sam, "Technology use as a support tool by secondary students with autism," *Autism*, vol. 22, no. 1, pp. 70–79, Jan. 2018.

[50] K. Gillespie-Lynch, S. K. Kapp, C. Shane-Simpson, D. S. Smith, and T. Hutman, "Intersections Between the Autism Spectrum and the Internet: Perceived Benefits and Preferred Functions of Computer-Mediated Communication," *Intellect. Dev. Disabil.*, vol. 52, no. 6, pp. 456–469, Nov. 2014.

[51] M. Mori, K. F. MacDorman, and N. Kageki, "The uncanny valley [from the field]," *IEEE RAM*, vol. 19, no. 2, pp. 98–100, 2012.

[52] Y. Ueyama, "A bayesian model of the uncanny valley effect for explaining the effects of therapeutic robots in autism spectrum disorder," *PloS One*, vol. 10, no. 9, 2015.

[53] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting Linear Mixed-Effects Models Using lme4," *J. Stat. Softw.*, vol. 67, no. 1, pp. 1–48, Oct. 2015.