INTERSPEECH 2019
September 15–19, 2019, Graz, Austria

Individual variation in cognitive processing style predicts differences in phonetic imitation of device and human voices

Cathryn Snyder, Michelle Cohn, and Georgia Zellou

Department of Linguistics, University of California, Davis

gzellou@ucdavis.edu

Abstract

Phonetic imitation, or implicitly matching the acoustic-phonetic patterns of another speaker, has been empirically associated with natural tendencies to promote successful social communication, as well as individual differences in personality and cognitive processing style. The present study explores whether individual differences in cognitive processing style, as indexed by self-reported scored from the Autism-Spectrum Quotient (AQ) questionnaire, are linked to the way people imitate the vocal productions by two digital device voices (i.e., Apple's Siri) and two human voices. Subjects first performed a word shadowing task of human and device voices and then completed the self-administered AQ. We assessed imitation of two acoustic properties: f0 and vowel duration. We find that the attention to detail and the imagination subscale scores on the AQ mediated degree of imitation of f0 and vowel duration, respectively. The findings yield new insight to speech production and perception mechanisms and how it interacts with individual cognitive processing style differences.

Index Terms: speech alignment, human-device interaction, individual differences

1. Introduction

Voice-AI systems, such as Apple's Siri, are now a prevalent type of interlocutor [1]. Humans engaging with the devices vary along multiple dimensions: by their age, gender, and background [2]. Yet, little work has explored variation across individuals' communication patterns with modern voice-AI systems. Prior work has observed individual differences in people's' behavior patterns during human-computer and human-robot interaction [3]. Some studies have begun outlining factors that might predict such individual variation in behavior during device interactions, such as the degree to which users anthropomorphize non-human entities [4] level or experience with the device, or cognitive/personality traits (cf. [5]). Recent work has demonstrated differences in speech patterns towards voice-AI devices based on the gender of the individual [6]. Understanding individual variation in human-device interaction is important for several reasons. For one, it can inform our scientific understanding of human-computer interaction and developing more inclusive models of human behavior toward devices. Furthermore, it is relevant for those interested in designing systems that can accommodate as many users as possible (cf., [3], [7]).

As reviewed in [5], several aspects of users' personalities and cognitive styles have been associated with their behavior during interactions with technology, such as spatial ability [8] and locus of control [9]. A relevant factor that has been understudied in individual variation human-device interaction is cognitive processing style, broadly defined as the constellation of psychological dimensions that reflect consistencies in the processing of sensory information that vary across individuals [10], [11]. Within a population of individuals of the same gender, age, and background (e.g., education level, socioeconomic status, etc.), there is diversity in cognitive processing style and this has been linked to a variety of behaviors in various domains, including identify formation [12], learning [13] and processing of sensory information [14]. Differences in cognitive processing style have also been linked to variation in individuals' speech production and perception patterns (see [15] for review). In this paper, we ask whether there is systematic variation in individuals' speech alignment toward voice-AI systems and humans as a function of their cognitive processing style.

One way to measure individual differences in cognitive processing style is with the Autism-Spectrum Quotient (AQ) questionnaire: a self-administered non-clinical assessment used to quantify degree of various "autistic" traits in neurotypical adults of normal intelligence [16]. The AQ assesses five dimensions of cognitive processing: social skills, attention switching, attention to detail, communication, and imagination. These AQ sub-scales provide insight into the multidimensional factors that influences individuals' cognitive processing styles and personality, in general. Several studies have found a link between individuals' scores on various AQ subscales and their performance on a range of language processing tasks. For example, higher communication AQ scores, i.e., more difficulty in social communication, relate to difficulties in using prosodic information to disambiguate distinct pragmatic meanings during sentence processing [17].

Since humans are interacting more and more with devices using *speech*, examining how individual variation in cognitive processing style might be linked to their phonetic patterns is an open question. Thus, of particular interest are studies showing a link between AQ scores and patterns of speech production and perception behavior. For example, women with high AQ attention switching scores, i.e., inflexibility in new situations, poor task switching and multitasking skills, tend to perceptually compensate, i.e., show less veridical acoustic perception, for /s/ following a rounded vowel relative to women with low AQ attention switching scores [18]. Additional research also shows that people with overall higher AQ scores (across subscales) are more likely to show sensitivity to fine-grained acoustic differences and are less likely to be influenced by higher-level lexical knowledge during the perception of sibilants [19]. Greater sensitivity to phonetic details in individuals with higher AQ scores, and subscale scores, has also been linked to differences in speech production. Yu et al. [20] showed that a higher score in the AQ attention switching subscale positively correlates with degree of phonetic imitation of voice onset time (VOT); this is

http://dx.doi.org/10.21437/Interspeech.2019-2669

thought to be driven by their heightened attention and phonetic sensitivity to each individual word (Yu et al., 2013).

Some prior work has linked subjects' AQ scores to their sensitivity to robot behavior: for example, individuals with fewer autistic traits were more accurate in detecting whether robot behavior was programmed or human-controlled [21] and in interpreting a robot's facial expression [22]. Thus, there is evidence to suggest that individual variation in AQ scores will predict patterns of human-computer communication in the domain of speech.

1.1. Current study

Little prior work has examined variation across individuals' cognitive processing styles (i.e., AQ scores) in interactions with voice-AI. In the current study, we tested whether individual variation in AQ subscores predicts patterns of phonetic imitation on a single-word shadowing task of four different interlocutors: two device voices (Apple's Siri) and two human voices. We focus on acoustic measures of imitation that have been shown to be sensitive to individual differences from prior work, specifically duration [23] pitch [24].

Our predictions about how individual variation in AQ sub-scores predict patterns of imitation toward device and human voices can be framed in terms of different perspectives on the motivations of phonetic imitation, in general. For one, Communication Accommodation Theory (CAT) [25], [26] proposes that imitation is a means by which people express social closeness. This has been empirically supported for example, if the interlocutor is perceived to have more positive social attributes, e.g. attitude, perceived attractiveness, similar ideologies, and social closeness, greater phonetic imitation is observed [27]–[29]. We predict that if imitation is more socially driven, as CAT suggests, differences in imitating humans and digital devices will be borne out in differences related to the AQ social skills sub-score, which measures flexibility, comfort, and comprehension of social cues during social encounters.

Another prediction is that individuals' phonetic imitation patterns may vary based on the imagination AQ subscale, which relates to the ability to comprehend fictional events and attribute human-like characteristics onto non-human objects. One possibility is that imitation patterns toward the human versus device voices are mediated by extent of attributing anthropomorphism to the voice-AI system, similar to past work demonstrating variation in personifying non-human entities [4]. This relates to theoretical frameworks of computer personification, such as the "Computers as Social Actors" (CASA) [30], [31], which posits that humans treat a computer as they would a human as soon as any degree of humanity can be detected. We ask whether AQ imagination score is related to the "degree of humanity" speakers detect from Siri voices.

Another perspective that might be relevant to phonetic imitation is that interactions with computer systems are driven by functional pressures [32], [33]. This is supported by findings that humans align with computers in ways that seem motivated to improve mutual intelligibility and communicative success, i.e., by choosing lexemes and speaking at a rate that they believe the computer will understand, e.g. [34], [35]. Thus, if imitation is more functionally driven, differences in imitating humans and digital devices will be borne out in differences related to the communication subcategory of the AQ, which assesses conversational competence and fluency skills.

Meanwhile, if imitation is more dependent on attentional mechanisms, as argued by Gambi and Pickering [32], we might expect that differences in imitating humans and digital devices will be borne out in a link between degree of imitation and scores in the AQ subscales that relate to attention (i.e., attention switching, attention to detail). This would align with findings from Yu's [23], [36] studies reporting higher AQ attention switching scores correlated with greater phonetic imitation.

2. Methods

2.1. Stimuli

Stimuli consisted of 12 low frequency CVN target words: *bomb, sewn, vine, pun, yawn, shun, chime, shone, wane, tame, wren, hem* (mean log frequency: 1.6, range: 1.1-2.1, taken from SUBTLEX [37]) produced by 2 real human talkers (1 female and 1 male, both native English speakers from California) and 2 Siri voices (American female, American male). The Siri voices were created on the Terminal on a Mac computer, while the human voices were recorded in a sound-attenuated booth.

2.2. Participants and procedure

A total of 43 female subjects participated in the experiment. We recruited only female subjects since [23] report that the association between AQ subscores and imitation was greatest for females. All subjects were native English speakers and reported no hearing impairment. All participants except for 3 reported that they have experience using Siri at least once a week.

While in the lab, subjects completed the 50-question Autism Quotient [16] that assesses individuals' self-reported autistic-like characteristics. The questionnaire is administered as a pen-and-paper survey. The test consists of items that fall into five subscales, consisting of 10 questions each: social skill (e.g., "I find social situations easy."), imagination (e.g., "If I try to imagine something, I find it very easy to create a picture in my mind."), attention to detail (e.g., "I often notice small sounds when others do not."), attention switching (e.g., "I find it easy to do more than one thing at once."), and communication (e.g. "Other people frequently tell me that what I've said is impolite, even though I think it is polite."). Questions are worded so that half would elicit an "agree" response and half would elicit a "disagree" response. Participants respond to each question on a 4-point scale ("strongly agree", "agree", "disagree", "strongly disagree"). Each response is then converted to a numeric score (1-4). Scoring can also be tabulated using a 0-1 value (grouping the slightly and strongly agree or disagree responses together). The total AQ score is calculated by summing the scores for all 50 questions. Each of the subscale scores is calculated by summing the scores for the 10 questions that correspond to each subscale trait. Table 1 presents the descriptive statistics for the Total AQ and subscale scores, as well as basic demographic characteristics of the participants. A high value denotes more ''autistic'' like traits, i.e. lower imagination, lower social skills, greater difficulty in attention switching, higher attention to detail, and lower ability to communicate.

Table 1: *Descriptive statistics of participant variables.*

	Mean	*SD*	*Min*	*Max*
Age	19.6	1.7	18	24
AQ total	108.2	13.4	83	133
Social skill (AQSS)	20.4	4.2	13	29
Imagination (AQIM)	19.6	3.9	13	29
Attention switching (AQAS)	24.4	4.2	17	32
Attention to detail (AQAD)	25.7	4.5	14	33
Communication (AQCM)	18.1	4.1	13	27

The study began with a pre-exposure phase, where subjects read each of the target words in isolation (presented randomly, 4 repetitions) to get their baseline productions. In the shadowing phase, subjects were introduced to the four interlocutors by name and picture: the two device voices (Siri, device female; Alex, device male) and the two human voices (Melissa, female; Carl, male). Images for the devices were two iPhones showing different home screens (e.g., "How can I help you today?"), while the images for the human interlocutors were stock images. Next, subjects were told that they would hear each of the four talkers say the word and that their task was to simply repeat the word. Subjects were not told explicitly to imitate. Word and model talker were randomized. In total, subjects shadowed 96 tokens (12 words * 4 model talkers * 2 repetitions).

3. Analysis

3.1. Acoustic assessment of phonetic imitation (DID)

We measured several acoustic properties of interest for each token produced by the subjects, as well as the productions by the model talkers: vowel duration (logged) and vowel mean f0 (mean, in semitones, ST). We then calculated a difference in distance (DID) measure [38] to quantify degree of acoustic convergence for each feature toward the model talker's production of that word. DID = |baseline - model| - |shadowed - model|. A positive DID value indicates change toward to direction of the model talker after exposure; a negative value indicates divergence from the model talker's speech.

3.2. Statistical analyses

We modeled DID values for vowel duration (logged) and mean f0 (ST) in two separate linear mixed effects models with the *lme4* R package [39]. Estimates for degrees of freedom, F-statistics, and p-values were computed using Satterthwaite approximation with *anova()* function in the *lmerTest* package (Kuznetsova et al., 2015). Both models had identical fixed and random effects structure. Fixed effects included Model Humanness (2 levels: human vs. device), Model Gender (2 levels: female vs. male). Exposure (2 levels: first vs. second repetition) was also included as a fixed effect predictor. The model also included participants' scores for each of the five AQ subscales (ranging from 5-40, logged) as fixed effects. An analysis of multicollinearity using the *ggpairs()* function in the GGally package [40] indicated high correlation between individuals' AQSS scores and their scores on the AQIM, AQCM, and AQAS subscores. Therefore, prior to model fitting, AQIM, AQCM, and AQAS were residualized for the effect of AQSS. The model included all possible two- and three-way interactions between Model Humanness x Model Gender x each AQ subscale variable. Random effects structure for each model included random intercepts for Participant and Lexical Item. In addition, each model included by-Participant random slopes for Model Gender, Model Humanness, and the interaction between these factors. Each of the discrete predictors were sum-coded, all continuous variables were centered and scaled.

4. Results

4.1. Phonetic imitation of vowel duration

The model run on DID duration scores computed significant main effects of Model Gender [F(1, 37)=23.9, *p*<.001]: participants converged in duration to male model talkers ($\bar{x}_{DID}$=.03), while there was significantly less alignment to female model talkers ($\bar{x}_{DID}$= -.002) (*β*=-0.01, *t*=-4.8, *p*<.001). We additionally observed a main effect of Model Humanness [F(1, 38)=22.8, *p*<.001]: participants showed greater convergence to human model talkers' durations ($\bar{x}_{DID}$= .02), but very small alignment to device voices, overall ($\bar{x}_{DID}$= .002) (*β*=-0.009, *t*=-4.8, *p*<.001). There was also a significant interaction between Model Gender and Model Humanness [F(1, 3575.4)=61.7, *p*<.001]: Participants converged in duration toward the human male voice most robustly ($\bar{x}_{DID}$= .04), followed by the male device voice ($\bar{x}_{DID}$= .006). The female human voice and the device female voice showed the smallest mean DID values (-.003, -.002, respectively).

There was a trend towards significance for the main effect of the residualized AQ Imagination subscale (AQIM) on degree of duration imitation [F(1, 32.7)=3.6, *p*=.06], with a negative coefficient value (*β*= -.008), indicating that listeners with more autistic-like imagination traits, i.e., poorer imagination, tended to display less convergence in vowel duration.

The model also revealed a significant three-way interaction between Model Gender, Model Humanness, and the residualized AQIM subscale [F(1, 3574.7)=5.3, *p*<.05]. The three-way interaction is illustrated in Figure 1: at lower AQIM scores, i.e., individuals with greater imagination skills, there is imitation toward the human male model talker, yet little to no convergence toward the other model talkers. Yet, as AQIM increases, signaling *poorer* imagination skills, participants display less convergence toward all the model talkers *except* the female device voice, which does not change across AQIM scores.

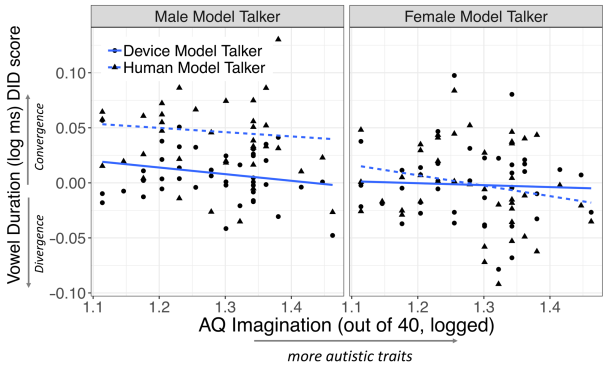

Figure 1: *DID duration means and standard errors by AQIM score, by Gender and Humanness of the Model Talker.*

4.2. Phonetic imitation of mean vowel f0

The model run on DID mean f0 values revealed a significant two-way interaction between Model Gender and AQ subscale of Attention to Detail (AQAD) [F(1, 36.9)=5.4, *p*<.05]. As observed in the right panel of Figure 2, we see less imitation of f0 with increasing AQAD scores for female model talkers.

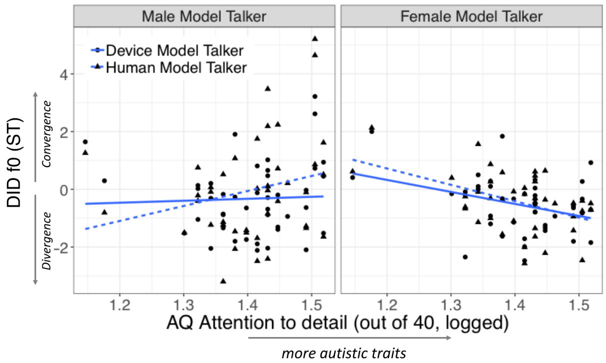

Figure 2: *DID mean f0 means and standard errors by AQAD score, by Gender and Humanness of the Model Talker.*

There was also a significant three-way interaction between Model Gender, Model Humanness, and AQAD [$F(1, 3541.2)=6.9$, $p<.01$]: as AQAD increases, there is greater pitch imitation for the male model talkers (relative to the female model talkers), and we see even *greater* imitation for the human male talker. In Figure 2, we additionally see that the slopes are steeper for pitch imitation of human voices than for device voices, on the basis of increasing AQAD score (though in opposite directions). No other effects or interactions were significant in the mean f0 DID model.

5. Discussion

In this study, we tested whether there is individual variation in shadowed productions of digital device (e.g. Apple's Siri) and human voices. This study was designed to address a gap in exploring individual variation in human-computer interactions, cf. [3], specifically in considering subjects' cognitive processing style [10], as measured by the self-reported Autism Quotient (AQ) [16]. Given prior work establishing links between cognitive processing style and speech behavior such as phonetic imitation [17], [23] we used a shadowing paradigm [38].

Overall, our results reveal individual variation in self-reported AQ subscale scores predict differences in patterns of imitation toward human and device voices. Specifically, for duration, we find that scores on the Imagination subscale of the AQ predict differences in imitation of device and human model talkers. Individuals with higher imagination ability, i.e., lower AQIM scores, showed greater imitation of duration, in general. In addition, this was mediated by both gender and humanity of the model talker: individuals with lower imagination ability display less convergence in duration toward all the model talkers, excluding the female device voice. Meanwhile, individuals with greater imagination skills display robust imitation toward the human male model talker, and some imitation of the male device voice. Conversely, individuals with high imagination scores show little to no convergence toward the female human and device model talkers.

Together, these findings suggest that imitation toward human/device voices is socially mediated, in that we see different patterns based on the apparent gender of the voices, in line with CAT [25], [26]; but in ways that relate to an individual's cognitive processing style. Furthermore, we see overlap in subjects' imitation patterns on the basis of model talker humanity: imitation patterns for the human and device male model talkers are in parallel, suggesting that across individual variation, people treat devices and humans differently, contra CASA [30], [31]..

One aspect that supports this is the difference in how individuals varying in imagination ability displayed imitation toward the male device voice: individuals with higher imagination displayed positive DID values toward the male device voice, while individuals with lower imagination showed divergence away from the male device voice. In other words, individuals with higher imagination were most likely to imitate a device voice. We do see similarities in imitation of female device and human voices. Yet, prior work has shown greater convergence toward male talkers [41], thus differences in imitation toward device and human voices could be realized more strongly since the ceiling is higher.

Our results additionally reveal differences in mean pitch (f0) imitation on the basis of AQ subscales. In particular, we see a relationship between the Attention to Detail subscale (AQAD), indicating *greater* attention to detail, and degree of imitation by the model talker humanity/gender. Subjects show greater f0 imitation for male talkers on the basis of increasing attention to detail, which even further increased when shadowing the *human male talker*. That we see differences based on an *attentional* measure is in line with proposals that imitation is attentionally mediated [32]. Our results additionally support proposals that imitation is socially mediated, in line with the duration patterns, in that we see different patterns of f0 imitation based on the gender of the model talker, supporting a CAT model. This was also observed for imitation for duration. In general, then, that we see imitation is mediated by gender of the model talker is in line with CAT theories of socially-mediated phonetic imitation.

At the same time, we see similar patterns of f0 imitation *within* genders for the male and female voices: with a general decline for female talkers and general incline for male talkers (on the basis of increasing AQAD score). These similarities are supportive of theories of computer personification, e.g. CASA, in that speech production patterns toward human and device voices are similar. Yet, we observe differences in the *steepness* of the function within these categories: subjects produce a steeper decline/incline for *human* model talkers.

In sum, an individual's cognitive processing style seems to influence how they treat humans and digital devices. Furthermore, these findings suggest that there is variation in human-device interaction. The results of this study have implications for models of human-device interaction. In particular, our findings raise questions as to the "automaticity" by which we treat computers as people, as proposed by CASA. We see individuals with different cognitive profiles demonstrating variation in the extent to which they "personify" digital device voices, suggesting that human-computer communication patterns are more complex than originally theorized. We propose an extension to the CASA theoretical framework, wherein the degree of computer personification is mediated by *social-cognitive mechanisms,* including speaker characteristics.

6. Acknowledgments

This work was partially funded by an Amazon Faculty Research Award and an Artificial Intelligence and Healthcare Seed Grant from UCD to GZ.

7. References

[1] M. B. Hoy, “Alexa, Siri, Cortana, and More: An Introduction to Voice Assistants,” *Med. Ref. Serv. Q.*, vol. 37, no. 1, pp. 81–88, 2018.

[2] D. C. Plummer *et al.*, *’Top strategic predictions for 2017 and beyond: Surviving the storm winds of digital disruption’. Gartner report G00315910. Gartner*. Inc, 2016.

[3] D. E. Egan, “Chapter 24 - Individual Differences In Human-Computer Interaction,” in *Handbook of Human-Computer Interaction*, M. Helander, Ed. Amsterdam: North-Holland, 1988, pp. 543–568.

[4] A. Waytz, J. Cacioppo, and N. Epley, “Who Sees Human?: The Stability and Importance of Individual Differences in Anthropomorphism,” *Perspect. Psychol. Sci.*, vol. 5, no. 3, pp. 219–232, May 2010.

[5] N. M. Aykin and T. Aykin, “Individual differences in human-computer interaction,” *Comput. Ind. Eng.*, vol. 20, no. 3, pp. 373–379, Jan. 1991.

[6] Anonymous, “Imitating Siri: Socially-mediated alignment to device and human voices,” *Proc. ICPhS*, to appear.

[7] M. Schmettow and J. Havinga, “Are Users More Diverse Than Designs?: Testing and Extending a 25 Years Old Claim,” in *Proceedings of the 27th International BCS Human Computer Interaction Conference*, Swinton, UK, UK, 2013, pp. 40:1–40:5.

[8] K. J. Vicente and R. C. Williges, “Accommodating individual differences in searching a hierarchical file system,” *Int. J. Man-Mach. Stud.*, vol. 29, no. 6, pp. 647–668, Jan. 1988.

[9] M. D. Coovert and M. Goldstein, “Locus of Control as a Predictor of Users’ Attitude toward Computers,” *Psychol. Rep.*, vol. 47, no. 3_suppl, pp. 1167–1173, Dec. 1980.

[10] 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.

[11] S. Messick and et al, *Individuality in learning*. Oxford, England: Jossey-Bass, 1976.

[12] M. D. Berzonsky, “Identity formation: The role of identity processing style and cognitive processes,” *Personal. Individ. Differ.*, vol. 44, no. 3, pp. 645–655, Feb. 2008.

[13] L. Price, “Individual Differences in Learning: Cognitive control, cognitive style, and learning style,” *Educ. Psychol.*, vol. 24, no. 5, pp. 681–698, Sep. 2004.

[14] R. J. Riding, “On the Nature of Cognitive Style,” *Educ. Psychol.*, vol. 17, no. 1–2, pp. 29–49, Mar. 1997.

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

[16] 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.

[17] S.-A. Jun and J. Bishop, “Priming Implicit Prosody: Prosodic Boundaries and Individual Differences,” *Lang. Speech*, vol. 58, no. 4, pp. 459–473, Dec. 2015.

[18] Z. Yu, “Attention and engagement aware multimodal conversational systems,” in *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, 2015, pp. 593–597.

[19] A. Yu, J. Grove, M. Martinovic, and M. Sonderegger, “Effects of Working Memory Capacity and ‘Autistic’ Traits on Phonotactic Effects in Speech Perception,” *Proc. Int. Congr. Phon. Sci. XVII*, pp. 2236–2239, 2011.

[20] Z. Yu, D. Gerritsen, A. Ogan, A. Black, and J. Cassell, “Automatic prediction of friendship via multi-model dyadic features,” in *Proceedings of the SIGDIAL 2013 Conference*, 2013, pp. 51–60.

[21] 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.

[22] P. E. McKenna, A. Ghosh, R. Aylett, F. Broz, I. Keller, and G. Rajendran, “Robot Expressive Behaviour and Autistic Traits,” in *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*, Richland, SC, 2018, pp. 2239–2241.

[23] 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.

[24] A. Bonnel, L. Mottron, I. Peretz, M. Trudel, E. Gallun, and A.-M. Bonnel, “Enhanced Pitch Sensitivity in Individuals with Autism: A Signal Detection Analysis,” *J. Cogn. Neurosci.*, vol. 15, no. 2, pp. 226–235, Feb. 2003.

[25] H. Giles, N. Coupland, and I. Coupland, “1. Accommodation theory: Communication, context, and,” *Contexts Accommod. Dev. Appl. Socioling.*, vol. 1, 1991.

[26] H. Giles and S. C. Baker, “Communication accommodation theory,” *Int. Encycl. Commun.*, 2008.

[27] M. Babel, “Evidence for phonetic and social selectivity in spontaneous phonetic imitation,” *J. Phon.*, vol. 40, no. 1, pp. 177–189, 2012.

[28] J. S. Pardo, “On phonetic convergence during conversational interaction,” *J. Acoust. Soc. Am.*, vol. 119, no. 4, pp. 2382–2393, 2006.

[29] M. Natale, “Convergence of mean vocal intensity in dyadic communication as a function of social desirability,” *J. Pers. Soc. Psychol.*, vol. 32, no. 5, pp. 790–804, 1975.

[30] C. Nass, J. Steuer, and E. R. Tauber, “Computers are social actors,” in *Proceedings of the SIGCHI conference on Human factors in computing systems*, 1994, pp. 72–78.

[31] C. Nass and Y. Moon, “Machines and mindlessness: Social responses to computers,” *J. Soc. Issues*, vol. 56, no. 1, pp. 81–103, 2000.

[32] C. Gambi and M. J. Pickering, “Prediction and imitation in speech,” *Front. Psychol.*, vol. 4, 2013.

[33] M. J. Pickering and S. Garrod, “Alignment as the basis for successful communication,” *Res. Lang. Comput.*, vol. 4, no. 2–3, pp. 203–228, 2006.

[34] H. P. Branigan, M. J. Pickering, J. Pearson, J. F. McLean, and A. Brown, “The role of beliefs in lexical alignment: Evidence from dialogs with humans and computers,” *Cognition*, vol. 121, no. 1, pp. 41–57, 2011.

[35] L. Bell, “Linguistic Adaptations in Spoken Human-Computer Dialogues-Empirical Studies of User Behavior,” Institutionen för talöverföring och musikakustik, 2003.

[36] A. C. L. Yu, “Perceptual Compensation Is Correlated with Individuals’ ‘Autistic’ Traits: Implications for Models of Sound Change,” *PLOS ONE*, vol. 5, no. 8, p. e11950, Aug. 2010.

[37] M. Brysbaert and B. New, “Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English,” *Behav. Res. Methods*, vol. 41, no. 4, pp. 977–990, 2009.

[38] J. S. Pardo, K. Jordan, R. Mallari, C. Scanlon, and E. Lewandowski, “Phonetic convergence in shadowed speech: The relation between acoustic and perceptual measures,” *J. Mem. Lang.*, vol. 69, no. 3, pp. 183–195, 2013.

[39] 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.

[40] B. Schloerke *et al.*, *Ggally: Extension to ggplot2. R package version 0.5. 0*. 2014.

[41] J. S. Pardo, R. Gibbons, A. Suppes, and R. M. Krauss, “Phonetic convergence in college roommates,” *J. Phon.*, vol. 40, no. 1, pp. 190–197, 2012.