Finding its Voice: The Influence of Robot Voice on Fit, Social Attributes, and Willingness to Use Among Older Adults in the U.S. and Japan

Long-Jing Hsu[1], Weslie Khoo[2], Natasha Randall[1], Waki Kamino[1], Swapna Joshi[1], Hiroki Sato[2], David J Crandall[2], Katherine M Tsui[3], and Selma Šabanović[1]

Abstract— **Robots may be able to significantly assist older adults through making activity recommendations. Prior research suggests that gender and age of a robot's voice may affect how people respond to such recommendations, but few studies have explored how a robot's voice is perceived by older adults, and whether their perceptions differ across cultures. We conducted a survey study with older adult participants (aged 65+) in the U.S. (N=225) and Japan (N=466), asking them to evaluate a humanoid robot speaking with three different voices (male, female, child). After seeing a video of a robot making recommendations, participants rated the fit of the voice to the robot, its sociality (via the Robotic Social Attributes Scale - RoSAS), and their willingness to use the robot in various contexts. We discovered that robot's social attributes and participants' culture impacted willingness to use the robot in both countries. Having positive social attributes and lower negative attributes increases willingness to use the robot. The U.S. older adults preferred the adult robot voices, had more positive social attributes, less negative social attributes, and were more likely to accept lifestyle recommendations than Japanese older adults. This study contributes to our understanding of older adults' perceptions of robot voice and provides design implications for robots that make recommendations to older adults.**

I. Introduction

When robots give reminders and suggestions to users, the type of voice they use can be very important. Voice serves as a social cue that affects how humans perceive and respond to the robot [1], just as oral communication has long affected humans' perception of each other [2]. Choosing an appropriate robot voice enables the Human-Robot Interaction (HRI) designers to strengthen the relationship between the intended user and robot [3], making the robot's suggestions more persuasive to users [4]. The robot voice can be designed to represent specific genders and/or ages [1], [5], leading people to perceive the robot as a member of a specific social group or as having specific robot characteristics.

While there have been many studies on users' perceptions of robot voices, most of this work has studied younger populations, whose preferences could be significantly different from those of older adults (OAs) [6]. OAs grew up in a different sociohistorical context than younger populations, which could influence their perceptions and decision-making [7]. Moreover, OAs' hearing abilities may decline due to aging-related changes in the ear, cognitive decline, and other factors [8]. They may also find it difficult to recognize speech, hear higher pitches, locate sound sources, and pay attention to certain information [9]. It thus remains an open question how OAs perceive robot voices mimicking different genders and ages.

HRI researchers have long been interested in OAs' perception of robots, including social attributes [10] and willingness to listen to recommendations [11]. Often measured through the Robotic Social Attributes Scale (RoSAS), a robot's social attributes relate to how humans perceive their human-like characteristics such as warmth, competence, and tendency to produce discomfort [10]. The social identity of the robot as perceived by the person may also affect how they evaluate a robot based on the age and gender-related identity it presents through its design [12] and significantly influence how the user interacts with it [10].

If robots were capable of making positive and meaningful recommendations, such as suggesting activities related to the meaning and purpose of OAs' lives [13]–[15], improving the quality of existing human connections [16], or reminding them about essential daily activities [17], it could potentially help bring happiness [13] to their lives. The *robot voice* and users' *cultural background* are crucial elements that may affect how OAs accept recommendations. While HRI designers have been trying to understand how robot recommendations are perceived, the robot's voice and the cultural background of users are factors that have not been much explored.

The United States (U.S.) and Japan are leaders in the production and consumption of robots [18]. U.S. and Japan were the world's largest manufacturer in the early 2000s because of their marketing strategies, diverse demand, and speed of production and export [18]. Several studies have explored differences across U.S. and Japanese users' perceptions of robots [19], [20]. People's different cultural backgrounds can affect their perception of the robot [21], including social attributes and willingness to use. However, few studies have looked at how OAs perceive robots, comparing the U.S. and Japan cultures, and specific concerns of the OA population.

The long-term goal of our research is to develop a social robot that can support OAs' ikigai, or meaning and purpose in one's life, by giving activity recommendations and prompting self-reflection [13]–[15]. To work towards this goal, in this paper, we study four important questions related to how the choice of the robot's voice could impact the effectiveness of a robot for : *(1) How do the age and gender ascribed by the robot's voice affect OAs' perception*

*This work was supported by Toyota Research Institute

[1]Informatics Department, Indiana University, Bloomington, IN 47408, USA hsulon@iu.edu; selmas@indiana.edu

[2]Computer Science Department, Indiana University, Bloomington, IN 47408, USA djcran@indiana.edu

[3]Toyota Research Institute, Cambridge, MA 02139, USA kate.tsui@tri.global

of how well the chosen voice fits the robot, as a function of the user's country (the United States or Japan)? (2) How do the age and gender ascribed by the robot's voice affect OAs' perception of the robot's social attributes, as a function of the user's country? (3) How do the age and gender ascribed by the robot's voice affect OA's perception of willingness to use the robot in different contexts, as a function of the user's country? (4) How do the robot's social attributes relate to OAs' willingness to use the robot in different contexts? This article contributes to the HRI community's understanding of how OAs in the U.S. and Japan perceive the voice of robots of different genders and ages. The findings will be valuable for HRI designers seeking to improve their designs for OAs.

II. Related work

Voice is complex and has multiple aspects to it. Pitch, in particular, is a significant indicator of age and gender, with higher-pitched voices associated with children or adult females, and lower-pitched voices with adult males [22]. In the section below, we discuss how a robot's voice can affect its perception, including its gender and age. We consider how robot voices can influence recommendations, especially for OAs, and how cultural differences may play a role.

A. Robot voice, gender, and age

There has been an ongoing discussion on how robot voices of different perceived genders affect users' interactions with the robot. HRI researchers can easily manipulate the gender implied by a robot's voice by simply modifying its pitch, with the implicit assumption that higher-pitched voices correspond to adult females and lower-pitched voices to adult males [22]. In a study carried out in the United States involving participants from Mechanical Turk, it was found that participants showed a higher level of acceptance towards male robots that rejected participants' commands compared to female robots [23]. Similar research was conducted on a SOTA, a small humanoid robot, to see how participants' behavior and attitudes change based on the robot's voice. The study discovered that the higher-pitched female voice robot was more persuasive than a male robot voice [24]. Some studies about robot voice and gender were conducted in Japan. In a scenario where Japanese university students were asked to rate robot guides with different apparent genders (implied by voice and name), the participants preferred female guides over male guides [25]. Existing studies do not agree on one preferred perceived gender in different human-robot interaction tasks, which suggests a potential impact of context of use on evaluations.

There is less research related to robot voice and the implied age of the robot. Within a limited study, HRI researchers modulated a robot's voice to sound more like a child by increasing the pitch [3], as children typically have higher-pitched voices [22]. When asked to rate the robot with a voice of a child or an adult, participants tended to associate the child robot voice with positive emotions such as "joy" and "anticipation," compared to other adult voice counterparts [3]. Previous research has attempted to link robot voice to robot appearance, where participants have to match a voice to an image of the robot. When a child-like robot, such as iCub, participants often associated it with a female voice where a child's voice was not available for selection [5]. In a series of studies, teenagers and young adults watched videos of a robot and evaluated its personality. The participants associated a robot child voice with extroverted, passionate, and relaxed [26] and warm [1] personalities.

The inconsistency of findings related to robot voice and gender, as well as the lack of study on robot voice and age, highlight the need for further studies to explore the effects of perceived robot gender in various contexts, including different cultural contexts.

B. Robots for providing recommendations

Social robots capable of verbal conversation are being used in a number of contexts, from receptionists [27], food service [28], story-telling [29], to supporting the well-being of OAs [13]–[15]. Many of these robots are designed to make recommendations to users. For example, a receptionist needs to recommend restaurants or attractions, story-telling robots may give suggestions on how the user should direct the story, and a robot that supports well-being may recommend social activities or exercises to maintain people's mental and physical health. In each of these cases, making sure the user complies with recommendations made by the robot is crucial for ensuring that the robot fulfills its intended purpose and provides a positive experience for its users. The robot's voice is a significant factor in this acceptance, as studies have shown that different voices can affect the persuasiveness of the robot's recommendations [28]. However, it is less understood how robot voice affects OAs' robot perception.

C. Cultural differences in perceptions of voice

Previous research exploring cultural differences between the U.S. and Japan showed inconsistent results. One study looked at Sociable Trash Box and found that Japanese participants have more positive evaluations on the robot than the U.S. participants [30]. Nomura et al. conducted a survey on university students' attitudes toward robots and found that, in comparison to Japanese students, U.S. participants had a more positive outlook on tasks that are challenging for humans to perform, but displayed a more negative attitude towards tasks related to nursing, social work, and education [19]. A study comparing Japanese and U.S. participants' attitudes towards Aibo robots found that Japanese participants were less positive than U.S. participants in their evaluations [20]. These findings suggest that there are inconsistencies among different cultures for robot, which highlights the need for culture- and task-specific recommendations on robot design and implementation.

III. Methods

Our goal was to understand how OAs in the U.S. and Japan perceive the voice type of a recommendation robot. To do so, we designed and deployed a large-scale survey for OAs in U.S. and Japan in 2022.

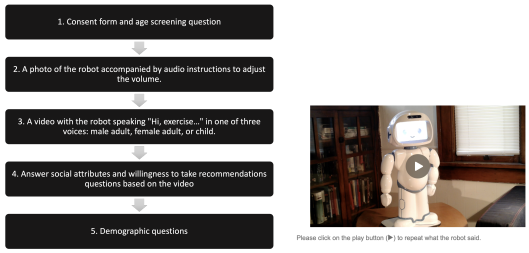

Fig. 1. (Left) A flowchart illustrating the survey process for the participants, and (Right) A screenshot from the video shown to the survey participants, featuring the QT robot with a voice-over.

Participants in the U.S. were recruited through Amazon Mechanical Turk and received $1.75 for completing the survey. In Japan, participants were recruited by Carter JMRN and Macromill, with compensation not exceeding ¥500.

1) Survey Design: Both surveys in the U.S. and Japan had identical content and formatting; the survey was first developed in English and then translated into Japanese. It was approved by Indiana University's Institutional Review Board (IRB).

Our survey began by showing the consent form and an age screening question. Only adults who were 65 years of age or older at the time of the study were eligible.

Next, a photo of the robot was presented with an audio prompt, instructing participants to adjust their computer speaker's volume as indicated in the on-screen instructions. No information about the robot's function or its recommendation task was provided to prevent any potential influence on participants' perception. The survey employed a between-participants design, meaning that each participant only experienced one of the three available robot voice options.

The participants then proceed to the next page where they click the "play" button to listen to a video of the robot speaking. In the video, the robot said, *"Hi, exercise always cheers me up. How about doing some exercises together?*" Participants were given the opportunity to listen to the video clip multiple times (on average, they played it 1.9 times). As an attention check, we then asked them to type what the robot said. Participants then viewed the same video clip with the same option to play it as many times as they liked (on average, they played it 1.15 times) as they evaluated it for the social attributes, willingness to take recommendations (see Section III-.3), and answered demographic questions. The whole survey took participants about 10 minutes.

2) Robot and Robot Voice Modulation: We selected the QT robot by LuxAI as the platform for this research; see Fig. 1. It is 64 cm (25 in) tall, and its arms and heads are articulated. It has a camera, microphone array, and speaker in its head. As described on LuxAI's website, QT is an "Expressive Humanoid Social Robot" [31] without mention of the gender or age of the robot. Additionally, QT is white and has a screen in its head, which allows for rich facial expressions. There are a number of language packages available including English and Japanese.

It was not feasible to use the same voice in both countries as one voice can only speak one language. With the limitation of the Sakura voice for the Japanese version, we chose the Lily voice in the U.S., as it was the closest match. Both are in the range of 165 Hz to 255 Hz [32], similar to a human female voice. Both voices were provided by the Acapela Group [33] that provides Text To Speech (TTS) service. For each country, we used the same voice but three different pitches: the child was adjusted to a pitch of 125% of the original voice (206.25 Hz to 318.75 Hz), the female was the original voice, and the male was lowered to a pitch of 75% of the original (123 Hz to 191 Hz). This produced a child's voice similar in range to a human child's voice (200 Hz to 400 Hz) and a male voice similar to that of an adult male's voice (85 Hz to 155 Hz).

The rationale behind associating particular pitches with certain ages and gender were based on previous research findings [22] that children have higher-pitched voices, females experience a drop in pitch of about one-third when they go through puberty [22], and males experience a drop in pitch of about one octave during puberty [22]. We chose to test only three voices (male, female, and child) to limit the number of variables and ensure we could obtain more meaningful and general results.

3) Measures: In the survey, we had two main measures. First, we asked participants to rate the degree of association of the robot with the adjectives of the RoSAS scale [10], with a 6-point Likert scale between "definitely not associated" and "definitely associated." RoSAS consists of 18-word descriptions that measure people's perceived robot attributes [10]. Specifically, there were three domains: warmth, competence, and discomfort. The descriptions that comprised the warmth domain were happy, feeling, social, organic, compassionate, and emotional. The descriptions that comprised the competence domain were capable, responsive, interactive, competent, and knowledgeable. The descriptions that comprised the discomfort domain were scary, strange, awkward, dangerous, awful, and aggressive.

Then, we asked participants about their *willingness to use the robot* in different scenarios, based on possible contexts and situations. There were five themes on the contexts for using the robot: lifestyle suggestions, listening to the robot, sharing with the robot, exploring artistic expression, and reaching out to the user's community. We asked the participants to rate from 1 (strongly disagree) to 6 (strongly agree). Lastly, we also asked participants if *'this voice fits the robot'* on the same 6-point Likert scale.

A. Survey analysis

First, we removed 10 participants from the analysis because they did not pass the "attention check" questions that were designed to confirm a good-faith effort to answer the survey questions. We conducted the statistical analyses using the software GraphPad Prism (San Diego, CA, USA) and Minitab (State College, PA, USA). The resulting plots display the mean ± standard deviation (SD) of the survey questions. For social attributes and willingness, comparisons of country and type of robot voice were made using two-way

TABLE I

DISTRIBUTION OF SURVEY AND DEMOGRAPHICS

	U.S.	Japan
Gender of participants		
Male	93	264
Female	132	202
Not disclosed	1	0
Type of voice		
Child	78	154
Female	78	159
Male	69	153
Age		
65-69	134	129
70-79	90	242
80-89	2	95

ANOVA with Tukey's post-test, and statistical differences are visualized by letters [34]. Means that do not share the same letter are significantly different. For instance, in Fig. 2, Child×U.S. with the letter B is significantly different from Female×U.S. with the letter A. However, Child×U.S. with letter B is not significantly different from Child×JP with letter BC, since they shared a common letter B.

The results reached statistical significance with a p-value of less than 0.05. A correlogram (Fig. 7) was used to visually compare the Pearson correlation coefficients between each pair of variables from social attributes and willingness. Minitab uses a color gradient to indicate low and high correlation coefficient values, blue to red respectively.

IV. RESULTS FROM THE SURVEY

A total of 692 surveys were collected, in which each participant heard a single robot voice. The three voice conditions were equally distributed. The distribution of the survey and demographics is listed in Table I.

A. Fit of voice to robot

U.S. OAs rated the fit of female and male robot voices to the QT robot significantly higher than the fit of the child voice (Fig. 2). Japanese OAs rated the fit of the child voice significantly higher than the fit of a male robot voice, but not significantly different from the fit of a female robot voice. There were also no differences between the fit of female and male robot voices rated by Japanese OAs. Comparing across countries, OAs in U.S. rated the fit of female and male robot voices as significantly higher compared to the ratings by Japanese OAs.

B. Views of social attribute of different voices (RoSAS)

Warmth: Items *happy* (4.08±1.50) and *social* (4.00±1.55) were rated significantly higher for all types of voices by U.S. OAs compared to Japanese OAs (*happy*: 3.03±0.94, *social*: 3.11±0.97); see Fig. 3. The female-robot voice was rated higher by U.S. OAs compared to all voice types rated by Japanese OAs for items *feeling* and *organic*. There were no significant differences between the type of voice and country for items *compassionate* and *emotional*.

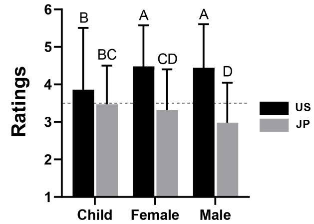

Fig. 2. Ratings on the **fit of voice** (mean ± standard deviation). Different letters (A, B, and C) mean that they are different by two-way ANOVA with Tukey's post-test analyses.

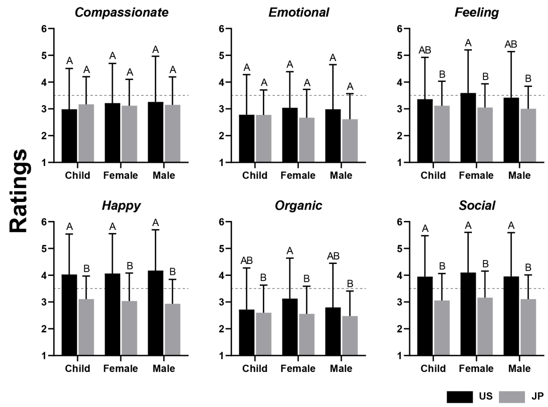

Fig. 3. RoSAS items ratings in the **warmth** domain (mean ± standard deviation). Values are shown as mean ± standard deviation. Means that share a common letter are not statistically different ($p < 0.05$) by two-way ANOVA with Tukey's post-test.

Competence: All items (*capable* (4.59±1.00), *competent* (4.49±1.12), *interactive* (4.54±1.16), *knowledgeable* (4.39±1.21), *reliable* (4.40±1.12), and *responsive* (4.36±1.20)) were rated significantly higher for all types of voices by U.S. OAs compared to Japanese OAs (*capable*: 3.38±0.97, *competent*: 3.28±0.92, *interactive*: 3.29±1.00, *knowledgeable*: 3.12±0.99, *reliable*: 3.33±0.92, *responsive*: 3.20±0.94); see Fig. 4. Within U.S. OAs for item *knowledgeable*, they rated the female-robot voice significantly higher than the child-robot voice. Within Japenese OAs for item *responsive*, they rated the female-robot voice significantly higher than male-robot voice.

Discomfort: Japanese OAs rated item *awkward* significantly higher than U.S. OAs within each type of robot voice (Fig. 5). Items *strange*, *dangerous*, and *awful* were rated significantly higher by Japanese OAs than U.S. OAs for female-robot voice. There were no significant differences within each type of voice and country for items *scary* and *aggressive*.

C. Willingness to listen for different voices

After rating the robot's social attributes, the participants rated on a scale of 1 (strongly disagree) to 6 (strongly agree), on how much would they agree with the statement for different use contexts. All use contexts were presented to the participants.

For the context of *lifestyle suggestions* (Fig. 6A), participants were to rate in statements regarding "I would listen to the robot's suggestion on exercise" and "I would listen to the robot's suggestions on meals." We found that the U.S. OAs rated themselves as significantly more willing (*exercise*:

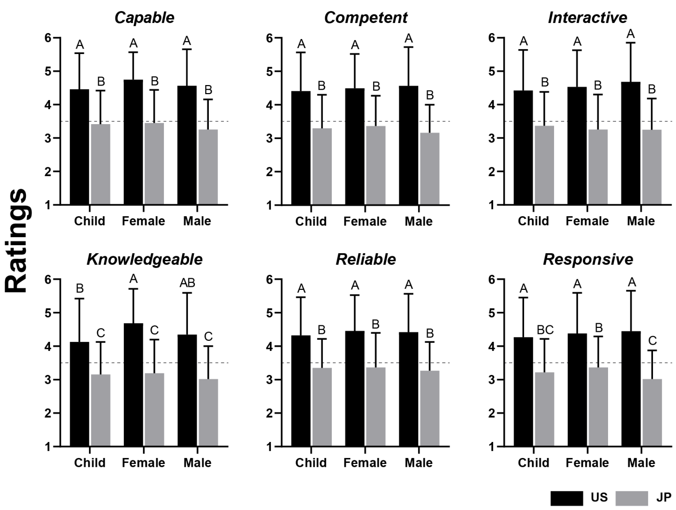

Fig. 4. Rating on RoSAS items in **competence** domain (mean ± standard deviation). Means that share a common letter are not statistically different ($p < 0.05$) by two-way ANOVA with Tukey's post-test.

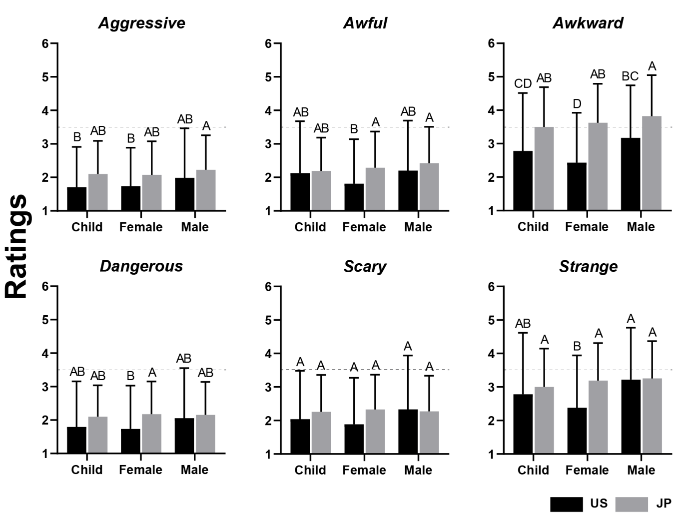

Fig. 5. RoSAS items ratings in the **discomfort** domain (mean ± standard deviation). Means that share a common letter are not statistically different ($p < 0.05$) by two-way ANOVA with Tukey's post-test.

4.23±1.33, *meals*: 4.03±1.35) than Japanese OAs (*exercise*: 3.28±1.11, *meals*: 3.10±1.04) to listen to the robot's suggestions on exercise and meals.

On the *listening to robot* (Fig. 6B) questions, the participants rated their agreement on: "I would believe the news the robot shared," and "I would listen to the robot's stories." There were no significant differences between any voice types within U.S. OAs and Japanese OAs. For both believing the news the robot shared and listening to the robot's stories, U.S. OAs rated both female-robot and male-robot voices significantly higher than Japanese OAs.

As for *sharing with robot* (Fig. 6C), participants were given the following statements: "I would share with the robot about what happened during my day" and "I would share my happiness with the robot." We found no significant differences in voice types between U.S. OAs and Japanese

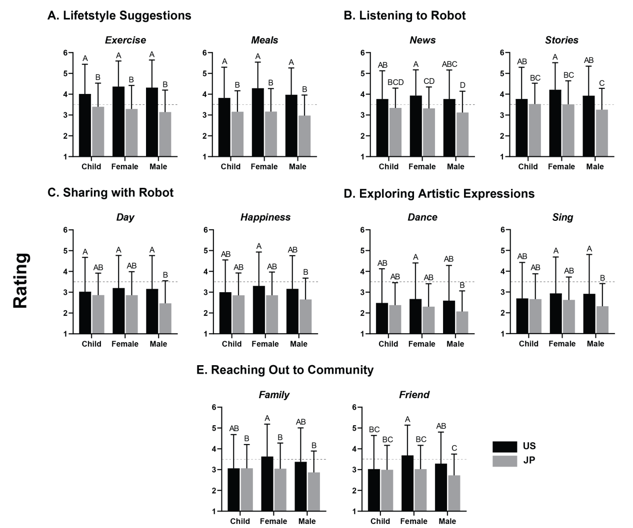

Fig. 6. Rating on questions of **willingness to use** the robot in different contexts. Values are shown as mean ± standard deviation. Means that share a common letter are not statistically different ($p < 0.05$) by two-way ANOVA with Tukey's post-test.

OAs. On the other hand, the male-robot voice was rated significantly higher by U.S. OAs than Japanese OAs regarding sharing what happened during their day.

When we asked participants for *exploring artistic expressions* (Fig. 6D)– "I would dance with the robot" and "I would sing with the robot" – we found no significant differences between any voice types for both countries. However, the male-robot voice was rated significantly higher by U.S. OAs than Japanese OAs regarding singing with the robot.

Last, for the context of *reaching out to community* (Fig. 6E), we asked participants: "I would listen to the robot's suggestion on calling my family," and "I would listen to the robot's suggestion on calling my friend." For the suggestion of calling family, the female-robot voice was rated significantly higher than all types of voice rated by Japanese OAs. There were no significant differences between voice types within U.S. OAs and Japanese OAs for the suggestion of calling family. For calling a friend, U.S OAs rated female voice significantly higher than child-voice, and significantly higher than all voice types rated by Japanese OAs. There were no significant differences between voice types within Japanese OAs for the suggestion of calling friends.

D. Correlation between social attribute and robot voice

For U.S. OAs: Participants' willingness to use the robot in the context of sharing with the robot, reaching out to the community, and exploring artistic expression, was positively correlated to the *warmth* social domain of the robot (Fig. 7). For the theme of lifestyle suggestions, willingness to accept the robot's recommendations was positively correlated to the *competence* social domain of the robot and negatively correlated to the items *strange* and *awkward* of the *discomfort* social domain. The willingness of listening to the robot's stories was negatively correlated with evaluations of the robot as *strange* and *awkward* in the *discomfort* social domain.

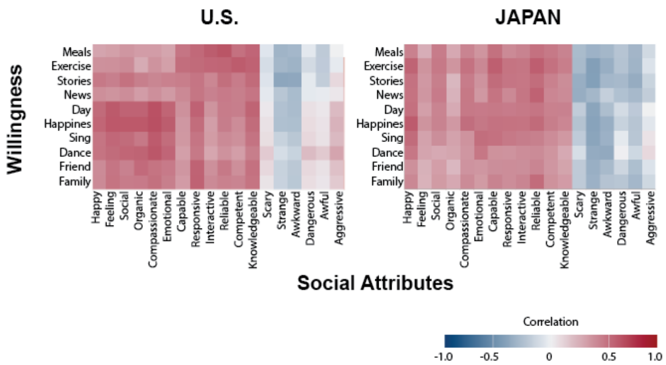

Fig. 7. Correlation between questions of **willingness** and **social attributes**.

For Japanese OAs: Willingness to share happiness with the robot was positively correlated with item *happy* from the *warmth* social domain (Fig. 7). Willingness to listen to the robot's suggestion on exercise was positively correlated with item *happy* from the *warmth* social domain, items *capable* and *reliable* from *competence* social domain, and was negatively correlated with item *strange* from *discomfort* social domain. Willingness to believe the news the robot shared was positively correlated to item *reliable* from *competence* social domain. Item *strange* from the *discomfort* social domain was negatively related to willingness on themes of sharing, listening to robot stories and suggestions on exercise. Themes of artistic expressions were negatively correlated to items *strange* and *awkward* from *discomfort* social domain.

V. Discussion

To enhance the effectiveness of suggestions provided by robots to older adults (OAs), it is crucial to consider factors such as the robot's voice, social attributes ascribed to the robot, country-specific characteristics, and the context in which the interaction takes place. Below, we discuss our findings and suggestions for robot design in detail.

A. Speaking in an adult voice fits a recommendation robot for U.S. OAs, while Japanese OAs were ambivalent

The U.S. perceived that the a female and male adult robot voice is a better fit for the robot (RQ1), which implies that U.S. OAs favored adult voices when receiving recommendations from the robot that we showed them. This could be because people generally perceive childlike voices as less reliable and competent, making them hesitant to follow the suggestions of robots with childlike voices [1] or that high pitch of the childlike voice was difficult for OAs to hear [9]. To prevent any confusion or resistance, we recommend using adult voices for robots and avoiding childlike voices when designing recommendation robots for OAs in the U.S. In other cases where we want to ensure that the robots are seen as reliable, competent, and trustworthy by the U.S. OAs, using an adult voice could be more appropriate.

On the other hand, the Japanese OAs did not have differential preferences regarding the voices of the robot (RQ1), which is congruent to previous studies on how voices had little affect on attitudes toward the robot [29]. This implies that there could be a better fit of robot voice, and possibly there could be more focus on other attributes such as physical appearance or cultural appropriateness, opening up an opportunity for future research. Nevertheless, this difference in voice fit can be attributed to cultural variations.

B. Robot voice fit, social attributes, and willingness differ across cultures: U.S. were more optimistic

The U.S. OAs rated the robot voice fit and the robot itself higher on social attributes such as *happy*, *social*, *capable*, and *interactive*, and lower in *strange* and *awkward*, compared to the Japanese OAs (RQ2). The OAs in the U.S. were also more willing to listen to the robot telling stories, news, lifestyle suggestions, and were also favorable to singing with the male-voice recommendation robot (RQ3).

All of these differences in ratings imply that cultural differences exist in OA's perception of robots, and echo prior cross-cultural research showing U.S. participants have more positive perceptions of robots with tasks that were difficult for people [19] or in companionship functions compared to Japan, Mexico, China, and other countries [20].

There are several potential reasons for the cultural differences observed in our recommended robot. One possible factor is the influence of media exposure and high expectations of robots in Japan, which may have affected the way our recommendation robot was received by Japanese OAs. People living in Japan are exposed to positive depictions of robots in the popular media [20] and in public spaces [35]. In popular media, robots are often portrayed as heroic and helpful [36], as seen in characters like Doraemon and Astro Boy. Another factor could be lifestyle-related cultural differences. Similar patterns have been observed in cross-cultural eating habits, where the U.S. population is more responsive to media-related healthy eating messages that associate food with health [37]. Preexisting biases or preconceived notions in Japan may also influence perceptions of robots, as evidenced by the tendency of older Japanese adults to prefer female appearances over male appearances for caretaking robots [38]. This preference may be attributed to the stereotype that guide robots are typically associated with female gender [25], which has been reinforced by the prevalence of female singing robots like HRP-4C Cyborg and Hatsune Miku in Japan [39], [40]. Overall, the cultural differences observed in our study may be attributed to a combination of factors, underscoring the importance of considering cultural aspects when designing and implementing recommendation robots.

C. Robot's social attributes matters: Higher positive and lower negative attributes relates to higher willingness to use

Our study also revealed a correlation between social attributes, willingness to engage with robots, and lifestyle suggestions in both countries, which addressed our fourth research question. When OAs rated higher on positive social attributes (i.e., *warmth* and *competence*) and lower on negative social attributes (i.e., *discomfort*), they have higher preferences in willingness to use the robot in different contexts

(RQ4). This thus indicates that if individuals perceived robots as being less *happy*, *feeling*, *social*, and *organic*, and more *strange* and *awkward*, they were less likely to follow lifestyle suggestions from robots. This correlation echoes previous research that examined how the personalities of robots can influence human perceptions of their interactions [41]. Specifically, individuals who perceived robots as optimistic were more likely to engage with them [41]. We suggest that HRI designers who are interested in robots providing lifestyle suggestions create robots that older adults will perceive as *warm* and *competent*. Additionally, HRI designers should not only consider robot's voice but also other factors, such as how the robot speaks and act, to increase the possibility that OA users may accept the robot's suggestions.

D. Entering the world of OAs might be challenging: Ambivalence on the willingness to use the robot in certain contexts

Another factor for HRI designers of recommendation robots to consider is OAs' willingness to engage in certain contexts. Our findings showed that OAs of both countries are hesitant to share what happened during the day and the happiness in their lives, as well as their willigness to sing or dance with the robot, regardless of the robot's voice (RQ3). These findings illustrate the need for designers to consider which suggestions might be given to OAs directly, and which might not be given at all or might be implicitly provided as topics for OAs to reflect on and make their minds up on.

There could be many reasons why OAs were hesitant about sharing personal information and singing and dancing. It could be due to OAs' privacy concerns [42] or because some OAs may feel embarrassed when singing children's songs or dancing in front of others [43], including robots. While previous research has demonstrated the benefits of using robots to engage OAs, such as in singing and dancing [44], HRI researchers and designers need to consider what kind of recommendation to give, what kind of data robots should access, and how it will be handled.

Despite the uncertainty about what types of information OAs were comfortable sharing with the robot, it is also important to acknowledge data collection can aid in personalizing the robot to individuals' needs, which could then enhance the robot's effectiveness [14]. Therefore, designers should evaluate acceptable methods of data collection that respect OAs' autonomy while providing appropriate recommendations. To address this, direct requests from the robot may not be the best approach. Instead, the robot could perform the activity and invite the OA to voluntarily join the activities it intends to suggest.

VI. Future research and limitations

Our study was not without limitations, which could provide the potential for future research. The survey used a short video clip of the robot suggesting an activity, rather than in person interaction, which may limit context and rely on participants' prior experience or imagination of robots. Perceptions of the robot can change over time [45] and may vary from robots in different forms (i.e., video vs. actual, humanoid vs. animal form). As robots increasingly become prominent in our society, future research could explore their perception and willingness to use recommendations in various contexts by testing different robot forms in different scenarios. Next, we did not use the exact same robot voice in Japan and the US, as the TTS company provided slightly different voices in the two languages. Therefore, some additional aspects of the voice aside from those that we considered – age and gender – may have affected our results. Lastly, willingness to use could be affected by participants' demographic characteristics, such as the OAs' gender [29]; however, we did not explore this topic in the paper due to space constraints.

VII. Conclusion

Our research explored how OAs in the U.S. and Japan perceive robots that make recommendations. By surveying OAs after watching a video of robots giving recommendations, we showed that in the U.S., OAs preferred adult robot voices, both female and male, over child-like voices when it came to receiving lifestyle recommendations.

We also demonstrated how important it is for the robot to portray positive social attributes and how cultural context matters. Compared to U.S. OAs, Japanese OAs rated the robot lower in fit of adult robot voice, social attributes, and were less willing to accept its recommendations, suggesting there is a need to design based on the cultural context. Moreover, we showed that OAs in both countries were ambivalent to recommendations relating to certain topics, such as sharing personal experiences and doing artistic expression with the robot. We should therefore carefully consider our approach if we want to design robots to provide recommendations for these topics. We thus propose that HRI researchers, designers, and engineers should consider voice, social attributes, culture, and recommendation topics when they design the robot, to can help it *"find its voice."*

Acknowledgement

This work was supported by Toyota Research Institute. We thank Donald S. Williamson, Grace (Yuying) Li, and Carter Japanese Market Research Network for their valuable input on the survey design.

References

[1] X. Dou, L. Yan, K. Wu, and J. Niu, "Effects of voice and lighting color on the social perception of home healthcare robots," *Applied Sciences*, vol. 12, no. 23, p. 12191, 2022.

[2] K. Kühne, M. H. Fischer, and Y. Zhou, "The human takes it all: Humanlike synthesized voices are perceived as less eerie and more likable. evidence from a subjective ratings study," *Frontiers in Neurorobotics*, vol. 14, p. 105, 2020.

[3] E. Nichols, S. R. Siskind, W. Kamino, S. Šabanović, and R. Gomez, "Iterative design of an emotive voice for the tabletop robot Haru," in *Proc. of the 13th Intl. Conf. on Social Robotics (ICSR)*. Springer, Nov 2021, pp. 362–374.

[4] M. Siegel, C. Breazeal, and M. I. Norton, "Persuasive robotics: The influence of robot gender on human behavior," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems*. IEEE, 2009, pp. 2563–2568.

[5] C. McGinn and I. Torre, "Can you tell the robot by the voice? An exploratory study on the role of voice in the perception of robots," in *ACM/IEEE Intl. Conf. on Human-Robot Interaction*, 2019, pp. 211–221.

[6] S.-E. Chien, L. Chu, H.-H. Lee, C.-C. Yang, F.-H. Lin, P.-L. Yang, T.-M. Wang, and S.-L. Yeh, “Age difference in perceived ease of use, curiosity, and implicit negative attitude toward robots,” *ACM Transactions on Human-Robot Interaction (THRI)*, vol. 8, no. 2, pp. 1–19, 2019.

[7] M. Gilligan, A. Karraker, and A. Jasper, “Linked lives and cumulative inequality: A multigenerational family life course framework,” *J. of Family Theory & Review*, vol. 10, no. 1, pp. 111–125, 2018.

[8] Y. Uchida, S. Sugiura, Y. Nishita, N. Saji, M. Sone, and H. Ueda, “Age-related hearing loss and cognitive decline—the potential mechanisms linking the two,” *Auris Nasus Larynx*, vol. 46, no. 1, pp. 1–9, 2019.

[9] M. A. Farage, K. W. Miller, F. Ajayi, and D. Hutchins, “Design principles to accommodate older adults,” *Global J. of Health Science*, vol. 4, no. 2, p. 2, 2012.

[10] C. M. Carpinella, A. B. Wyman, M. A. Perez, and S. J. Stroessner, “The robotic social attributes scale (RoSAS) development and validation,” in *Proc. of ACM/IEEE Intl. Conf. on Human-Robot Interaction*, 2017, pp. 254–262.

[11] X. Wang, Z. Lu, and M. Yin, “Will you accept the AI recommendation? Predicting human behavior in AI-assisted decision making,” in *Proc. of the ACM Web Conf.*, 2022, pp. 1697–1708.

[12] C. Edwards, A. Edwards, B. Stoll, X. Lin, and N. Massey, “Evaluations of an artificial intelligence instructor’s voice: Social identity theory in human-robot interactions,” *Computers in Human Behavior*, vol. 90, pp. 357–362, 2019.

[13] N. Randall, S. Joshi, W. Kamino, L.-J. Hsu, A. Agnihotri, G. Li, D. Williamson, K. M. Tsui, and S. Šabanović, “Finding ikigai: How robots can support meaning in later life,” *Frontiers in Robotics and AI*, vol. 9, 2022.

[14] W. Khoo, L.-J. Hsu, K. J. Amon, P. V. Chakilam, W.-C. Chen, Z. Kaufman, A. Lungu, H. Sato, E. Seliger, M. Swaminathan, K. M. Tsui, D. J. Crandall, and S. Sabanović, “Spill the tea: When robot conversation agents support well-being for older adults,” in *Companion of ACM/IEEE Intl. Conf. on Human-Robot Interaction*, 2023, pp. 178–182.

[15] N. Randall, W. Kamino, A. Karim, W.-C. Chen, L.-J. Hsu, K. M. Tsui, and S. Sabanovic, “‘ikigai’ robots: Designing for direct benefits to older adults and indirect benefits to caregivers,” in *Companion of ACM/IEEE Intl. Conf. on Human-Robot Interaction*, 2023, pp. 701–705.

[16] M. R. Fraune, T. Komatsu, H. R. Preusse, D. K. Langlois, R. H. Au, K. Ling, S. Suda, K. Nakamura, and K. M. Tsui, “Socially facilitative robots for older adults to alleviate social isolation: A participatory design workshop approach in the US and Japan,” *Frontiers in Psychology*, p. 6256, 2022.

[17] S. Rossi, M. Staffa, and A. Tamburro, “Socially assistive robot for providing recommendations: Comparing a humanoid robot with a mobile application,” *Intl. J. of Soc Robotics*, vol. 10, pp. 265–278, 2018.

[18] S. J. Shim and A. Kumar, “Development of the robotics industry in the United States and Japan: a comparison of competitive factors,” *Competitiveness Review: An Intl. Business Journal*, vol. 16, no. 1, pp. 57–66, 2006.

[19] T. Nomura, T. Suzuki, T. Kanda, J. Han, N. Shin, J. Burke, and K. Kato, “What people assume about humanoid and animal-type robots: Cross-cultural analysis between Japan, Korea, and the United States,” *Intl. J. of Humanoid Robotics*, vol. 5, no. 01, pp. 25–46, 2008.

[20] C. Bartneck, T. Suzuki, T. Kanda, and T. Nomura, “The influence of people’s culture and prior experiences with Aibo on their attitude towards robots,” *AI & Society*, vol. 21, pp. 217–230, 2007.

[21] P. P. Rau, Y. Li, and D. Li, “Effects of communication style and culture on ability to accept recommendations from robots,” *Computers in Human Behavior*, vol. 25, no. 2, pp. 587–595, 2009.

[22] J. Abitbol, P. Abitbol, and B. Abitbol, “Sex hormones and the female voice,” *J. of Voice*, vol. 13, no. 3, pp. 424–446, 1999.

[23] R. B. Jackson, T. Williams, and N. Smith, “Exploring the role of gender in perceptions of robotic noncompliance,” in *Proc. of ACM/IEEE Intl. Conf. on Human-Robot Interaction*, 2020, pp. 559–567.

[24] S. Song, J. Baba, J. Nakanishi, Y. Yoshikawa, and H. Ishiguro, “Mind the voice!: Effect of robot voice pitch, robot voice gender, and user gender on user perception of teleoperated robots,” in *Extended Abstracts of CHI Conf. on Human Factors in Computing Systems*, 2020, pp. 1–8.

[25] T. Nomura and Y. Kinoshita, “Gender stereotypes in cultures: Experimental investigation of a possibility of reproduction by robots in Japan,” in *2015 Intl. Conf. on Culture and Computing (Culture Computing)*. IEEE, 2015, pp. 195–196.

[26] X. Dou, C.-F. Wu, K.-C. Lin, and T.-M. Tseng, “The effects of robot voice and gesture types on the perceived robot personalities,” in *Human-Computer Interaction. Perspectives on Design: Thematic Area, HCI 2019, Held as Part of the 21st HCI Intl. Conf., HCII 2019, Proceedings, Part I 21*. Orlando, FL, USA: Springer, July 2019, pp. 299–309.

[27] M. P. Michalowski, S. Sabanovic, and R. Simmons, “A spatial model of engagement for a social robot,” in *9th IEEE Intl. Workshop on Advanced Motion Control*. IEEE, 2006, pp. 762–767.

[28] E. Hughson, P. Tuttösí, A. Matsufuji, C. Zhang, and A. Lim, “I’m a robot, hear me speak!” in *Companion of ACM/IEEE Intl. Conf. on Human-Robot Interaction*, 2023, pp. 909–911.

[29] S. C. Steinhaeusser, P. Schaper, O. Bediako Akuffo, P. Friedrich, J. Ön, and B. Lugrin, “Anthropomorphize me! effects of robot gender on listeners’ perception of the social robot NAO in a storytelling use case,” in *Companion of ACM/IEEE Intl. Conf. on Human-Robot Interaction*, 2021, pp. 529–534.

[30] M. R. Fraune, S. Kawakami, S. Sabanovic, P. R. S. De Silva, and M. Okada, “Three’s company, or a crowd?: The effects of robot number and behavior on HRI in Japan and the USA,” in *Robotics: Science and Systems*, vol. 10, 2015.

[31] LuxAI, “QTrobot V2 humanoid social robot for human AI research & teaching,” Last Accessed December 7, 2022. [Online]. Available: https://luxai.com/qtrobot-for-research/

[32] R. J. Baken and R. F. Orlikoff, *Clinical measurement of speech and voice*. Cengage Learning, 2000.

[33] [Online]. Available: https://www.acapela-group.com/

[34] W. Y. Khoo, B. J. Chrisfield, S. Sae-Tan, and J. D. Lambert, “Mitigation of nonalcoholic fatty liver disease in high-fat-fed mice by the combination of decaffeinated green tea extract and voluntary exercise,” *J. of Nutritional Biochemistry*, vol. 76, p. 108262, 2020.

[35] W. Kamino and S. Sabanovic, “Coffee, tea, robots? The performative staging of service robots in ‘robot cafes’ in Japan,” in *ACM/IEEE Intl. Conf. on Human-Robot Interaction*, 2023, pp. 183–191.

[36] F. Diana, M. Kawahara, I. Saccardi, R. Hortensius, A. Tanaka, and M. E. Kret, “A cross-cultural comparison on implicit and explicit attitudes towards artificial agents,” *Intl. J. of Soc Robotics*, pp. 1–17, 2022.

[37] S. R. Hawks, H. N. Madanat, R. M. Merrill, M. B. Goudy, and T. Miyagawa, “A cross-cultural analysis of ‘motivation for eating’as a potential factor in the emergence of global obesity: Japan and the United States,” *Health Promotion International*, vol. 18, no. 2, pp. 153–162, 2003.

[38] T. Suzuki and T. Nomura, “Gender preferences for robots and gender equality orientation in communication situations,” *AI & Society*, pp. 1–10, 2022.

[39] K. Kaneko, F. Kanehiro, M. Morisawa, K. Miura, S. Nakaoka, and S. Kajita, “Cybernetic human hrp-4c,” in *9th IEEE-RAS Intl.Conf. on Humanoid Robots*. IEEE, 2009, pp. 7–14.

[40] Y. Sone and Y. Sone, “Hatsune miku, virtual machine-woman,” *Japanese Robot Culture: Performance, Imagination, and Modernity*, pp. 139–166, 2017.

[41] M. Paetzel, G. Perugia, and G. Castellano, “The influence of robot personality on the development of uncanny feelings towards a social robot,” *Comput. Hum. Behav.*, 2020.

[42] H. Draper and T. Sorell, “Ethical values and social care robots for older people: an international qualitative study,” *Ethics and Information Technology*, vol. 19, no. 1, pp. 49–68, 2017.

[43] W. R. Crozier, “Embarrassment,” in *Blushing and the social emotions: The self unmasked*. Springer, 2006, pp. 97–114.

[44] R. Khosla, K. Nguyen, and M.-T. Chu, “Human robot engagement and acceptability in residential aged care,” *Intl. J. of Human–Computer Interaction*, vol. 33, no. 6, pp. 510–522, 2017.

[45] C. Bartneck, “Who like androids more: Japanese or US Americans?” in *17th IEEE Intl. Symp. on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 2008, pp. 553–557.