Iterative Design of an Emotive Voice for the Tabletop Robot Haru

Eric Nichols[1(✉)] , Sarah Rose Siskind[2] , Waki Kamino[3] ,
Selma Šabanović[3] , and Randy Gomez[1]

[1] Honda Research Institute Japan Co., Ltd., Wako, Japan
{e.nichols,r.gomez}@jp.honda-ri.com
[2] Hello SciCom, New York, USA
sarah@hellosci.com
[3] Luddy School of Informatics, Computing, and Engineering,
Indiana University Bloomington, Bloomington, USA
wkamino@iu.edu, selmas@indiana.edu

Abstract. Designing a voice for a social robot is particularly challenging because the voice needs to convincingly convey a target personality while maintaining rich, emotive capabilities in order to foster the development of bonds with humans. In this paper, we describe the ongoing design and implementation process of a voice for a social robot. To aid in our design and analysis, we identify three desirable characteristics for its voice: 1. *convincingness*, 2. *emotiveness*, and 3. *consistency*. In this paper, we present a preliminary study that investigates convincingness by comparing samples taken from human voice talents and eliciting human judgements about their appropriateness. This study compares human judgements, elicited through surveys, on a range of characteristics related to *convincingness*, emotions conveyed, and impressions of the overall consistency of the voice. Finally, we discuss the implications of the survey findings for designing a voice for a social robot.

Keywords: Social robot interaction · TTS voice design · Behavior design

1 Introduction

Designers regularly use various robot characteristics, such as appearance, speech, and behavior, to inspire people to engage with robots as social actors. Robot appearance, including the use of familiar human- and animal-like forms, can visually represent a robot's social capabilities. Similarly, the content of verbal utterances (e.g. saying "Hello") can explicitly invite people to engage with the robot socially. More subtle cues to a robot's social abilities can be provided through the nonverbal characteristics of the robot's voice – implicit age, gender, emotional expression, markers of cultural origin, and individual vocal quirks can evoke a certain type of character and unique social presence for a robot.

Studies in social robotics and human-robot interaction have explored how various aspects of a robot's voice can affect people's perceptions of the robot's

Fig. 1. The social robot Haru.

capabilities, personality, and appropriateness for different tasks. People deem certain voices as more or less appropriate for specific robots [15]. Robots with higher levels of vocal expressiveness are perceived as having more social presence [12]. The perceived age of a robot's voice can affect people's perceptions of its credibility and social presence [8]. Robots with child-like voices can be seen as more extroverted and relaxed [7]. Users also respond to the perceived gender of human-like robot voices, reporting more positive attitudes about robots with which they share a gender [9]. Additionally, the delivery of verbal content as manipulated through voice pitch, empathy, and humor, has been found to not only affect people's perceptions of and attitudes towards the robot, but also their enjoyment of the human-robot interaction task [16].

While prior research has identified particular aspects of robot voice that affect users' interaction experiences and robot evaluations in short term scenarios, long term interaction will also require users to see the robot as a believable, relatable, and cohesive social agent. Based on the above-mentioned prior research, a combination of various vocal aspects (e.g. age, gender, vocal style/genre) will need to be employed to create a convincing social presence for robots. With this idea of long-term, companionable social interaction with robots in mind, this paper describes the process of designing an expressive voice for the Haru robot through the selection of diverse voice talent. It then explores how users evaluate different text-to-speech voices, meant to be used with the Haru social robot, in relation to their appropriateness for the robot, as well as their convincingness, expressiveness, and cohesiveness. The broader aim of this work is to inform the design of engaging and persuasive social characters for companion social robots.

As our target robot, we selected the social robot Haru [10,11], shown in Fig. 1. Haru is an experimental tabletop robot for multimodal communication that uses verbal and non-verbal channels for interactions. Haru's design is centered on its potential to communicate emotions through richness in expressivity [10].
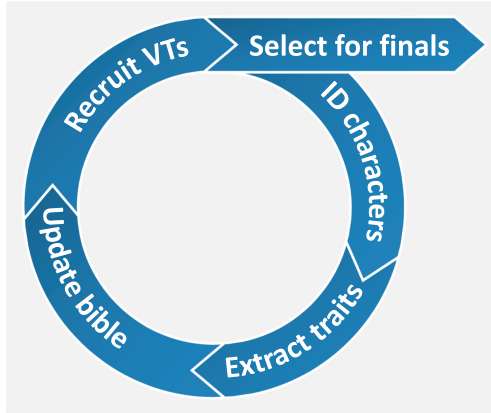
Fig. 2. Iterative refinement design process for Haru's personality and voice concept.

Haru has five motion degrees of freedom (namely base rotation, neck leaning, eye stroke, eye rotation and eyes tilt) that allow it to perform expressive motions. Furthermore, each eye includes a 3-inch TFT screen display in which the robot eyes are displayed. Inside the body there is an addressable LED matrix (the mouth). Haru can communicate via text-to-speech (TTS)—albeit currently with off-the-shelf voices—through animated routines, projected screen, etc. Haru's range of communicative strategies positions the robot as a potent embodied communication agent that can support long-term interaction with people.

This paper is organized as follows: in Sect. 2, we describe our iterative design process of refining our social robot's personality description while auditioning voice talents; in Sect. 3, we present an elicitation survey that evaluates a select number of voice talent finalists; in Sect. 4, we discuss the findings of our survey; in Sect. 5, we outline relevant work in social robots; and, finally, in Sect. 6, we recap our findings and discuss future work.

2 Design Process

One of our core research topics is the development of a long-term robotic companion, which can lead to the forging of a bond between a human and a social robot similar to the bond shared by other social creatures [11]. But this goal requires a persuasive character with rich expressivity that is beyond the capability of conventional TTS systems. We identify three characteristics that an ideal TTS voice for a social robot should have:

Convincingness. The voice should fit the robot's character, physical appearance, and application scenarios.

Emotiveness. The voice should be capable of conveying a wide range of emotions and vocal delivery styles.
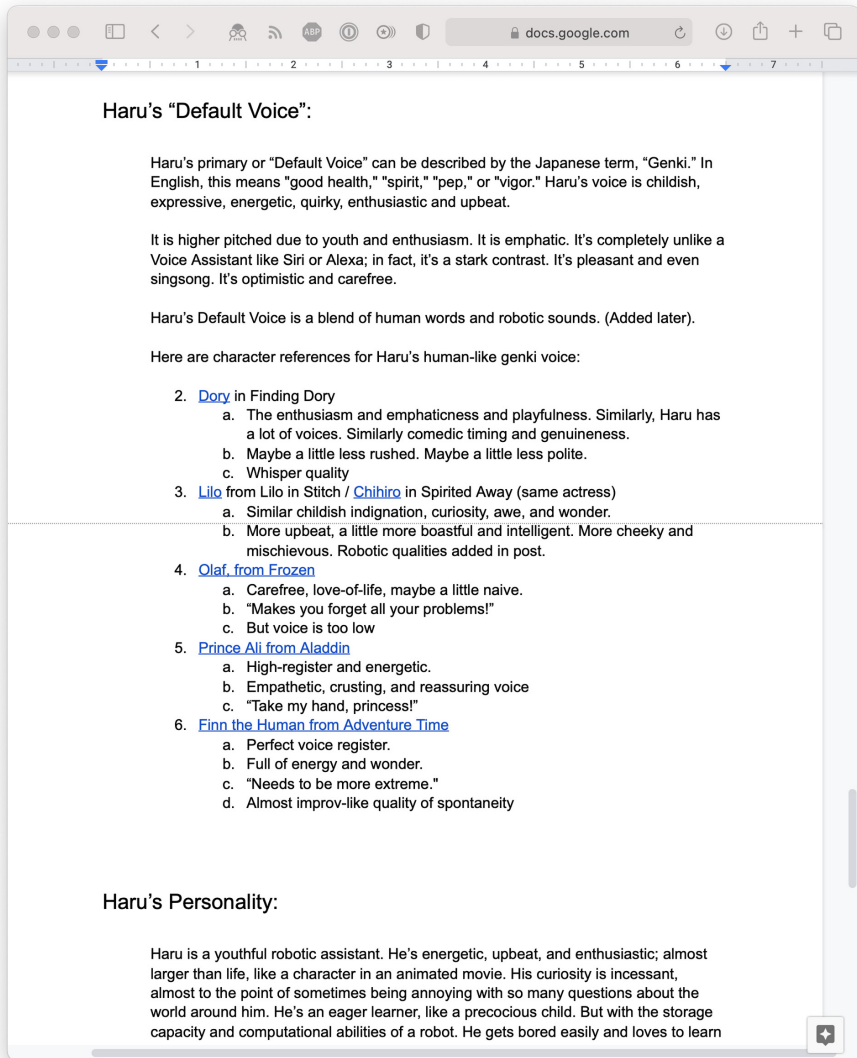
Fig. 3. Haru's personality bible.

Consistency. Throughout its application, the voice should sound like it seamlessly belongs to a single entity.

We adopt a holistic approach to designing Haru's personality and voice concept based on a process of iterative refinement, where updates to the personality definition feed into recruitment and evaluation of voice talent, and their evaluation informs refinement of the personality definitions. To aid us in organizing

---

HARU SELF-INTRODUCTION

HARU (UPBEAT)
Hi there! I'm Haru, your friendly robot companion. Would you like
to go find some buried treasure with me?

HARU (SAD)
Nobody else wanted to go on an adventure with me. And I don't want
to go alone.

HARU (SERIOUS)
But I promise you it's real and it's somewhere in this room. And
we gotta find it right now.

HARU (WHINY)
Because I don't want to wait anymore. So let's go right now!

HARU (CHEEKY)
Unless you're scared, that is. I think you're just being chicken.

Fig. 4. Haru's self-introduction script.

Haru's personality definition, we borrow a practice from screenwriting and keep a *personality bible* [13,14] for Haru, recording important personality traits, reference characters, and other relevant background information. The personality bible is a reference document for writers and engineers to check in order to keep Haru's personality consistent. It outlines information about Haru's background (his fear of social isolation and magnets, for example). It also outlines how Haru speaks (enthusiastically and informally). An excerpt is shown in Fig. 3. The iterative refinement process is shown in Fig. 2, and we describe it below.

Identify Reference Characters. Based on our existing vision of Haru's personality, we brainstorm reference characters that effectively convey some aspect of Haru's personality. Some examples include *Prince Ali* from *Aladdin*, *Dory* from *Finding Dory*, and *Finn the Human* from *Adventure Time*.

Extract personality traits. We summarize the relevant personality traits of the reference characters from Step 1 into keywords. We consider behavioral traits as well as vocal characteristics and speaking mannerisms. For example, Haru has the enthusiasm and empathy of Dory, the trusting and reassuring cadence of Prince Ali, and the energy and childish sense of wonder of Finn the Human.

Update Personality Bible. We update Haru's personality bible with reference characters from Step 1. and relevant personality trait keywords from Step 2.

Recruit and Audition voice talents. We recruit voice talents online and audition them through interactive table reads using scenario scripts showcasing Haru's personality. Voice talents are shown Haru's personality bible and

Table 1. Voice talent search finalists selected for the elicitation study.

Alias	Nationality	Age	Gender
Voice A	United States	child	Male
Voice B	Australia	20 s	Female
Voice D	Philippines	20 s	Male
Voice G	United States	50 s	Male
Voice M	United States	40 s	Male

coached to convey our vision of Haru while encouraging creativity in their portrayal.

Select voice talents for Finals. Finally, we analyze the results of the audition in Step 4., considering both the quality of their portrayal and how Haru's personality bible could be refined. If the voice talent is deemed satisfactory, we select them for the final evaluation. We then go back to Step 1. and repeat the process until we have gathered enough voice talents for the final evaluation.

2.1 Voice TalentSearch

We searched for playful, energetic, curious voices from adult males, adult females and young children. The search took place over several months. Roughly 30 voices were researched closely and 10 voice talents were recorded, of which three children and two women were actively considered. While a mastery of English was required, the talent was sourced from all over the world. Given the diversity of Haru's intended audience, we wanted voice talent from diverse backgrounds to counteract any regional idiosyncrasies.

The desirability of the voice talent was measured across several criteria. The first was a youthful quality. This quality is impossible to achieve by simply raising the pitch of an adult voice. The effect of having smaller vocal chords produces a slightly raspy, at times even nasal quality that is completely unique to young children. The second criteria was emotiveness. The voice talent had to be capable of conveying a slightly exaggerated degree of emotion. This exaggeration is important because much of the nuance of a performance is 'lost in translation' in the voice capture process. The third criteria is technical ability. The voice talent (VT) needs to be of a certain technical reading level to get through the material required of the voice capture process. Given that the VT may be a child, however, a certain amount of stumbling and coaching is expected. The fourth criteria is coachability. The VT needs to be able to take direction well in order to calibrate a performance correctly.

2.2 Audition Process

Each voice talent set aside an hour to go through specially designed audition scripts (see Fig. 4 for an example) to test the talent's range of emotion, techni-

Table 2. Target characteristics in the voice talent elicitation study.

Characteristic	Question
Suitability	On a scale of 1 (*poor*) to 5 (*great*), how well does this voice fit Haru?
Expressiveness	On a scale of 1 (*expressionless*) to 5 (*expressive*), how expressive is this?
Naturalness	On a scale of 1 (*artificial*) to 5 (*natural*), how natural is this voice?
Friendliness	On a scale of 1 (*unfriendly*) to 5 (*friendly*), how friendly is this voice?
Interestingness	On a scale of 1 (*boring*) to 5 (*interesting*), how interesting is this voice?
Energy	On a scale of 1 (*unenergetic*) to 5 (*energetic*), how energetic is this?
Enthusiasm	On a scale of 1 (*unenthusiastic*) to 5 (*enthusiastic*), how enthusiastic is this voice?
Curiosity	On a scale of 1 (*uninterested*) to 5 (*curious*), how curious is this voice?
Empathy	On a scale of 1 (*uncaring*) to 5 (*caring*), how empathetic is this voice?
Youthfulness	On a scale of 1 (*old*) to 5 (*young*), how youthful is this voice?
Gender	On a scale of 1 (*masculine*) to 3 (*neutral*) to 5 (*feminine*), what gender is this voice?

cal reading ability, endurance, and ability to take direction. The auditions were interactive with the writer playing Randy, the human, and the voice talent playing Haru. The audio of the dialogues was recorded for later comparison. Through the video call, the writer was able to give directions to the talent. For example: smile during an upbeat performance or slightly grit your teeth to convey seriousness. It was often required for the writer to give a line reading for the talent to imitate. This also had the effect of keeping the performances relatively consistent across all the different auditions. Finalists are summarized in Table 1.

3 Elicitation Study

To evaluate the convincingness of our voice talent finalists, we conducted an online survey where participants evaluated them over a variety of characteristics.

3.1 Online Survey

To familiarize themselves with Haru's appearance and personality, survey participants first watched a video of Haru non-verbally interacting with an off-screen human. Non-verbal interaction was selected to avoid preconceptions about Haru's voice. Next, for each voice talent, participants listened to a short clip of them from a short Haru self-introduction script designed to be representative of Haru's personality and desired emotive range (see Fig. 4). Then, they were asked to rate each voice for a series of target characteristics on a scale of 1–5 (see Table 2). Participants also selected all emotions conveyed from a list 8 of emotions from Plutchik's circumplex model [17]. Finally, they rated the overall suitability of the voice for Haru, and were asked for their free-form opinions on appropriateness of the voice and about Haru. The survey was conducted over Google Forms, and a sample form with synthetic responses can be seen at this link.

Table 3. Average scores for each VT by characteristic. The highest score for each characteristic is shown in **bold**. Standard deviations are given in (parentheses).

Char./Voice	A	B	D	G	M
Suitability	**3.89** (± 1.32)	3.12 (± 1.15)	3.04 (± 1.46)	2.75 (± 1.37)	3.39 (± 1.29)
Expressiveness	**4.65** (± 0.64)	4.00 (± 0.91)	4.26 (± 1.08)	4.05 (± 1.11)	4.32 (± 0.93)
Naturalness	**4.07** (± 1.02)	3.72 (± 1.24)	3.74 (± 1.32)	3.49 (± 1.45)	4.05 (± 1.12)
Friendliness	**4.42** (± 0.78)	3.75 (± 0.96)	3.44 (± 1.30)	3.51 (± 1.12)	4.09 (± 1.02)
Interestingness	**4.21** (± 0.92)	3.63 (± 1.14)	3.55 (± 1.25)	3.54 (± 1.28)	3.79 (± 1.25)
Energy	**4.77** (± 0.42)	3.73 (± 1.07)	4.09 (± 0.98)	3.93 (± 1.08)	3.82 (± 1.27)
Enthusiasm	**4.72** (± 0.56)	3.89 (± 1.10)	4.00 (± 1.16)	3.75 (± 1.14)	3.70 (± 1.19)
Curiosity	**4.16** (± 0.92)	3.73 (± 1.12)	3.64 (± 1.27)	3.81 (± 1.13)	3.70 (± 1.14)
Empathy	**3.93** (± 1.00)	3.46 (± 1.20)	3.23 (± 1.30)	3.34 (± 1.13)	3.61 (± 1.11)
Youthfulness	**4.74** (± 0.64)	3.60 (± 1.15)	3.30 (± 1.22)	2.12 (± 1.09)	2.44 (± 1.23)
Gender	2.28 (± 1.16)	**3.40** (± 1.35)	2.04 (± 1.19)	1.21 (± 0.49)	1.21 (± 0.62)

3.2 Protocol

We posted an advertisement and recruited participants on the US-based crowd-sourcing platform Upwork over three days in the month of July 2021. The platform allows interested workers to send a 'proposal' to the client who has posted the job. We recruited participants from those who submitted the proposal on a-first-come-first-served basis, with a preference given to users with higher ratings and consideration for geographical diversity. We initially collected survey responses from 61 participants, of which 57 were analyzed after filtering the data using a comprehension-check question requiring the completion of a brief video to answer correctly. The responses of any participant who answered incorrectly (n = 4) were discarded. The survey was expected to take approximately 30 min to complete, and those who completed the task were offered a fixed amount of $20 for their participation.

The demographics of the participants was as follows. We had more female participants (n = 34) than male participants (n = 23) and the most common age group was 18–30 (n = 37), followed by 30–40 (n = 16); 40–50 (n = 2); above 50 (n = 2). In terms of geographic location, 57 participants participated in the study from 27 different countries, with the largest number of participants from Asia (n = 26), followed by Europe (n = 12); Africa (n = 12); North America (n = 4); and South America (n = 3).

We used Upwork to recruit participants for the following reasons: 1) access to a diverse population of participants [2,4,6], and 2) an expected level of response quality, based on the platform's profile-oriented nature that reveals names and location of the workers, and the mutual-rating system between clients and workers.)

4 Discussion

Results. Average scores for characteristics are summarized in Table 3. Voice talent preferences for each characteristic are given as orderings annotated with statistical significance in Table 4. Select comments on voice talent suitability are given in Table 5. Finally, recognized emotions are summarized in Table 6.

Table 4. VT preference orderings, where $\gg$: $p < 0.05$, and $>$: $0.05 \leq p \leq 0.45$, and $\approx$: $p > 0.45$, as measured via a single-tail t-test. *Masculinity* and *femininity* are derived from the *gender* scores.

Characteristic	Ordering								
Suitability	A	$\gg$	M	$>$	B	$>$	D	$\gg$	G
Expressiveness	A	$\gg$	M	$>$	D	$>$	G	$\gg$	B
Naturalness	A	$\approx$	M	$>$	D	$\approx$	B	$\gg$	G
Friendliness	A	$\gg$	M	$\gg$	B	$>$	G	$\gg$	D
Interestingness	A	$\gg$	M	$>$	B	$>$	D	$\gg$	G
Energy	A	$\gg$	D	$>$	G	$>$	M	$\gg$	B
Enthusiasm	A	$\gg$	D	$>$	B	$>$	G	$\gg$	M
Curiosity	A	$\gg$	G	$>$	B	$>$	M	$\gg$	D
Empathy	A	$>$	M	$>$	B	$>$	G	$\gg$	D
Youthfulness	A	$\gg$	B	$>$	D	$\gg$	M	$\gg$	G
Masculinity	M	$\approx$	G	$\gg$	D	$>$	A	$\gg$	B
Femininity	B	$\gg$	A	$>$	D	$\gg$	G	$\approx$	M

Characteristics. We find that with the exception of overall *suitability* and demographic-related characteristics (*youthfulness*, *gender*), all were ranked positively, confirming their importance.

Demographics. Survey participants ranked voices by *youthfulness* in the same order as the voice talent's ages. *Gender* exhibited a similar trend: the single female voice talentwas ranked as most feminine, followed by the child voice talent. This is understandable, as children have higher-pitched voices than adults, and are often perceived as more feminine. Likewise, the voices of the men over 40 were ranked as most masculine.

Emotions. Voices with higher *acceptability* tend to have more emotions detected, supporting our theory that expressive voices are preferred for social robots. We also note that overall positive emotions (e.g. surprise, joy, anticipation, trust) are recognized more than negative ones (e.g. anger, fear, sadness, disgust), although this may be due to the contents of Haru's self-introduction script.

Overall. Survey participants overwhelmingly preferred Voice A to all others across all characteristics with statistical significance[1]. This supports our intuition that young, energetic voices are ideal and provides confirmation of our design direction for Haru's voice.

5　Related Research

Much research on expressive TTS has focused on evaluating emotion conveyance with robots or virtual avatars. Breazeal evaluated the expression of emotion in

---

[1] $p < 0.05$ as measured by a single-tail t-test.

Table 5. Select comments on voice talent suitability from study participants.

Voice A	
1.	This voice fits Haru because of how it attracts your attention the moment you hear it, and this is because of how young, cheerful, and interesting the voice is. It's always pleasing to hear the voice/sound of a young child talking. It makes Haru more relatable & adorable
2.	From the previous video, Haru seems to be cheerful and spontaneous, as well as being quite charismatic and expressive. That's why the voice of an 8–10 year old with these characteristics is ideal for him
3.	The robot is able to express a number of emotions. You can sense it being empathetic and also emotional. There is also fear in its voice as well as anticipation
4.	He doesn't sound like a robot/AI which is good, making the robot more human and a more realistic social companion
Voice B	
1.	For me, the voice is good but it just doesn't fit Haru. It sounded like a female's voice but I see Haru more masculine
2.	It sounds like an adult female that is trying to sound like a child
3.	I think for Haru a female voice is better
4.	It doesn't sound natural, but the voice is still great
Voice D	
1.	I find this voice is good because it was full of emotion in this voice. It does make me feel annoying at the last second of the video which make it real
2.	i can feel anger and sadness in the voice
3.	He sounds like a teenage boy, a little bit aggressive but that's okay
4.	The voice reminds me of a character from Digimon
Voice G	
1.	I feel this voice is in a haste/hurry and I can't really relate it and what he's saying. Although the voice is quite energetic and all but the relation/friendliness is not there for me
2.	Its sort of urgent adventure, perfectly fit tone of voice for a teen
3.	Not a great voice, looks like an old person pretending to be a young person
4.	The change in tone of voice, from loud to whisper, tells that there is something hidden in his offer
Voice M	
1.	The expressiveness of this voice is clear. It also has a more human voice to it which makes it relatable as well
2.	This sounds like artificial, although it is better than the others except the first one but again the voice age doesn't suits the content as it is better suited like a child voice
3.	I feel that the voice could fit well if it was a bit younger. It's a bit coarse but interesting to listen to.
4.	This voice fits Haru. You just start smiling the moment you listen to it

Table 6. Emotions recognized by study participants.

Emotion/Voice	A	B	D	G	M
Surprise	**23**	16	16	19	20
Joy	**41**	21	26	21	28
Anticipation	**43**	31	32	31	25
Anger	6	8	**23**	10	5
Fear	9	13	**19**	18	14
Trust	16	14	9	16	**17**
Sadness	25	25	22	**26**	21
Disgust	4	7	**17**	7	7
None of the above	0	2	0	3	4

TTS for anthropomorphic robots with an analysis of vocal affect [5]. Tang et al. evaluated emotive TTS with a 3-D virtual avatar [19]. Roehling and authors present a summary of research on vocal correlates in expressive speech and examine available TTS to use in their robotic project by applying the findings. Authors suggest that factors including pitch, duration, loudness, spectral energy structure, and voice quality are crucial for an expressive speech synthesis [18] Barnes et al. [3] compared the effectiveness of human voices and synthesized voices for use with humanoid and dinosaur robots and found that monotone synthesized voices were unsuitable for emotion-rich interactions. Through an online evaluation of TTS for three social robots, Alonso-Martin and authors suggest the correlation between intelligibility and expressiveness of TTS systems, as well as an inverse correlation between these two factors and artificiality [1].

6   Conclusion

In this paper, we described an iterative refinement process for developing a social robot's personality while auditioning voice talents. Through this process, we selected five finalist voice talents and evaluated them through an online elicitation study. The preferences exhibited by participants toward young, energetic, and expressive voices provided supporting evidence for our design direction.

In future work, we plan to continue to refine our definition of Haru's personality and to conduct a more detailed survey that includes evaluation of voice appropriateness given the context of specific applications for Haru. We will also limit the survey to the three regions where Haru is being considered for deployment: the US & Canada, Europe, and Japan in order to account for differences in cultural perception. Finally, we plan to finalize our voice talent selection and carry out the development and evaluation of an expressive TTS for Haru.

References

1. Alonso-Martín, F., Malfaz, M., Castro-González, Á., Castillo, J.C., Salichs, M.A., et al.: Online evaluation of text to speech systems for three social robots. In: Salichs, M.A. (ed.) ICSR 2019. LNCS (LNAI), vol. 11876, pp. 155–164. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-35888-4_15

2. Barchard, K.A., Williams, J.: Practical advice for conducting ethical online experiments and questionnaires for united states psychologists. Behav. Res. Methods **40**(4), 1111–1128 (2008)

3. Barnes, J., Richie, E., Lin, Q., Jeon, M., Park, C.H.: Emotive voice acceptance in human-robot interaction. In: Proceedings of the 24th International Conference on Auditory Display (2018)

4. Behrend, T.S., Sharek, D.J., Meade, A.W., Wiebe, E.N.: The viability of crowd-sourcing for survey research. Behav. Res. Methods **43**(3), 800–813 (2011)

5. Breazeal, C.: Emotive qualities in robot speech. In: Proceedings 2001 IEEE/RSJ International Conference on Intelligent Robots and Systems. Expanding the Societal Role of Robotics in the the Next Millennium (Cat. No.01CH37180), vol. 3, pp. 1388–1394, October 2001. https://doi.org/10.1109/IROS.2001.977175

6. Dandurand, F., Shultz, T.R., Onishi, K.H.: Comparing online and lab methods in a problem-solving experiment. Behav. Res. Methods **40**(2), 428–434 (2008)

7. Dou, X., Wu, C.-F., Lin, K.-C., Tseng, T.-M.: The effects of robot voice and gesture types on the perceived robot personalities. In: Kurosu, M. (ed.) HCII 2019. LNCS, vol. 11566, pp. 299–309. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-22646-6_21

8. Edwards, C., Edwards, A., Stoll, B., Lin, X., Massey, N.: Evaluations of an artificial intelligence instructor's voice: social identity theory in human-robot interactions. Comput. Hum. Behav. **90**, 357–362 (2019)

9. Eyssel, F., De Ruiter, L., Kuchenbrandt, D., Bobinger, S., Hegel, F.: 'If you sound like me, you must be more human': on the interplay of robot and user features on human-robot acceptance and anthropomorphism. In: 2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp. 125–126 (2012)

10. Gomez, R., Nakamura, K., Szapiro, D., Merino, L.: A holistic approach in designing tabletop robot's expressivity. In: Proceedings of the International Conference on Robotics and Automation (2020)

11. Gomez, R., Szapiro, D., Galindo, K., Nakamura, K.: Haru: hardware design of an experimental tabletop robot assistant. In: Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction, February 2018

12. Heerink, M., Kröse, B., Evers, V., Wielinga, B.: Relating conversational expressiveness to social presence and acceptance of an assistive social robot. Virtual Reality **14**(1), 77–84 (2010)

13. Levy, D.B.: Animation Development: From Pitch to Prod. Simon& Schuster (2010)

14. Macdonald, I.W.: Tablets of stone or DNA? TV series bibles. J. Screenwriting **9**(1), 3–23 (2018)

15. McGinn, C., Torre, I.: Can you tell the robot by the voice? an exploratory study on the role of voice in the perception of robots. In: 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pp. 211–221. IEEE (2019)

16. Niculescu, A., van Dijk, B., Nijholt, A., Li, H., See, S.L.: Making social robots more attractive: the effects of voice pitch, humor and empathy. Int. J. Soc. Robot. **5**(2), 171–191 (2013)

17. Plutchik, R.: A psychoevolutionary theory of emotions. Soc. Sci. Inf. **21**(4–5), 529–553 (1982). https://doi.org/10.1177/053901882021004003
18. Roehling, S., MacDonald, B., Watson, C.: Towards expressive speech synthesis in English on a robotic platform. In: Proceedings of the Australasian International Conference on Speech Science and Technology, pp. 130–135. Citeseer (2006)
19. Tang, H., Fu, Y., Tu, J., Hasegawa-Johnson, M., Huang, T.S.: Humanoid audio-visual avatar with emotive text-to-speech synthesis. IEEE Trans. Multimedia **10**(6), 969–981 (2008). https://doi.org/10.1109/TMM.2008.2001355