Kawaii Game Vocalics: A Preliminary Model

Katie Seaborn
seaborn.k.aa@m.titech.ac.jp
Tokyo Institute of Technology
Tokyo, Japan

Katja Rogers
University of Amsterdam
Amsterdam, The Netherlands
k.s.rogers@uva.nl

Somang Nam
Toronto Metropolitan University
Toronto, Canada
somang.nam@torontomu.ca

Miu Kojima
Tokyo Institute of Technology
Tokyo, Japan
kojima.m.ap@m.titech.ac.jp

ABSTRACT

Kawaii is the Japanese concept of cute++, a global export with local characteristics. Recent work has explored kawaii as a feature of user experience (UX) with social robots, virtual characters, and voice assistants, i.e., kawaii vocalics. Games have a long history of incorporating characters that use voice as a means of expressing kawaii. Nevertheless, no work to date has evaluated kawaii game voices or mapped out a model of kawaii game vocalics. In this work, we explored whether and how a model of kawaii vocalics maps onto game character voices. We conducted an online perceptions study (N=157) using 18 voices from kawaii characters in Japanese games. We replicated the results for computer voice and discovered nuanced relationships between gender and age, especially youthfulness, agelessness, gender ambiguity, and gender neutrality. We provide our initial model and advocate for future work on character visuals and within play contexts.

CCS CONCEPTS

• **Human-centered computing** → **Human computer interaction (HCI)**; • **Applied computing** → **Computer games**; *Psychology*; • **Hardware** → **Sound-based input / output**.

KEYWORDS

kawaii vocalics, kawaii, cuteness, vocalics, voice UX, game audio, user experience

1 INTRODUCTION

Kawaii is the Japanese term for "cute," but it is also much more than that. Kawaii refers to impressions of cuteness, endearment, and even pity. As a psychosocial factor of perception and experience, kawaii has been linked to cognition, behaviour, and emotion [37] within and beyond its native Japan [9, 42]. As a global phenomenon, kawaii has been adopted across a range of media around the world. Within the games space, kawaii can be attributed to series such as Pokémon, Animal Crossing, Kirby, and Super Mario Bros., as well as a wealth of other Japanese and Japanese-adjacent games [21, 48, 53]. Perhaps the most famous kawaii game characters include Pikachu from Pokémon, the titular character of the Kirby series, and Toad from the Super Mario Bros. games. The appeal of these characters is near-universal, speaking to the transnational and cross-cultural characteristics of kawaii.

Many kawaii game characters are judged based on their appearance and behaviour: the visuals. Yet, character *voice* may also be a key feature of their kawaii appeal and lasting value. From Pikachu's eponymous "pika pika" to Toad's vocal bursts, these vocalizations and speech patterns are instantly recognizable. Indeed, recent work in human-computer interaction (HCI) has pointed to the importance and dearth of work on kawaii sound [60], especially for interactive characters like voice assistants [51]. Moreover, the game context, with its complex audiovisual stimuli, may have special implications for kawaii vocalics, as there are many different attributes to consider. Voice can be direct intelligible speech or unintelligible gibberish (e.g., Simlish), presented in tandem with character animations. Characters can grunt or bark out lines (brief exclamations) or let out non-linguistic vocal bursts. Non-voicing sound techniques–type-writer or chalkboard sounds accompanying dialogue boxes[1]–are also used. Voices and vocal expressions can appear to originate from within the game world (i.e., a visible character, often with speech animation), or present without a clear source in the game world (e.g., in menu interfaces or for characters not visible in the scene); while a matter of interpretation, this can be seen as a diegetic/non-diegetic distinction [8]. Further, perception of vocalics (kawaii or otherwise) could interplay with dynamic and static visual character attributes, speech content (if applicable), and other auditory elements in the game (e.g., music or ambient noises). A model of kawaii game vocalics would offer a map of specific vocal attributes distinct from, but related to, other aspects of the game character known (or theorized) to invoke perceptions of kawaii, notably visual appearance and behaviour. This would increase our understanding of the relationship between vocal attributes or acoustical features and other well-studied game character features, and could be used by designers, researchers, critics, and even game fans who may be experimenting with new voice-generating technologies in gaming spaces on YouTube and Twitch [23]. As yet, kawaii game character voices are virtually unexplored.

In this preliminary work, we propose an initial model of kawaii game vocalics: vocal features that lead to perceptions of kawaii in or through voices. Our goal was to motivate kawaii game vocalics as a new area of study in game user research by way of an initial model and empirical evidence of kawaii perceptions of in-game character voices. To this end, we conducted an online perceptions study using a range of stimuli sourced from popular kawaii game characters in Japan. We asked two foundational research questions to elucidate the presence and perception of kawaii game vocalics. First, we asked RQ1: *Are the voices of game characters deemed*

[1] E.g., https://www.youtube.com/watch?v=B3S9QCCMPRw

kawaii (implicitly by way of their visual appearance) also perceived as kawaii? Second, we asked RQ2: *Do these perceptions depend on the "girlish" qualities of the voices?* This question was motivated by two models of kawaii: the initial model of kawaii vocalics for computer voice [51], comprised of agedness (young), genderedness (girlish or ambiguous), language fluency, and anthropomorphism, and the original two-layer model of kawaii [38], comprised of visual cuteness and reactions to cuteness as a foundation alongside other perceptual attributes of the kawaii source, such as friendliness, harmlessness, and prettiness. Our contributions are (i) empirical evidence of kawaii voice perceptions in game characters based on audio stimuli alone; (ii) confirmation of the initial model of kawaii vocalics for game characters; and (iii) a data set of kawaii vocalics ratings for game characters. This work sets the stage for a new research agenda within the broader area of game sound and contributes to the games research community's understanding of a popular mode of expression in people and characters that, being associated with sexist and ageist attitudes [51, 54], is still policed, discredited, and sidelined.

2 THEORETICAL BACKGROUND

2.1 A Psychosocial Model of Kawaii for Visuals and Voice

Kawaii and cuteness as research phenomena have a relatively recent history with modern foundations[2]. Perhaps the first instance of "cute" being taken seriously as an object of study was by Lorenze in the earlier part of the last century [22, 37]. Lorenze proposed that perceptions of cuteness were linked to "baby schema" or Kindchenschema, referring to our innate reaction to the appearance of young animals, humans and otherwise: the roundness of features, large eyes, small mouths and noses, and so on. Nittono and colleagues [37–42] later developed a two-layer model of kawaii as (i) *positive emotions or affective responses* that stem from (ii) *features of the socio-cultural environment*. Kawaii illustrates how phenomena that may originally relate to innate responses, i.e., Kindchenschema, are culturally influenced and socially mediated. The particular features of the Japanese sociocultural context, notably amae, or a desire for acceptance, and chizimi shikou, or a preference for small things, highlight his interplay. Kawaii is now a decades-long cultural trend, firmly established in Japan but also embraced around the world: a cross-cultural phenomenon.

The two-layer model of kawaii is made up of a *stimulus*, such as a kawaii object, which has *attributes* deemed kawaii, such as Kindchenschema, smiles, and colours, that are *perceived* in certain ways, such as cute, pretty, and gentle, leading to *cognitive appraisals* of the stimuli as kawaii. These appraisals are revealed through *emotional or affective responses*, which can *manifest* subjectively (e.g., attitudes, thoughts), behaviourally, or physiologically. Thus we can chart a kawaii response to a given stimuli. Nevertheless, kawaii has essentially been explored as a visual phenomenon [51], with only a few exceptions on nonverbal sounds and melodies [60], touch [43], behavioural conduct [24, 25], and voice [24, 51]. As Seaborn et al. [51] point out, this leaves open the question of whether and how other modalities, notably sound, fit into the two-layer model of kawaii. We attempt to explore this within the context of games that feature kawaii voices.

The notion of kawaii has shifted over time, from its origins as pity for the weak to a non-threatening and endearing quality of people, non-human animals, and even objects [38]. This has implications for games, as many characters–especially kawaii characters–are not human or even particularly humanlike, e.g., Pikachu, Kirby, Toad. Nevertheless, voice is an anthropomorphic, or at least biomorphic, feature not typically attributed to objects and generally associated with human-likeness [50]. Indeed, the initial model of kawaii vocalics provided evidence that human-likeness is key for kawaii perceptions in voice assistants. As such, we would expect that the game voices deemed most kawaii would also be the most humanlike and least artificial or machinelike. We thus hypothesized:

> H1. Perceptions of voice kawaiiness link to perceptions of low artificiality and high anthropomorphism.

Kawaii has also been conceptualized as a "girlish" phenomenon or feature of expression. This has led to its dismissal as a serious topic by some [17, 38] and warnings of underlying sexism by others [51, 54]. At the same time, Nittono and colleagues have generally approached kawaii in a gender-neutral way, as well as provided some evidence of agelessness in the form of smiling elders deemed kawaii [40]. Seaborn et al. [51], focusing on voice phenomena, found preliminary evidence that gender ambiguity may be perceived as most kawaii for certain individuals. However, given that the voices deemed gender ambiguous were also deemed young, it is not clear to what extent gender, age, or the combination of the two explain perceptions of ambiguity. We attempt to further explore and distinguish the intersection of age and gender by including voices from human and non-human game characters. We thus hypothesized:

> H2. Perceptions of voice gender will link to kawaii by way of femininity, i.e., gendered feminine.

> H3. Perceptions of voice age will link to kawaii by way of youthfulness, i.e., aged young.

> H4. Perceptions of voice "girlishness" (feminine, young) will link to high kawaii ratings.

2.2 Theorizing Kawaii Game Vocalics within Game Audio Studies

We approach kawaii game vocalics as a matter of perceiving auditory objects that are meaningful and may result in a variety of perceptions, cognitions, behaviours, and other reactions [13]. Voice in games is multifarious and can be situated within a range of game audio taxonomies. A common distinction concerns *audio diegesis* [18, 20, 31, 56]. For example, a voice's source of origin within the game world (tied to a visible character vs. disembodied) can influence whether it is considered diegetic or non-diegetic. Liljedahl [20] distinguishes between speech and dialogue, sound effects, and music. Some vocal audio in games would fit into the speech and dialogue category, but others (e.g., non-linguistic vocal bursts) could fit into the sound effects category, which, according to Friberg and Gärdenfors, refers to character, avatar, ambient, and ornamental

[2] A history of kawaii is out of scope for this paper; we refer the reader to Shiokawa [54] and Seaborn et al. [51], in ascending order of currency.

sounds [10]. All can involve vocal elements (e.g., general background noise of a crowd speaking—termed *walla* if unintelligible). In a master's thesis, Holmes [16, p.30] categorized dialogue as main dialogue, reaction & guidance lines (extending the former), ambient dialogue (background dialogue), walla (unintelligible ambient dialogue), combat barks (orders, commands, reactions b/w combatants), and emotes ("non-linguistic vocalisations such as screams, effort grunts, or breathing"), although it is unclear how this categorization was sourced or derived.

Much work has been done on voice interaction in audio research (e.g., in games [1], voice assistants [5, 57], and robots [27, 59]). Still, more work is needed on design guidelines for voice interface design [29]. Indeed, the vocal design of game characters and how voice impacts impression formation is a significant research gap. Voices have certain characteristics that can convey emotions [30, 47], impact trust in the speaker [59], and influence perceptions of the personality [4, 26]. This impression formation is driven by characteristics like frequency or pitch [26], harmonic-to-noise ratio [26], and speaking rate or intonation [4]. However, we know very little about how this works for game characters, and how acoustical features in the voices of game characters contribute to players' impression of those characters.

Voice designers and voice actors in the games industry have a lot of training and experience in how to elicit certain impressions through vocalics. Indeed, such audio design and production roles are of growing importance[3]. For example, professional voice (and motion capture) actors employ a full range of dramatic repertoire for character motion and vocal characteristics, such as intensity through physical exertion (e.g., by doing push-ups before or holding weights while recording[4]). But non-linguistic speech (gibberish or vocal bursts) can also portray emotions (e.g., Bastion's melodic beeping in Overwatch [3]). Developing an empirical understanding of the vocal factors that can yield specific impressions could help (voice) audio practitioners hone their craft more quickly and contribute to basic knowledge.

3 METHODS

We carried out an online user perceptions study using the research designs on visual [40] and voice-based [2] phenomena, aiming to replicate the kawaii vocalics study by Seaborn et al. [51] with voice stimuli from game characters. Our protocol was registered before data collection on May 11th, 2023 via OSF[5] and approved by the research ethics board at the first author's institution.

3.1 Participants and Recruitment

Respondents (N=162) were recruited through Yahoo! Crowdsourcing Japan on June 12th, 2023. Five incomplete responses were removed. The final set of participants (N=157) included women (n=76, one transgender), men (n=73), and none of another gender, while eight preferred not to say. Most were aged 45-54 (n=49) or 35-44 (n=42), with some older (55-64 n=29, 65-74 n=11, 75+ n=2) and younger (18-34 n=16); eight preferred not to say. Most played games every day (n=49, 31%), while others played games multiple times a week (n=38, 24%), others never did (n=24, 15%), some played once a week (n=14, 9%) or once a month (n=13, 8%), and others used to play games but stopped (n=11, 7%), with eight preferring not to say. Participants were paid at ~1200 yen per hour, equating to ~300 yen for the study duration.

3.2 Procedure

Participants were given a link to a SurveyMonkey questionnaire containing the consent form, stimuli and rating scales, and demographic questions. After consenting on the first page, they were presented with the clips of game character voices, one voice per page (cf. Section 3.3). They were asked to rate these stimuli based on various vocal and social qualities related to kawaii (cf. Section 3.4). The stimuli were presented in a random order to avoid novelty and order effects [49]. Demographics were collected on the final page to avoid priming effects [15]. The study took ~16 min.

3.3 Materials

We selected 18 voices using a three-step process. Initially, we attempted to identify game character *voices* deemed kawaii by experts and novices on magazines, websites, and social media. However, many focused on the voice actors, who were known beyond games and may have been judged based on their status as idols and celebrities in Japan. We also recognized that people may not associate kawaii with voice, given the lack of work on non-visual kawaii so far. We then instead identified game characters deemed kawaii *in general*, using the same methods. Finally, we relied on our expertise to manually add characters that were missed. All authors decided on a final set of voice stimuli from all three steps that: (i) reflected a range of ages, genders, and vocal types for verifying our hypotheses; (ii) were diverse in terms of game of origin, i.e., representing a range of game types and particular game series or offerings; (iii) were similar in terms of game origin, e.g., Toad and Toadette; and (iv) where we could isolate the voice against any background sound or music. The first author recorded 2- to 20-minute voice clips from videos containing the original voices. From these, a diverse sample of vocal expressions were chosen and arranged into ~8-second clips. The final set is in Table 1.

3.4 Instruments and Measures

All instruments used a 5-point Likert scale unless noted. The item order was randomized to curtail order effects [49]. All items were translated into Japanese by a native speaker, and then back-translated into English and checked with an advanced speaker and native speaker. The full questionnaire was pilot tested with six native Japanese speakers.

3.4.1 Kawaii Perceptions. In the absence of a validated measure, we used the one-item scale created by Seaborn et al. [51], which asks for an agreement rating of kawaii-ness. Given that it is a one-item scale, it was operationalized as a mean greater than 3.5 (skewed towards agreement on kawaii-ness) *and* a median of 4 or above (nominal agreement). Marginal cases, where only one or the other of these metrics were met, are noted.

3.4.2 Perceptions of Anthropomorphism, Artificiality, and Fluency. We used the one-item humanlikeness scale from Baird et al. [2].

[3] https://blog.audiokinetic.com/en/a-speed-run-through-the-world-of-voice-design/
[4] https://www.youtube.com/watch?v=8sZMSBtSFrg
[5] https://osf.io/7tr6b

Following Seaborn et al. [51], we divided the item into two items based on the poles (humanlike and artificial). We also added the item on language fluency, key for the kawaii vocalics model [51].

3.4.3 Age Perceptions. Agedness was captured in a nominal scale: infant/baby (0-2 years), child (3-12 years), teenaged (13-19 years), adult (20-39 years), middle-aged (40-64 years), older adult (65+ years), and ageless.

3.4.4 Gender Perceptions. Genderedness was captured in a nominal scale: feminine, masculine, aspects of both (reported as ambiguous), and neither (reported as neutral). Participants could also enter another option (free text).

3.4.5 Demographics. We collected gender, age, education, and game use frequency.

3.5 Data Analysis

We used descriptive statistics to group voices by perceived age (baby, child, teen, adult, ageless) and gender (masculine, feminine, gender ambiguous, gender neutral). When Shapiro-Wilks tests showed non-normal distributions, we used non-parametric statistics, e.g., Kendall's tau correlations. We also used Bonferroni corrections.

4 RESULTS

Descriptive statistics are summarized in Table 1. We now report on the hypotheses and full descriptive results.

4.1 H1. Perceptions of voice kawaiiness link to perceptions of low artificiality and high anthropomorphism.

A strong, negative correlation was found between Artificial and Humanlike ratings (τb = -.725, p < .05). A moderate positive correlation was found between Kawaii and Humanlike ratings (τb = .356, p < .05). A moderate, positive correlation was found between Humanlike and Fluent ratings (τb = .571, p <.05), and a moderate, negative correlation was found between Artificial and Fluent ratings (τb = -.480, p < .05). No others were found. As before [51], we can **accept the hypothesis that the most kawaii voices were also the most fluent, humanlike, and least artificial**.

4.2 H2. Perceptions of voice gender will link to kawaii by way of femininity, i.e., gendered feminine.

A Chi-Square test found a statistically significant association between Perceived Gender and Kawaii ratings across all voices, χ^2(16, 2826) = 489.66, p < .05, ϕ = .208. Follow-up Chi-Square test of independence (Bonferroni corrected) showed relationships between Perceived Gender and Kawaii ratings for all categories: Feminine (M=3.8, SD=.8, MD=4, IQR=1), Ambiguous (M=3.4, SD=1, MD=4, IQR=1), Masculine (M=3.0, SD=.9, MD=3, IQR=2), Neutral (M=2.7, SD=1.1, MD=3, IQR=2), p < .05. No voice was classified as 'Other' across participants. A Kruskal-Wallis test indicated a significant difference by gender category, χ^2(4) = 365.05, p < .05, with a Dunn's test (Bonferroni corrected) revealing significant differences across all gender categories. As with computer voices [51], we can **accept the hypothesis that voices deemed feminine were perceived as the most kawaii**.

4.3 H3. Perceptions of voice age will link to kawaii by way of youthfulness, i.e., aged young.

A Chi-Square test found a significant relationship between Perceived Age and Kawaii ratings across all voices, χ^2(24, 2826) = 404.52, p < .05, ϕ = .189. A Kruskal-Wallis test indicated a significant difference by age category, χ^2(5) = 226.91, p < .05, with a Dunn's test (Bonferroni corrected) revealing significant differences across all age categories except for Baby-Child and Teen-Adult: Teen (M=3.8, SD=.8, MD=4, IQR=1), Adult (M=3.7, SD=.9, MD=4, IQR=1), Child (M=3.5, SD=1.0, MD=4, IQR=1), Baby (M=3.4, SD=1.0, MD=4, IQR=1), and Ageless (M=2.7, SD=1.1, MD=3, IQR=2). As before [51], we can **accept the hypothesis that kawaii is an age-based phenomenon, linked to voice youthfulness**.

4.4 H4. Perceptions of voice "girlishness" (feminine, young) will link to high kawaii ratings.

Chi-Square tests found a significant relationship between Perceived Age and Kawaii rating for feminine voices, χ^2(24, 1256) = 121.84, p < .05, ϕ = .156, gender ambiguous voices, χ^2(24, 942) = 116.46, p < .05, ϕ = .176, and gender neutral voices, χ^2(20, 314) = 37.99, p < .05, ϕ = .174, but not masculine voices. A Kruskal-Wallis test for voice and gender classifications revealed significant differences in Kawaii ratings, χ^2(7) = 369.174, p < .05. A Dunn's test (Bonferroni corrected) indicated this for all pairs except Feminine-Child and Feminine-Teen, Feminine-Child and Feminine-Adult, Feminine-Teen and Feminine-Adult, Ambiguous-Baby and Ambiguous-Child: Masculine-Child (M=3.0, SD=.9, MD=3, IQR=2), Feminine-Child (M=3.8, SD=.9, MD=4, IQR=1), Feminine-Teen (M=3.8, SD=.8, MD=4, IQR=1), Feminine-Adult (M=3.7, SD=.9, MD=4, IQR=1), Ambiguous-Baby (M=3.4, SD=1.0, MD=4, IQR=1), Ambiguous-Child (M=3.4, SD=1.1, MD=4, IQR=1), Neither-Ageless (M=2.7, SD=1.1, MD=3, IQR=2). We can **accept the hypothesis of kawaii as girlish but also gender ambiguous and gender neutral**, confirming previous work [51] and demarcating **gender neutrality**.

5 DISCUSSION

5.1 Kawaii Beyond Game Character Visual Appearance (RQ1)

The results indicate that game character voices, even in the absence of other cues, especially visuals, can invoke a feeling of kawaii. In short, people can "hear cute." This gives weight to the importance of vocal design when creating game characters, especially if aiming for a kawaii impression. Yet, the level of kawaii varied across characters and voices deemed kawaii in ways that relate to the factors in the general and vocalics kawaii models, to which we now turn.

Table 1: Descriptive statistics for perceptions of kawaii, age, and gender in game character voice, in order of kawaii ratings

Character	Game or Series	Kawaiiness	Age Group	Gender Group
Barbara	Genshin Impact [28]	Y: M=4.2, SD=.7, MD=4	Child (MD=2, 60%)	Fem. (MD=2, 97%)
Pikachu	The Pokémon series, e.g., [12]	Y: M=4.2, SD=.7, MD=4	Child (MD=2, 50%)	Amb. (MD=2, 97%)
Edea	Bravely Default [55]	Y: M=3.9, SD=.7, MD=4	Teen (MD=3, 48%)	Fem. (MD=2, 87%)
Ayaka	Genshin Impact [28]	Y: M=3.9, SD=.9, MD=4	Adult (MD=4, 87%)	Fem. (MD=2, 99%)
QiQi	Genshin Impact [28]	Y: M=3.9, SD=.8, MD=4	Child (MD=2, 58%)	Fem. (MD=2, 94%)
Peach	The Super Mario series, e.g., [36]	Y: M=3.7, SD=.8, MD=4	Teen (MD=3, 43%)	Fem. (MD=2, 92%)
Kirby	The Kirby series, e.g., [14]	Y: M=3.7, SD=.8, MD=4	Child (MD=3, 69%)	Amb. (MD=3, 51%)
Ashley	The Wario series, e.g., [35]	Y: M=3.7, SD=.8, MD=4	Teen (MD=3, 68%)	Fem. (MD=2, 96%)
Zelda	TLoZ series, e.g., [34]	Y: M=3.6, SD=.8, SD=4	Adult (MD=4, 63%)	Fem. (MD=2, 98%)
Jigglypuff	The Pokémon series, e.g., [11]	M: M=3.4, SD=1, MD=4	Child (MD=2, 45%)	Amb. (MD=3, 43%)
Toadette	The Super Mario series, e.g., [36]	M: M=3.4, SD=.9, MD=4	Child (MD=2, 59%)	Fem. (MD=2, 55%)
Yoshi	The Super Mario series, e.g., [36]	M: M=3.3, SD=1, MD=4	Baby (MD=2, 57%)	Amb. (MD=3, 53%)
Young Link	TLoZ series, e.g., [32]	N: M=3.2, SD=.9, MD=3	Child (MD=2, 62%)	Masc. (MD=1, 55%)
Baby Bowser	The Super Mario series, e.g., [36]	N: M=2.9, SD=1, MD=3	Child (MD=2, 35%)	Amb. (MD=3, 45%)
Toad	The Super Mario series, e.g., [36]	N: M=2.9, SD=.9, MD=3	Child (MD=2, 55%)	Masc. (MD=1, 54%
Inkling Girl	Splatoon 3 [7]	N: M=2.7, SD=1, MD=3	Child (MD=2, 41%)	Amb. (MD=3, 41%)
Luma	Super Mario Galaxy [6]	N: M=2.7, SD=1, MD=3	Ageless (MD=7, 76%)	Neu. (MD=4, 69%)
Shizue[a]	Animal Crossing, e.g., [33]	N: M=2.6, SD=1, MD=3	Ageless (MD=7, 70%)	Neu. (MD=4, 57%)

Y: Yes. M: Marginal. N: No. TLoZ: The Legend of Zelda. Fem.: Feminine. Masc.: Masculine. Amb.: Ambiguous. Neu.: Neutral. [a]Isabella.

5.2 Is Kawaii "Girlish"? Confirming and Advancing Models of Kawaii (RQ2)

We found the expected patterns based on the kawaii vocalics model [51] for this novel stimuli. We can confirm the extension to the two-factor model of kawaii for visual phenomena and social behaviour [38] by including the social identity [58] factors of gender and age, the computer agent factors of anthropomorphism and artificiality [2], and the voice UX factors of vocal expressivity and fluency [50]. We also confirm the results on age and gender ambiguity found for computer voice [51]. We also selected samples that allowed us to explore gender neutrality. While not well distinguished [52], *neutrality* speaks to an absence of gender, while *ambiguity* speaks to the potential of pluralistic gender/ing. These *gender liminal* perceptions may reflect generational shifts in attitudes towards the genderedness of kawaii: less about "girl" and "boy" varieties [44] and more about "individuality" [54]. The results for anthropomorphism in combination with the factors in Table 1 also hint at patterns of liminiality for the non-human(oid) characters. The marginal cases of Jigglypuff (Child-Ambiguous), Toadette (Child-Feminine), and Yoshi (Baby-Ambiguous) juxtapose the low kawaii Luma (Ageless-Neutral) and Shizue (Ageless-Neutral). Toadette aside, all clips were vocal bursts or gibberish. Future work can explore biomorphism and vocal types against these factors in kawaii voice phenomena.

5.3 Limitations

We acknowledge several limitations. We conducted an online study with a relatively small set of vocal stimuli. More stimuli from characters deemed kawaii or not need to be explored. Also, we did not capture whether respondents knew the character belonging to the voice. We thus cannot rule out the possibility that knowledge of the character and preexisting impressions of kawaii towards that character influenced results, which future work can tease out.

5.4 Research Agenda

As with the source phenomenon, kawaii game vocalics likely have repercussions for cognition, behaviour, and emotion that vary by game character and context, as well as by individual player. Here are some trajectories for future work:

5.4.1 Game Visuals: Of Character and Context. An ongoing issue in research on game audio is that findings on audio perception captured *outside* of games do not necessarily translate *within* or *across* games [46]. Research on computer agents has long indicated that voice and body can and often do intersect in ways that may be influenced by other features of the interactive context [50]. The visual appearance of the character, including animations and behaviours, as well as the visual context, i.e., the visual game environment, will need to be teased out in terms of whether and how perceptions of kawaii in character vocalizations are influenced by relevant features of the game's visual modalities.

5.4.2 Towards a Taxonomy of Game Voice. Game voice–of characters and beyond–is complex and multifaceted. One of the challenges we encountered was finding a rigorously developed framework by which to select and categorize voice phenomena for data analysis. This is a gap that needs addressing. The type of voice phenomena–from vocal bursts to gibberish–and the presence or absence of speech may affect kawaii perceptions. We relied on a preliminary taxonomy [16] and our general knowledge of how voice phenomena is operationalized. The next step is to develop a game context-oriented taxonomy using rigorous methods like type-building analysis [19].

5.4.3 The Kawaii Vocalics Data Set: An Open Science Endeavour. As a "first" next step, we have started and offer an open data set. Like the ABOT (Anthropomorphic Robot Database) project [45], we aim to provide a database of psychosocial factors attributed to specific voices. We provide our initial data set here: https://bit.ly/kawaiigamevocalics

6 CONCLUSION

Kawaii game vocalics brings a fresh perspective to the fields of game audio research and voice interaction. Moreover, existing models of kawaii premised on visuals need to be expanded to include voice UX factors. Experimental work on a variety of non/humanoid characters and contexts will be needed–and there is a wealth of game material to draw on.

ACKNOWLEDGMENTS

This work was supported by department funds at Tokyo Institute of Technology.

REFERENCES

[1] Fraser Allison, Marcus Carter, Martin Gibbs, and Wally Smith. 2018. Design Patterns for Voice Interaction in Games. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play* (Melbourne, VIC, Australia) *(CHI PLAY '18)*. Association for Computing Machinery, New York, NY, USA, 5–17. https://doi.org/10.1145/3242671.3242712

[2] Alice Baird, Stina Jørgensen, Emilia Parada-Cabaleiro, Nicholas Cummings, Simone Hantke, and Björn Schuller. 2018. The perception of vocal traits in synthesized voices: Age, gender, and human likeness. *Journal of the Audio Engineering Society* 66 (April 2018), 277–285. https://doi.org/10.17743/jaes.2018.0023

[3] Blizzard Entertainment. 2016. *Overwatch.* Game [Windows]. Blizzard Entertainment, Irvine, California, United States..

[4] Bruce L. Brown, William J. Strong, and Alvin C. Rencher. 1973. Perceptions of personality from speech: effects of manipulations of acoustical parameters. *The Journal of the Acoustical Society of America* 54, 1 (July 1973), 29–35. https://doi.org/10.1121/1.1913571

[5] Julia Cambre and Chinmay Kulkarni. 2019. One Voice Fits All? Social Implications and Research Challenges of Designing Voices for Smart Devices. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 223 (nov 2019), 19 pages. https://doi.org/10.1145/3359325

[6] Nintendo Entertainment Planning & Development. 2007. *Super Mario Galaxy.* Game [Wii, Wii U, Nvidia Shield TV].

[7] Nintendo Entertainment Planning & Development. 2022. *Splatoon.* Game [Nintendo Switch].

[8] Inger Ekman. 2005. Meaningful noise: Understanding sound effects in computer games. *Proc. Digital Arts and Cultures* 17 (2005), 4 pages.

[9] Angela Y.H. Fan, Chen Ji, Ella Dagan, Samir Ghosh, Yuhui Wang, and Katherine Isbister. 2023. The Cuteness Factor: An Interpretive Framework for Artists, Designers and Engineers. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (Pittsburgh, PA, USA) *(DIS '23)*. Association for Computing Machinery, New York, NY, USA, 2509–2521. https://doi.org/10.1145/3563657.3596035

[10] Johnny Friberg and Dan Gärdenfors. 2004. Audio games: new perspectives on game audio. In *Proceedings of the 2004 ACM SIGCHI International Conference on Advances in computer entertainment technology.* ACM, New York, NY, USA, 148–154.

[11] Game Freak. 1998. *Pokémon Red and Blue.* Game [Game Boy]. Nintendo, Kyoto, Japan..

[12] Game Freak. 2022. *Pokemon Legends: Arceus.* Game [Nintendo Switch]. Nintendo, Kyoto, Japan..

[13] Timothy D. Griffiths and Jason D. Warren. 2004. What is an auditory object? *Nature Reviews Neuroscience* 5, 11 (Nov. 2004), 887–892. https://doi.org/10.1038/nrn1538 Number: 11 Publisher: Nature Publishing Group.

[14] HAL Laboratory and Vanpool. 2023. *Kirby's Return to Dream Land Deluxe.* Game [Nintendo Switch]. Nintendo, Kyoto, Japan..

[15] Thomas C Head, Ricky W Griffin, Thomas S Bateman, Lynn Lohman, and Valerie L Yates. 1988. The priming effect in task design research. *Journal of Management* 14, 1 (1988), 33–39.

[16] Thomas J. Holmes. 2021. *Defining Voice Design in Video Games.* Master's thesis. Aalto University.

[17] Yomota Inuhiko. 2006. *"Kawaii" ron [The Theory of Kawaii].* Number 578 in Chikuma Shinsho. Chikuma Shobō, Tokyo, Japan. https://www.worthsharing.jpf.go.jp/en/vol_1/the-theory-of-kawaii/

[18] Kristine Jørgensen. 2011. Sound for Fantasy and Freedom. In *Time for new terminology? Diegetic and nondiegetic sounds in computer games revisited*, Mark Grimshaw (Ed.). IGI Global, Hershey, PA, USA, 78–97. https://doi.org/10.4018/978-1-61692-828-5.ch005

[19] Udo Kuckartz. 2013. Qualitative Text Analysis: A Guide to Methods, Practice and Using Software. *Qualitative Text Analysis* (2013), 1–192.

[20] Mats Liljedahl. 2011. Sound for Fantasy and Freedom. In *Game Sound Technology and Player Interaction*, Mark Grimshaw (Ed.). IGI Global, Hershey, PA, USA, 22–43. https://doi.org/10.4018/978-1-61692-828-5.ch002

[21] Chaolan Lin, Travis Faas, Lynn Dombrowski, and Erin Brady. 2017. Beyond cute: Exploring user types and design opportunities of virtual reality pet games. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology (VRST '17)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3139131.3139132

[22] Konrad Lorenz. 1943. Die angeborenen Formen möglicher Erfahrung. *Zeitschrift für Tierpsychologie* 5, 2 (1943), 235–409. https://doi.org/10.1111/j.1439-0310.1943.tb00655.x _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1439-0310.1943.tb00655.x.

[23] Zhicong Lu, Chenxinran Shen, Jiannan Li, Hong Shen, and Daniel Wigdor. 2021. More kawaii than a real-person live streamer: Understanding how the otaku community engages with and perceives virtual YouTubers. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems.* ACM, New York, NY, USA, 1–14.

[24] Xingyang Lv, Yue Liu, Jingjing Luo, Yuqing Liu, and Chunxiao Li. 2021. Does a cute artificial intelligence assistant soften the blow? The impact of cuteness on customer tolerance of assistant service failure. *Annals of Tourism Research* 87 (March 2021), 103114. https://doi.org/10.1016/j.annals.2020.103114

[25] Aaron Marcus, Ayako Hashizume, Masaaki Kurosu, and Xiaojuan Ma. 2017. *Cuteness Engineering: Designing Adorable Products and Services* (1st ed. 2017 ed.). Springer International Publishing : Imprint: Springer, Cham. https://doi.org/10.1007/978-3-319-61961-3

[26] Phil McAleer, Alexander Todorov, and Pascal Belin. 2014. How Do You Say 'Hello'? Personality Impressions from Brief Novel Voices. *PLoS ONE* 9, 3 (March 2014), e90779. https://doi.org/10.1371/journal.pone.0090779

[27] Conor McGinn and Ilaria Torre. 2019. 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)*. IEEE, New York, NY, USA, 211–221. https://doi.org/10.1109/hri.2019.8673305

[28] miHoYo. 2021. *Genshin Impact.* Game [Mobile, PC].

[29] Christine Murad and Cosmin Munteanu. 2020. Designing Voice Interfaces: Back to the (Curriculum) Basics. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376522

[30] Iain R. Murray and John L. Arnott. 1993. Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *The Journal of the Acoustical Society of America* 93, 2 (Feb. 1993), 1097–1108. https://doi.org/10.1121/1.405558

[31] David Neumeyer. 2009. Diegetic/Nondiegetic: A Theoretical Model. *Music and the Moving Image* 2, 1 (April 2009), 26–39. https://doi.org/10.5406/musimoviimag.2.1.0026

[32] Nintendo EAD. 2000. *The Legend of Zelda: Majora's Mask.* Game [Nintendo 64]. Nintendo, Kyoto, Japan..

[33] Nintendo EAD. 2012. *Animal Crossing: New Leaf.* Game [Nintendo 3DS]. Nintendo, Kyoto, Japan..

[34] Nintendo EPD. 2023. *Super Mario Bros.* Game [Nintendo Switch]. Nintendo, Kyoto, Japan..

[35] Intelligent Systems Nintendo EPD. 2018. *WarioWare Gold.* Game [Nintendo 3DS].

[36] Nintendo R&D4. 1985. *Super Mario Bros.* Game [NES]. Nintendo, Kyoto, Japan..

[37] Hiroshi Nittono. 2010. A behavioral science framework for understanding kawaii. In *Proceedings of The Third International Workshop on Kansei.* Editorial comittee of the third international workshop on Kansei, Fukuoka, Japan, 80–83.

[38] Hiroshi Nittono. 2016. The two-layer model of 'kawaii': A behavioural science framework for understanding kawaii and cuteness. *East Asian Journal of Popular Culture* 2, 1 (April 2016), 79–95. https://doi.org/10.1386/eapc.2.1.79_1 Publisher: Intellect.

[39] Hiroshi Nittono. 2022. The psychology of "kawaii" and its implications for human-robot interaction. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, Sapporo, Japan, 3–3. https://doi.org/10.1109/HRI53351.2022.9889591

[40] Hiroshi Nittono, Michiko Fukushima, Akihiro Yano, and Hiroki Moriya. 2012. The power of kawaii: Viewing cute images promotes a careful behavior and narrows attentional focus. *PLOS ONE* 7, 9 (Sept. 2012), e46362. https://doi.org/10.1371/journal.pone.0046362 Publisher: Public Library of Science.

[41] Hiroshi Nittono and Namiha Ihara. 2017. Psychophysiological responses to kawaii pictures with or without baby schema. *SAGE Open* 7, 2 (April 2017), 2158244017709321. https://doi.org/10.1177/2158244017709321 Publisher: SAGE Publications.

[42] Hiroshi Nittono, Shiri Lieber-Milo, and Joshua P. Dale. 2021. Cross-cultural comparisons of the cute and related concepts in Japan, the United States, and Israel. *SAGE Open* 11, 1 (Jan. 2021), 2158244020988730. https://doi.org/10.1177/2158244020988730 Publisher: SAGE Publications.

[43] Yuka Okada, Mitsuhiko Kimoto, Takamasa Iio, Katsunori Shimohara, Hiroshi Nittono, and Masahiro Shiomi. 2020. Can a robot's touches express the feeling of kawaii toward an object?. In *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, Las Vegas, NV, USA, 11276–11283. https://doi.org/10.1109/IROS45743.2020.9340874 ISSN: 2153-0866.

[44] Keiko Ota. 2020. *For the Boys to Come*. Otsuki Bookstore, Otsuki, Japan.

[45] Elizabeth Phillips, Xuan Zhao, Daniel Ullman, and Bertram F Malle. 2018. What is human-like? Decomposing robots' human-like appearance using the Anthropomorphic Robot (ABOT) Database. In *Proceedings of the 2018 ACM/IEEE international conference on human-robot interaction*. New York, NY, USA, 105–113. https://doi.org/10.1145/3171221.3171268

[46] Katja Rogers, Maximilian Milo, Michael Weber, and Lennart E. Nacke. 2020. The Potential Disconnect between Time Perception and Immersion: Effects of Music on VR Player Experience. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play* (Virtual Event, Canada) *(CHI PLAY '20)*. Association for Computing Machinery, New York, NY, USA, 414–426. https://doi.org/10.1145/3410404.3414246

[47] K Scherer. 2003. Vocal communication of emotion: A review of research paradigms. *Speech Communication* 40, 1-2 (April 2003), 227–256. https://doi.org/10.1016/s0167-6393(02)00084-5

[48] Douglas Schules. 2015. Kawaii Japan: Defining JRPGs through the Cultural Media Mix. (2015).

[49] Howard Schuman and Stanley Presser. 1996. *Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context*. SAGE, Thousand Oaks, CA, USA. Google-Books-ID: Je640UKqNaYC.

[50] Katie Seaborn, Norihisa P. Miyake, Peter Pennefather, and Mihoko Otake-Matsuura. 2022. Voice in human–agent interaction: A survey. *Comput. Surveys* 54, 4 (May 2022), 1–43. https://doi.org/10.1145/3386867

[51] Katie Seaborn, Somang Nam, Julia Keckeis, and Tatsuya Itagaki. 2023. Can voice assistants sound cute? Towards a model of kawaii vocalics. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, 1–7. https://doi.org/10.1145/3544549.3585656

[52] Katie Seaborn and Peter Pennefather. 2022. Neither "hear" nor "their": Interrogating gender neutrality in robots. In *Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction (HRI '22)*. IEEE Press, Sapporo, Hokkaido, Japan, 1030–1034. https://doi.org/10.5555/3523760.3523929

[53] Akiko Shibuya, Hibiki Okura, Akiyo Shoun, and Naoko Asou. [n. d.]. Male and Female Game Players' Preferences for Game Characters and Real-world Personalities in Japan. ([n. d.]).

[54] Kanako Shiokawa. 1999. Cute but deadly: Women and violence in Japanese comics. In *Themes and Issues in Asian Cartooning: Cute, Cheap, Mad, and Sexy*, John A. Lent (Ed.). Bowling Green State University Popular Press, Bowling Green, OH, USA, 93–126. Google-Books-ID: b0EJmrszhyQC.

[55] Claytechworks Square Enix, Silicon Studio and Team Asano. 2012. *Bravely Default*. Game [Nintendo 3DS].

[56] Robynn J. Stilwell. 2007. The Fantastical Gap between Diegetic and Nondiegetic. In *Beyond the Soundtrack: Representing Music in Cinema*, Daniel Goldmark, Lawrence Kramer, and Richard Leppert (Eds.). University of California Press, Oakland, CA, USA, 184–202. https://doi.org/10.1525/j.ctt1ppgzx.15

[57] Selina Jeanne Sutton, Paul Foulkes, David Kirk, and Shaun Lawson. 2019. Voice as a Design Material: Sociophonetic Inspired Design Strategies in Human-Computer Interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–14. https://doi.org/10.1145/3290605.3300833

[58] Henri Tajfel and John C Turner. 2004. The Social Identity Theory of Iintergroup Behavior. In *Political Psychology*. Psychology Press, London, UK, 276–293.

[59] Ilaria Torre and Laurence White. 2020. Trust in Vocal Human–Robot Interaction: Implications for Robot Voice Design. In *Voice Attractiveness*. Springer Singapore, Singapore, 299–316. https://doi.org/10.1007/978-981-15-6627-1_16

[60] Brian J. Zhang, Knut Peterson, Christopher A. Sanchez, and Naomi T. Fitter. 2021. Exploring consequential robot sound: Should we make robots quiet and kawaii-et?. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. Prague, Czech Republic, 3056–3062. https://doi.org/10.1109/IROS51168.2021.9636365 ISSN: 2153-0866.

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009