

Known-Voice Cloning and the Limits of Personality Transfer for Character Design^{*}

Johanna Magdalena Kuch¹[0000-0001-5322-9552],
 Maud van Lier²[0000-0001-7951-7026],
 Daksitha Senel Withanage Don¹[0009-0003-8831-9795],
 Patrick Gebhard³[0000-0002-5566-4520],
 Tanja Schneeberger³[0000-0002-7247-6978], and
 Elisabeth André¹[0000-0002-2367-162X]

¹ Augsburg University, Human-Centered Artificial Intelligence, Augsburg, Germany
 {johanna.kuch,daksitha.withanage.don,elisabeth.andre}@uni-a.de

² Institute of Intelligent Systems and Robotics, CNRS, Sorbonne Université, Paris, France

³ German Research Center for Artificial Intelligence (DFKI), Affective Computing Group, Germany

Abstract. Recent advances in zero-shot voice cloning systems make it possible to reproduce a recognizable voice from a short recording and attach it to arbitrary virtual characters. Prior work suggests that familiar vocal identity can carry over associated personality impressions to new targets such as robots or virtual characters. However, these works have mainly examined familiarity and personalization effects rather than directly testing whether a cloned familiar voice systematically transfers personality impressions to an embodied virtual character. Such transference raises the concern that cloned voices could be used to manipulate people into engaging with virtual characters as capable social actors. Across two studies, we investigated whether personality impressions transfer when a virtual character speaks with the cloned voice of a familiar person. Study 1 examined this question in a real-world university setting using the cloned voice of a lecturer familiar to the participating students. Study 2 examined the same question online using voices of social media content creators and members of their communities. Drawing on the concepts of identity activation, trait transfer, and person reconstruction, we show that evidence of personality transfer was limited in both studies. Observed effects were selective and did not generalize across traits or study contexts. In the online study in particular, recognizable voice identity cues activated expectations about the speaker but did not reconstruct the broader person impression. Qualitative responses further

^{*} This preprint has been accepted for publication in *Connected Futures – Nature, Robots, and Society* (Robophilosophy 2026), in the *Lecture Notes in Computer Science* (LNCS) series, volume 17299 (expected publication: December 2026).

indicated that judgments were frequently shaped by perceived artificiality and by whether voice and appearance fit together. Our findings moderate common concerns that cloned voices straightforwardly manipulate people into perceiving artificial characters as capable social actors.

Keywords: voice cloning · virtual characters · personality attribution · human–robot interaction · social perception

1 Introduction

From short speech samples, listeners infer personality traits and social qualities such as trustworthiness, dominance, and competence [25]. When a voice is familiar, perception extends beyond acoustic inference: Familiar voices are stored in memory with sufficient specificity for robust identity recognition and activate stored person-specific knowledge, including trait attributions and relational associations [14].

Research has now shown that this stored knowledge can be activated on the basis of particular cues alone, leading us to project person-based attributions from familiar individuals onto new targets, such as when we feel an instant liking for a new person who resembles a dear friend in some way [15,40]. Since hearing a familiar voice can trigger person-specific representations or associated attributions even without explicit identification [4,20], it is likely that familiar voices can form such cues as well. If they can, then it is possible that the attachment of familiar voices to new targets can bias our judgments of these targets.

From a social robotics and AI perspective, this is an interesting possibility, especially now that recent work suggests that we evaluate synthetic voices along similar social dimensions as we do natural human voices [39]. If we already infer personality traits and social qualities from synthetic voices, what would happen when we provide robots or virtual characters not just with human voices, but with voices of humans that we are already familiar with? Would it activate our stored person-knowledge, biasing our evaluation of the capacities of that system to function as a social actor, thereby increasing our willingness to trust and engage with that system?

Prior work has demonstrated that voice design can shape the appeal, trustworthiness, and perceived believability of virtual agents and characters [7,12]. And, even when just a synthetic voice is used, similarity-based voice design has been shown to increase robot likability (both with and without conscious awareness of the similarity) [16]. In addition, related work suggests that character-inspired voices can shape user engagement and evaluations of conversational AI systems [9]. Based on these findings, the hypothesis can be drawn that the use of a familiar voice is likely to further increase our willingness to engage with an artificial system. What this seems to depend on, however, and what the aforementioned findings do not yet establish, is whether a cloned familiar voice can systematically transfer personality impressions to an embodied artificial character and whether the coupling of a known voice with a visible body amplifies such transfer.

This is an important question to answer, especially given recent advances in zero-shot voice cloning. Such systems generate realistic speech from short reference recordings [6,29], making it possible for recognizable voices to be reproduced and attached to arbitrary characters across a variety of contexts. Besides the ethical concerns raised by the fact that there are not yet adequate policies in place to protect us from having our voices cloned and used without consent,⁴ there is also the worry that the implementation of familiar voices in Human-Robot and AI interaction can constitute a *dark pattern* [37].⁵ After all, if familiar vocal identity indeed carries over associated personality impressions in a reliable way, then designers could strategically activate character traits through voice selection alone, thereby manipulating users to engage with their artificial systems in ways that are not necessarily in their best interest. Given these pressing ethical concerns, it is thus important that we find out whether voice-based familiarity provides a plausible pathway for attribution transfer.

In this paper, two studies are presented that provide initial insights into the possible effects that the use of familiar voices can have on our evaluation of an artificial system as a capable social actor. We thereby make three contributions to existing research: (1) We provide empirical evidence across two contexts that familiar voice cloning yields only small and trait-selective alignment effects rather than robust person-consistent evaluations. (2) We show, through qualitative analysis, that judgments are often dominated by perceived artificiality and voice–appearance mismatch. (3) We derive practical guidance for responsible character design by identifying boundary conditions under which familiar voice cloning is unlikely to support social evaluation and may backfire.

In the following, we first introduce our main concepts (section 2) and present the two studies and their results (sections 3–4). Based on these results, we argue that familiar voices can influence evaluations of artificial systems as social actors, but that this influence is selective, context-dependent, and insufficient to support straightforward assumptions of robust personality transfer or person reconstruction. We end by discussing limitations of the present studies and outline directions for future work (sections 5–6).

2 Conceptual Framework

To clarify the possible effects of familiar voices in robots and virtual characters, it is useful to distinguish between three conceptually different outcomes.

⁴ Voice cloning detaches vocal identity from the embodied person and makes its reuse scalable across platforms. Public debates about deepfakes and voice cloning highlight perceived threats to identity and trust. Beyond authentication risks [41], qualitative research documents fears of identity displacement, fragmentation of selfhood, and loss of control over vocal representation [8,22,23].

⁵ We adopt here the definition of a dark pattern proposed by Shamsudhin & Jotterand [37], who define a dark pattern as a persuasive design of which “(1) the intent to persuade must be necessary, (2) the resulting user action is not in the user’s best interest and is in the interest of the persuader, and (3) the methods employed are deceptive and covert” (p. 96).

Identity activation occurs when a familiar voice triggers recognition or retrieval of a stored representation of the original speaker. Hearing the voice can activate associated person knowledge, expectations, and affective responses linked to that individual. Such activation may influence how the artificial character is perceived or evaluated. However, the influence remains indirect and does not necessarily result in consistent attribution of the original speaker’s personality traits to the new target [21].

Trait transfer (as used in this paper) refers to a pattern in which evaluations of an artificial character converge toward trait and social judgments associated with a familiar speaker when the character uses a cloned version of that speaker’s voice. Importantly, this is not meant as *projection* in the classic sense of attributing one’s own characteristics to others. Rather, it describes an *attributional carryover* in which activated person knowledge about the familiar speaker systematically biases judgments of the new target.

Person reconstruction represents the strongest possible form of transfer. In this case, the familiar voice cue would reinstate a broader person impression, such that listeners evaluate the artificial character in ways that closely resemble their internal model of the original individual across multiple traits and social judgments [15,32].

These outcomes differ in strength and theoretical implication. *Identity activation* concerns the retrieval of person-related expectations, *trait transfer* concerns selective attribution of characteristics, and *person reconstruction* would imply that an isolated perceptual cue can sustain a coherent person representation. The studies presented in this paper examine whether familiar voice cloning produces evidence for *trait transfer* and whether it approaches the stronger case of *person reconstruction* in evaluations of embodied artificial characters.

3 Method

Table 1 summarizes the main design characteristics of the two studies. We use the empirical studies as a case to examine how voice-based identity cues operate in embodied characters.

Table 1. Overview of the two studies.

	Study 1	Study 2
Context	University lecture	Online communities
Participants (<i>n</i>)	17	65
Known-voice owner	Lecturer	Content creator
Design	Within-subject	Within-subject
Stimulus length	~14 s	~43 s
Speech content	Oldenburg sentences [18]	Trivia statements [28]
Personality measure	BFI-K (21 items) [31]	BFI-10 [30]
Likability & Competence	Items from [5,11]; German adaptation [42]	
Trust	—	[33]

Participants first rated a familiar real person from memory to capture their baseline internal person model. They then evaluated the same virtual character in two voice conditions: a *known-voice clone* and a *cloned unknown voice*. Visual appearance and spoken content were held constant, making voice the only manipulated factor. The unknown voice was unfamiliar to participants and served as control voice for the same character and speech content.

If the *known-voice clone* activates stored person knowledge and supports attribution transfer, character ratings should be closer to real-person ratings in the *known-voice clone* condition than in the *unknown voice* condition. We operationalized this as smaller absolute distances between known-person and character ratings across personality traits and social evaluation measures, indicating greater alignment between the remembered person model and the evaluated character.

The order of the two character conditions was randomized in Study 1 and Study 2; the memory-based familiar-person rating always preceded the character evaluations.

Study 1 was an exploratory study using a university lecturer as the familiar person. Participants were recruited in a university teaching context and completed the study during a lecture or tutorial using their own smartphones and headphones. The familiar person was the lecturer whose course they attended. Of 18 completed questionnaires, one participant was excluded for not knowing the lecturer, leaving 17 valid responses. Participants had a mean age of 22.19 years ($SD = 2.99$; range: 20–32).

Study 2 tested the same design online using the voices of two social media content creators. Participants were recruited through creator-related online communities, primarily Reddit, supplemented by mailing lists and social media posts. Of 208 participants who started the survey, 65 valid responses remained after applying preregistered exclusion criteria, including incomplete responses, failed attention checks, and insufficient creator recognition. Participants had a mean age of 27.9 years ($SD = 6.38$; range: 19–52).

At the beginning of Study 2, participants completed a familiarity check for the two selected creators and indicated which one they knew better. Actual familiarity was verified by requiring correct identification of both the creator’s real name and artist name, as well as content consumption within the previous three months. Participants who failed these preregistered criteria were excluded. After rating the selected creator from memory, participants evaluated the virtual character in the known-voice and unknown-voice conditions.

Study 2 was preregistered.⁶ All participants took part voluntarily and provided informed consent before participation. Voices were cloned with permission from the respective voice owners or from materials explicitly authorized for the study.

⁶ <https://aspredicted.org/zy8j9m.pdf>

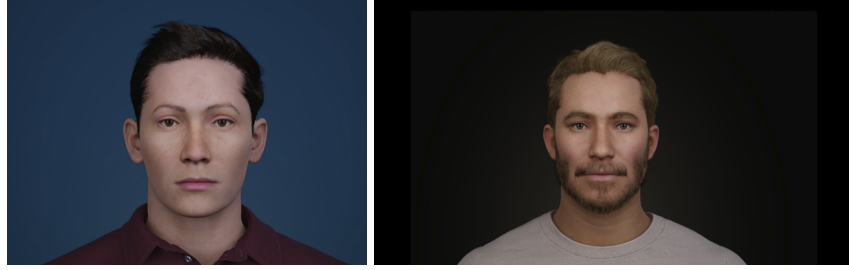


Fig. 1. Virtual characters used in the studies. Left: Character used in Study 1 (exploratory real-world study). Right: Character used in Study 2 (online social media study). In both studies, visual appearance and spoken content were held constant across voice conditions.

3.1 Measures

Personality. Study 1 assessed Big Five other-ratings using the German Big Five Inventory adapted for rating others [31,43] (21 items; five-point Likert scales; negatively keyed items reverse-coded). Study 2 used brief Big Five scales adapted for other-ratings [30].

Social evaluation. Likability was measured with a multi-item scale [11,42], and competence with established person-perception items [5]. Study 2 additionally measured trust [33]. Likability and competence used seven-point Likert scales in Study 1; Study 2 followed the instruments’ original response formats.

3.2 Stimuli

Two different virtual characters were used across the studies (Figure 1). Within each study, visual appearance and spoken content were held constant across voice conditions. The order of voice conditions was randomized.

Voice clones were generated from short reference recordings using zero-shot voice cloning systems [6,29]: OpenVoice in Study 1 [29] and XTTSv2 in Study 2 [6].

The virtual character and lip synchronization were created with MeLaX [44]. Stimulus duration differed between studies (Table 1): Study 1 used shorter, phonetically balanced sentences [18], whereas Study 2 used longer statements [28] to support more reliable personality judgments following participant feedback from Study 1.

3.3 Analysis

Study 1 was exploratory and used paired t-tests or Wilcoxon signed-rank tests depending on the distribution of the difference scores. Study 2 followed the preregistered directional analysis plan, using one-tailed paired t-tests or one-tailed Wilcoxon signed-rank tests to test whether distances were smaller in the

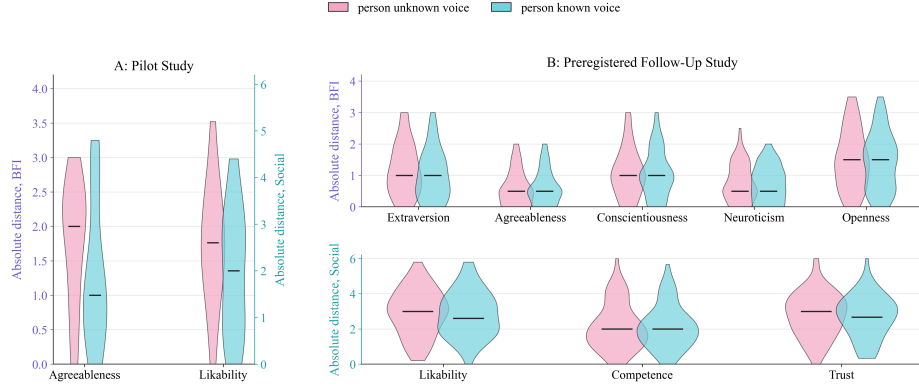


Fig. 2. Absolute distances between known-person ratings and character ratings for the unknown-voice and known-voice conditions. Panel A shows Study 1 ($n = 17$); Panel B shows Study 2 ($n = 65$). Violins show distance distributions, and horizontal lines indicate medians. In Panel A, BFI and social evaluation distances are shown on separate y-axes because the underlying response scales differed.

known-voice than in the unknown-voice condition. Distances were calculated as absolute differences between person and character ratings on each scale.

Open-ended responses from Study 2 were analyzed using qualitative content analysis [24], focusing on voice recognition, voice–appearance fit, perceived artificiality, and emotional tone.

Given its exploratory nature, Study 1 used no correction for multiple comparisons. Study 2 tested the directional hypothesis separately for each outcome according to the preregistered analysis plan,⁶ also without correction for multiple comparisons.

4 Results

4.1 Study 1 (Lecturer, $n = 17$)

Wilcoxon signed-rank tests showed significantly smaller distances in the known-voice condition for likability ($W = 11.5$, $p = .010$, rank-biserial $r = -.75$) and agreeableness ($W = 16.0$, $p = .038$, rank-biserial $r = -.59$). Likability distances were lower in the known-voice condition ($M = 1.96$, $SD = 1.35$, $Mdn = 2.00$) than in the unknown-voice condition ($M = 2.56$, $SD = 1.32$, $Mdn = 2.60$); the same pattern was found for agreeableness ($M = 1.28$, $SD = 1.05$, $Mdn = 1.00$ vs. $M = 1.74$, $SD = 0.92$, $Mdn = 2.00$). Competence and the remaining Big Five dimensions were not significant (Figure 2).

4.2 Study 2 (Content creators, $n = 65$, preregistered)

Only extraversion showed a significant effect in the predicted direction ($W = 588.0$, $p = .041$), with descriptively lower distances in the known-voice condi-

tion ($M = 1.02$, $SD = 0.81$, $Mdn = 1.00$) than in the unknown-voice condition ($M = 1.19$, $SD = 0.84$, $Mdn = 1.00$). All other outcomes, including likability, competence, and trust, were not significant (Figure 2). Thus, the Study 1 effects for likability and agreeableness did not reappear in Study 2; instead, only extraversion showed the predicted effect.

4.3 Qualitative Findings

Open-ended responses from Study 2 were analyzed using qualitative content analysis [24]. Figure 3 shows the frequency of coded themes by voice condition.

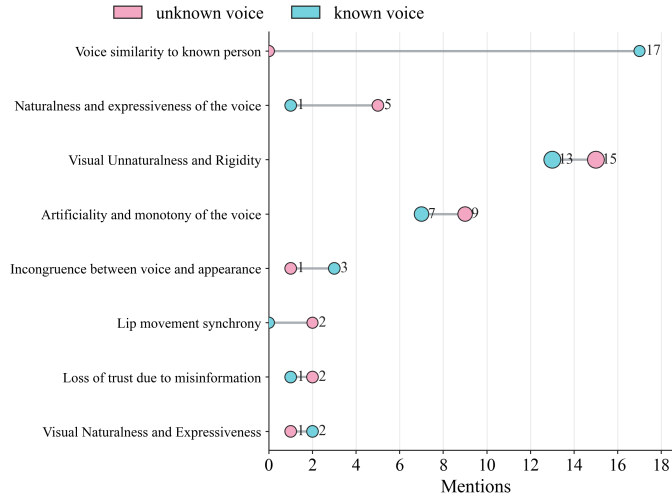


Fig. 3. Frequency of coded themes in open-ended responses by voice condition. Numbers indicate mentions.

Recognition of the cloned voice was salient in the known-voice condition (17 mentions). Participants often acknowledged similarity to the familiar person while emphasizing deviations from the original voice, indicating imperfect perceived fidelity.

Across both conditions, the dominant theme was visual rigidity and artificial facial movement (unknown: 15; known: 13 mentions). Participants described limited expression, stiffness, and, in some cases, an uncanny or unsettling impression, independent of voice familiarity.

Comments describing the voice as comparatively natural or pleasant were more frequent in the unknown-voice condition (5 vs. 1 mention), whereas voice–appearance mismatch was reported more often in the known-voice condition (3 vs. 1 mention).

Several participants questioned whether personality judgments were possible under the given conditions, suggesting perceived limits of trait attribution. In isolated cases, familiarity slightly attenuated negative impressions without reversing them. Distrust-related comments primarily referred to perceived misinformation rather than to the use of a familiar voice.

5 Discussion and Ethical Implications

Across both studies, familiar voice cloning produced only limited alignment between character ratings and memory-based ratings of the original speaker. Effects were selective, context-sensitive, and did not generalize across traits or study contexts. Interpreted through the conceptual distinction introduced earlier, the findings are therefore most consistent with *identity activation*, provide only weak evidence for *trait transfer*, and provide little evidence for *person reconstruction*.

The following sections discuss why familiar voices appear to function less as direct carriers of personality than as partial identity cues whose effects depend on the broader multimodal context.

5.1 A weak transfer channel under cue competition

Effects varied clearly by trait. Only extraversion⁷ showed the predicted uncorrected directional effect in Study 2, whereas agreeableness and likability effects from Study 1 did not replicate. This difference may relate to the nature of familiarity with the speaker. In Study 1, participants knew the lecturer from direct interaction, likely resulting in richer and more stable person representations that extend beyond vocal cues. In contrast, familiarity with social media creators is typically mediated and may rely more strongly on cues such as voice and visual appearance. Under such conditions, vocal cues may play a comparatively larger role in shaping trait impressions, which may partly explain why extraversion showed the clearest difference in the online context. Extraversion is closely tied to acoustic parameters such as vocal effort, intensity, and dynamic variability, making it comparatively voice-diagnostic [27,36]. Traits such as agreeableness or likability may rely more strongly on broader behavioral cues and may therefore be less recoverable from voice alone.

At the same time, overall effect sizes were small. Qualitative responses indicate that perceived vocal artificiality and visual rigidity often shaped impression formation. Familiar listeners are known to hold relatively robust mental representations of familiar voices and can detect subtle deviations in prosody, timing, and accent [1,14,32]. When a cloned voice diverges from this stored prototype, attention may shift toward evaluating the authenticity of the voice rather than forming person-based impressions. This may also affect social evaluations such as perceived warmth or credibility.

⁷ In personality psychology, extraversion refers to a tendency toward sociability, talkativeness, energetic behavior, and outward emotional expressiveness. In the Big Five framework, it is typically contrasted with introversion [13].

Prior work suggests that technical imperfections draw attention to a voice’s artificiality [32]. While it has been shown that familiar voices in emotionally charged contexts can retain persuasive effects despite such deficits [38], these findings were based on cases where listeners were exposed to audio cues only, so without a visual carrier that embodied the voice. Our studies suggest that coupling a familiar voice with an embodied character produces mismatch awareness that influences these relational effects.

5.2 Using clones of familiar voices configures expectations more than it transfers traits

Interpreting the results of Studies 1 and 2 through the conceptual distinction introduced in section 2 helps clarify the mechanisms at work. Participants frequently recognized or commented on the familiar voice, indicating that *identity activation* occurred. The cloned voice triggered comparisons with the original speaker and shaped expectations about how the character should appear or behave.

Evidence for *trait transfer* was considerably weaker. Only a small number of traits showed significant alignment effects, and these effects were inconsistent across contexts. Rather than systematically projecting personality characteristics from the original speaker to the artificial character, participants appeared to rely primarily on cues present in the stimulus itself.

Most importantly, the data provide little to no evidence for *person reconstruction*. The familiar voice did not reinstate a broader person impression corresponding to the remembered individual. Instead, the cloned voice functioned as a recognizable cue that activated expectations while remaining embedded in a multimodal environment where other cues such as visual appearance and perceived artificiality strongly shaped evaluation.

Importantly, participants often recognized the speaker identity in the familiar voice condition. Their comments focused on appearance fit and comparison with the original person. This aligns with evidence that cloned familiar voices can pass categorical identity thresholds while still deviating from the richer stored prototype [10,14,19,32]. Recognition can be high even when synthetic origin is detected only modestly above chance. *Identity activation* does not require reliable authenticity detection [1]. Our data suggest that such activation mainly configures expectations about appearance and behavior rather than transferring the speaker’s broader trait profile.

In self-voice paradigms, similarity-based mechanisms increase likability and behavioral trust, often via perceived homophily and personification [16,35]. Third-party familiar voices introduce a stronger reference standard and invite prototype comparison. This amplifies scrutiny and highlights deviations in prosody, style, or interactional fit [10,14]. Under these conditions, a familiar voice may heighten mismatch salience without producing a reliable positive shift in applied evaluations.

5.3 Boundary conditions suggested by the broader evidence

In the introduction, we raised the ethical concern that cloned voices could be used to strategically shape how people evaluate artificial characters. Prior research suggests, however, that such effects depend strongly on the surrounding context. When audiovisual cues are inconsistent, trust and believability decrease and uncanny impressions increase [12,34]. In contrast, when voice, appearance, and interaction quality are coherent, similarity cues have more opportunity to shape evaluation because they are less likely to be overshadowed by other concerns [3,14,32].

Our stimuli were intentionally controlled and relatively brief in order to isolate the effect of the voice cue. Even under these conditions, participants repeatedly reported that the available information was insufficient for reliable personality judgments. This suggests that short voice exposures alone may not provide enough information to support stable person-based projection. Rather than being inert, the effectiveness of known-voice cloning appears to depend on the broader cue environment. In particular, the visual appearance and behavioral coherence of the system deploying the voice may strongly constrain how the voice is interpreted. Alignment across multiple cues may therefore be required for familiar voices to meaningfully shape social evaluation.

5.4 Ethical implications

Ethically, these findings qualify rather than eliminate concerns about familiar voice cloning. The data do not support the assumption that a familiar cloned voice alone is sufficient to make an artificial character appear socially continuous with the original speaker. However, recognizable voices can still activate identity-based expectations and can make mismatches between voice, appearance, and behavior more salient.

This raises the question under what conditions cloned familiar voices should be used in social robotics and AI, if at all. What is it that a familiar voice could add to an artificial system that would enhance its capacities to function in a social context? One suggestion is that this use could increase the willingness to engage with systems meant to assist in educational and health practices. As we saw from our studies, however, the use of a familiar voice heightens awareness of the mismatch between the identity cue and the carrier of the voice. What is more, some participants stated that a mismatch between the voice and the carrier created uncanny or unsettling impressions (see section 4.3). Both of these findings make it likely that the use of a familiar voice would reduce the social functionality of the system rather than stimulate it. A further factor to consider is that especially in education and healthcare, users will be vulnerable. It might be that these users have more difficulty distinguishing reality from simulation, making it easier to manipulate them, or will be more affected by the uncanny or unsettling impressions. Given these factors, we therefore discourage the default use of familiar voices in social robotics, as well as their use without explicit consent.

At the same time, we acknowledge that voice cloning can have beneficial uses as well, like providing autonomy to people with speech impediments or voice loss by allowing them to communicate in a voice that they identify with [26]. We can even imagine scenarios in which users have the option to give a particular voice to a robot that they personally own. In both of these cases, the autonomy of the user is protected by turning the use of a familiar cloned voice into a *choice*. Even so, this should only be allowed under the condition that voice cloning itself is better regulated. To protect the autonomy of voice-owners, this entails developing transparent mechanisms to give consent for voice cloning, as well as the possibility to keep track of, update, and retract this consent. Additionally, the platforms that provide access to these voice cloning technologies should be properly monitored to ensure that they have these mechanisms in place. To protect the autonomy of the listeners, it is necessary to develop better mechanisms to detect impersonations of a person’s voice, also known as *deepfakes* [2]. Here, both non-technical mechanisms such as *family/friend codewords* [38] and technical mechanisms such as deepfake detectors that differentiate between human and synthetic voices [2] are promising candidates.

Finally, research into the effects of familiar voice cloning on personality transfer is bound to continue. We caution those pursuing this line of research to keep the following in mind:

- **Treat a cloned familiar voice as a weak identity cue.** Our results show no robust transfer of perceived personality, only small and selective alignment effects.
- **Multimodal coherence matters.** Perceived artificiality and voice–appearance mismatch dominate judgments and override familiarity effects.
- **Consider selective bias.** Even without broad personality transfer, recognizable voices can shift expectations and influence specific trait attributions.

Even when broad *trait transfer* fails, recognizable voices can still activate expectations and subtly bias evaluations. The use of familiar voices in Human-Robot and AI interaction therefore requires transparency, consent, and careful attention to multimodal coherence.

6 Limitations and Future Directions

Several limitations should be considered. First, both studies relied on brief, single-exposure stimuli, approximately 14 s in Study 1 and 43 s in Study 2. This helped isolate the voice manipulation, but may have limited the formation of stable person-based impressions.

Second, participants sometimes perceived the cloned voices and virtual characters as artificial. Such cues may have drawn attention away from personality attribution and toward authenticity or audiovisual fit.

Third, generalizability is limited by the sample sizes and study contexts. Study 1 involved a small university sample evaluating a single lecturer, while

Study 2 examined online communities around two content creators. The observed trait-selective effects therefore require replication.

Fourth, the studies focused on relatively specific forms of familiarity. Effects may differ for closer or more emotionally significant relationships, such as family members, close friends, or long-term colleagues. The findings are also limited to German-speaking contexts and explicit self-report measures.

Future work should examine familiar voice cloning in longer and more naturalistic interactions, across different forms of embodiment, and in voice-only settings. It should also compare different relationship types, include cross-linguistic and cross-cultural replications, and complement self-report ratings with behavioral or implicit measures. These extensions would help clarify when familiar voice cues meaningfully shape social evaluation and when their influence remains constrained by the broader interaction context.

7 Conclusion

Voice cloning enables embodied characters to speak with the recognizable voice of a familiar person. Across two studies, we examined whether this familiarity leads listeners to transfer parts of their internal person model from the original speaker to the artificial character.

Our findings provide little evidence for robust personality transfer. Effects were selective and did not generalize across traits or study contexts. Viewed through the conceptual framework introduced in this paper, the implementation of a familiar voice in an embodied artificial character appears to produce *identity activation* and, at most, limited *trait transfer*, while providing little evidence for *person reconstruction*.

These results suggest that familiar voice cloning functions less as a mechanism for reproducing personality and more as a context-dependent identity cue. Recognizable voices can shape expectations and selectively influence evaluations, yet their effects remain constrained by factors such as perceived artificiality and voice–appearance coherence.

From an ethical perspective, this moderates concerns that cloned familiar voices straightforwardly make artificial characters appear socially continuous with the individuals they resemble. At the same time, recognizable vocal identities can still activate trust-related expectations and subtly bias perception. Responsible deployment therefore requires clarity about the artificial system’s status and should avoid implying personal continuity or endorsement by the original speaker.

Acknowledgments. This research was funded by the European Union’s Horizon Europe research and innovation programme under the CAVAA project (Grant Agreement No. 101071178), which deals with counterfactual assessment and valuation for an artificial awareness architecture, and the PILLAR-Robots project (Grant Agreement No. 101070381), which aims to develop purposeful intrinsically motivated lifelong learning autonomous robots.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

Prior Version. A previous non-archival version of this work was presented as an extended abstract in the context of a non-archival poster [17]. That version already reported both empirical studies, including methodological information. The present paper builds on that empirical foundation but revises the framing and emphasis. While the data and core findings remain the same, the contribution is recontextualized for the Robophilosophy audience by focusing more explicitly on the limits of identity cue transfer, the conceptual implications of voice-based person reference, and the ethical positioning of embodied artificial characters.

References

1. Barrington, S., Cooper, E.A., Farid, H.: People are poorly equipped to detect ai-powered voice clones. *Scientific reports* **15**(1), 11004 (2025). <https://doi.org/10.1038/s41598-025-94170-3>
2. Blue, L., Warren, K., Abdullah, H., Gibson, C., Vargas, L., O'Dell, J., Butler, K., Traynor, P.: Who are you (i really wanna know)? detecting audio {DeepFakes} through vocal tract reconstruction. In: 31st USENIX Security Symposium (USENIX Security 22). pp. 2691–2708 (2022)
3. Bodenhausen, G.V., Todd, A.R.: Social cognition. *Wiley interdisciplinary reviews. Cognitive science* **1**(2), 160–171 (2010). <https://doi.org/10.1002/wcs.28>
4. Campeanu, S., Craik, F.I.M., Alain, C.: Speaker's voice as a memory cue. *International journal of psychophysiology : official journal of the International Organization of Psychophysiology* **95**(2), 167–174 (2015). <https://doi.org/10.1016/j.ijpsycho.2014.08.988>
5. Carli, L.L., LaFleur, S.J., Loeber, C.C.: Nonverbal behavior, gender, and influence. *Journal of Personality and Social Psychology* **68**(6), 1030–1041 (1995). <https://doi.org/10.1037/0022-3514.68.6.1030>
6. Casanova, E., Davis, K., Gölge, E., Gökner, G., Gulea, I., Hart, L., Aljafari, A., Meyer, J., Morais, R., Olayemi, S., Weber, J.: Xtts: a massively multilingual zero-shot text-to-speech model. In: *Interspeech 2024*. pp. 4978–4982. ISCA (2024). <https://doi.org/10.21437/Interspeech.2024-2016>
7. Chiou, E.K., Schroeder, N.L., Craig, S.D.: How we trust, perceive, and learn from virtual humans: The influence of voice quality. *Computers & Education* **146**, 103756 (2020). <https://doi.org/10.1016/j.compedu.2019.103756>
8. Du, J., Huang, H., Zou, X., Yin, S., Gao, B., Fu, X.: The social dynamics of voice cloning: Trust, privacy, and ethical tensions when sharing your ai voice replica. In: Pinatti de Carvalho, A.F., Vitak, J., Ma, X., Page, X., Rossitto, C., Su, N.M., Barkhuus, L., Divitini, M., Hughes, A.L., Huh-Yoo, J., Ignat, C., Kelly, R., Luther, K., McCrickard, S., McDonald, D., Pater, J., Tellioglu, H., Vivacqua, A., Wang, H.C., Cai, J., Jacobs, M., Stigberg, S.K. (eds.) *Companion Publication of the 2025 Conference on Computer-Supported Cooperative Work and Social Computing*. pp. 307–311. ACM Digital Library, Association for Computing Machinery (2025). <https://doi.org/10.1145/3715070.3749244>
9. Figueroa, D.A., Alaoja, M.S., Li, C., Chrysostomou, D.: Beyond words: The influence of character-inspired voices on human-ai dialogue system. *International Journal of Human-Computer Interaction* **42**(1), 541–552 (2026). <https://doi.org/10.1080/10447318.2025.2508315>

10. Fontaine, M., Love, S.A., Latinus, M.: Familiarity and voice representation: From acoustic-based representation to voice averages. *Frontiers in Psychology* **8**, 1180 (2017). <https://doi.org/10.3389/fpsyg.2017.01180>
11. Hale, J., Hamilton, A.F.d.C.: Cognitive mechanisms for responding to mimicry from others. *Neuroscience and biobehavioral reviews* **63**, 106–123 (2016). <https://doi.org/10.1016/j.neubiorev.2016.02.006>
12. Higgins, D., Zibrek, K., Cabral, J., Egan, D., McDonnell, R.: Sympathy for the digital: Influence of synthetic voice on affinity, social presence and empathy for photorealistic virtual humans. *Computers & Graphics* **104**, 116–128 (2022). <https://doi.org/10.1016/j.cag.2022.03.009>
13. John, O.P., Srivastava, S.: The big five trait taxonomy: History, measurement, and theoretical perspectives. In: Pervin, L.A., John, O.P. (eds.) *Handbook of personality*, pp. 102–138. Guilford Press, New York, NY (1999)
14. Kanber, E., Lavan, N., McGettigan, C.: Highly accurate and robust identity perception from personally familiar voices. *Journal of experimental psychology. General* **151**(4), 897–911 (2022). <https://doi.org/10.1037/xge0001112>
15. Kraus, M.W., Chen, S.: Facial-feature resemblance elicits the transference effect. *Psychological science* **21**(4), 518–522 (2010). <https://doi.org/10.1177/0956797610364949>
16. Kuch, J.M., Heisler, M., Klein, S., Mertes, S., Eing, L., André, E., Becker-Asano, C.: Your robot, my voice: Enhancing android robot likability through personalization by cloning the user’s voice. In: *2025 34th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. pp. 192–198. IEEE (2025). <https://doi.org/10.1109/RO-MAN63969.2025.11217611>
17. Kuch, J.M., Withanage Don, D.S., Heimerl, N., Gebhard, P., André, E., Schneeberger, T.: What happens when a virtual character speaks with the cloned voice of a person you know? In: *Extended Abstracts of the 2026 CHI Conference on Human Factors in Computing Systems. CHI EA ’26*, ACM, New York, NY, USA (2026). <https://doi.org/10.1145/3772363.3799332>
18. Kuehnelt, V., Kollmeier, B., Wagener, K.: Entwicklung und evaluation eines satztests für die deutsche sprache i: Design des oldenburger satztests. *Zeitschrift für Audiologie* **38**, 4–15 (1999)
19. Latinus, M., Belin, P.: Human voice perception. *Current biology : CB* **21**(4), R143–5 (2011). <https://doi.org/10.1016/j.cub.2010.12.033>
20. Lavan, N.: Time after time: Voice perception from first impressions to identity recognition. *Quarterly Journal of Experimental Psychology* **78**(12), 2583–2593 (2025). <https://doi.org/10.1177/17470218251379036>
21. Lavan, N., McGettigan, C.: A model for person perception from familiar and unfamiliar voices. *Communications psychology* **1**(1), 1 (2023). <https://doi.org/10.1038/s44271-023-00001-4>
22. Lee, P.Y.K., Ma, N.F., Kim, I.J., Yoon, D.: Speculating on risks of ai clones to selfhood and relationships: Doppelgänger-phobia, identity fragmentation, and living memories. *Proceedings of the ACM on Human-Computer Interaction* **7**(CSCW1), 1–28 (2023). <https://doi.org/10.1145/3579524>
23. Leuenberger, M.: The role of the voice for identity and implications for voice cloning technology. *Philosophy & Technology* **38**(4) (2025). <https://doi.org/10.1007/s13347-025-00969-3>
24. Mayring, P.: Qualitative content analysis: Theoretical background and procedures. In: Bikner-Ahsbals, A., Knipping, C., Presmeg, N. (eds.) *Approaches to qualitative research in mathematics education*, pp. 365–380. *Advances in Mathematics*

- Education, Springer, Dordrecht and Heidelberg and New York and London (2015). https://doi.org/10.1007/978-94-017-9181-6_13
25. McAleer, P., Todorov, A., Belin, P.: How do you say 'hello'? personality impressions from brief novel voices. *PloS one* **9**(3), e90779 (2014). <https://doi.org/10.1371/journal.pone.0090779>
 26. Mills, T., Bunnell, H.T., Patel, R.: Towards personalized speech synthesis for augmentative and alternative communication. *Augmentative and Alternative Communication* **30**(3), 226–236 (2014)
 27. Mohammadi, G., Vinciarelli, A.: Automatic personality perception: Prediction of trait attribution based on prosodic features. *IEEE Transactions on Affective Computing* **3**(3), 273–284 (2012). <https://doi.org/10.1109/T-AFFC.2012.5>
 28. Nadarevic, L., Erdfelder, E.: Initial judgment task and delay of the final validity-rating task moderate the truth effect. *Consciousness and Cognition* **23**, 74–84 (2014). <https://doi.org/10.1016/j.concog.2013.12.002>
 29. Qin, Z., Zhao, W., Yu, X., Sun, X.: Openvoice: Versatile instant voice cloning. *arXiv preprint arXiv:2312.01479* (2023), <https://arxiv.org/abs/2312.01479>
 30. Rammstedt, B., Kemper, C.J., Klein, M.C., Beierlein, C., Kovaleva, A.: Big five inventory (bfi-10). <https://doi.org/10.6102/ZIS76>
 31. Rammstedt, B., John, O.P.: Kurzversion des big five inventory (bfi-k). *Diagnostica* **51**(4), 195–206 (2005). <https://doi.org/10.1026/0012-1924.51.4.195>
 32. Rosi, V., Soopramanien, E., McGettigan, C.: Perception and social evaluation of cloned and recorded voices: Effects of familiarity and self-relevance. *Computers in Human Behavior: Artificial Humans* **4**, 100143 (2025). <https://doi.org/10.1016/j.chbah.2025.100143>
 33. Roth, D., Mal, D., Purps, C.F., Kullmann, P., Latoschik, M.E.: Injecting nonverbal mimicry with hybrid avatar-agent technologies. In: *Proceedings of the Symposium on Spatial User Interaction*. pp. 69–73. ACM Conferences, ACM, New York, NY (2018). <https://doi.org/10.1145/3267782.3267791>
 34. Ruijten, P.A.M., Midden, C.J.H., Ham, J.: Ambiguous agents: The influence of consistency of an artificial agent's social cues on emotion recognition, recall, and persuasiveness. *International Journal of Human–Computer Interaction* **32**(9), 734–744 (2016). <https://doi.org/10.1080/10447318.2016.1193350>
 35. Schanke, S., Burtch, G., Ray, G.: Dynamic voice clones elicit consumer trust. In: Bui, T. (ed.) *Proceedings of the 55th Hawaii International Conference on System Sciences*. *Proceedings of the Annual Hawaii International Conference on System Sciences*, Hawaii International Conference on System Sciences (2022). <https://doi.org/10.24251/HICSS.2022.538>
 36. Scherer, K.R.: Personality inference from voice quality: The loud voice of extroversion. *European Journal of Social Psychology* **8**(4), 467–487 (1978). <https://doi.org/10.1002/ejsp.2420080405>
 37. Shamsudhin, N., Jotterand, F.: Social robots and dark patterns: Where does persuasion end and deception begin? In: Jotterand, F., Ienca, M. (eds.) *Artificial Intelligence in Brain and Mental Health: Philosophical, Ethical & Policy Issues*, pp. 89–110. *Advances in Neuroethics*, Springer International Publishing, Cham (2021). https://doi.org/10.1007/978-3-030-74188-4_7
 38. Sharevski, F., Vander Loop, J., Evans, B., Ponticello, A.: "you creep! it really worked!": An empirical study of telephone scams with cloned familiar voices and trusted caller ids. In: *Proceedings of the 2025 New Security Paradigms Workshop*. pp. 92–107. ACM, New York, NY, USA (2025). <https://doi.org/10.1145/3774761.3774918>

39. Shiramizu, V.K.M., Lee, A.J., Altenburg, D., Feinberg, D.R., Jones, B.C.: The role of valence, dominance, and pitch in perceptions of artificial intelligence (ai) conversational agents' voices. *Scientific reports* **12**(1), 22479 (2022). <https://doi.org/10.1038/s41598-022-27124-8>
40. Tal-Or, N., Sela, S., Igumnov, I., Milwidsky, H.D., Rafaeli, B., Sanilevich, M.: Does what we know about actors' real lives: Influence our reactions to the characters they play? *Journal of Media Psychology* **33**(3), 155–164 (2021). <https://doi.org/10.1027/1864-1105/a000293>
41. Wang, K., Chen, M., Lu, L., Feng, J., Chen, Q., Ba, Z., Ren, K., Chen, C.: From one stolen utterance: Assessing the risks of voice cloning in the aigc era. In: Blanton, M., Enck, W., Nita-Rotaru, C. (eds.) *46th IEEE Symposium on Security and Privacy*. pp. 4663–4681. IEEE, Piscataway, NJ (2025). <https://doi.org/10.1109/SP61157.2025.00238>
42. Wessler, J., Schneeberger, T., Christidis, L., Gebhard, P.: Virtual backlash: non-verbal expression of dominance leads to less liking of dominant female versus male agents. In: *Proceedings of the 22nd ACM International Conference on Intelligent Virtual Agents. IVA '22*, Association for Computing Machinery, New York, NY, USA (2022). <https://doi.org/10.1145/3514197.3549682>
43. Wiechers, Y., Kandler, C.: Validierung des deutschen big five inventar-2. *Diagnostica* **71**(2), 64–75 (2025). <https://doi.org/10.1026/0012-1924/a000344>
44. Withanage Don, D.S., Kiderle, T., Mertes, S., Schiller, D., Ritschel, H., André, E.: Melax: Conversations with generative ai in socially interactive agents. In: Li, T., Paternò, F., Väänänen, K., Leiva, L., Spano, D., Verbert, K. (eds.) *Companion Proceedings of the 30th International Conference on Intelligent User Interfaces*. pp. 163–166. ACM Digital Library, Association for Computing Machinery (2025). <https://doi.org/10.1145/3708557.3716363>