



Synthetic speech and affective experience in virtual reality: A scoping review

Mateusz Dubiel, Jean Botev*

Department of Computer Science, University of Luxembourg, L-4364 Esch-sur-Alzette, Luxembourg.

***Correspondence to:** Jean Botev, Department of Computer Science, University of Luxembourg, L-4364 Esch-sur-Alzette, Luxembourg.
E-mail: jean.botev@uni.lu

Received: June 06, 2025 **Accepted:** September 28, 2025 **Published:** September 29, 2025

Cite this article: Dubiel M, Botev J. Synthetic speech and affective experience in virtual reality: A scoping review. *Empath Comput.* 2025;1:202513. <https://doi.org/10.70401/ec.2025.0011>

Abstract

Aims: This scoping review systematically maps the existing literature at the intersection of virtual reality (VR), synthetic speech, and affective computing. As immersive and voice-based technologies gain traction in education, mental health, and entertainment, it is critical to understand how synthetic speech shapes emotional experiences in VR environments. The review clarifies how these concepts are defined, how they contribute to empathic computing, and identifies common applications, methodological approaches, research gaps, and ethical considerations.

Methods: A comprehensive search across multiple databases (e.g., ACM Digital Library, IEEE Xplore, ScienceDirect) was conducted following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) framework. Eligible studies investigated synthetic or computer-generated speech in VR or comparable immersive 3D settings and assessed emotional responses or related outcomes. Data were extracted on study characteristics, applied technologies, affect-related measures, and reported effects.

Results: The findings reveal a growing interdisciplinary body of research at the convergence of synthetic speech technologies, embodied virtual agents, and affective data processing in immersive environments. Interest in this area has accelerated with the development of advanced text-to-speech models, suggesting increasing relevance and expansion of this topic.

Conclusion: This review underscores a rapidly expanding yet fragmented research landscape. It highlights conceptual and methodological gaps, stressing the need for clearer definitions, standardized evaluation measures, and ethically informed design of synthetic speech in VR. The results provide a foundation for advancing research and applications in emotionally responsive virtual environments.

Keywords: Virtual reality, synthetic speech, prosody, perception, affective computing, emotion recognition, empathy

1. Introduction

Advances in immersive technologies and artificial intelligence are enabling increasingly naturalistic forms of human-machine interaction. Virtual reality (VR), whether fully or semi-immersive, provides a controlled and low-confounding environment that has become a dynamic platform for education^[1], therapy^[2], training^[3], and social interaction^[4]. The effectiveness of VR experiences relies heavily on the realism and expressiveness of virtual agents, which are embodied characters capable of communicating through speech, gestures, and behavior. Synthetic speech plays a pivotal role in supporting these interactions and enabling the creation of scalable and adaptable voice interfaces.

However, while synthetic speech allows for control and customization, it also presents challenges in conveying emotional authenticity. Unlike natural human voices, synthetic voices often lack subtle prosodic features and affective nuances that convey emotion. They may also fail to align with the visual representation of an avatar, which can hinder emotional engagement, particularly in contexts that require empathy, trust, or affective resonance.

These challenges highlight the importance of synthetic speech in interdisciplinary domains such as affective computing and empathic computing. Affective computing emphasizes systems that can recognize and interpret human emotions, while empathic computing focuses on systems capable of responding in ways that mirror human emotion-sharing. Within VR, these paradigms are used not only to detect and adapt to users' emotional states but also to design agents that elicit meaningful, emotionally appropriate, and socially intelligent responses.



Emotional and expressive speech synthesis refers to the generation of artificial speech that conveys affective states such as joy, sadness, or anger through prosodic features including pitch, rhythm, stress, and intonation^[5]. Unlike traditional text-to-speech (TTS) systems, expressive approaches extend beyond intelligibility and pronunciation to approximate the emotional nuances of human speech, thereby enhancing realism and user engagement in applications such as virtual agents. As highlighted by Pitrelli *et al.*, expressive speech enables a system to “reinforce the message with paralinguistic cues, to communicate moods and other content beyond the text itself,” allowing it to “say the same words differently and appropriately for different situations”^[6].

Initially, the research on expressive synthetic speech focused on understanding and modeling prosody (e.g., intonation, rhythm, stress duration) to achieve the expressive effect^[7]. Early expressive TTS systems relied on formant synthesis^[8] and unit selection^[9,10]. However, their limited flexibility, dependence on speech databases, and constraints in computational power restricted both expressiveness and overall quality of the speech produced. In recent years, the advent of deep learning has transformed speech synthesis, enabling more natural and expressive outputs. The Tacotron 2 model^[11], based on an end-to-end neural architecture, was a breakthrough that allowed direct mapping from text to speech waveforms without the need for hand-crafted linguistic features or manual alignment rules. The introduction of variational autoencoders, which disentangle content, speaker identity, and style, provided greater control over the synthesis of specific emotions^[12]. More recently, diffusion models have further advanced expressive TTS by offering finer control over prosody^[13]. Mori *et al.*^[14] proposed that as the likeness of a robot or artificial entity to a human increases, so does our affinity and empathy, but only up to a certain threshold. Beyond this point, if the entity appears almost human yet not entirely convincing, positive feelings shift abruptly to unease or discomfort, a phenomenon known as the “uncanny valley”. When the likeness continues to improve to the point of being indistinguishable from a human, affinity is restored. Although extensively studied in the context of graphics and visual realism, the uncanny valley remains underexplored in the auditory domain. A recent study by Ross *et al.*^[15] reported a positive correlation between perceived voice realism and approval ratings, with no evidence of negative reactions to highly human-like synthetic voices.

Early theoretical frameworks for affective interaction via voice in VR also informed the development of emotion-aware systems employing synthetic speech. One example is MAUI^[16], a multimodal affective user interface proposed by Lisetti *et al.* This framework leverages advances in computer vision, prosody recognition, and biofeedback for real-time recognition of affect, leveraging multimodal inputs to create a “multimodal perceived user’s state”. MAUI further introduced the concept of “multicultural animated avatars” that users could select or create, designed to mirror facial expressions in real time. The purpose of such a mirroring feature is to provide feedback on users’ emotional states and can be applied to self-awareness or adaptive teaching strategies. Potential applications envisioned by the authors include driver safety, military training, and tele-home care.

Another illustrative example of adaptive intelligent agents is *AutoTutor*^[17], a system designed to adjust pedagogical strategies based on learners’ emotional states to enhance the learning process. This work emphasizes the intrinsic link between emotions and learning, particularly when students face challenging material that may trigger states such as confusion or frustration, which can eventually transition into more constructive states such as hope or enthusiasm. The study further envisioned the integration of multiple sensor channels into a unified emotion classifier, potentially achieved through sensor fusion techniques. The overarching goal of *AutoTutor* is to respond intelligently to learners’ emotions, fostering a “virtuous cycle” in which periods of confusion are resolved into productive engagement while avoiding boredom or prolonged frustration.

Only with recent advances in artificial intelligence has development of systems of this sophistication become feasible and increasingly relevant. Building on current research that integrates synthetic speech, affective experience, and VR, this article seeks to synthesize existing findings and contribute to the field by:

- Mapping the state of the art in VR-based applications of synthetic speech and affective experience.
- Developing a taxonomy and overview of existing research efforts and trends.
- Identifying common applications and methodological approaches.
- Highlighting research gaps and ethical considerations.

2. Method

To meet the research objectives and provide a transparent and rigorous overview of the topic, we conducted a scoping review following Arksey and O’Malley’s framework^[18] and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) guidelines^[19]. The PRISMA-ScR flow diagram is presented in [Figure 1](#), and the completed checklist is provided in [Table S1](#).

Relevant literature was identified by querying a combination of the terms “synthetic speech”, “virtual reality”, and “affective/emotional response”. Search terms were adapted according to database requirements, and the full list of queries is shown in [Table 1](#). The asterisks (*) used in terms such as affect, emotion, and response denotes wildcard characters, enabling retrieval of variations such as affective and affect. All searches were completed in May 2025.

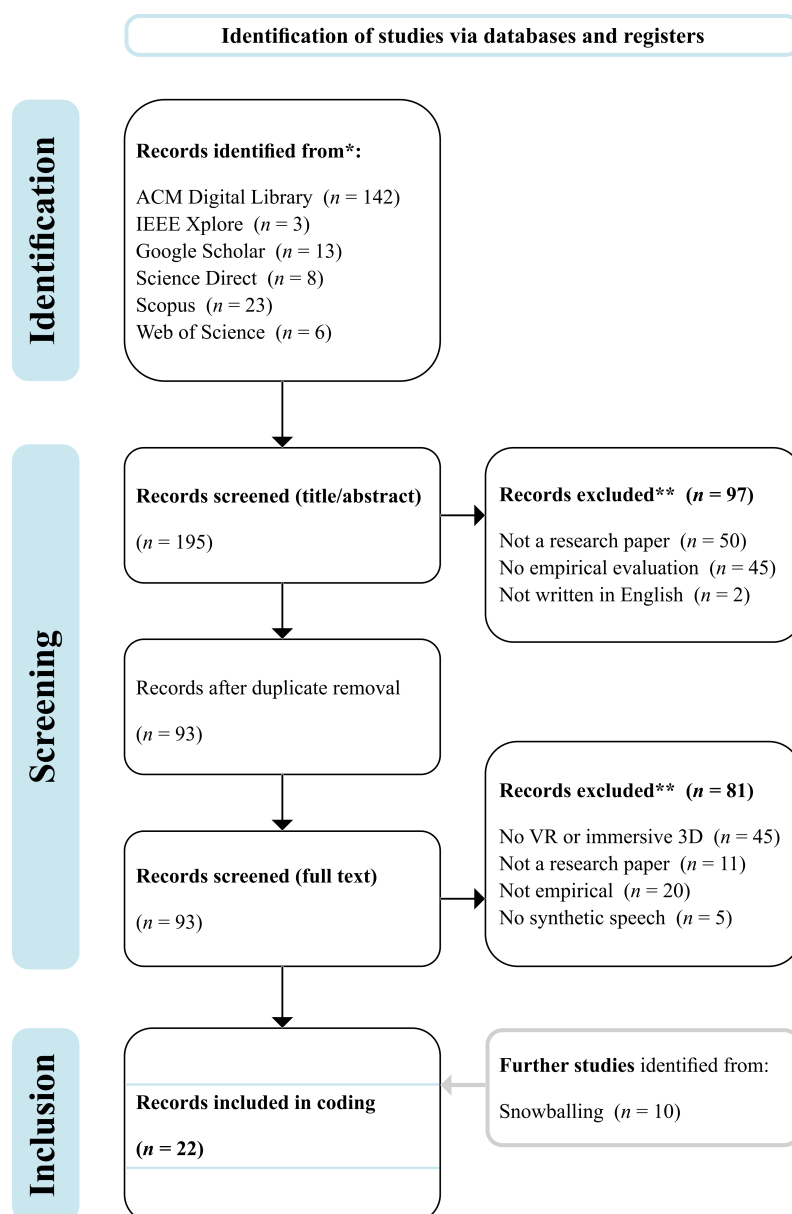


Figure 1. PRISMA-ScR approach for selecting relevant papers. PRISMA-ScR: Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews.

Table 1. Search strategy applied to databases.

Database	Query
ACM Digital Library	("synthetic speech") AND ("virtual reality" OR "vr")
IEEE Xplore	("synthetic speech") AND ("virtual reality" OR "vr")
Web of Science	("synthetic speech") AND ("virtual reality" OR "vr")
Google Scholar	("synthetic speech") AND ("virtual reality" OR "vr") AND ("affect* response" OR "emotion response" OR "affect feedback")
Science Direct	("synthetic speech") AND ("virtual reality" OR "vr") AND ("affective response" OR "emotional response" OR "affective feedback")
Scopus	("synthetic speech") AND ("virtual reality" OR "vr") AND ("affect* response" OR "emotion response" OR "affect feedback")

2.1 Search and selection

We conducted a manual search across six databases: ACM Digital Library, IEEE Xplore, Google Scholar, ScienceDirect, Scopus, and Web of Science. This process yielded 98 relevant publications from leading conferences and journals in the areas of human-computer interaction (HCI), affective computing, virtual agents, and immersive technologies.

In HCI and user experience, key sources included the *International Journal of Human-Computer Interaction*, the *Journal of Computer Assisted Learning*, and the ACM International Conference on Intelligent User Interfaces. For affective and multimodal systems, research was drawn from the *International Conference on Affective Computing and Intelligent Interaction*, as well as journals such as *Multimodal Technologies and Interaction* and *Expert Systems with Applications*. Core venues for virtual agents and autonomous systems included the ACM International Conference on Intelligent Virtual Agents and the International Conference on Autonomous Agents and Multiagent Systems. Foundational contributions in virtual reality and immersive graphics were published in the *IEEE Transactions on Visualization and Computer Graphics*, the IEEE Conference on Virtual Reality and 3D User Interfaces, and the IEEE International Symposium on Mixed and Augmented Reality. Additional contributions came from graphics-focused conferences such as *SIGGRAPH Motion, Interaction and Games* and journals such as *Computers & Graphics* and *Proceedings of the ACM on Computer Graphics and Interactive Techniques*. Speech-specific research was represented by the Annual Conference of the International Speech Communication Association and the *ACM Transactions on Applied Perception*. Bibliographic information, including author names, abstracts, publication year, and venue, was imported into Zotero. After removing five duplicates, 93 publications remained for screening. Both authors independently reviewed all articles using predefined inclusion and exclusion criteria. Consensus was reached to include 12 studies and exclude 81.

- **Inclusion:** Studies that examined the use of synthetic speech in virtual reality (fully or semi-immersive, i.e., immersive 3D environments) to elicit and measure users' affective experiences and emotional responses (e.g., anxiety, comfort, empathy, trust).

- **Exclusion:** Studies that were not peer reviewed (e.g., doctoral theses, preprints), lacked empirical evaluation, or were works in progress without reported findings.

We subsequently applied forward and backward snowballing to the 12 included publications, which yielded an additional 10 papers that met the inclusion criteria. In total, 22 studies were included in the review. These studies are listed in [Table 2](#) in reverse chronological order, based on their first appearance in a journal or conference.

Table 2. Studies included in the review, grouped by publication and year.

Conference or Journal Name	Year	Reference
Proceedings of the ACM on Computer Graphics and Interactive Techniques (PACMCGIT, ACM)	2024	[20]
Journal of Computer Assisted Learning (JCAL, Wiley)	2024	[21]
International Journal of Human-Computer Interaction (IJHCI, Taylor & Francis)	2024	[22]
Expert Systems with Application (ESWA, Elsevier)	2024	[23]
IEEE International Symposium on Mixed and Augmented Reality (ISMAR)	2023	[24]
ACM International Conference on Intelligent User Interfaces (ACM IUI)	2023	[25]
Autonomous Agents and Multi-Agent Systems (AAMAS, Springer)	2022	[26]
SIGCHI Conference on Human Factors in Computing Systems (CHI)	2022	[27]
Computers & Graphics (CG, Elsevier)	2022	[28]
Multimodal Technologies and Interaction (MTI, MDPI)	2022	[29]
ACM Transactions on Applied Perception (TAP)	2021	[30]
ACM International Conference on Intelligent Virtual Agents (IVA)	2021, 2023	[31,32]
ACM SIGGRAPH Conference on Motion, Interaction, and Games (MIG)	2021	[33]
IEEE Conference on Virtual Reality and 3D User Interfaces (IEEE VR)	2020, 2025	[34,35]
International Conference on Autonomous Agents and Multiagent Systems (AAMAS)	2020, 2023	[36,37]
International Conference on Affective Computing and Intelligent Interaction (ACII)	2019	[38]

IEEE Transactions on Visualization and Computer Graphics (TVCG, IEEE)	2018, 2025	[39,40]
Annual Conference of the International Speech Communication Association (InterSpeech)	2017	[41]

2.2 Data extraction

Key data were extracted from the included publications regarding relational aspects (Table 3), intrinsic characteristics, and interaction outcomes and impact factors (Table 4). Relevant factors were initially identified through an individual semantic review and subsequently finalized in a joint discussion. Individual items are presented in alphabetical order from left to right, while publications are arranged in reverse chronological order from top to bottom. A filled dot indicates presence, a circle indicates partial presence, and an empty space indicates absence of the respective factor in a given study.

Table 3. Relational aspects.

	Acceptance	Affect	Affinity	Appeal	Communication	Conversation Naturalness	Credibility	Empathy	Engagement	Likability	Perceived Comfort	Perceived Discomfort	Perceived Safety	Rapport	Speaking Anxiety Reduction	Trust	Understandability	Warmth
Bargum <i>et al.</i> , 2025 [35]	○				●			●	●	●					○			
Gao <i>et al.</i> , 2025 [40]								●		○					●			
Guo <i>et al.</i> , 2024 [20]	○			○	○				○	●					○	●	○	
Dai <i>et al.</i> , 2024 [21]				○					●							○		
Zhu <i>et al.</i> , 2024 [22]		○			●	●	○		○		●	●			●		●	
Llanes-Jurado <i>et al.</i> , 2024 [23]	○				●			○	○									
Obremski <i>et al.</i> , 2023 [37]					○			●									●	
Maxim <i>et al.</i> , 2023 [32]					●		○		○	●			●		●			
Choi <i>et al.</i> , 2023 [24]				○	○	○			○	●		●	○		○			
Chang <i>et al.</i> , 2023 [25]		●			○	○		○										
Do <i>et al.</i> , 2022 [27]	○			○	○	○	●		○						●			
Parmar <i>et al.</i> , 2022 [26]	●	●	○	○	○	●	●	●	○	●	○		○		●		●	
Higgins <i>et al.</i> , 2022 [28]		●	●	●	○	●	○	●		●	●		○		○	○		
Abdulrahman <i>et al.</i> , 2022 [29]	○	○				●		○	○	●		○			●	●	○	
Ehret <i>et al.</i> , 2021 [30]					○	●												
Zibrek <i>et al.</i> , 2021 [33]			●	●	○			○			●	●						
Murali <i>et al.</i> , 2021 [31]	○	○			●						○		○	●	○		○	
Morotti <i>et al.</i> , 2020 [34]	●				○	●			●		●				○			
Parmar <i>et al.</i> , 2020 [36]	●			○	○	●	●		○		○				●	○	○	
Kimani <i>et al.</i> , 2019 [38]					●				○					●	○			
Zibrek <i>et al.</i> , 2018 [39]			●	●	○			●	○		●	●			○			
Cabral <i>et al.</i> , 2017 [41]				●			●								○	●		

●: present; ○: partial.

Table 4. Intrinsic characteristics (left), interaction outcomes and impact factors (right).

	Aliveness	Anthropomorphism	Appropriateness	Believability	Expressiveness	Intelligence	Naturalness	Realism	Social Presence	Behavioral Change	Eeriness	Emotional Response	Information Recall	Persuasion	Presentation Experience	Racial Bias
Bargum <i>et al.</i> , 2025 ^[35]				●	●			●				○				
Gao <i>et al.</i> , 2025 ^[40]		○			○			●	●	●		○				
Guo <i>et al.</i> , 2024 ^[20]	●	○		●		●	○	○	●		●	○				
Dai <i>et al.</i> , 2024 ^[21]					●		●	○			○		●			
Zhu <i>et al.</i> , 2024 ^[22]								○	○							
Llanes-Jurado <i>et al.</i> , 2024 ^[23]					●	○	●	●	○	○		●				
Obremski <i>et al.</i> , 2023 ^[37]												●			●	
Maxim <i>et al.</i> , 2023 ^[32]	●	○							○					●		
Choi <i>et al.</i> , 2023 ^[24]		●		○		●	○	○	○	○	○	○				
Chang <i>et al.</i> , 2023 ^[25]				○	●			○				●				
Do <i>et al.</i> , 2022 ^[27]							●	●	○		○			●		
Parmar <i>et al.</i> , 2022 ^[26]			●	○			●	●			○	○		●		
Higgins <i>et al.</i> , 2022 ^[28]	○	○	○	●	●		●	●	●		●	●				
Abdulrahman <i>et al.</i> , 2022 ^[29]		●					○	○	●	●	●	●				
Ehret <i>et al.</i> , 2021 ^[30]	●	○					●	○	●			○				
Zibrek <i>et al.</i> , 2021 ^[33]				○	●		●	●	●		●	●				
Murali <i>et al.</i> , 2021 ^[31]			○				○	○		○		○			●	
Morotti <i>et al.</i> , 2020 ^[34]							●									
Parmar <i>et al.</i> , 2020 ^[36]			●				●	●			○	○		●		
Kimani <i>et al.</i> , 2019 ^[38]										●					●	
Zibrek <i>et al.</i> , 2018 ^[39]		○	○				○	●			●	○				
Cabral <i>et al.</i> , 2017 ^[41]					●		●	○								

●: present; ○: partial.

Relational aspects encompass affect and affinity, empathy, engagement, likeability, rapport, trust, and other related dimensions. Intrinsic characteristics include attributes such as believability, expressiveness, realism, and social presence. Interaction outcomes and impact factors primarily concern the observed and intended effects, including behavioral change, emotional response, and persuasive influence.

3. Results

The 22 included publications span the period from 2017 to 2025 (Table 2) and cover a range of application areas, including virtual counseling (Section 3.4), education (Section 3.3), public speaking support (Section 3.2), retail experiences (Section 3.5), and empathy elicitation (Section 3.1). As shown in Figure 2, most studies were initially published at conferences, with an increasing number appearing in journals over time. This trend reflects the growing interdisciplinary interest in the convergence of synthetic speech technologies, embodied virtual agents, and affective computing within immersive environments.

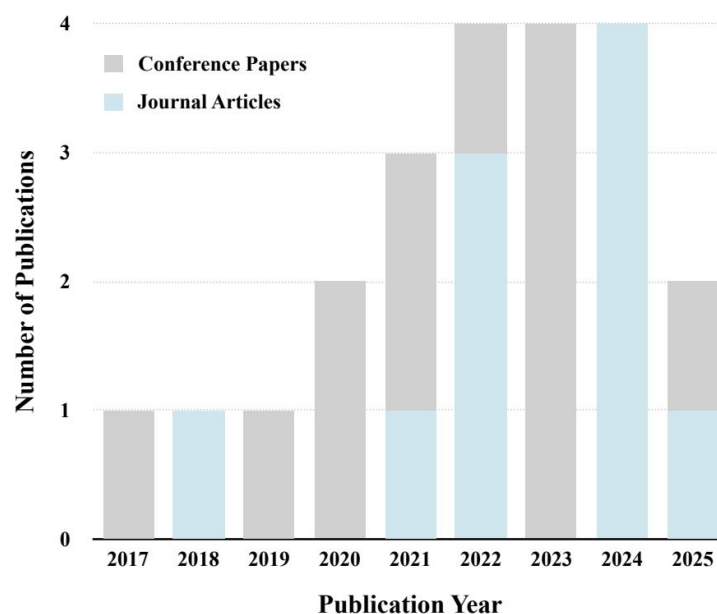


Figure 2. Number of publications per year and type.

Most studies employed between-subjects experimental designs, with participant sample sizes ranging from 28 to over 1,000. Affective experiences were primarily assessed through self-report measures (e.g., Likert scales evaluating trust, empathy, warmth, presence, eeriness, or affect) and behavioral data (e.g., eye tracking, interaction duration, or task compliance). Physiological measures, such as heart rate and skin conductance, were occasionally included but used less frequently. Synthetic voices were generated using TTS engines with varying degrees of expressiveness, while a few studies implemented custom or controlled emotional voice synthesis pipelines. Table 3 and Table 4 provide a detailed overview of the qualitative and methodological characteristics of these studies, and the following sections further synthesize and categorize the findings.

3.1 Emotional engagement and empathy

A substantial body of research has investigated how synthetic speech, particularly prosody and affective expression, influences users' perceptions of virtual agents. Evidence indicates that emotionally congruent speech enhances the perceived naturalness, warmth, and believability of agents, especially when aligned with avatar embodiment and visual realism^[30,35,41]. Related studies have explored how agent personality, conveyed through synthetic speech, impacts perceptions of trust, sympathy, and social presence^[25,28,33,39,40].

Cabral *et al.*^[44] examined how the quality of synthetic voices affects evaluations of virtual characters. Participants judged characters speaking with high-quality, emotionally expressive synthetic voices as more appealing and believable than those using neutral or flat voices, highlighting the role of vocal prosody in fostering social connection with digital agents. Ehret *et al.*^[30] systematically manipulated both prosody and embodiment in a conversational agent to assess their effects on perceived naturalness. They found that emotionally rich prosody significantly increased the perceived naturalness and human-likeness of the agent's speech, particularly when paired with a humanlike visual form. This emphasizes the importance of multimodal affective congruence in eliciting empathy and engagement. Bargum *et al.*^[35] investigated the impact of AI-generated speech on avatar perception in a VR environment. Their results showed that expressively modulated synthetic speech enhanced agents' lifelikeness and emotional engagement, especially when facial expressions and lip movements were well-synchronized. These findings highlight the need for alignment between vocal expressiveness and visual behavior in emotionally credible agents.

Chang *et al.*^[25] demonstrated that synthetic speech emotionally aligned with an agent's facial expressions fosters stronger rapport and greater emotional coherence, whereas mismatched affective cues disrupt the sense of natural interaction and reduce engagement. Similarly, Higgins *et al.*^[28] found that sympathy conveyed via voice and facial behavior enhanced users' emotional attunement. Eye-tracking data revealed heightened attention to agents whose vocal expressions matched visual affect, reinforcing the role of vocal prosody in empathic perception. Gao *et al.*^[40] argued that trust in stylized agents depends more on emotional congruence than on voice realism. Agents using affectively appropriate synthetic voices were perceived as more trustworthy and competent, even when their appearance was highly stylized or abstract. Finally, the two studies by Zibrek *et al.*^[33,39] provide a nuanced perspective on voice-visual interactions. Their first study indicated that the appeal of photorealistic avatars depended on congruent vocal delivery, while the second showed that synthetic voices lacking emotional depth or alignment undermined the credibility and naturalness of realistic characters. Collectively, these findings highlight that both vocal realism and emotional expression are crucial for maintaining social believability in high-fidelity avatars.

3.2 Affective regulation

Several studies have examined how virtual agents that use synthetic speech can support individuals in emotionally challenging situations, such as public speaking or mental health counseling. These studies indicate that affectively congruent synthetic speech can reduce anxiety, enhance self-confidence, and increase user comfort^[26,31,32,36,38].

Murali *et al.*^[31] introduced *Friendly Face*, a virtual agent designed to reduce public speaking anxiety by acting as a supportive audience member. The agent senses the speaker's behavior, including motion, speech, and prosody, and provides both emotional and instrumental support. Empirical evaluation against a text-only system showed that *Friendly Face* decreased public speaking anxiety, increased speaker confidence, and improved overall satisfaction. The approach of using a virtual agent as a calming and supportive audience member was well received and demonstrated considerable potential to enhance the public speaking experience. Similarly, Kimani *et al.*^[38] developed *You'll Be Great*, a virtual coaching system aimed at reducing public speaking anxiety through automated Cognitive Behavioral Therapy (CBT) with a focus on cognitive restructuring. The system significantly reduced maladaptive thoughts, alleviated cognitive and somatic anxiety, and improved participants' presentation experience. Users reported feeling a strong connection and understanding from the CBT agent and expressed willingness to continue using it for future presentations. Another line of research investigated affective regulation in health counseling contexts. Parmar *et al.*^[36] examined how animation quality, speech quality, and rendering style influence user perceptions and persuasive outcomes. They hypothesized that agents with consistent quality across animation, voice, and rendering would have the greatest impact, rather than a simple additive effect of high quality in each channel. The study found that while natural animations and recorded human voices generally enhance acceptance, trust, and credibility, they do not necessarily maximize persuasion. For tasks such as health counseling, minimal agent animation (using a static character) may be more effective for persuasive outcomes. This challenges the initial hypothesis and highlights the complex interplay between agent design and task goals. In a follow-up study, Parmar *et al.*^[26] explored simulated empathy as a design variable and its interaction with realism factors. High and consistent levels of empathy improved agent perception, but excessively high empathy could interfere with the formation of a trusting bond. In the mental health domain, Maxim *et al.*^[32] investigated how manipulating a virtual human's vocal properties to convey different levels of extroversion affects user perceptions, rapport, and persuasion. They found that matching a virtual human's vocal extroversion to the user's personality does not reliably produce a similarity-attraction effect or improve persuasion. Instead, virtual humans exhibiting low extroversion traits, such as slower speech, lower pitch, and reduced volume, were generally more persuasive for promoting mental wellness, independent of the user's personality. The study also emphasized that the user's own personality plays a significant role in shaping perceptions and interactions with virtual agents.

3.3 Educational and cognitive engagement

Another prominent area of research concerns the role of synthetic voices in VR in supporting learning, cognitive engagement, and attention. Studies in educational and collaborative learning contexts have shown that expressive or emotionally stable voices can enhance student engagement, retention, and motivation during tasks^[20,21,24,29].

Dai *et al.*^[21] investigated the use of TTS agents in VR primary school classrooms, comparing emotionally expressive and neutral voices. The study found that emotionally congruent synthetic speech significantly increased students' enjoyment, attentiveness, and task performance, suggesting that affectively rich voices can enhance engagement and comprehension in immersive educational settings. Abdulrahman *et al.*^[29] evaluated the impact of natural versus synthetic speech in VR-based English language learning. Learners reported higher motivation and perceived helpfulness when guided by a synthetic agent with more natural-sounding prosody, providing further evidence that voice realism can facilitate second-language acquisition in immersive environments. Guo *et al.*^[20] explored the use of self-similar virtual agents in a puzzle-solving task, manipulating both voice and appearance to match participants' gender and ethnicity. Agents with synthetic voices resembling the user led to increased perceived collaboration and social cohesion, suggesting that voice personalization can foster affective alignment and improve group performance in VR teamwork scenarios. Choi *et al.*^[24] examined team performance dynamics in a collaborative VR game, focusing on how the emotional stability of agents' voices affected users' experience. Agents with emotionally stable synthetic speech were perceived as more competent and reassuring, resulting in improved coordination and higher player confidence. These findings highlight the importance of voice tone in shaping cooperative dynamics within educational simulations.

3.4 Persuasion, empathy, and bias reduction

Three studies have examined the impact of synthetic speech on persuasive effectiveness, interpersonal empathy, and bias-related responses. These studies focused on agents designed to reduce implicit or explicit racial bias, promote empathic reasoning, and persuade users to modify their attitudes or decisions^[23,27,37].

Obremski *et al.*^[37] investigated a VR intervention using Intelligent Virtual Agents (IVAs) to reduce implicit and explicit racial bias. The scenario involved a pub quiz in which a mixed-cultural IVA acted as a supportive teammate and positive out-group exemplar, while a mono-cultural IVA served as a rude, uncooperative in-group exemplar. Interaction with the virtual agents resulted in significant decreases in both implicit and explicit racial bias scores among participants. Llanes *et al.*^[23] developed virtual human (VH) characters

designed to elicit social emotions through conversation. The system combined realistic avatars with a large language model (LLM) to generate contextually relevant responses, incorporating personality, mood, and attitudes. The VHs exhibited synchronized lip movements, voice synthesis, and facial expressions to reflect the intended emotion. In a study with 64 participants interacting with six different VHs, each targeting a specific basic emotion, the system successfully generated the intended emotional valence. Participants rated the conversations as relatively natural and realistic. While arousal was less effectively evoked, it was still detectable in the VHs. Do *et al.*^[27] examined how speech fidelity and listener gender influence social perceptions of a virtual human speaker. They compared a standard concatenative TTS, an advanced neural TTS, and recorded human speech. Results showed that neural TTS led to the VH being perceived as significantly less trustworthy by both male and female listeners compared with human speech. Male listeners also judged standard TTS as less trustworthy than human speech. These findings suggest that neural TTS may not be optimal for persuasive virtual humans and that listener gender plays a role in shaping perceptions of synthetic speech.

3.5 Commercial and aesthetic appeal

Several studies^[22,34] have investigated how synthetic voice characteristics, including tone, pacing, and affect, influence consumer trust, purchase intent, and perceived agent professionalism in commercial applications such as VR shopping. Zhu *et al.*^[22] examined the effects of speech modality and avatar realism on user perceptions in a VR shopping environment. Their findings indicated that integrating synthetic speech with highly anthropomorphic avatars enhanced perceived realism, trust, and shopping intent. The study highlights the importance of aligning voice characteristics with visual design to improve the consumer experience in VR retail. Morotti *et al.*^[34] studied voice assistants in fashion retail VR, demonstrating that affective tone and prosody influenced perceptions of professionalism and brand coherence. Expressive voices that matched the store's aesthetic increased user satisfaction and contributed to a more immersive shopping experience.

4. Discussion

As with any scoping review, this study is limited by the selection criteria and scope of the included publications. Some relevant work may have been excluded due to language, indexing, or publication type. Nevertheless, the retained studies indicate a growing interest in using emotionally expressive synthetic voices to influence user perception and behavior in immersive environments. Synthetic speech was found to modulate core affective outcomes, including empathy, trust, discomfort, and anxiety regulation, across diverse applications ranging from emotional tutoring and social coaching to mental health support and retail interactions.

A key theme emerging from the studies is that prosody and emotional congruence between a synthetic voice and a virtual agent's visual or behavioral design strongly shape users' affective responses. Monotone or generic TTS voices often hinder emotional connection, whereas emotionally expressive synthetic voices, particularly when aligned with an agent's visual embodiment, enhance engagement, empathy perception, and social presence. These effects were particularly pronounced in healthcare and counseling contexts, where affective alignment between voice and visual cues is critical for establishing trust and rapport. For example, Chang *et al.*^[25] demonstrated that aligning affective vocal tone with agent behavior substantially improves perceptions of naturalness and emotional consistency. Similarly, Parmar *et al.*^[26] found that agents using emotionally appropriate synthetic speech elicited higher empathy and interpersonal warmth ratings. Studies employing synthetic voices to reduce anxiety or support emotional regulation also reported positive outcomes. Murali *et al.*^[31] showed that a friendly synthetic voice in a supportive VR agent helped participants manage public speaking anxiety, enhancing both confidence and calmness. Obremski *et al.*^[37] similarly reported that interaction with emotionally aware virtual agents increased participants' emotional openness.

However, the benefits of synthetic speech are not universal. Several studies examining trust and social presence found that flat or unnatural prosody undermines perceptions of authenticity and credibility. Some of these effects can be mitigated through affective tuning or multimodal alignment. Gao *et al.*^[40] found that stylized, emotionally congruent voices improved trust in stylized agents, even when speech was not fully human-like. Conversely, incongruent or low-fidelity synthetic voices can induce discomfort, particularly when paired with photorealistic avatars, supporting the uncanny valley hypothesis in vocal perception. Zibrek *et al.*^[33] and Higgins *et al.*^[28] demonstrated that mismatches between high visual realism and low vocal expressiveness increased user discomfort and reduced believability. This evidence highlights that incongruent multimodal cues, especially in emotionally salient contexts, can disrupt affective resonance. As noted in [Section 1](#), there is a trade-off between fidelity and perceived pleasantness^[15]. Moreover, even advanced expressive TTS models often struggle to replicate the subtle affective nuances of natural speech, limiting their effectiveness in emotionally sensitive scenarios.

4.1 Ethical considerations

High-fidelity voice replication has become increasingly feasible thanks to large-scale neural TTS systems trained on thousands of hours of speech^[42]. As a result, virtual agents with high-quality synthetic voices are becoming more widespread, raising important ethical questions regarding their impact on users.

One key area of concern is trust and user agency. LLM-based agents that communicate through speech are often anthropomorphized and perceived as more trustworthy than non-verbal agents^[43]. Szekely *et al.* caution that increasing the realism, expressivity, and

human-likeness of synthetic voices can have socio-affective effects, potentially influencing user perceptions and behavior^[44]. Emotional proximity with highly anthropomorphic agents may impair human judgment and affect willingness to follow their advice^[45]. Dubiel *et al.*^[46] provide an example of this influence, showing that expressive synthetic speech can subtly steer users' choices, thereby undermining their agency.

Another area of concern involves relationships and well-being. Research indicates that users tend to interact with highly anthropomorphic AI companions as if they were in human-like relationships^[47], which can lead to dependence and emotional attachment^[48]. The long-term consequences of such interactions remain unclear. In a large-scale study ($n = 981$), Phang *et al.*^[49] investigated the impact of ChatGPT's Advanced Voice Mode (AVM) on users' emotional well-being and reported mixed outcomes. While AVM showed short-term benefits, prolonged use was associated with increased loneliness, greater emotional dependence, and compulsive usage. In response, Raedler *et al.*^[50] proposed actionable strategies to mitigate potential harm from AI companions, including encouraging real-world socialization, setting usage limits, and screening users for risk factors.

Given these potential risks, the use of expressive, human-like voices in virtual reality warrants careful investigation and broader ethical discussion. This is particularly important in immersive environments, where the perceived realism and emotional engagement of synthetic voices may amplify their impact on users.

4.2 Integration with affective and empathic computing

The findings highlight the central role of affective computing as a foundation for emotionally intelligent interactions. In this framework, synthetic speech functions both as a channel for emotional expression and as a cue for emotional inference. The reviewed studies indicate that affective congruence in voice is essential for empathic computing, enabling systems to not only detect user emotions but also respond in socially intelligent and emotionally appropriate ways. Consequently, voice serves not merely as a mode of interaction but as a critical component for eliciting perceived empathy in artificial agents.

4.3 Limitations of the Literature

Despite promising findings, the reviewed literature exhibits several limitations. First, most studies focused on short-term emotional responses, providing limited insight into the longitudinal effects of synthetic voices on user emotion or rapport. Second, sample populations were often restricted to university students or online participants, raising questions about the cultural and age-related generalizability of the results. Third, the heterogeneity of measurement approaches, ranging from single-item Likert scales to biometric data, complicates cross-study comparisons. Finally, many studies relied on off-the-shelf TTS engines with limited emotional control. Only a few employed state-of-the-art expressive TTS systems, such as emotion-conditioned models, while most used commercial platforms that constrain the granularity of experimental manipulations.

4.4 Implications for design and research

Designing emotionally credible virtual agents requires treating voice, appearance, and behavior as interconnected affective cues. Misalignment among these cues can undermine trust and perceived realism, whereas affective congruence enhances emotional credibility and social presence. Future research should focus on multimodal affect modeling, real-time adaptive voice synthesis, and cross-cultural validation to develop agents that are not only functional but also emotionally responsive and capable of supporting empathic computing. Additionally, transitioning from fully controlled VR environments to augmented reality (AR), which involves background noise, unpredictable acoustics, and dynamic user contexts, presents challenges for evaluating synthetic speech technologies. This shift necessitates careful examination of agent performance in less controlled and more ecologically valid conditions.

5. Conclusion

This scoping review demonstrates that although research on synthetic speech in VR is growing, its application for emotional and affective purposes remains underdeveloped. Existing studies indicate that features such as prosody, affective congruence, and voice-avatar alignment can significantly influence perceptions of realism, trust, empathy, and user engagement across diverse domains, including education, mental health, collaboration, retail, and social bias mitigation.

The review provides a structured synthesis of 22 peer-reviewed studies, organizing the literature around key affective outcomes and thematic application areas. It identifies patterns such as the importance of emotional alignment across modalities (e.g., voice and embodiment), the role of affectively expressive speech in enhancing learning and persuasion, and the potential of synthetic voices to shape social dynamics in immersive contexts. Nevertheless, the field remains fragmented, with considerable methodological variation and limited consensus on evaluating affective responses or comparing results across studies.

Notably, many studies rely primarily on subjective measures, with few incorporating physiological or behavioral data. Comparisons between synthetic and natural speech are rare, and the affective qualities of voices are often under described or inconsistently manipulated. Moreover, synthetic voice design seldom considers cultural, linguistic, or individual differences, despite evidence that perceived similarity and contextual appropriateness can influence emotional outcomes.

By identifying these trends and limitations, this review highlights both the potential and current boundaries of affective synthetic speech in VR, providing guidance for future research aimed at advancing synthetic speech technologies and enhancing affective experience in immersive environments for empathic computing.

Supplementary materials

The supplementary material for this article is available at: [Supplementary materials](#).

Declarations

Acknowledgements

AI (Gemini Flash 2.0) was used exclusively to support language editing and polishing. The authors are fully responsible for the accuracy, integrity, and originality of the final content.

Authors contribution

Mateusz Dubiel: Writing-review & editing, writing-original draft, methodology, investigation, formal analysis.

Jean Botev: Writing-review & editing, writing-original draft, visualization, methodology, investigation, formal analysis, funding acquisition.

Conflicts of interest

Jean Botev is an Editorial Board Member of *Empathic Computing*. Another author has no conflicts of interest to declare.

Ethical approval

Not applicable.

Consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

Not applicable.

Funding

None.

Copyright

©The Author(s) 2025.

References

1. Kavanagh S, Luxton-Reilly A, Wuensche B, Plimmer B. A systematic review of Virtual Reality in education. *Themes Sci Technol Educ*. 2017;10(2):85-119. Available from: <https://files.eric.ed.gov/fulltext/EJ1165633.pdf>
2. Powers MB, Emmelkamp PMG. Virtual reality exposure therapy for anxiety disorders: A meta-analysis. *J Anxiety Disord*. 2008;22(3):561-569. [DOI]
3. Xie B, Liu H, Alghofaili R, Zhang Y, Jiang Y, Lobo FD, et al. A review on virtual reality skill training applications. *Front Virtual Real*. 2021;2:645153. [DOI]
4. Sun N, Botev J. Digital partnerships: Understanding delegation and interaction with virtual agents. In: Chen Y, Blasch E, editors. *Digital Frontiers - Healthcare, Education, and Society in the Metaverse Era*. London: IntechOpen; 2024. [DOI]
5. Schröder M. Expressive speech synthesis: Past, present, and possible futures. In: Tao J, Tan T, editors. *Affective information processing*. London: Springer; 2009. p. 111-126. [DOI]
6. Pitrelli JF, Bakis R, Eide EM, Fernandez R, Hamza W, Picheny MA. The IBM expressive text-to-speech synthesis system for American English. *IEEE Trans Audio Speech Lang Process*. 2006;14(4):1099-1108. [DOI]
7. Grosz BJ, Hirschberg J. Some intonational characteristics of discourse structure. In: 2nd International Conference on Spoken Language Processing (ICSLP 1992); 1992 Oct 13-16; Alberta, Canada. International Speech Communication Association; 1992. p. 429-432. [DOI]

8. Cahn JE. The generation of affect in synthesized speech. *J Am Voice I/O Soc.* 1990;8(1):1-19. Available from: <https://www.cs.columbia.edu/~julia/papers/cahn90.pdf>
9. Hunt AJ, Black AW. Unit selection in a concatenative speech synthesis system using a large speech database. In: 1996 IEEE international conference on acoustics, speech, and signal processing conference proceedings; 1996 May 9; Atlanta, USA. Piscataway: IEEE. 1996. p. 373-376. [DOI]
10. Iida A, Campbell N. Speech database design for a concatenative text-to-speech synthesis system for individuals with communication disorders. *Int J Speech Technol.* 2003;6(4):379-392. [DOI]
11. Shen J, Pang R, Weiss RJ, Schuster M, Jaitly N, Yang Z, et al. Natural tts synthesis by conditioning wavenet on mel spectrogram predictions. In: 2018 IEEE international conference on acoustics, speech and signal processing (ICASSP); 2018. p. 4779-4783; Calgary, Canada. Piscataway: IEEE; 2018. p. 4779-4783. [DOI]
12. Zhang YJ, Pan S, He L, Ling ZH. Learning latent representations for style control and transfer in end-to-end speech synthesis. In: ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); Brighton, UK; 2019 May 12; Piscataway: IEEE; 2019. p. 6945-6949. [DOI]
13. Huang R, Zhang C, Ren Y, Zhao Z, Yu D. Prosody-tts: Improving prosody with masked autoencoder and conditional diffusion model for expressive text-to-speech. In: Rogers A, Boyd-Graber J, Okazaki N, editors. Findings of the Association for Computational Linguistics: ACL 2023. Toronto: Association for Computational Linguistics; 2023. p. 8018-8034. [DOI]
14. Mori M, MacDorman KF, Kageki N. The uncanny valley [from the field]. *IEEE Robot Autom Mag.* 2012;19(2):98-100. [DOI]
15. Ross A, Corley M, Lai C. Is there an uncanny valley for speech? Investigating listeners' evaluations of realistic synthesised voices. In: Speech Prosody 2024 2024; 2024 Jul 2-5; Leiden, The Netherlands. (pp. 1115-1119). International Speech Communication Association (ISCA); 2024. p. 1115-1119. Available from: https://www.isca-archive.org/speechprosody_2024/ross24_speechprosody.pdf
16. Lisetti CL, Nasoz F. MAUI: a multimodal affective user interface. In: Proceedings of the tenth ACM international conference on Multimedia; 2002 Dec 1; France. New York: Association for Computing Machinery; 2002. p. 161-170. [DOI]
17. D'Mello S, Picard RW, Graesser A. Toward an affect-sensitive AutoTutor. *IEEE Intell Syst.* 2007;22(4):53-61. [DOI]
18. Arksey H, O'malley L. Scoping studies: towards a methodological framework. *Int J Soc Res Methodol.* 2005;8(1):19-32. [DOI]
19. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. *Ann Intern Med.* 2018;169(7):467-473. [DOI]
20. Guo S, Choi M, Kao D, Mousas C. Collaborating with my doppelgänger: The effects of self-similar appearance and voice of a virtual character during a jigsaw puzzle co-solving task. *Proc ACM Comput Graph Interact Tech.* 2024;7(1):1-23. [DOI]
21. Dai L, Kritskaia V, van der Velden E, Vervoort R, Blankendaal M, Jung MM, et al. Text-to-speech and virtual reality agents in primary school classroom environments. *J Comput Assist Lear.* 2024;40(6):2964-2984. [DOI]
22. Zhu S, Hu W, Li W, Dong Y. Virtual Agents in Immersive Virtual Reality Environments: Impact of Humanoid Avatars and Output Modalities on Shopping Experience. *Int J Hum-Comput Interact.* 2023;40(19):5771-5793. [DOI]
23. Llanes-Jurado J, Gómez-Zaragozá L, Minissi ME, Alcañiz M, Marín-Morales J. Developing conversational virtual humans for social emotion elicitation based on large language models. *Expert Syst Appl.* 2024;246:123261. [DOI]
24. Choi M, Koiliias A, Volonte M, Kao D, Mousas C. Exploring the appearance and voice mismatch of virtual characters. In: 2023 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct); Sydney, Australia; 2023 Oct 16. Piscataway: IEEE; 2023. p. 555-560. [DOI]
25. Chang CJ, Sohn SS, Zhang S, Jayashankar R, Usman M, Kapadia M. The importance of multimodal emotion conditioning and affect consistency for embodied conversational agents. In: Proceedings of the 28th International Conference on Intelligent User Interfaces; 2023 Mar 27; Sydney, Australia. New York: Association for Computing Machinery; 2023. p. 790-801. [DOI]
26. Parmar D, Olafsson S, Utami D, Murali P, Bickmore T. Designing empathic virtual agents: manipulating animation, voice, rendering, and empathy to create persuasive agents. *Auton Agent Multi-Agent Syst.* 2022;36(1):17. [DOI]
27. Do TD, McMahan RP, Wisniewski PJ. A new uncanny valley? The effects of speech fidelity and human listener gender on social perceptions of a virtual-human speaker. In: Proceedings of the 2022 CHI conference on human factors in computing systems; 2022 Apr 29; New Orleans, USA. New York: Association for Computing Machinery; 2022. p. 1-11. [DOI]
28. 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. *Comput Graph.* 2022;104:116-128. [DOI]
29. Abdulrahman A, Richards D. Is natural necessary? Human voice versus synthetic voice for intelligent virtual agents. *Multimodal Technol Interact.* 2022;6(7):51. [DOI]
30. Ehret J, Bönsch A, Aspöck L, Röhr CT, Baumann S, Grice M, et al. Do prosody and embodiment influence the perceived naturalness of conversational agents' speech? *ACM Trans Appl Percept.* 2021;18(4):1-15. [DOI]
31. Murali P, Trinh H, Ring L, Bickmore T. A friendly face in the crowd: Reducing public speaking anxiety with an emotional support agent in the audience. In: Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents; 2021 Sep 14; Japan. New York: Association for Computing Machinery; 2021. p. 156-163. [DOI]
32. Maxim A, Zalake M, Lok B. The impact of virtual human vocal personality on establishing rapport: A study on promoting mental wellness through extroversion and vocalics. In: Proceedings of the 23rd ACM International Conference on Intelligent Virtual Agents; 2023 Sep 19-22; Würzburg Germany. New York: Association for Computing Machinery; 2023. p. 1-8. [DOI]
33. Zibrek K, Cabral J, McDonnell R. Does synthetic voice alter social response to a photorealistic character in virtual reality? In: Proceedings

- of the 14th ACM SIGGRAPH Conference on Motion, Interaction and Games; 2021 Nov 10-12; Switzerland. New York: Association for Computing Machinery; 2021. p. 1-6. [DOI]
34. Morotti E, Donatiello L, Marfia G. Fostering fashion retail experiences through virtual reality and voice assistants. In: 2020 IEEE conference on virtual reality and 3D user interfaces abstracts and workshops (VRW); 2020 Mar 22; Atlanta, USA. Piscataway: IEEE; 2020. p. 338-342. [DOI]
 35. Bargum AR, Hansen ES, Erkut C, Serafin S. Exploring the impact of AI-generated speech on avatar perception and realism in virtual reality environments. In: 2025 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW); 2025 Mar 8; Saint Malo, France. Piscataway: IEEE; 2025. p. 647-652. [DOI]
 36. Parmar D, Ólafsson S, Utami D, Murali P, Bickmore T. Navigating the combinatorics of virtual agent design space to maximize persuasion. In: Proceedings of the 19th International Conference on Autonomous Agents and MultiAgent Systems; 2020 May 9-13; Auckland, New Zealand. Richland: International Foundation for Autonomous Agents and Multiagent Systems. 2020. p. 1010-1018. Available from: <https://www.ifaamas.org/Proceedings/aamas2020/pdfs/p1010.pdf>
 37. Obremski D, Akuffo OB, Lücke L, Semineth M, Tomiczek S, Weichert HF, et al. Reducing racial bias by interacting with virtual agents: An intervention in virtual reality. In: Proceedings of the 2023 international conference on autonomous agents and multiagent systems; 2023 May 29; London, United Kingdom. Richland: International Foundation for Autonomous Agents and Multiagent Systems; 2023. p. 747-755. Available from: <https://www.ifaamas.org/Proceedings/aamas2023/pdfs/p747.pdf>
 38. Kimani E, Bickmore T, Trinh H, Pedrelli P. You'll be great: virtual agent-based cognitive restructuring to reduce public speaking anxiety. In: 2019 8th international conference on affective computing and intelligent interaction (ACII); 2019 Sep 3; Cambridge, UK. Piscataway: IEEE; 2019. p. 641-647. [DOI]
 39. Zibrek K, Kokkinara E, McDonnell R. The effect of realistic appearance of virtual characters in immersive environments-does the character's personality play a role? IEEE Trans Visual Comput Graphics. 2018;24(4):1681-1690. [DOI]
 40. Gao Y, Dai Y, Zhang G, Guo H, Mostajeran F, Zheng B, et al. Trust in Virtual Agents: Exploring the Role of Stylization and Voice. IEEE Trans Visual Comput Graphics. 2025;31(5):3623-3633. [DOI]
 41. Gao Y, Dai Y, Zhang G, Guo H, Mostajeran F, Zheng B, Yu T. Trust in virtual agents: Exploring the role of stylization and voice. IEEE Trans Vis Comput Graph. 2025;31(5):3623-3633. [DOI]
 42. Casanova E, Davis K, Gölge E, Gökner G, Gulea I, Hart L, et al. Xtts: a massively multilingual zero-shot text-to-speech model. arXiv: 2406.04904 [Preprint]. 2024. [DOI]
 43. Cohn M, Pushkarna M, Olanubi GO, Moran JM, Padgett D, Mengesha Z, et al. Believing anthropomorphism: examining the role of anthropomorphic cues on trust in large language models. In: Extended Abstracts of the CHI Conference on Human Factors in Computing Systems; 2024 May 11-16; Honolulu, USA. New York: Association for Computing Machinery; 2024. p. 1-15. [DOI]
 44. Székely É, Miniota J. Will AI shape the way we speak? The emerging sociolinguistic influence of synthetic voices. arXiv: 2504.10650 [Preprint]. 2025. [DOI]
 45. Kirk HR, Gabriel I, Summerfield C, Vidgen B, Hale SA. Why human-AI relationships need socioaffective alignment. Humanit Soc Sci Commun. 2025;12(1):728. [DOI]
 46. Dubiel M, Sergeeva A, Leiva LA. Impact of voice fidelity on decision making: A potential dark pattern? In: Proceedings of the 29th International Conference on Intelligent User Interfaces; 2024 Mar 18-21; Greenville, USA. New York: Association for Computing Machinery; 2024. p. 181-194. [DOI]
 47. Marriott HR, Pitardi V. One is the loneliest number... Two can be as bad as one. The influence of AI Friendship Apps on users' well-being and addiction. Psychol Mark. 2023;41(1):86-101. [DOI]
 48. Banks J. Deletion, departure, death: Experiences of AI companion loss. J Soc Pers Relatsh. 2024;41(12):3547-3572. Available from: https://www.researchgate.net/publication/385243512_Deletion_departure_death_Experiences_of_AI_companion_loss
 49. Phang J, Lampe M, Ahmad L, Agarwal S, Fang CM, Liu AR, et al. Investigating affective use and emotional well-being on ChatGPT. arXiv:2504.03888 [Preprint]. 2025. [DOI]
 50. Raedler JB, Swaroop S, Pan W. AI companions are not the solution to loneliness: Design choices and their drawbacks. In: ICLR 2025 Workshop on Human-AI Coevolution; 2025 Apr 27; Singapore. Available from: <https://openreview.net/pdf?id=nFkL7qAirE>

Publisher's Note

Science Exploration remains a neutral stance on jurisdictional claims in published maps and institutional affiliations. The views expressed in this article are solely those of the author(s) and do not reflect the opinions of the Editors or the publisher.