Whispers in the Dark: Effects of Voice Type and Dialog Context in Likability and Tension

Jeongmin Lee[1], Eunsol Kim[1] , Jisoo Han[1], Jinsun Suhr[2], Youngjae Yoo[3*]

[1]Graduate School of Information, Master's Student, Yonsei University, Seoul, Korea

[2]Department of Psychology, Master's Student, Yonsei University, Seoul, Korea

[3]Graduate School of Information, Lecturer, Yonsei University, Seoul, Korea

Abstract

Background There is a lack of research on nighttime interactions involving auditory interfaces in traditional human–computer interaction (HCI). Most commercial speech agents rely on task-oriented features and uniform voice tones, which require a contextualized design approach. This study investigates the use of whispering voice in nighttime contexts and explores the effects of voice characteristics (non-whispering vs. whispering) and dialog contexts (task-oriented vs. social-oriented) on likability and tension levels during nighttime interactions with voice agents (VAs).

Methods A controlled laboratory experiment was conducted with 31 participants who interacted with VAs employing two different voices. The participants were tested in a darkened environment, and their experiences were evaluated in terms of tension and likability after exposure to all four experimental conditions.

Results Voice type had a significant effect on likability, and a significant interaction between voice type and dialog context was observed for both likability and tension.

Conclusions The findings reveal that nighttime voice UX is shaped not by a single optimal voice type, but by the congruence between vocal delivery and dialog intent. Whisper-style delivery enhanced likability and reduced tension when aligned with socially oriented dialog, whereas mismatches, particularly a standard conversational voice in emotionally sensitive interactions, increased tension. These results reframe whispering as a context-sensitive design resource rather than a universal nighttime solution, and offer transferable design principles for adaptive, context-aware voice agents in low-stimulation settings.

Keywords Whispering Voice, Voice Interaction, Context-Aware Design, Context of Dialog

This work was supported by ICONS(Institute of Convergence Science) at Yonsei University.

*Corresponding author: Youngjae Yoo (yooyoungjae@yonsei.ac.kr)

Citation: Lee, J., Kim, E., Han, J., Suhr, J., & Yoo, Y. (2026). Whispers in the Dark: Effects of Voice Type and Dialog Context in Likability and Tension. *Archives of Design Research, 39*(2), 161-183.

http://dx.doi.org/10.15187/adr.2026.05.39.2.161

Received : Nov. 10. 2025 ; **Reviewed :** Mar. 01. 2026 ; **Accepted :** Apr. 18. 2026
pISSN 1226-8046 **eISSN** 2288-2987

1. Introduction

Voice agents (VAs) have evolved rapidly, yet most commercial systems rely on uniform, task-oriented tones that assume a "one-size-fits-all" daytime productivity context (Porcheron et al., 2017, Sciuto et al., 2018, Choi & Drumwright, 2021). While these standardized voices are efficient for daytime productivity, they often fail to account for the nuanced psychological and physiological shifts users experience at night. Nighttime interaction, in particular, constitutes a distinct context characterized by reduced alertness and heightened emotional sensitivity (Thayer, 1986), in which users tend to seek reassurance, intimacy, and psychological safety rather than efficiency or productivity. In such low-arousal states, loud or high-intensity voices may induce unnecessary tension rather than support comfort. While generative AI now enables VAs to adjust their vocal styles dynamically, the field lacks empirical design principles that guide these adaptations for nighttime contexts.

Interactions in dim environments before sleep require distinct design strategies because auditory perception becomes the primary sensory channel once visual stimuli are reduced. In this context, even familiar sounds can trigger disproportionate physiological arousal that disrupts sleep onset (March et al., 2008). Consequently, vocal strategies optimized for daytime productivity may fail to align with these shifted expectations in nocturnal contexts.

Despite the prevalence of late-evening VA use for domestic routines (Porcheron et al., 2017; Sciuto et al., 2018), most HCI research remains predominantly 'daytime-oriented,' focusing on well-lit, high-productivity environments. This gap has left the distinct cognitive and emotional conditions of nighttime—such as the need for low-stimulation and emotional reassurance—relatively unexplored. Beyond this contextual gap, little attention has been paid to how vocal characteristics should adapt to nighttime conditions. In particular, there remains an empirical gap regarding how vocal strategies should adapt to contexts characterized by low stimulation, heightened emotional sensitivity, and expectations of calm—such as nighttime interactions.

Prior research emphasizes that users trust and prefer voice interfaces when vocal characteristics align with conversational content (Nass & Lee, 2001; Cambre & Kulkarni, 2019). This principle of voice–content congruence—whereby how something is said shapes users' interpretations, expectations, and affective responses—has rarely been examined in low-stimulation contexts. Critically, it remains unclear whether the appropriateness of a given vocal style varies depending on conversational goal (e.g., task-oriented versus socially oriented dialog), and how this interaction may differ at night. This gap limits our understanding of how voice agents should adapt their vocal strategies to nighttime use, where users may prioritize calmness and emotional attunement over efficiency.

In light of these observations, whispering emerges as a uniquely suitable focal strategy for nighttime interaction. Rather than functioning merely as low-volume speech, whispering operates as a socially meaningful vocal register—historically associated with privacy and emotional attunement in human communication (Wilkinson, 1982). By leveraging these

cultural and psychological affordances, whispering presents a theoretically grounded approach to minimize tension and foster a calming user experience after dark. However, the appropriateness of whispering likely depends on the nature of the conversation. A whispered instruction for a task-oriented interaction (e.g., setting an alarm) may feel incongruent or even ineffective, whereas a whispered social exchange may enhance intimacy and reassurance. To address this, the current study examines how both vocal style and conversational context jointly influence user experience.

This study investigates how whispering functions as a vocal strategy in after-dark interactions and how its effectiveness depends on the alignment between vocal style and conversational context. Based on this approach, the following research questions are proposed.

RQ1: How do users perceive different voice characteristics of VAs in nighttime interaction contexts?

RQ2: How do the effects of voice characteristics vary depending on the type of conversation with VAs?

To address these research questions, we conducted a controlled 2×2 within-subject factorial experiment in a simulated low-light environment that emulates nighttime conditions. We systematically manipulated voice type (whispering vs. non-whispering) and dialog context (task-oriented vs. social-oriented) to examine their effects on user experience, with particular attention to affective responses such as likability and tension.

Our study makes three primary contributions: (1) it provides empirical evidence on how whispering interaction functions in after-dark contexts; (2) it identifies the moderating role of dialog context in these interactions; and (3) it offers practical design implications for context-adaptive voice agents in low-stimulation settings.

2. Literature Review

2. 1. User Interface in the Nighttime Context

While a substantial body of HCI literature has examined interface design in nighttime or presleep contexts, most of this work has concentrated on visual modalities. Common strategies—such as screen dimming, blue-light filtering, and simplified layouts—are typically intended to lower physiological arousal and facilitate relaxation (Oshima et al., 2020; Nagare et al., 2019). For example, Chang et al. (2015) demonstrated that evening exposure to blue light from electronic devices can suppress melatonin production, delay circadian rhythms, and reduce alertness the following morning. Such findings underscore the physiological and cognitive rationale for minimizing sensory stimulation in night-oriented UX design. However, in practice, most voice agent interactions still show limited adaptation to the sensory constraints and expectations of low light or nighttime environments. As visual cues

diminish in low light environments, the auditory system assumes a more prominent role in spatial and situational awareness (Zheng et al., 2024), potentially increasing users' sensitivity to sound intensity and quality during voice based interaction.

Although these strategies were developed for visual interfaces, their underlying goals are potentially transferable to auditory modalities, particularly in dark or low-light environments. Auditory design becomes particularly relevant in nighttime UX due to a combination of environmental and physiological factors. From an environmental perspective, domestic nighttime settings impose unique social constraints, such as the need to minimize acoustic leakage to adjacent spaces (e.g., inter-floor noise) or to avoid disturbing cohabitants. From a psychological perspective, users preparing for sleep typically exhibit reduced tolerance for sensory stimulation and heightened sensitivity to auditory cues, making abrupt or salient sounds more likely to disrupt relaxation and sleep readiness.

Despite these characteristics, existing nighttime UX research has paid limited attention to how such constraints translate to auditory or voice-based interaction. While prior work has begun to explore nighttime UX more broadly, far fewer studies have examined how principles established in visual nighttime interface design could inform the design of auditory interfaces or shape the behavior of conversational agents in nocturnal settings (Nißen et al., 2022).

Moreover, current research on nighttime auditory interfaces remains narrow in scope, focusing largely on biometric tracking or clinical sleep interventions (Karlgren et al., 2022; Aarts et al., 2022; Semertzidis et al., 2023; Karlgren & McMillan, 2022). As a result, a significant knowledge gap persists regarding how conversational agents should adapt their social and acoustic behaviors to harmonize with everyday domestic nighttime environments. Addressing this gap, the present study extends context-sensitive design principles established in visual nighttime UI research to the auditory domain. Focusing on domestic after-dark interactions, we examine whispering as a context-aware vocal strategy that responds to both the environmental constraints (e.g., low stimulation, noise sensitivity) and the psychological expectations (e.g., calmness, reassurance) characteristic of nighttime interaction.

2. 2. Voice Characteristics of the VAs

Individuals can anthropomorphize VAs using the "voice" element, thereby attributing human-like characteristics and behaviors to them (Appel et al., 2012; Wagner & Schramm-Klein, 2019). This serves as a core feature that distinguishes computers (Nass et al., 1994) and allows all devices sharing a common voice to be considered to possess shared intelligence (Nass & Brave, 2005). Due to these characteristics, significant research has been conducted in the field of VAs with a focus on voice. Despite the central role of voice in user experience, most commercial VAs continue to utilize standardized voices (Cambre & Kulkarni, 2019), and academic inquiry has focused predominantly on factors such as gender (McGinn & Torre, 2019) and pitch (Tolmeijer et al., 2021; Goodman & Mayhorn, 2023).

More recent work has begun to acknowledge the role of usage context(Cambre & Kulkarni, 2019), but this line of inquiry has largely concentrated on spatial settings such as automobiles or public environments (Park & Lee, 2022; Jestin et al., 2022). Temporal context—

particularly nighttime use—remains comparatively underexplored, despite its distinct social and physiological constraints.

Auditory interfaces must be calibrated for the distinct socio-environmental constraints of nighttime. In interpersonal communication, individuals naturally modify their speech patterns after dark, adopting softer tones and slower pacing to maintain quietness and avoid disturbing cohabitants. While this behavior is partly a response to environmental factors like noise sensitivity, it also reflects a shift toward a lower-arousal communicative state. Instead of merely reducing volume, humans often utilize whispering—a specialized vocal register—to facilitate private and intimate sharing in the dark.

Nighttime interaction therefore calls for a vocal strategy that goes beyond incremental stylistic adjustments within voiced speech.
While parameters such as pitch, prosody, or speech rate modulate expressiveness along a continuous dimension, whispering constitutes a distinct phonatory mode characterized by the absence of vocal fold vibration (Wilkinson, 1982). Because this shift represents a categorical change in vocal production rather than a gradual prosodic adjustment, whispering carries acoustic and social meanings that are difficult to achieve through volume or pitch modulation alone. For this reason, the present study treats whispering as a theoretically distinct vocal strategy, rather than as an extreme point on a continuum of vocal intensity. Importantly, this nighttime shift in speech is not limited to quantitative adjustments such as reduced volume or slower speech rate, but often involves the adoption of whispering as a socially learned communicative convention.

Prior research suggests that whispering is commonly used to reduce speech audibility (Li, 2011) and to signal intimacy by decreasing perceived interpersonal distance (Andersen, 2015). In mediated contexts, studies on ASMR and related auditory content indicate that whispering can effectively reduce physiological arousal, such as anxiety and heart rate (Barratt & Davis, 2015; Seifzadeh et al., 2023; Hardian et al., 2020). However, whispering may also evoke negative impressions, such as being perceived as "creepy," when deployed without contextual appropriateness (Parviainen & Søndergaard, 2020). This underscores the critical importance of alignment between vocal register and interaction context.

It is worth noting that commercial voice assistants have already begun to incorporate whispering as a design feature. For example, Amazon Alexa's 'whisper mode' enables the system to respond softly when it detects a whispered input (Erika, R., 2023). However, such implementations remain largely heuristic and reactive, relying on simple acoustic triggers rather than a principled understanding of when and why whispering is most appropriate. To date, little empirical research has examined how whispered interaction modulates conversational context, user affect, or perceived social appropriateness, particularly during the sensitive nighttime period. By systematically testing these variables, the present study seeks to provide the empirical evidence necessary to move beyond reactive features toward context-aware, human-centered vocal design.

At the same time, whispering is not universally appropriate. When misaligned with conversational intent or situational expectations, it may evoke discomfort or be perceived as eerie or inappropriate (Parviainen & Søndergaard, 2020; Li, 2011). This dual nature underscores the importance of contextual alignment in vocal design. Accordingly, this study examines whispering as a theoretically grounded yet context-sensitive vocal strategy, comparing a whispering voice agent with a conventional voice agent to investigate how this phonation mode shapes user experience in nighttime interactions.

H1: Interactions with a whispering VA in a nighttime context will lead to lower tension than interactions with a normal VA in the same context.

The voice of a VA is closely tied to its likability, which refers to how users perceive the system as useful, enjoyable, and friendly (Hone & Graham, 2000). Likability has long been recognized as a key factor in motivating technology adoption (Davis et al., 1992). Prior research also shows that human-like traits in AI VAs can significantly enhance user perceptions and interactions (Kühne et al., 2020). Building on these findings, we expect that whispering voices that align more closely with nighttime contexts will further enhance user likability.
H2: Interactions with a whispering VA in a nighttime context will lead to higher likability than interactions with a normal VA in the same context.

2. 3. Dialog Context in VAs

Building on the established importance of nighttime-appropriate voice design, interactions with VAs in these contexts are shaped not only by voice characteristics but also by the purpose and content of conversation. Cassell (1999) emphasized that VAs should integrate both task- oriented and social aspects, and subsequent work has stressed the importance of strategies that go beyond task fulfillment to enrich user experience (Sciuto et al., 2018; Gaver et al., 2008; Barber et al., 2014).

Despite research on voice characteristics, the role of conversational content in voice-based interactions remains underexplored. Conversational behavior is often classified as social- or task-oriented, depending on purpose and relationship (Chattaraman et al., 2019; Keeling et al., 2010). In task-oriented conversations, dialog is goal-driven and formal; however, in social-oriented conversations, it is characterized by relational interactions, including exclamatory feedback and small talk (Chattaraman et al., 2019; Bickmore & Cassell, 2001).

Other work has highlighted how voice interacts with dialog type. Female voices, for instance, are often rated more positively in social interactions (Lee & Park, 2019), and preferences in counseling contexts vary depending on user demographics (Kim et al., 2021). Moreover, while social small talk with VAs has been effective in domains such as customer service, it has proven less effective in serious domains like military training (Bickmore & Cassell, 2005).

These findings collectively suggest that both dialog type and context shape how users evaluate VA voices. However, few studies have explored how these factors interact in nighttime settings. To address this gap, we examine whether whispering voices, which are

often linked to intimacy and relaxation, enhance social nighttime interactions compared to task- oriented.

H3: There will be an interaction effect between voice type and dialog context on the user's level of tension toward VAs.

H4: There will be an interaction effect between voice type and dialog context on user's likability evaluation of VAs.

3. Method

3. 1. Study Design

3. 1. 1. Experimental design

We conducted the experiment in a controlled laboratory setting to simulate after dark interaction conditions while minimizing external distractions. While ecological methods such as the Experience Sampling Method (ESM) or Ecological Momentary Assessment (EMA) are widely used to capture in-situ experiences and offer high ecological validity, such approaches may not be methodologically appropriate for pre-sleep or nighttime contexts. Repeated prompts or real-time self-reporting can introduce additional cognitive or physiological arousal, which may interfere with the low-arousal state that characterizes nighttime interactions prior to sleep. Similarly, although continuous physiological measurements (e.g., electroencephalography (EEG) or galvanic skin response (GSR)) could provide real-time indicators of arousal, the use of unfamiliar sensing equipment may disrupt natural behavior and introduce reactivity effects, such as the Hawthorne effect (Franke, 1978).

Therefore, we adopted a controlled laboratory setting to prioritize internal validity and ensure standardized delivery of auditory stimuli while minimizing sources of unintended arousal. This approach allowed us to systematically examine the effects of voice type and dialog context by reducing unpredictable real-world factors—such as ambient noise, domestic interruptions, and device variability—that could otherwise confound the assessment of perceived tension. This methodological choice aligns with prior laboratory-based approaches in pre-sleep research, which emphasize the importance of maintaining experimental control when assessing subjective pre-sleep experiences (Semertzidis et al., 2019). Specifically, following methodological practices used in prior studies on pre-sleep arousal (Tang & Harvey, 2004), we utilized a controlled environment and post-exposure self-report measures to capture participants' perceived experiences immediately following the interactions.

A 2 × 2 within-subjects factorial design was employed, with two independent variables: voice type (whispering vs. non-whispering) and dialog context (task-oriented vs. social-oriented). To control for order and learning effects, we employed a Latin Square design to determine the presentation order of the four experimental conditions. Participants were randomly assigned to one of four counterbalanced sequences, ensuring that each condition appeared exactly once in each ordinal position across participants. This approach minimized potential bias

from fatigue, familiarity, and first-order carryover effects between conditions. After each condition, participants completed questionnaires assessing likability and tension to evaluate the impact of voice type and dialog style on user experience in dark environments.

3. 1. 2. Experimental environment

The experiment was conducted in a quiet, noise-controlled laboratory space on campus that was deliberately configured to approximate an after-dark, pre-sleep interaction context. The room was dimmed to a low-light level to reflect nighttime conditions, and external light sources were blocked using blinds and opaque materials. Partitions were installed behind and beside the participant to reduce visual distractions and create a stable, enclosed listening environment. Although the space was not a fully furnished bedroom, the setup was designed to reproduce key perceptual characteristics of nighttime in-home voice interaction. The laboratory environment remained quiet throughout all sessions, minimizing potential interference from ambient noise.

To support the intended nighttime framing, the physical environment was complemented by contextual instructions that encouraged participants to mentally situate themselves in a pre-sleep interaction context. Prior to each trial, participants were provided with a brief overview of the upcoming scenario and asked to imagine interacting with the voice agent at home before going to sleep (e.g., lying in bed or preparing to sleep in a quiet room). This cognitive framing was intended to compensate for the limitations of a laboratory setting by approximating psychological characteristics of nighttime interaction in conjunction with the low-light physical environment.

A Kakao Mini speaker was used to deliver whispering and non-whispering voice stimuli at consistent volume levels. Participants were provided with a brief adaptation period and a short practice interaction before the main trials to acclimate to the environment and reduce potential novelty- or laboratory-induced tension.

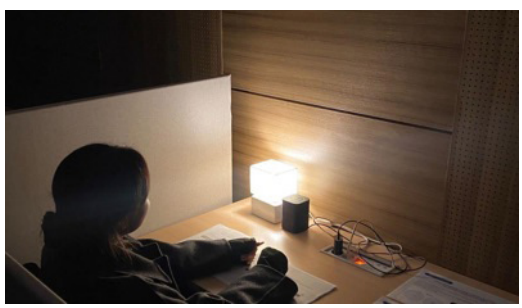

Figure 1 Experimental Environments

3. 2. Participants

We recruited participants with experience using VAs through an online advertisement. A total of 31 participants were recruited via online advertisements, with a mean age of 25.5 years (*SD*= 3.82, Male (n = 12), Female (n = 19)). All participants were native Korean speakers without hearing problems. Additionally, all participants had experience with VAs. Each

participant received $10 as compensation. There were instances where a participant provided double responses to a single survey question, for which the average of the two responses was used in the data analysis. The study was approved by the University's IRB (IRB No. 7001988-202404-HR-2074-04).

3. 3. Stimuli

To understand the user experiences, we created the treatments (voice and scenarios) required for this experiment.

3. 3. 1. Voice design

The voice was classified into two categories: whispering and non-whispering. Both voices were generated using Typecast's text-to-speech technology, an AI-based voice synthesis service. To ensure consistency, both voices were selected as female voices with similar tonal characteristics. The whispering voice was created using Ari at 0.8x playback speed, type B. Ari is classified as a whispering voice in Typecast's voice library and was specifically designed to exhibit natural whispering characteristics for Korean speech synthesis. We adjusted the tone and speed of Ari's voice to closely resemble Ji-yoon's non-whispering voice. To reflect the unique characteristics of whispering, Ari's voice was controlled to play at a lower volume (38 dB) than that of the non-whispering voice (55 dB). The non-whispering voice was created using Jiyoon at 0.8x playback speed and type A. Both voices are female, and we chose voices that are like each other.

In this study, the whispering condition was operationalized as an 'intimacy-forward, low-intensity voice delivery.' This approach intentionally bundled whispering phonation with reduced loudness to reflect the ecological reality of nighttime interactions, where lower intensity supports privacy and comfort. Volume levels were pre-set to fixed output settings for each condition prior to data collection. Sound pressure levels were measured at a distance of approximately 0.8 meters from the speaker using a digital sound level meter. This procedure ensured stable and replicable output levels within each condition across all sessions.

3. 3. 2. Scenario design

Four dialog scenarios were developed to examine user experiences across different conversational contexts: two task-oriented (weather and news recommendations) and two social-oriented (concerns about the future and stress from work). Scenario topics were selected to reflect after-dark, in-home voice-agent use and routines. Prior research on in-home conversational agents reports that the bedroom is a frequent device location and that users describe routines that include "getting ready for bed," alongside clear time-of-day patterns with usage peaking again in the late evening. In addition, information-seeking requests such as weather checks are among the most popular voice commands in everyday usage (Sciuto et al., 2018). Based on this evidence, we designed task-oriented scenarios to represent routine information-seeking and planning interactions in after-dark home settings. For social-oriented scenarios, we focused on topics such as future anxiety and work stress. This choice was motivated by psychological findings that repetitive negative thinking (e.g., rumination and worry) is closely associated with sleep-related difficulties(Takano et

al., 2012, Takano et al., 2014). Accordingly, these topics were selected to reflect emotionally supportive conversations that can occur near bedtime.

The initial drafts of the experimental scenarios were generated using ChatGPT-4o and were subsequently revised and restructured by the authors to align with the study's objectives. To ensure methodological consistency, all scripts were standardized using a fixed four-turn exchange format (User–Agent–User–Agent) with agent responses strictly limited to 2–3 sentences to maintain consistent information density and interaction duration. Average interaction times were as follows: weather recommendation (M = 03:18, SD = 00:13), news recommendation (M = 03:10, SD = 00:20), worries about the uncertain future (M = 03:12, SD = 00:22), and stress about work (M = 03:32, SD = 00:20). Participants experienced all four conditions in a counterbalanced order, as determined by the Latin Square design described above.

To support internal validity and comparability across conditions, we employed scripted scenarios rather than open-ended dialog. This design choice enabled the standardization of interaction duration, information density, and conversational structure across participants, thereby reducing variability associated with personal conversational content. To prevent specific topics (e.g., weather vs. news) from confounding the results, the four scenarios were decoupled from the voice types and counterbalanced across participants via the Latin Square design described above. For instance, the 'weather' scenario was presented in a whispering voice for some participants and in a non-whispering voice for others. This cross-assignment ensured that the observed differences in user experience were attributable to the voice type and dialog context, rather than the inherent nature or emotional load of the specific topics.

Table 1 presents one task–oriented and one social–oriented dialog as representative examples of the four scenarios used in this study.

Task–oriented (1): weather recommendation	User: Nua, hello! Agent: Hello, how can I assist you today? User: Tell me tomorrow's weather. Agent: Sure, let me tell you about tomorrow's weather. It is expected to be colder than today. The high temperature will be 3˚C, and the low temperature will be 5˚C. Overall, the temperature will be low, and with a slight wind, the perceived temperature will feel even colder. Additionally, strong winds are expected along the coast, and the winds will be strong with high waves in most sea areas. Would you like me to recommend what to wear tomorrow? If you'd like clothing recommendations, just say, "Recommend clothes for tomorrow." User: Yes, recommend clothes for tomorrow. Agent: Sure, let me recommend clothing suitable for tomorrow's weather. Since the perceived temperature will be low, I recommend wearing warm, highly insulated clothes. For the top, fleece or knit materials would be ideal. For the bottom, choose materials that can block out the wind. Would you like a more specific clothing recommendation? If you'd like a more specific outfit suggestion, just say, "Recommend specific clothes." User: Yes, please recommend specific clothes. Agent: Sure. Since the wind will be strong, I recommend wearing a padded jacket or a thick wool coat tomorrow. A longer design will provide better insulation. Additionally, a scarf and beanie can protect your neck and head, helping to maintain body heat. Since the perceived temperature will be low, I suggest considering items like scarves and beanies to help keep your body warm. Lastly, I recommend boots and gloves to protect your feet and hands from the cold. These will help increase warmth and keep you comfortable. By choosing the suggested outfit, you will stay warm and protected from the cold.

Social oriented (1): worry about the uncertain future	User: Nua, hello! Agent: Hello, how can I assist you today? User: I've been feeling really worried about how uncertain the future is these days. Agent: I see. I understand how you feel. Uncertainty about the future is something everyone experiences, so try not to worry too much. While I can't directly solve your problems, I'm a good listener. Would you like to talk more about what is troubling you? If you'd like to share more, just say, "I'm feeling anxious about my future choices." User: I'm feeling anxious about my future choices right now. Agent: That must be tough, but just remember, not every choice has to be perfect. The important thing is that you're doing your best right now. It's okay to adjust your plans along the way, so don't be too hard on yourself or worry too much! It is natural to have concerns about the future, but if you look at it positively and keep a mindset open to challenges, you will grow from it. I can share a story about an actor's journey if you are interested. Just say, "Tell me the story," if you'd like to hear more. User: Really? Then, tell me the story. Agent: Sure, let me share a story that might help ease your worries. There was an actor who struggled with anxiety during his unknown years. To overcome his anxiety, he wrote in a journal every morning, noting the difficulties he had overcome each day. By doing this, he gained courage from his past achievements. He never gave up on his dream of becoming an actor and eventually won the best actor award at a prestigious film festival. Just like him, think about the challenges you have already overcome. You've made it through so many obstacles, and you can do it again. I'll be here cheering you on whenever you need me!

3. 4. Procedure

The experiment was conducted in a dimly lit on-campus laboratory designed to simulate nighttime conditions. Upon arrival, participants were briefed on the study, provided informed consent, and given a five-minute rest period to adapt to the environment. They then completed a baseline activation questionnaire before the experiment began. An introductory voiceover presented the agent's voice, followed by a short practice dialog to ensure familiarity with the system. Before each scenario, participants were instructed to imagine that they were interacting with the voice agent at home after dark, in a quiet, low-light environment as part of their bedtime routine.

Participants engaged in four scenario-based conversations with the voice agent "Nua" each followed by a paper-based questionnaire assessing their experience. A two-minute break was given between sessions. The entire procedure lasted approximately 50 minutes(Figure 2).

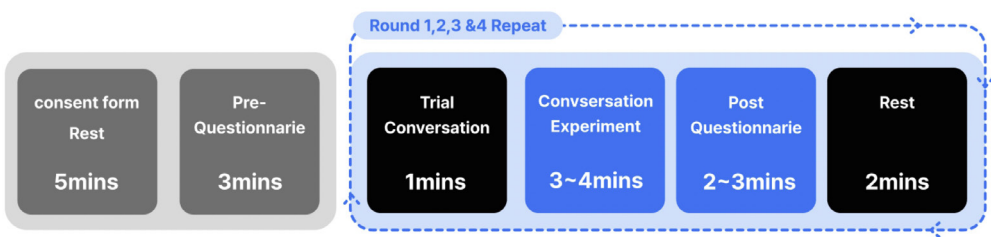

Figure 2 Experimental Procedure

3. 5. Measurement

To evaluate user experiences, we collected quantitative data on likability and tension using the Activation-Deactivation Adjective Checklist (AD-ACL) scale (Thayer, 1986).

The tension subscale of AD-ACL was used to evaluate participants' activation or arousal levels, capturing their momentary mood and arousal states. AD-ACL is a self-reported scale

that provides a subjective assessment of arousal, serving as an alternative to physiological measurements by reflecting the participant's immediate perception of their overall activation state. In neuropsychological contexts, arousal is linked to awareness of emotional experiences (Bhandari et al., 2017). In this study, we measured the element of stability using tension.

The AD-ACL scale comprises four primary categories: energy (general activation), tension(high activation), tiredness (general deactivation), and calmness (deactivation-sleep). Among these, tension was selected to examine psychological comfort in the nighttime context, particularly after establishing the dark setting of the experiment. Tension represents tense arousal, which ranges from tense to calm, and is a key indicator of external stress responses (Thayer, 1986; Bhandari et al., 2017). This subscale included five descriptive terms that participants could rate based on their immediate mood, measured on a 4-point Likert scale (1-Strongly Disagree, 4-Strongly Agree).

Likability refers to a positive emotional response toward an object. In this study, likability is defined as an interpersonal liking that motivates participant to continue engaging with VA(Kang et al., 2021). The subjective assessment of speech system interfaces scale was based on the translated version used by Kang et al. (2021) to assess the degree to which the agent exerts influence on the participant (Hone & Graham 2000). All measurements were conducted using a 5-point Likert scale.

3. 6. Statistical Analysis Method

First, paired t-tests were performed to analyze differences in tension scores between the baseline and each condition. A two-way repeated measures ANOVA was then conducted to examine the main and interaction effects of voice type and dialog context on tension. Finally, likability was also analyzed using a two-way repeated measures ANOVA.

4. Results

4. 1. Descriptive Statistics and Measurement Validation

The reliability of the measurement tools was verified since the data collected in the experiment were based on the participants' subjective measurements. To assess the reliability of the survey items for the two dependent variables, likability and tension, Cronbach's α coefficients were calculated. The results showed that likability (α = .83) and tension (α = .72) demonstrated Cronbach's α values above 0.7, thus confirming the reliability of the dependent variable(Fornell & Larcker, 1981).The means and standard deviations (*SDs*) of the dependent variables for each condition are presented in Table 2.

Table 2 Mean and SD scores (type of voice x context of dialog, N = 31).

Dependent Variables	Baseline	Social-oriented		Task-oriented	
		Non-whispering	Whispering	Non-whispering	Whispering
Likability	-	3.419 (1.007)	3.991 (0.708)	3.596 (0.848)	3.572 (0.834)
Tension	8.580 (2.754)	8.483 (3.150)	7.225 (1.927)	7.548 (2.461)	7.354 (2.274)

4. 2. Hypothesis Testing

Baseline–Compare Tension for each condition

To investigate the effects of whispering and non-whispering voice conditions on tension, we conducted paired t-tests using the baseline as the reference for each condition. In the whispering voice conditions, we observed a significant difference in tension between the baseline and task-oriented whispering (TW) condition ($t(30) = 2.295$, $p < .05$, $SD = 0.412$). The tension score for the TW condition ($M = 7.35$, $SD = 2.274$) was significantly lower than that of the baseline ($M = 8.58$, $SD = 2.754$). We also observed a significant difference in tension between the baseline and social-oriented whispering (SW) conditions ($t(30) = 2.730$, $p < .05$, $SD = 0.366$). The tension score in the SW condition ($M = 7.23$, $SD = 1.927$) was significantly lower than that in the baseline.

In the non-whispering voice conditions, we observed a significant difference in tension between the baseline and task-oriented non-whispering (TNW) conditions ($t(30) = 2.062$, $p < .05$, $SD = 0.370$). The tension score in the TNW condition ($M = 7.55$, $SD = 2.461$) was significantly lower than that in the baseline. However, no significant difference was observed between baseline and social-oriented non-whispering (SNW) conditions ($t(30) = 0.172$, $p = .864$, $M = 8.48$, $SD = 3.150$). Therefore, the whispering condition showed lower levels of tension than the other conditions, with the SW condition exhibiting the lowest scores overall. Hypothesis 1 predicted that interactions with a whispering voice would result in lower tension compared to interactions with a non-whispering voice. Paired t-tests supported this prediction: both task-oriented and social-oriented whispering conditions showed significantly lower tension than the baseline.

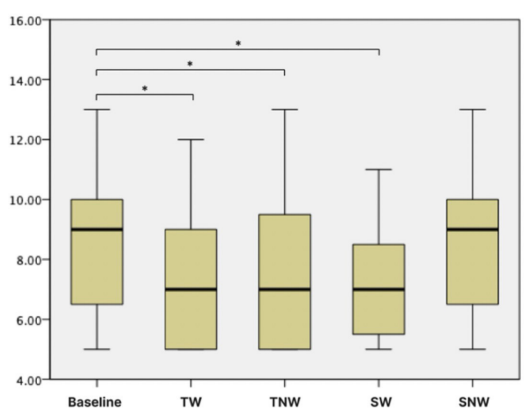

Figure 3 Comparison of tension levels between baseline and experimental conditions
** p 〈 .05, ** p 〈 .01, TW (task + whispering), SW (social + whispering), TNW (task + non-whispering), SNW(social + non-whispering)*

Within-Subjects Analysis of the Tension

A two-way repeated measures ANOVA revealed that tension did not have a significant effect of either voice type or dialog context (Table 3). However, the interaction effect of voice type and dialog context ($F(1, 30) = 4.430$ $p < .05$, $\eta_p^2 = .129$) was significant (Figure 4). In other words, non-whispering voice and social-oriented dialogs had the highest tension levels, and whispering voice and social-oriented dialogs had the lowest tension levels. Thus, Hypothesis 3, which stated that there would be an interaction effect between the voice type and dialog context on the user's level of arousal toward VA, was supported. However, the repeated-measures ANOVA revealed no significant main effect of voice type on tension. Instead, a significant interaction between voice type and dialog context was found, indicating that the effect of whispering voices on tension was context-dependent rather than uniform across all dialog types. Thus, H1 was partially supported.

Table 3 Results of Repeated Measures Two-Way ANOVA

Dependent Variables		Mean square	*F*	*p*	Partial Eta Squared
Tension	Voice Type	16.331	2.935	0.097	0.089
	Dialog Context	5.040	2.123	0.155	0.066
	Voice Type * Dialog Context	8.782	4.430	0.044*	0.129

*$p < .05$

Within-Subjects Analysis of Likability

Repeated measures two-way ANOVA results showed that the main effect of voice type ($F(1,30) = 4.358$, $p < .05$, $\eta_p^2 = .127$) was significant; however, no significant effect was found for the dialog context (Table 4). According to Bonferroni's post-hoc analysis results, the whispering voice had a higher mean ($M = 3.782$, $SD = .119$) than the non-whispering voice ($M = 3.508$ $SD = 0.151$), suggesting that the whispering voice was more likely to be favorable in a nighttime context after dark. The interaction effect of voice and dialog type ($F(1, 30) = 6.731$, $p < .05$, $\eta_p^2 = .183$) was significant (Figure 5). Whispering voice and social-oriented dialog had the highest likability ratings. Thus, Hypotheses H2 and H4 were supported.

Table 4 Results of Repeated Measures Two-Way ANOVA

Dependent Variables		Mean square	*F*	*p*	Partial Eta Squared
Likability	Voice Type	2.331	4.358	0.045*	0.127
	Dialog Context	0.454	1.914	0.177	0.060
	Voice Type * Dialog Context	2.760	6.731	0.015*	0.183

*$p < .05$

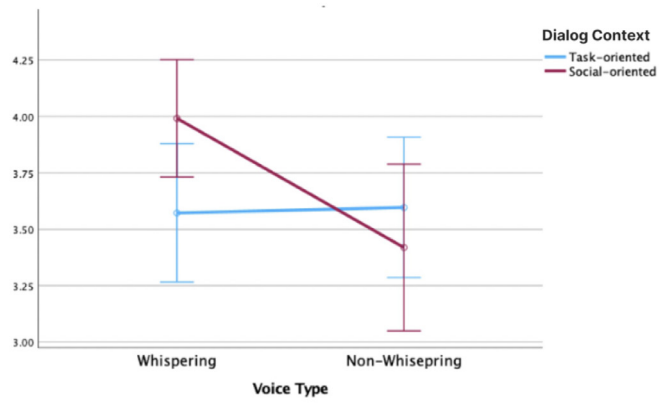

Figure 5 Interaction effect between voice type and dialog context under likability

4. 3. Summary of Hypotheses Testing Results

To provide a clear overview of the experimental findings, summarizes the results of the hypothesis testing. Below, a detailed interpretation of each hypothesis in relation to the observed data is provided.

• H1: Partially supported. Whispering voices reduced tension in t-tests relative to baseline, but no main effect emerged in ANOVA.
• H2: Supported. Whispering voices were generally rated as more likable than non-whispering voices.
• H3: Supported. A significant interaction revealed that the calming effect of whispering was strongest in social-oriented dialogs.
• H4: Supported. Likability was highest when whispering voices were paired with social-oriented dialogs.

H1, which predicted that whispering voices would reduce user tension compared to non-whispering voices in nighttime interactions, was partially supported. While paired t-tests showed that both task-oriented and social-oriented whispering conditions significantly reduced tension relative to the baseline, the repeated-measures ANOVA revealed no significant main effect of voice type on tension. Instead, the effect of whispering on tension was contingent on dialog context. H2 was supported, as whispering voices were generally perceived as more likable than non-whispering voices, evidenced by a significant main effect of voice type on likability. H3, which proposed an interaction effect between voice type and dialog context, was supported for both tension and likability. Specifically, whispering voices were most effective in reducing tension and enhancing likability in social-oriented dialogs, highlighting the importance of contextual congruence in nighttime voice interactions. Finally, H4 was supported, showing that likability was highest when whispering voices were paired with social-oriented dialogs.

5. Discussion

Voice agents are increasingly embedded in domestic routines, yet most design guidelines assume daytime or productivity-oriented use. Our findings challenge this assumption by positioning nighttime interaction as a distinct design context characterized by reduced arousal and heightened emotional sensitivity. Within this setting, vocal strategies do not operate uniformly. Rather than functioning as a universally calming feature, whisper-style delivery becomes effective when it aligns with dialog intent. This reframes whispering from a standalone acoustic enhancement to a context-sensitive design resource shaped by vocal–contextual congruence.

5. 1. Central Contribution: Vocal–Contextual Congruence in Nighttime UX

The central contribution of this study is the articulation of vocal–contextual congruence as a primary driver of nighttime voice interaction UX. Rather than positioning whispering as a universally calming strategy, our findings demonstrate that its effectiveness depends on the alignment between vocal delivery (the "how") and dialog intent (the "what") within low-arousal nighttime contexts. Across both affective and evaluative measures—tension and likability—the interaction between voice type and dialog context consistently shaped user responses. In line with this pattern, Hypotheses 2, 3, and 4 were fully supported, demonstrating that whispering voices—specifically in social-oriented dialog contexts—significantly enhanced likability and reduced tension. User experience in nighttime interactions was shaped by the degree of congruence between vocal delivery and dialog intent: when these were aligned (SW condition), both likability and tension improved; when mismatched (SNW condition), tension increased and likability declined.

Importantly, Hypothesis 1 was only partially supported, as whispering did not produce a uniform main effect across all dialog types. Although whispering generally reduced tension relative to baseline conditions, the absence of a broad main effect in the ANOVA reinforces a nuanced insight: nighttime voice UX is driven less by a single "optimal" voice type and more by achieving harmony between vocal tone and conversational purpose. For instance, the significant interaction effect reveals that psychological relief is optimized only through contextual alignment. This suggests that tension arises from vocal-contextual incongruity rather than from the whispering voice per se, effectively countering concerns that novel whispering tones might provoke "uncanny valley" discomfort (Moore, 2017; Carr et al., 2017). Together, these results extend prior work on the calming and intimate qualities of whispered speech (Barratt & Davis, 2015; Poerio et al., 2018; Zarazaga & Malisz, 2023) by demonstrating that likability and tension reduction are fundamentally optimized through situational appropriateness.

Ultimately, this study demonstrates that voice is not merely a medium of information delivery but a dynamic design resource that must be calibrated to the user's temporal and social context. Our findings underscore the necessity of moving beyond "one-size-fits-all" vocal models toward multidimensional context-awareness that considers when, where, and why an interaction occurs (Sciuto et al., 2018; Luger & Sellen, 2016). The core takeaway is that

nighttime voice UX is not driven by a single optimal voice type, but by the harmony between vocal intensity and conversational dimensions. By prioritizing situational appropriateness over uniform vocal strategies, designers can transition toward adaptive, context-sensitive voice assistants that better align with the user's emotional and temporal state (Purington et al., 2017).

5. 2. Design Principles for After–Dark Voice Interaction

Based on the empirical evidence of this study, we propose two core design principles for developing voice agents in after-dark routines. These principles move beyond our specific implementation and articulate reusable design knowledge at a general level.

- Principle 1: Prioritize Tone–Intent Congruence. Intimacy-forward, low-intensity vocal delivery should be selectively aligned with supportive or socially oriented dialog in low-arousal contexts. The benefits of whisper-style delivery emerge most clearly when it resonates with users' emotional expectations and conversational goals.
- Principle 2: Treat Vocal–Intent Mismatch as a Potential Source of Tension. In nighttime settings, the absence of vocal adaptation is not a neutral design choice. A high-intensity, standard conversational tone used in emotionally sensitive interactions may function as a mismatch that increases tension or diminishes likability. Designers should therefore evaluate vocal strategies based on contextual appropriateness rather than acoustic clarity alone.

These principles are expected to generalize across low-arousal, domestic nighttime contexts where users seek emotional attunement over task efficiency. However, they may not directly apply to daytime productivity interactions, high-urgency alert systems, or contexts where acoustic clarity takes precedence over intimacy. The boundary conditions of these principles should be tested in future work across diverse temporal and functional interaction contexts.

5. 3. Theoretical and Practical Implications

Academic contributions. Prior research on voice assistants has largely focused on daytime or productivity-related contexts, leaving nighttime interactions underexplored. By examining how whispering voices shape user experiences in nighttime settings, this study fills an important

gap in the literature on context-aware voice interaction. In doing so, it extends earlier conceptual work on nighttime UX (e.g., Parviainen & Søndergaard, 2020) with empirical findings that go beyond theoretical propositions. Critically, whereas prior voice interaction research has treated vocal characteristics as context-independent design variables, this study introduces vocal–contextual congruence as an empirically grounded construct—demonstrating that the perceived intimacy, comfort, and social engagement afforded by whispering are contingent on dialog intent rather than inherent to the voice itself. This reframes how researchers and designers should theorize the role of voice in emotionally sensitive, low-arousal interaction contexts.

Practical implications. Our findings suggest that whispering may be more effective as a situational choice rather than a default nighttime voice, with its benefits emerging most clearly when the vocal tone aligns with socially oriented interaction goals. In socially

oriented dialogs, whispering voices enhanced comfort, reduced tension, and increased likability, indicating that vocal strategies should be selected in relation to both dialog type and temporal context rather than applied uniformly. By aligning voice tone with the appropriateness of context, designers can foster more engaging, supportive, and emotionally resonant interactions, especially in intimate settings such as the home. From a design perspective, whispering Vas can be applied in everyday contexts such as supporting bedtime routines, guiding stress management practices, or serving as sleep companions that provide reassurance during nighttime. In healthcare or elder-care settings, whispering voices could deliver unobtrusive nighttime support, enhancing comfort without disturbing rest.

Our study focused on user-initiated interactions in which participants explicitly called the agent to initiate a conversation. In this context, whispering was associated with enhanced social presence during socially oriented nighttime interactions. However, system-initiated interactions (e.g., proactive notifications) may require more cautious use of whispering. Because whispering functions as a strong signal of intimacy, an unsolicited whispered response from an agent could be perceived as intrusive or eerie, particularly when the user has not yet engaged in a social dialog. These findings suggest that interaction initiative should be considered as a key factor when applying whispering as a nighttime vocal strategy.

In addition to temporal context, our observations during the study suggest that physical and environmental factors may also shape nighttime voice interactions. For example, several participants spontaneously adjusted their posture or moved the device closer in order to better hear whispering voices in the low-light setting. While not systematically measured, this behavior highlights how vocal intensity can interact with spatial proximity and listening effort. From a design perspective, this suggests that whispering voices may benefit from complementary cues—such as adaptive volume scaling, proximity awareness, or clarity enhancement—to maintain intimacy without increasing cognitive or physical effort for users. Designers should therefore consider the embodied conditions of nighttime use—where listening effort and physical comfort are paramount.

Finally, adaptive voice agents may benefit from incorporating user-specific signals. Rather than relying solely on time-of-day heuristics, future systems could adjust vocal intensity based on users' recent interaction patterns or speech characteristics. Treating whispering as a situational response—rather than a fixed design choice—allows voice agents to better accommodate individual preferences and evolving nighttime contexts.

5. 4. Limitations and Future works

Limitations. This study focused on voice-based interactions in after-dark contexts, examining how whispering voices shape user experiences in low-arousal environments. While this scope enabled a detailed exploration of nighttime-specific design needs, it limits the ability to determine whether the effects observed are unique to nighttime or reflect broader user preferences. Our single-context approach aligns with prior studies that examined nighttime or dark settings without daytime comparisons, such as Oshima et al. (2020) on smartphone use in dark environments and Himschoot et al. (2024) on feelings of safety during nighttime outdoor recreation under different artificial lighting conditions.

Additionally, the controlled laboratory environment minimized distractions but may not fully capture the complexity of real-life bedroom contexts, including the situational nuances and emotional states that can shape users' experiences and affective responses during bedtime routines. Furthermore, the use of predetermined dialog scenarios, while necessary to ensure experimental control and comparability across conditions, may have constrained participants' sense of personal relevance and spontaneity compared to fully user-initiated interactions. As a result, the observed responses reflect reactions to standardized interaction structures rather than participants' own naturally occurring nighttime conversations.

Furthermore, a methodological limitation concerns the operationalization of the whispering condition. In this study, whispering was implemented as a low-intensity vocal delivery consistent with how whispering naturally occurs in nighttime contexts. As such, phonatory characteristics were inherently accompanied by reduced loudness, making it difficult to attribute the observed effects exclusively to whisper phonation alone. The findings should therefore be interpreted as reflecting an intimacy-forward, low-intensity voice configuration rather than an isolated phonatory manipulation. Future research should systematically disentangle phonation, loudness, and vocal identity to clarify their independent contributions to user experience.

Moreover, this study focused on short-term interactions, limiting our findings to initial interactions. Long-term studies are needed to explore how ongoing dialog with whispering agents affects user experience and relationship-building over time. Finally, our sample was restricted to adults aged 20 to 30, who were experienced voice-agent users, which limits the generalizability of the findings. Experienced users in their 20s may be more receptive to novel voice tones and social-oriented conversations than other demographic groups or novice users.

Future research. Future research should (1) include daytime and other contextual comparisons to test the boundary conditions of whisper-style voice effects, (2) replicate the study in more naturalistic settings and through in-the-wild and longitudinal designs to capture everyday routines and contextual variability (e.g., ambient lighting and background noise), (3) recruit broader demographic groups and users with varying levels of VA experience, and (4) disentangle additional vocal parameters (e.g., pitch, prosody, speech rate) to identify which acoustic components drive comfort and trust in low-stimulation contexts. In summary, this study establishes that voice is not merely a communication tool but a contextual cue that fundamentally shapes user perceptions and experiences. By situating voice design within nighttime contexts, it advances both academic discourse on voice interaction and the practical development of adaptive, context-aware VAs that can support richer and more natural interactions.

6. Conclusion

This study examined how voice type and dialog context shape user experiences with voice agents in nighttime environments. The central finding is that nighttime voice UX is not determined by a single optimal voice type, but by the congruence between vocal delivery and dialog intent. When whisper-style delivery was aligned with socially oriented dialog, users reported lower tension and higher likability; conversely, mismatches—particularly a standard conversational voice in emotionally sensitive nighttime interactions—increased tension and diminished likability. These results reframe whispering not as a universally beneficial feature, but as a context-sensitive design resource whose effectiveness depends on situational appropriateness. Overall, the findings offer transferable design guidance: voice agents operating in after-dark contexts should prioritize tone–intent congruence over uniform vocal strategies, moving toward adaptive, contextually calibrated interaction design.

References

1. Aarts, T., Markopoulos, P., Giling, L., Vacaretu, T., & Pillen, S. (2022, June). Snoozy: a chatbot-based sleep diary for children aged eight to twelve. In *Proceedings of the 21st Annual ACM Interaction Design and Children Conference* (pp. 297–307). https://doi.org/10.1145/3501712.3529718
2. Andersen, J. (2015). Now you've got the shiveries: Affect, intimacy, and the ASMR whisper community. *Television & New Media, 16*(8), 683–700. https://doi.org/10.1177/1527476414556184
3. Appel, J., Von der Pütten, A., Krämer, N. C., & Gratch, J. (2012). Does humanity matter? Analyzing the importance of social cues and perceived agency of a computer system for the emergence of social reactions during human-computer interaction. *Advances in Human-Computer Interaction, 2012*(1), 324694. https://doi.org/10.1155/2012/324694
4. Barber, L. K., Rupprecht, E. A., & Munz, D. C. (2014). Sleep habits may undermine well-being through the stressor appraisal process. *Journal of Happiness Studies, 15*(2), 285–299. https://doi.org/10.1007/s10902-013-9422-2
5. Barratt, E. L., & Davis, N. J. (2015). Autonomous Sensory Meridian Response (ASMR): a flow-like mental state. *PeerJ, 3*, e851. https://doi.org/10.7717/peerj.851
6. Bhandari, U., Neben, T., Chang, K., & Chua, W. Y. (2017). Effects of interface design factors on affective responses and quality evaluations in mobile applications. *Computers in Human Behavior, 72*, 525–534. https://doi.org/10.1016/j.chb.2017.02.044
7. Bickmore, T., & Cassell, J. (2001, March). Relational agents: a model and implementation of building user trust. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 396–403). https://doi.org/10.1145/365024.365304
8. Bickmore, T., & Cassell, J. (2005). Social dialogue with embodied conversational agents. *Advances in natural multimodal dialogue systems, 30*, 23–54. https://doi.org/10.1007/1-4020-3933-6
9. Cambre, J., & Kulkarni, C. (2019). One voice fits all? Social implications and research challenges of designing voices for smart devices. *Proceedings of the ACM on human-computer interaction, 3*(CSCW), 1–19. https://doi.org/10.1145/3359325
10. Cassell, J., Bickmore, T., Billinghurst, M., Campbell, L., Chang, K., Vilhjálmsson, H., & Yan, H. (1999, May). Embodiment in conversational interfaces: Rea. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (pp. 520–527). https://doi.org/10.1145/302979.303150
11. Carr, E. W., Hofree, G., Sheldon, K., Saygin, A. P., & Winkielman, P. (2017). Is that a human? Categorization (dis) fluency drives evaluations of agents ambiguous on human-likeness. *Journal of Experimental Psychology: Human Perception and Performance, 43*(4), 651. https://doi.org/10.1037/xhp0000304

12. Chattaraman, V., Kwon, W. S., Gilbert, J. E., & Ross, K. (2019). Should AI-Based, conversational digital assistants employ social-or task-oriented interaction style? A task-competency and reciprocity perspective for older adults. *Computers in human behavior, 90*, 315-330. https://doi.org/10.1016/j.chb.2018.08.048
13. Chang, A. M., Aeschbach, D., Duffy, J. F., & Czeisler, C. A. (2015). Evening use of light-emitting eReaders negatively affects sleep, circadian timing, and next-morning alertness. *Proceedings of the National Academy of Sciences, 112*(4), 1232-1237. https://doi.org/10.1073/pnas.1418490112
14. Choi, T. R., & Drumwright, M. E. (2021). "OK, Google, why do I use you?" Motivations, post-consumption evaluations, and perceptions of voice AI assistants. *Telematics and Informatics, 62*, 101628. https://doi.org/10.1016/j.tele.2021.101628
15. Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1992). Extrinsic and intrinsic motivation to use computers in the workplace 1. *Journal of applied social psychology, 22*(14), 1111-1132. https://doi.org/10.1111/j.1559-1816.1992.tb00945.x
16. Erika, R. (2023). How to enable whisper mode on Alexa. Retrieved January 2, 2024 from https://www.digitaltrends.com/home/how-to-enable-whisper-mode-on-alexa/. Html
17. Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of marketing research, 18*(1), 39-50. https://doi.org/10.1177/002224378101800104
18. Franke, R. H., & Kaul, J. D. (1978). The Hawthorne experiments: First statistical interpretation. *American sociological review*, 623-643. https://doi.org/10.2307/2094540
19. Gaver, W., Boucher, A., Law, A., Pennington, S., Bowers, J., Beaver, J., ... & Wilkie, A. (2008, April). Threshold devices: looking out from the home. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1429-1438). https://doi.org/10.1145/1357054.1357278
20. Goodman, K. L., & Mayhorn, C. B. (2023). It's not what you say but how you say it: Examining the influence of perceived voice assistant gender and pitch on trust and reliance. *Applied Ergonomics, 106*, 103864. https://doi.org/10.1016/j.apergo.2022.103864
21. Hardian, H., Febriani, S. S., Sumekar, T. A., Muniroh, M., Indraswari, D. A., Purwoko, Y., & Ambarwati, E. (2020). Improvement of Sleep Quality by Autonomous Sensory Meridian Response (ASMR) Stimulation Among Medical Students. *Malaysian Journal of Medicine & Health Sciences, 16*.
22. Himschoot, E. A., Crump, M. C., Buckley, S., Cai, C., Lawson, S., White, J., ... & Newman, P. (2024). Feelings of safety for visitors recreating outdoors at night in different artificial lighting conditions. *Journal of Environmental Psychology, 97*, 102374. https://doi.org/10.1016/j.jenvp.2024.102374
23. Hone, K. S., & Graham, R. (2000). Towards a tool for the subjective assessment of speech system interfaces (SASSI). *Natural Language Engineering, 6*(3-4), 287-303. https://doi.org/10.1017/S1351324900002497
24. Jestin, I., Fischer, J., Galvez Trigo, M. J., Large, D., & Burnett, G. (2022, July). Effects of wording and gendered voices on acceptability of voice assistants in future autonomous vehicles. In *Proceedings of the 4th conference on conversational user interfaces* (pp. 1-11). https://doi.org/10.1145/3543829.3543836
25. Kang, Y., Na, G., & Choi, J. (2021). The Effect of Preceding Utterance on the User Experience in the Voice Agent Interactions-Focus on the Conversational Types in the Smart Home Context. *The journal of the convergence on culture technology, 7*(1), 620-631.
26. Karlgren, K., & McMillan, D. (2022, October). Designing for Extreme Sleepers: Rethinking the Rhythms of Sleep Technology. In *Nordic Human-Computer Interaction Conference* (pp. 1-17). https://doi.org/10.1145/3546155.3546685
27. Karlgren, K., Brown, B., & McMillan, D. (2022). From self-tracking to sleep-hacking: online collaboration on changing sleep. *Proceedings of the ACM on human-computer interaction, 6*(CSCW2), 1-26. https://doi.org/10.1145/3555630

28. Keeling, K., McGoldrick, P., & Beatty, S. (2010). Avatars as salespeople: Communication style, trust, and intentions. *Journal of business research, 63*(8), 793-800. https://doi.org/10.1016/j.jbusres.2008.12.015
29. Kim, H. C., Cha, M. C., & Ji, Y. G. (2021). The impact of an agent's voice in psychological counseling: Session evaluation and counselor rating. *Applied Sciences, 11*(7), 2893. https://doi.org/10.3390/app11072893
30. Kühne, K., Fischer, M. H., & Zhou, Y. (2020). The human takes it all: Humanlike synthesized voices are perceived as less eerie and more likable. evidence from a subjective ratings study. *Frontiers in neurorobotics, 14*, 593732. https://doi.org/10.3389/fnbot.2020.593732
31. Lee, S., Ratan, R., & Park, T. (2019). The voice makes the car: Enhancing autonomous vehicle perceptions and adoption intention through voice agent gender and style. *Multimodal Technologies and Interaction, 3*(1), 20. https://doi.org/10.3390/mti3010020
32. Li, X. (2011). Whispering: the murmur of power in a lo-fi world. *Media, Culture & Society, 33*(1), 19-34. https://doi.org/10.1177/0163443710385498
33. Luger, E., & Sellen, A. (2016, May). "Like Having a Really Bad PA" The Gulf between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI conference on human factors in computing systems* (pp. 5286-5297). https://doi.org/10.1145/2858036.2858288
34. March, W., Nafus, D., Swan, L., & Taylor, A. S. (2008). Night and darkness: interaction after dark. In *CHI'08 Extended Abstracts on Human Factors in Computing Systems* (pp. 3985-3988). https://doi.org/10.1145/1358628.1358973
35. McGinn, C., & Torre, I. (2019, March). Can you tell the robot by the voice? An exploratory study on the role of voice in the perception of robots. In *2019 14th ACM/IEEE international Conference on human-robot interaction (HRI)* (pp. 211-221). IEEE. https://doi.org/10.1109/HRI.2019.8673305.
36. Moore, R. K. (2017, August). Appropriate voices for artefacts: some key insights. In *1st International workshop on vocal interactivity in-and-between humans, animals and robots*.
37. Nass, C., Steuer, J., & Tauber, E. R. (1994, April). Computers are social actors. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 72-78).
38. Nass, C., & Lee, K. M. (2000, April). Does computer-generated speech manifest personality? An experimental test of similarity-attraction. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (pp. 329-336). https://doi.org/10.1145/332040.332452
39. Nass, C. I., & Brave, S. (2005). *Wired for speech: How voice activates and advances the human-computer relationship* (p. 9). Cambridge, MA: MIT press.
40. Nagare, R., Plitnick, B., & Figueiro, M. G. (2019). Does the iPad Night Shift mode reduce melatonin suppression?. *Lighting Research & Technology, 51*(3), 373-383. https://doi.org/10.1177/1477153517748189
41. Nißen, M., Selimi, D., Janssen, A., Cardona, D. R., Breitner, M. H., Kowatsch, T., & von Wangenheim, F. (2022). See you soon again, chatbot? A design taxonomy to characterize user-chatbot relationships with different time horizons. *Computers in Human Behavior, 127*, 107043. https://doi.org/10.1016/j.chb.2021.107043
42. Oshima, H., Matoba, Y., & Siio, I. (2020, April). Smartphone Operation in a Dark Environment for Semi-awake Users. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-8). https://doi.org/10.1145/3334480.3382911
43. Park, S. H., & Lee, S. C. (2022, September). Which Voice Do You want To Hear From Your Automated Vehicle? User Preference on In-Vehicle Intelligent Agent Voice in Automated Vehicles. In *Adjunct Proceedings of the 14th International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 91-93). https://doi.org/10.1145/3544999.3552523
44. Parviainen, E., & Søndergaard, M. L. J. (2020, April). Experiential qualities of whispering with voice assistants. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1-13). https://doi.org/10.1145/3313831.3376187

45. Poerio, G. L., Blakey, E., Hostler, T. J., & Veltri, T. (2018). More than a feeling: Autonomous sensory meridian response (ASMR) is characterized by reliable changes in affect and physiology. *PloS one, 13*(6), e0196645. https://doi.org/10.1371/journal.pone.0196645
46. Porcheron, M., Fischer, J. E., & Sharples, S. (2017, February). "Do Animals Have Accents?" Talking with Agents in Multi-Party Conversation. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing* (pp. 207-219). https://doi.org/10.1145/2998181.2998298
47. Purington, A., Taft, J. G., Sannon, S., Bazarova, N. N., & Taylor, S. H. (2017, May). "Alexa is my new BFF" social roles, user satisfaction, and personification of the Amazon Echo. In *Proceedings of the 2017 CHI conference extended abstracts on human factors in computing systems* (pp. 2853-2859). https://doi.org/10.1145/3027063.3053246
48. Sciuto, A., Saini, A., Forlizzi, J., & Hong, J. I. (2018, June). "Hey Alexa, What's Up?" A Mixed-Methods Studies of In-Home Conversational Agent Usage. In *Proceedings of the 2018 designing interactive systems conference* (pp. 857-868). https://doi.org/10.1145/3196709.3196772
49. Seifzadeh, S., Asayesh, V., Nikjeh, M. T., Dehghani, M., Parsa, E. R., & Asgharianasl, F. (2023). The physiological effects of ASMR on anxiety. *Frontiers in Biomedical Technologies, 10*(4), 459-464. https://doi.org/10.18502/fbt.v10i4.13728
50. Semertzidis, N. A., Li Pin Hiung, A., Vranic-Peters, M. J., & Mueller, F. F. (2023, April). Dozer: Towards understanding the design of closed-loop wearables for sleep. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (pp. 1-14). https://doi.org/10.1145/3544548.3581044
51. Semertzidis, N. A., Sargeant, B., Dwyer, J., Mueller, F. F., & Zambetta, F. (2019, May). Towards understanding the design of positive pre-sleep through a neurofeedback artistic experience. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (pp. 1-14). https://doi.org/10.1145/3290605.3300804
52. Takano, K., Iijima, Y., & Tanno, Y. (2012). Repetitive thought and self-reported sleep disturbance. *Behavior therapy, 43*(4), 779-789. https://doi.org/10.1016/j.beth.2012.04.002
53. Takano, K., Sakamoto, S., & Tanno, Y. (2014). Repetitive thought impairs sleep quality: An experience sampling study. *Behavior therapy, 45*(1), 67-82. https://doi.org/10.1016/j.beth.2013.09.004
54. Tang, N. K., & Harvey, A. G. (2004). Effects of cognitive arousal and physiological arousal on sleep perception. *Sleep, 27*(1), 69-78. https://doi.org/10.1093/sleep/27.1.69
55. Tay, B., Jung, Y., & Park, T. (2014). When stereotypes meet robots: the double-edge sword of robot gender and personality in human-robot interaction. *Computers in Human Behavior, 38*, 75-84. https://doi.org/10.1016/j.chb.2014.05.014
56. Thayer, R. E. (1986). Activation-deactivation adjective check list: Current overview and structural analysis. *Psychological reports, 58*(2), 607-614. https://doi.org/10.2466/pr0.1986.58.2.607
57. Tolmeijer, S., Zierau, N., Janson, A., Wahdatehagh, J. S., Leimeister, J. M. M., & Bernstein, A. (2021, May). Female by default?-exploring the effect of voice assistant gender and pitch on trait and trust attribution. In *Extended abstracts of the 2021 CHI conference on human factors in computing systems* (pp. 1-7). https://doi.org/10.1145/3411763.3451623
58. Wagner, K., & Schramm-Klein, H. (2019). Alexa, are you human? Investigating anthropomorphism of digital voice assistants-a qualitative approach.
59. Wilkinson, L. C. (1982). Introduciton: A sociolinguistic approach to communicationg in the classroom. *Communicating in the clasroom*, 3-11.
60. Zarazaga, P. P., Henter, G. E., & Malisz, Z. (2023, June). A processing framework to access large quantities of whispered speech found in ASMR. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1-5). IEEE. https://doi.org/10.1109/ICASSP49357.2023.10095965.
61. Zheng, H., Luo, M., Wang, Y., & Wei, Y. (2024). Multi-sensory interaction and spatial perception in urban microgreen spaces: A focus on vision, auditory, and olfaction. *Sustainability, 16*(20), 8809. https://doi.org/10.3390/su16208809