# AI or Human Voiceover in Ads: How Pre-Exposure Affects Trust in the Advertisement, Perceived Voice Effectiveness and Mental Effort

Koosje Bogers

Snr 2125610

Master's Thesis

Communication and Information Sciences

School of Humanities and Digital Sciences

Tilburg University, Tilburg

Supervisor: Dr. J. Engelen

Second reader: Dr. N. Cohn

January 2025

**Technology statement**

1. Did you use any tools or services to paraphrase text from other sources? *Quillbot.*
2. Did you use any tools or services to check spelling or grammar? *Built-in Word spelling checker and Grammarly*
3. Did you use any tools or services to typeset the given text? *No*
4. Did you use any tools or services to generate part of the text? *No*
5. Did you use any generative AI tools or software for other aspects of your thesis? *Microsoft CoPilot was used to find relevant scientific sources and ChatGPT for paraphrasing my own text into correct academic English.*

By submitting my thesis for assessment, I hereby confirm that:

- The thesis is my own intellectual property and that ideas as well as language from other sources have been properly cited. All quotes and sourced information have been properly identified as such.
- I have disclosed any technology that I have used in the writing process.

**Abstract**

The increasing integration of artificial intelligence (AI) in advertising, particularly through AI voiceovers, raises questions about consumer perceptions compared to human voices. This study investigates the effects of voice type (human vs. AI) and pre-exposure (with vs. without) on trust in the advertisement, perceived voice effectiveness, and mental effort. A 2x2 between-subjects design was implemented, involving 161 Dutch-speaking participants who were randomly assigned to one of four conditions. Participants listened to a fictional radio advertisement either with or without pre-exposure to the voice. Results indicate that human voices significantly outperform AI voices in trust and voice effectiveness, emphasizing the importance of human-like natural prosody of human speech. Contrary to expectations, pre-exposure to AI voices did not significantly enhance trust, and voice effectiveness, or reduce mental effort, suggesting that brief exposure may be insufficient to overcome inherent biases toward artificial voices. Mediation analyses also revealed no significant indirect effects of mental effort on trust or perceived voice effectiveness, suggesting that cognitive voice processing difficulties alone do not fully explain the observed differences. These findings challenge theories like the Mere Exposure Effect and Processing Fluency, which predict that familiarity enhances preference and trust. Implications for advertisers include prioritizing human voiceovers for trust-building or investing in higher-quality AI voices with improved naturalness to bridge the perceptual gap. Future research should explore the role of extended or repeated exposure to AI voices, as well as the potential benefits of transparency in disclosing AI-generated content to consumers.

*Keywords:* AI voiceovers, human voice, advertising, trust, mere exposure effect, cognitive load, processing fluency, uncanny valley.

# AI or Human Voiceover in Ads: How Pre-Exposure Affects Trust in the Advertisement, Perceived Voice Effectiveness and Mental Effort

In recent years, artificial intelligence (AI) has impacted multiple industries, including the marketing and advertising sector (Davenport et al., 2019 ; Gao et al., 2023 ; Campbell et al., 2022). The market value of AI in marketing worldwide in 2024 is 36 billion US dollars, and is estimated to be 107.5 billion US dollars in 2028 (Statista, 2023). Speech technologies, in particular, have gained popularity as companies integrate AI voiceovers into their marketing strategies (Weitzman, 2023). Voiceovers for advertisements, commercials and virtual assistants are being implemented using AI technologies (Amit, 2024). For example, supermarket chain Aldi announced their future commercials in The Netherlands will be replaced with an AI-generated voiceover (Minczeles et al., 2024 ; Kaat, 2024). Besides the new Aldi commercial, many other forms of advertising are nowadays created through text-to-speech (TTS) (MarketResearch, 2023). With TTS, computers mimic human speech patterns by turning text into spoken words (Keim, 2024).

Next to AI voices being time and cost-efficient (Leif, 2024), it allows companies to reach a diverse audience and offer customized content to individuals. This is caused by the endless possibilities in languages, genders, accents, and pitch (voicetune.ai, 2024 ; Lovo.ai, 2024). Moreover, the usage of AI voices can enhance consumer engagement massively, due to the creation of more personalised and interactive experiences. This will make content more relatable and engaging for listeners (Martin, 2024).

This rapid technological advancement also has its downsides. First, AI-generated advertisements can come across as less authentic, credible, and human (Craig, 2023). AI voices may have unnatural sounding prosody (intonation, pitch, rhythm), which leads to consumers feeling high levels of eeriness (creepiness) towards the voiceover and therefore the advertisement (Schreibelmayr & Mara, 2022b). This may cause an uncanny valley effect,

which refers to the unpleasant feelings that people may experience towards human-like robots or audio-visual simulations that closely resemble humans, but are not quite realistic (Mori, 2023 ; Kendall, 2024). The theoretical framework will elaborate on this phenomenon. Second, there are ethical concerns, as voices can be mimicked without consent, resulting in identity theft (Saumya, 2024). Another concern entails the disappearance of voice actors, according to multiple Dutch and international news sites (Pols, 2023 ; Colomé, 2023 ; Taylor, 2024 ; Kaat, 2024). However, little research has been done yet into the downsides of using an AI-generated speech (voicetune.ai, 2024).

Looking at the general perception and attitude of consumers towards AI, literature shows varying results. A survey by tech firm Cognizant observed that only one-third of consumers trust AI technologies (Balaratnam, 2024). Also a KPMG survey reveals that 61% of individuals are either uncertain about or unwilling to trust AI (Mabbott et al., 2023). A study on consumer interpretation of AI in marketing found limited understanding of AI communication, likely due to its novelty or the seamlessness of AI interactions that go unnoticed (Kumar et al., 2019). Chen et al. (2021) found that consumers largely accepted AI-driven marketing as inevitable but felt neutral to mildly negative about it, with little impact on brand or product perceptions.

Perceptions of AI marketing remain cautious, with limited understanding and distrust (Balaratnam, 2024; Mabbott et al., 2023). Building trust in AI practices, such as AI-generated speech in ads, is crucial for consumer engagement and brand loyalty (Chen et al., 2022). Although this study did not disclose AI usage beforehand, research shows that transparency can boost trust (Kirkby et al., 2023). A Yahoo and Publicis Media survey found that ads explicitly stating AI use were 73% more trustworthy, with a 96% rise in overall company trust (Yahoo Inc., 2024).

When using AI voices in marketing, besides guarding consumer's trust in the advertisement, it is also important to use an effective voice. Voice effectiveness refers to the impact of the voice in an advertisement on the listener's perception and response. It entails six quality aspects; clarity, correctness, pleasantness, credibility, persuasiveness, and comprehensibility (Rodero, 2012). These aspects show how the advertisement effectively can communicate its message, influence the listener's feelings, and persuade them to respond to the ad's message. The effectiveness of an AI voice significantly influences the overall efficacy of advertisements (Wiener & Chartrand, 2014).

As consumers become more familiar with artificial voices, understanding how they process them and respond to them is crucial. In recent years, very few studies have been conducted on AI voices in advertising. A study by Emma Rodero (2017) examined how human and artificial voices are processed in advertising narratives. The research found that human voices are generally more effective than artificial voices in terms of achieving better attention, concentration, and recall. It suggests that the natural prosody (intonation, rhythm, and stress) of human speech plays a significant role in how voices are processed and understood by listeners. While the study did not specifically measure cognitive load, it implied that human voices with natural prosody are easier to process cognitively. The natural variations in intonation and rhythm make the information easier to follow and understand. It also enhances listener engagement and attentiveness to the narrative. The increased comprehensibility of human speech reduces the mental effort required for listeners, thereby allowing more cognitive resources to be allocated to accurate message encoding. In contrast, artificial voices, being less comprehensible, require more concentration and cognitive effort (Rodero, 2017).

A study on the effectiveness of AI voiceovers in short video ads examined participants' cognitive load using both subjective (self-report surveys) and objective (eye-tracking, dual-

task performance) measures. It found that a human voiceover leads to reduced cognitive load and boosts purchase intention more than an AI voiceover. It also found that subtitles reduce cognitive load and improve ad comprehension, regardless of whether the voiceover is human or AI, with the effect mediated by cognitive load (Wang et al., 2024). However, this study focused exclusively on advertisement evaluation in terms of ad conversion and purchase intention, without specifically assessing the voiceover itself.

In line with these findings, Gong's (2023) EEG-based study found that brain activity responses were greater when participants listened to human voices compared to AI voices. Human voices triggered more engagement in brain regions linked to processing, comprehension, and working memory, supporting the idea that AI voices increase cognitive load. The study suggests further research on familiarizing with AI voices to reduce cognitive load and explore long-term effects and adaptation for improved media effectiveness.

Although research within the advertising context remains limited, studies in other contexts have indicated that a human voiceover is not always explicitly preferred. A study examining intelligent virtual agents in the context of providing study stress tips found that the agent’s AI voice achieved similar levels of trust and likability to the agent’s human voice (Abdulrahman & Richards, 2022). This suggests that a synthetic voice could potentially offer benefits similar to those of a human voice in fostering trust and likability in interactions. Research on explainer videos for students investigated their perceptions of human and AI voiceovers, revealing a preference for the human-like voiceover while showing no significant negative perception of the AI voiceover (Leong et al., 2023).

Familiarization or pre-exposure to an AI voice in an advertisement may significantly shape how listeners perceive and evaluate the voice (Gong, 2023). Pre-exposure could play a critical role in determining how well a voice can be cognitively processed with the listener, potentially impacting trust in the advertisement and the effectiveness of the AI voice (Heald &

Nusbaum, 2014). A 2024 study compared human and AI voice advertising, considering ad duration (8 vs. 30 seconds). It found no significant impact of duration on voice effectiveness or attitudes toward the ad. However, human voiceovers led to higher effectiveness and more positive attitudes, with most listeners unaware of whether the voice was AI or human (Lu et al., 2024).

Since the difference in comprehension between human and synthetic speech may be mediated by the amount of exposure, Lauretta et al. (1990) investigated synthetic telephone communication, focusing on message length and structure. They found minimal difference in comprehension between human and synthetic speech, with human speech slightly better understood. The study suggests that a short introductory message can improve the comprehensibility of high-quality synthetic speech in some telephone applications.

Although the role of exposure in voice processing and comprehension has not been extensively explored (Lauretta et al., 1990), it raises important questions about whether familiarization with a voice significantly impacts consumer attitudes and advertisement effectiveness. While previous research has found that duration does not have a significant effect on attitudes or voice effectiveness (Lu et al., 2024), no studies have examined whether familiarization with AI-generated voices can positively influence consumer attitudes and, consequently, the advertisement. It has been suggested that further research is needed to explore the familiarization effects of artificial speech through pre-exposure to the voice, which may enhance comprehension, improve processing fluency, and reduce cognitive load (Lauretta et al., 1990; Gong, 2023).

To summarize, despite the growing use of AI speech, little is known about this rather new technology and how consumers perceive it compared to human speech, particularly in terms of trust and voice effectiveness (Balaratnam, 2024). Enhancing AI voices in advertising and learning more about consumer preferences to modify marketing tactics are areas of

significant attention (Gélinas-Chebat et al., 1996). Moreover, it remains unclear whether pre-exposure with the voice can influence these perceptions and may decrease the required cognitive load, presenting a gap in the literature that this study seeks to address. Knowing which type of voice and exposure is most appropriate in an advertisement is important for increasing the consumer experience. Therefore, this study contributes to the literature on further developments in AI-generated content. The following research question is formulated: "*To what extent do the voiceover type (AI versus human) in an advertisement and pre-exposure to the voiceover affect consumers' trust in the advertisement and perceived voice effectiveness, and how does mental effort mediate these relationships?".*

## Theoretical framework

### The uncanny valley effect with artificial voices

To explain why we can have negative feelings toward an artificial voice, the uncanny valley theory might be useful. The theory suggests that as robots or objects become more human-like, people's affection for them grows until they reach a point where they are almost, but not fully, human, which triggers feelings of eeriness and discomfort. This decline in emotional response is called the "uncanny valley". Once the objects achieve a completely human-like appearance, people's affection for them rises again (Mori, 2023). Mori proposed that not only visual features can be the cause of the uncanny valley effect. Also auditory features such as the voice can fall into the uncanny valley and have gained small attention from researchers in the past. There is a growing body of research on the audio uncanny valley, though it is not as extensive as research on the visual uncanny valley (Mitchell et al., 2011, Romportl, 2014).

A study examining how deviations from human voices cause uncanniness in artificial voices, suggests that artificial voices can sometimes escape the uncanny valley if they closely mimic human voices (Diel & Lewis, 2023). According to another study, the artificial voices did not consistently fall into the uncanny valley. In fact, higher ratings of human-likeness in the artificial voices were associated with reduced eeriness and increased likability, contradicting the typical uncanny valley expectation. However, this effect was not universal across all synthesized voices, indicating that certain voices may still evoke unease depending on specific features, like intonation, sound and emotion of the voice (Kühne et al., 2020).

An AI voice that sounds almost human but just not quite, may fall into the uncanny valley. The unnatural prosody of an AI voice, such as incorrect intonation, rhythm, or stress, can make it sound robotic and awkward (Staff, 2024). According to research, prosodic factors play a crucial role in the perceived attractiveness of AI voices, and manipulation of these

factors can improve the naturalness of an artificial voice. This includes a moderate and natural pace, speech patterns that mimic humans with rising and falling intonations, and an appropriate pitch (Wang et al., 2024 ; Hirose, 2019 ; Rodero, 2017).

**Processing artificial voices: cognitive and affective routes**

When processing a voice, whether human or artificial, listeners engage in both cognitive and affective routes of processing (Schirmer & Gunter, 2017).

*Cognitive route*

The cognitive process of hearing and interpreting a voice involves two key components: decoding the message and managing cognitive load. Decoding refers to recognizing speech sounds, understanding words, and extracting meaning, engaging brain areas like the auditory cortex (Sedivy, 2020 ; Herbert & Pisoni, 2023) Simultaneously, cognitive load reflects the mental effort required to process and retain the information, influenced by factors such as message complexity, voice clarity, and familiarity with the speech (Lang et al., 2013). Together, these processes enable the listener to understand and integrate auditory input effectively (Heald & Nusbaum, 2014).

A theoretical framework supporting this understanding is the Cognitive Load Theory (CLT), developed by John Sweller (1988, 2011). CLT explains how information is processed and stored in the human brain and categorizes cognitive load into three distinct types: intrinsic load, extraneous load, and germane load (Sweller, 1988, 2011). Intrinsic load refers to the inherent complexity of the material being processed. Auditory information, such as hearing a voice, is considered part of this load.

Next to consciously processing a new voice, humans can process it subconsciously, even when not actively paying attention. It is found that we can even process a new voice during our sleep (Ameen et al., 2022). This unconscious processing refers to the brain’s ability to

adapt and become familiar with a new voice without conscious effort. When it comes to artificial voices, processing them typically demands more cognitive effort than processing human voices. This is because artificial voices often exhibit robotic characteristics, which can draw attention due to their novelty and unfamiliarity (Schreibelmayr & Mara, 2022b ; Lu et al., 2024).

Voice quality can also influence cognitive load. For example, listening to a creaky voice has been shown to increase cognitive load compared to listening to a typical voice. This heightened cognitive effort leaves fewer mental resources available for secondary tasks. In one study, participants were tasked with reacting to a visual stimulus (secondary task) while simultaneously listening to a voice (primary task), demonstrating the impact of voice quality on cognitive resource allocation. (Schiller et al., 2023). Human speech is generally more comprehensible, which reduces cognitive load and allows listeners to allocate more mental resources to accurately encode the message. As a result, when the speech is less comprehensible, as is often the case with artificial voices, listeners must exercise more focus and mental effort to understand the content (Rodero, 2017).

Research has shown that hearing a novel, artificial voice requires more intrinsic cognitive load, leading to a longer processing time (Rodero, 2017; Wang et al., 2024). According to CLT, when intrinsic cognitive load is high, listeners are more likely to make mistakes, perform tasks less accurately, and become quickly overloaded with information (Paas & van Merriënboer, 2020). In the context of a voiceover in an audio advertisement, this increased cognitive load could disadvantage the use of AI voices.

Pre-exposing participants to the AI voice may reduce the difficulty of decoding the message and lower cognitive load, as they become familiar with the voice through prior exposure. Cognitively, repeated exposure to a new voice helps listeners to "get used to" it, reducing the mental effort required to process and understand the voice. Studies indicate that

the benefits of familiarity for speech comprehension can emerge quickly, sometimes within minutes of exposure (Weisman & Peña, 2021; Holmes et al., 2021). This reduction in cognitive load helps listeners decode the speech with greater ease, which aligns with the Processing Fluency Theory (Reber, 2012) explained in the next chapter. This forms the basis for the hypothesis that pre-exposure to the AI voice enhances trust in the advertisement and perceived voice effectiveness as a result of fluent processing and reduced cognitive load for participants.

*Affective route*

Pre-exposure to or familiarization with an artificial voice may not only facilitate easier cognitive processing but also generate more positive affective responses. The affective route of voice processing describes the listener's feelings toward the voice, which can significantly impact their attitudes toward the message. Human voices typically evoke positive feelings, while artificial voices can provoke unease if their characteristics appear unnatural (Mori, 2023 ; Wang et al., 2024 ; Hirose, 2019 ; Rodero, 2017). However, familiarization through pre-exposure to AI voices can improve affective evaluations, as listeners develop preferences for voices they find familiar (Zajonc, 1968; Nickerson, 2023).

The Mere Exposure Theory (Zajonc, 1968) suggests that repeated exposure to a stimulus increases liking. Attention to the stimulus is crucial for achieving the Mere Exposure effect, as research demonstrates that the effect is significant only when participants actively engage with the stimulus. Without focused attention, the Mere Exposure effect is not effectively triggered, highlighting the need for active engagement to enhance preference and liking for the exposed stimuli (Yagi et al., 2009 ; Huang & Hsieh, 2013). The theory implies that the more frequently listeners hear an AI voice, the more likely they are to develop a preference for it (Nickerson, 2023). The combination of cognitive ease and positive emotional responses

highlights the dual impact of familiarization on voice processing. Repeated exposure not only reduces cognitive effort but also enhances emotional responses, which can lead to more favourable evaluations of both the voice and the advertisement.

**Processing artificial voices: bridging routes**

The Processing Fluency Theory (Reber, 2012) connects the cognitive and affective routes of voice processing, suggesting that the ease with which information is processed in the human mind influences our preferences and judgements. Perceiving speech from a familiar talker (human voice) requires fewer cognitive resources compared to unfamiliar voices (artificial voice) (Ingvalson & Stoimenoff, 2021 ; McKenzie et al., 2020). This suggests more automatic and easier message decoding, resulting in a preference for hearing the easy-to-process human voice. When a voice is processed more fluently and is more easily decoded, due to familiarity, listeners tend to develop positive feelings toward it (Landwehr et al., 2017).

In this study, participants may find a human voice easier to process due to its inherent familiarity, leading to higher trust in the advertisement and enhanced perceived effectiveness of the voice. This study suggests that familiarizing individuals to the AI voice, using pre-exposure, can enhance listeners' evaluations of artificial voices, by increasing familiarity (Mere Exposure Theory), reducing cognitive barriers (CLT) and consequently enhancing the processing fluency.

This study specifically examines the effect of pre-exposure on the evaluation of AI-generated voices, without addressing the impact of pre-exposure on human voices. This focus is justified by the increasing adoption of AI-generated voices in advertising contexts (Weitzman, 2023 ; Amit, 2024). By investigating the role of pre-exposure exclusively on artificial voices, this study addresses a critical gap in the literature on artificial voice

processing, contributing to a deeper understanding of how familiarity influences perceptions of AI-generated speech.

**Trust in the advertisement**

Since trust in AI marketing activities is found to be minimal, consumer trust is crucial for companies to control, when using AI speech in their advertisements (Balaratnam, 2024; Mabbott et al., 2023). Fostering this trust is essential for companies, as it promotes consumer engagement, and enhances brand loyalty (Chen et al., 2022).

People are more likely to trust human voices, indicating a lack of confidence in voices that sound artificial and robotic (Audacy Inc., 2024). Research into human-robot interactions found that robotic voices can negatively affect trust. Artificial voices that sound robotic can create an impression that the speaker is not to be trusted, which affects the perceived credibility of the information provided (Torre & White, 2020b ; Schreibelmayr & Mara, 2022b). This is because humans have a tendency to trust familiar and easy-to-process information, linking back to the Processing Fluency Theory (Reber, 2012). When artificial voices deviate from this familiarity, they are perceived as less trustworthy, affecting the credibility of the information they convey. This distrust may therefore cause a decreased trust in the advertisement and in the company. Consequently, it challenges the brand loyalty and consumer engagement.

However, the quality of an artificial voice significantly influences trust. A study on the influence of voice quality concluded that higher-quality artificial voices were more likely to be trusted by participants. In the study, high-quality refers to voices that are clear, natural and human-like (Chiou et al., 2020). Research has shown that the tone of voice used in AI-generated advertisements plays a crucial role in shaping client trust and engagement. Formal

tones are more effective in fostering trust, whereas humorous tones are better at enhancing engagement (Shafiq et al., 2024).

**Perceived voice effectiveness**

Voices in audio advertising are a crucial aspect in delivering a message. It has been found that it influences how people respond to advertisements, particularly when the listeners are not fully attentive to the advertisement (Gelinas-Chebat & Chebat, 1992). When a voice works effectively, the message comes across more clearly. This in turn affects the ad efficacy (Wiener & Chartrand, 2014). The quality of the voice is found to be important in delivering effective audio advertisements. It increases voice effectiveness, recall of the content and attitude towards the advertisement (Lu et al., 2024). Using a high-quality voice enhances the chances of favourable reactions and evaluations form listeners (Dubey et al., 2018 ; Wiener & Chartrand, 2014). When companies start implementing AI voices in advertising contexts, it is therefore important to use a high-quality artificial voice.

Using a high-quality AI voice increases consumer trust and, at the same time, voice effectiveness. High-quality in this case describes a clear, natural, human-like voice (Chiou et al., 2020). A key feature that differentiates AI voices from human voices is the lack of naturalness. The importance of naturalness of an artificial voices has been noted more often in artificial voice studies (Duville et al., 2022 ; Rodero & Lucas, 2021). While human voices are inherently expressive, with varying pitch, rhythm, and emotional undertones, AI voices often struggle to replicate these subtleties, especially in conveying emotion or emphasis. This can make AI voices sound robotic and monotone, which may trigger a sense of artificiality for listeners. Additionally, AI voices may have slight irregularities in pronunciation or pacing, which can disrupt the smoothness and natural flow that human voices naturally provide. The absence of these characteristics makes it more apparent to listeners that the voice is artificial,

influencing their perception of the advertisement's trustworthiness and effectiveness (Lu et al., 2024 ; Wang et al., 2024 ; Hirose, 2019 ; Rodero, 2017). Based on the theoretical framework and the literature review, the following hypotheses are proposed:

H1: Participants exposed to a human voice will show higher levels of (a) trust in the advertisement and (b) perceived voice effectiveness, compared to participants exposed to an AI voice.

H2: Participants exposed to a human voice, will show lower levels of mental effort compared to participants exposed to an AI voice.

H3: Participants pre-exposed to an AI voice will show higher levels of (a) trust in the advertisement and (b) perceived voice effectiveness, compared to participants who did not get pre-exposed with the AI voice.

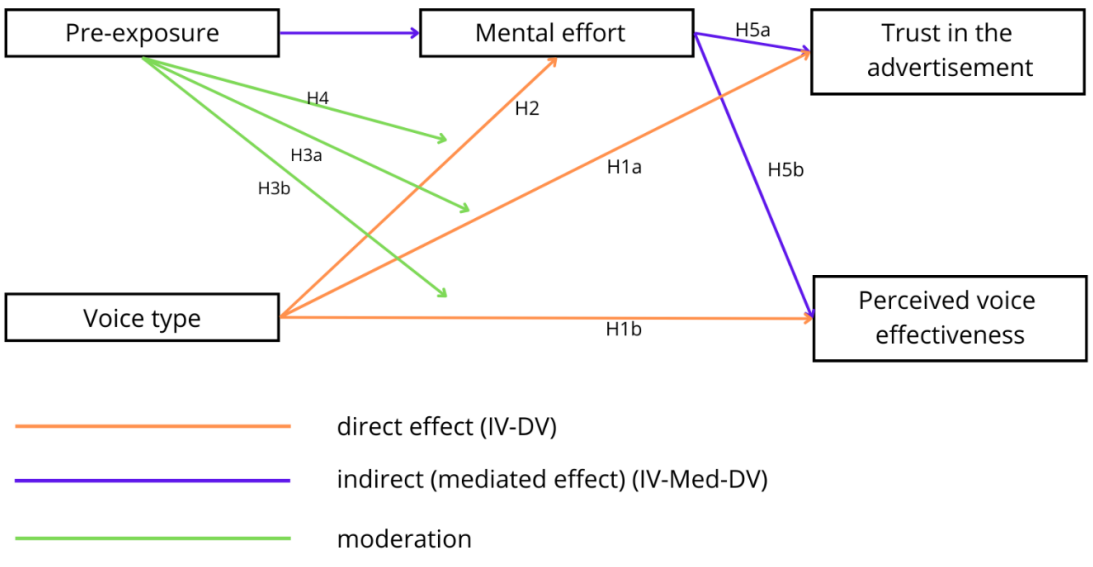
H4: Participants pre-exposed to an AI voice, will show lower levels of mental effort, compared to participants who did not get pre-exposed with the AI voice.

H5: Mental effort mediates the effect of pre-exposure on (a) trust in the advertisement and (b) perceived voice effectiveness.

**Conceptual model**

*Figure 1 visually represents the current study's conceptual model with the corresponding hypotheses.*

**Figure 1**



# Method

## Design

The experiment had a 2x2 between-subject design with voice type (human versus AI) and pre-exposure (with versus without) as independent variables. This resulted in four experimental conditions. Participants were randomly assigned to one of four groups and all listened to the same 25-second advertisement of a fictional aviation company, of which groups 1 and 3 heard a human voice and groups 2 and 4 heard an AI voice. The dependent variables are trust in the advertisement, perceived voice effectiveness, and mental effort. Moreover, mental effort served as a mediator, as it may explain the process through which the voice type and pre-exposure influence the outcomes of the dependent variables. Table 1 provides an overview of the experimental conditions.

**Table 1**

*Overview experimental conditions*

| Condition | Voice Type | Pre-Exposure |
|---|---|---|
| 1 | Human | Without |
| 2 | AI | Without |
| 3 | Human | With |
| 4 | AI | With |

## Materials

### *Stimuli*

Given the researcher's predominantly Dutch environment, this experiment was conducted in Dutch. The survey for the 'pre-exposure' condition in groups 3 and 4 included an introductory segment during which participants listened to an 18-second audio clip of the

voiceover. This segment served as the familiarization phase for this condition. In the audio, the voiceover briefly introduces himself to ensure that respondents become accustomed to the voice they will encounter again in the subsequent radio commercial.

Additionally, all participants were exposed to the same 25-second Dutch audio commercial of fictional aviation company ‘SkyWings’ (in an either human or AI voice depending on the condition group). The audio was presented after the pre-exposure and before completing the survey. The ad was created especially for this study to keep control over the duration and message content. The scripts of the introductory audio and the advertisement can be found in Appendix A.

The advertisement was recorded by a 25-year old male, who had a standard Dutch accent. To create the AI voiceover, text-to-speech program *NaturalReader* was used (NaturalReader, 2024). To minimize the risk of participants’ responses being influenced by vocal pitch, voices that sounded similar were utilized, despite one being human and the other artificial. The male AI voice in *NaturalReader* that most closely resembled the human voice was selected for this purpose.

*Pretest*

A pretest (see Appendix B) was conducted for the survey with a sample size of 16 participants. The purpose of the pretest was to test the functionality of the random assignment of the four condition groups. In addition, instruction and item comprehension, audio clarity, survey flow and duration were tested. During the pretest, a female voice was utilized. However, based on respondents' evaluations, it became evident that a male voice would be more effective for this study. Lastly, the pretest looked at the optimal length of the pre-exposure message. The researcher asked respondents whether they would have wanted to hear the voice in the pre-exposure conditions longer, in order to feel more familiar to the voice.

This was performed with a statement question with a scale ranging from totally disagree (0) to totally agree (4). 6 out of the 9 participants in the pre-exposure condition groups, indicated that a longer pre-exposure was not needed for better familiarization. The remaining three respondents felt neutral about it. After the pretest, the experimental procedure and survey questions were corrected.

**Procedure**

This experiment was conducted using the online survey platform Qualtrics (see Appendix C). All participants gave their informed consent to take part in the study. They completed the survey individually, in their private environment, and on their own electronic device. After consent, participants were asked demographic questions about their gender, age, and educational level. Next, they were asked control questions about their openness to technology, familiarity, and experience with artificial intelligence. Afterwards, pre-exposure conditions listened to the introductory audio and then continued to the radio commercial. Condition groups 1 and 2 did not listen to this audio fragment, as these groups do not have a familiarization phase. Afterward, participants of all condition groups listened to a 25-second radio commercial of fictional airplane company *SkyWings*. After listening to the audio, participants filled out the survey, which took about 5 minutes. Lastly, participants read the debriefing and were thanked for their time.

**Participants**

The required sample size of this study was 178 participants, allocating approximately 45 participants per condition group. This number was determined through a power analysis conducted using G*Power 3.1 software, aiming for a statistical power of 0.80, an effect size of 0.25, and a significance level of 0.05. Convenience sampling within the researcher's direct

environment was used for recruiting Dutch participants. Participation was voluntary and participants were not compensated financially.

In total, 201 participants took part in the study, however, not all the participants finished the survey. After removing these participants from the dataset, the final sample consisted of 161 participants. Out of the 161 participants, 80 were male and 81 female, and the age group of 35-44 years old was most prevalent. The distribution of the age and gender per condition are shown in table 2.

**Table 2**

*Age and gender distribution among participants per condition.*

| | Age | Males | Females |
|---|---|---|---|
| 1. human, without *(N=42)* | 37.8 | 24 | 18 |
| 2. ai, without *(N=43)* | 39.0 | 20 | 22 |
| 3. human, with *(N=42)* | 33.9 | 19 | 22 |
| 4. ai, with *(N=34)* | 36.6 | 16 | 18 |

*Note.* Age is expressed in Mean years using the category codes.

* N=34 for group 4 because relatively more respondents within this group did not complete the survey in full and therefore needed removal.

**Measurements**

The dependent variables in this research were trust in the advertisement, perceived voice effectiveness, and mental effort.

*Trust in the advertisement*

Trust in the advertisement was based on an existing ADTRUST scale made by Soh and colleagues. Originally, the 20-item Likert scale (1= strongly disagree, 7= strongly agree) is based on four distinct factors: reliability, usefulness, affect, and willingness to rely on advertising (Soh et al., 2009). In the original study participants completed the statement

"*information conveyed in national advertising is*…". In this study the statement was adjusted to "*information conveyed in this advertisement is…*", since it is about the one specific ad that participants are hearing and not asking about their general perception of advertising.

Usefulness and willingness to rely on are unrelated to this study, since these look at specific outcomes for the ad. As this study uses a fictional brand, there is no need to examine this. Therefore, the researcher decided to solely focus on reliability and affect. In the current study, the internal consistency of the variable trust in the advertisement (only containing factors 'reliability and 'affect') was measured using a reliability analysis within *Jamovi* (version 2.3.). The analysis found a Cronbach's alpha of .85, indicating a strong internal consistency. See table 3 for the modified construct, translated into Dutch.

**Table 3**

*Trust in the advertisement construct Dutch*

| Factor | Informatie in deze advertentie is… | Zeer mee oneens | Zeer mee eens |
|---|---|---|---|
| **Reliability** | 1. Eerlijk | | |
| | 2. Waarheidsgetrouw | | |
| | 3. Geloofwaardig | | |
| | 4. Betrouwbaar | | |
| | 5. Nauwkeurig | | |
| | 6. Feitelijk | | |
| | 7. Volledig | | |
| | 8. Duidelijk | | |
| **Affect** | 9. Sympathiek | | |
| | 10. Aangenaam | | |
| | 11. Positief | | |

*Perceived voice effectiveness*

Voice effectiveness is defined as the ability of the voice to attract attention and produce a better effect or result. Six items were adopted from the Rodero et al. (2012) study.

The scale had a Cronbach's alpha of .73, indicating an acceptable internal consistency. The scale asks participants how clear, correct, pleasant, credible, persuasive and comprehensible the voice in the advertisement was. The average of these variables is the effectiveness index. The scale is bi-polar with opposite pairs (1= lowest level, 7= highest). See table 4 for the construct translated into Dutch.

**Table 4**

*Perceived voice effectiveness construct in Dutch*

| In hoeverre vond je de stem van de advertentie… | | |
|---|---|---|
| Onduidelijk | | Duidelijk |
| Incorrect | | Correct |
| Onaangenaam | | Aangenaam |
| Niet geloofwaardig | 1 2 3 4 5 6 7 | Geloofwaardig |
| Niet overtuigend | | Overtuigend |
| Onbegrijpelijk | | Begrijpelijk |

*Mental effort*

To measure whether participants experience cognitive load when listening to the voiceover, the Paas Mental Effort Rating scale was used. The original scale presents a single question: "*How much mental effort did you invest in this task?".* Participants then respond using a nine-point Likert scale, where 1 represents "very, very low mental effort" and 9 represents "very, very high mental effort". Research suggests that among alternatives for measuring cognitive load, the Paas scale is mostly preferred (Ouwehand et al., 2021).

For this study, a single question was asked with a focus on the auditory information participants received, namely the fictional radio commercial. "*Hoeveel mentale inspanning heb je gedaan om naar de advertentie te luisteren?*". The question was accompanied by a brief explanation of the definition of mental effort (mental effort refers to the amount of

attention, concentration and energy you have to perform to listen to the commercial) to ensure that respondents could answer the question correctly.

*Manipulation check*

The survey asks respondents within AI voice-condition groups whether they recognized the voice as AI-generated, in order to check the implementation of the manipulation with the AI voice.

**Data analysis**

Data analysis was performed using the statistical software *Jamovi* (version 2.3.0). All data from Qualtrics was exported into a .csv file to be analysed in *Jamovi*. Both the scales trust in the advertisement and voice effectiveness were checked for reliability. In this study, a p-value of less than 0.05 was considered statistically significant. Skewness and kurtosis indicators with values between −1.96 and +1.96 were used to establish the normality of the data distribution. Three separate Factorial ANOVAs were used to test trust in the advertisement, perceived voice effectiveness and mental effort between the condition groups. Additionally, a mediation analysis was conducted utilizing the Generalized Linear Model (GLM) Mediation Model within the MEDMOD package in Jamovi. This analysis aimed to explore the total, direct, and indirect effects among the mediation effects of mental effort (H5).

# Results

## Manipulation check and control variables

To check whether people in the AI voice conditions quickly recognised the voice as AI-generated, they were asked whether they quickly realised this. 43 out of the 67 respondents within this condition said they immediately recognised the voice as AI. Moreover, to check the effectiveness of the pre-exposure for the AI voice, respondents in the AI, with pre-exposure condition were asked the question whether they would have preferred to hear the introduction longer, in order to get more used to the voice. It showed that 26% were neutral and 65% of the respondents within this condition disagreed with the statement 'I would have liked to hear the voice for longer to get more used to it'. Furthermore, all respondents were asked about their openness to technology, familiarity and experience with AI. Table 5 below shows that respondents mostly agreed with being open to new technologies. On average, respondents indicated agreement with being familiar with AI. Finally, they indicated that they averagely agreed with the statement of having experience with AI.

**Table 5**

*Descriptives statistics for control variables*

| | Mean | Median | SD |
|---|---|---|---|
| **Openness to technology** | 3.29 | 3 | 0.77 |
| **Familiarity with AI** | 3.08 | 3 | 0.80 |
| **Experience with AI** | 2.66 | 3 | 0.99 |

*Note:* The total N=161. All three variables ranged from 0 (totally disagree) to 4 (totally agree)

**Table 6**

*Descriptive statistics of the dependent variables per condition group*

| | Voice type | Pre-exposure | | | |
|---|---|---|---|---|---|
| | | Without | | With | |
| | | *N* | *M (SD)* | *N* | *M (SD)* |
| **Trust in the advertisement** | Human | 42 | 3.82 (0.90) | 43 | 3.64 (0.93) |
| | AI | 42 | 3.13 (0.73) | 34 | 3.38 (0.80) |
| **Voice effectiveness** | Human | 42 | 5.21 (0.99) | 43 | 5.39 (0.90) |
| | AI | 42 | 4.84 (0.75) | 34 | 5.16 (0.96) |
| **Mental effort** | Human | 42 | 2.74 (1.86) | 43 | 2.91 (1.41) |
| | AI | 42 | 3.38 (1.75) | 34 | 3.24 (1.54) |

*Note.* The total N=161. Trust in the advertisement ranges from 0-6, Voice effectiveness ranges from 0-6, Mental effort ranges from 0-8.

**Trust in the advertisement**

A Factorial ANOVA was used to test hypotheses H1a and H3a. There was a main effect of voice type: $F(1, 157) = 12.67$, $p < .001$, partial $\eta 2 = 0.08$. Trust ratings were higher for human voices (M=3.73, SE=0.09) than for AI voices (M=3.26, SE=0.10). These results support H1a, indicating that participants perceive human voices as more trustworthy. There was no main effect of pre-exposure: $F(1, 157) = 0.08$, $p = .78$, partial $\eta 2 = 0.00$. Participants pre-exposed to the voices, both AI and human (M=3.51, SE=0.10) did not show significantly higher trust ratings than those who were not pre-exposed (M=3.48, SE=0.09). Finally, no significant interaction effect was found between voice type and pre-exposure: $F(1, 157) = 2.48$, $p = .12$, partial $\eta 2 = 0.02$. Pre-exposure did not significantly moderate the effect of voice type on trust ratings.

Although the interaction between voice type and pre-exposure was not significant ($p = .12$), the simple main effect of pre-exposure, specifically with an AI voice on trust ratings in comparison with no-pre-exposure with the AI voice was tested. There was no significant effect of pre-exposure on trust in the advertisement within the AI voice condition, $F(1, 157) = 1.62$, $p = .205$, partial $\eta 2 = .01$. Participants who were pre-exposed to the AI voice (M=3.38, SE=0.15) did not significantly differ in trust levels compared to participants who were not

pre-exposed to the AI voice (M=3.13, SE=0.13). The estimated difference was b = 0.25, SE = 0.20, 95% CI [−0.14, 0.63], t(157) = 1.27, p = .205. Therefore H3a is not supported.

To conclude, the findings provide partial support for the hypotheses. H1a is supported, highlighting that human voices elicit greater trust in advertisements than AI voices. This aligns with prior research suggesting that humanized elements enhance perceived trustworthiness (Audacy Inc., 2024, Torre & White, 2020b ; Schreibelmayr & Mara, 2022b). In contrast, the analysis does not provide evidence to support H3a (pre-exposure to an AI voice increases trust in the advertisement). Although the mean trust score for participants with pre-exposure was slightly higher, this difference was not statistically significant. Additionally, the lack of a significant interaction effect implies that pre-exposure does not moderate the relationship between voice type and trust levels.

**Perceived voice effectiveness**

A Factorial ANOVA was used to test hypotheses H1b and H3b. There is no main effect of voice type: F(1, 157) = 4.46, p =.036, partial η2 =0.03. Human voices (M=5.30, SE=0.10) were rated as significantly more effective than AI voices (M=5.00, SE=0.10). This supports H1b, indicating that participants perceived human voices as more effective than AI voices. There was no main effect of pre-exposure: F(1, 157)= 3.02, p=.084, partial η2 =0.02. Participants pre-exposed to AI voices (M=5.27, SE=0.10) did not significantly differ in their ratings of voice effectiveness compared to those not pre-exposed (M=5.03, SE=0.10). Finally, there was no significant interaction between voice type and pre-exposure: F(1, 157)= 0.221, p=.639, partial η2 = 0.001. This indicates that pre-exposure did not moderate the effect of voice type on perceived voice effectiveness.

Although the interaction between voice type and pre-exposure was not significant (p=.639), the simple main effect of pre-exposure, specifically with an AI voice on voice

effectiveness ratings in comparison with no-pre-exposure with the AI voice was tested. There was no significant effect of pre-exposure on perceived voice effectiveness within the AI voice condition, $F(1, 157) = 2.30$, $p = .132$, $\eta^2p = .014$. Participants who were pre-exposed to the AI voice ($M= 5.16$, $SE=0.96$) did not significantly differ in perceived voice effectiveness compared to participants who were not pre-exposed to the AI voice ($M= 4.84$, $SE= 0.75$). The estimated difference was $b = 0.32$, $SE = 0.21$, 95% CI [−0.10, 0.73], $t(157) = 1.52$, $p = .132$. Therefore H3b is not supported.

To conclude, the results provide partial support for the hypotheses. Participants perceived human voices as more effective than AI voices, which supports H1b. However, the analysis does not provide evidence to support H3b (pre-exposure to an AI voice increases perceived voice effectiveness). Although the mean perceived effectiveness was slightly higher for pre-exposed participants, this difference was not statistically significant. Furthermore, the lack of a significant interaction effect indicates that pre-exposure does not moderate the influence of voice type on perceived voice effectiveness.

**Mental effort**

A Factorial ANOVA was used to test hypotheses H2 and H4. There is no main effect of voice type: $F(1, 157) = 3.44$, $p =.066$, partial $\eta2 =0.021$. Mental effort was marginally higher for AI voices ($M=3.31$, $SE=0.19$) than for human voices ($M=2.82$, $SE=0.18$). While the trend aligns with H2, the effect was not statistically significant. There is no main effect of pre-exposure: $F(1, 157)= 0.00$, $p=.965$, partial $\eta2 =0.00$. Mental effort ratings were similar for participants with pre-exposure ($M=3.07$, $SE=0.19$) and without pre-exposure ($M=3.06$, $SE=0.18$). Finally, there was no significant interaction effect found between voice type and pre-exposure: $F(1, 157)= 0.360$, $p=.549$, partial $\eta2 = 0.002$. Pre-exposure did not moderate the relationship between voice type and mental effort.

Although the interaction between voice type and pre-exposure was not significant (p=.549), the simple main effect of pre-exposure, specifically with an AI voice on mental effort ratings in comparison with no-pre-exposure with the AI voice was tested. There was no significant effect of pre-exposure on perceived voice effectiveness within the AI voice condition, $F(1, 157) = 0.15$, $p = .703$, $\eta^2p = .001$. Participants who were pre-exposed to the AI voice (M= 3.24, SE= 1.54) did not report significantly lower mental effort compared to participants who were not pre-exposed to the AI voice (M= 3.38, SE= 1.75). The estimated difference was $b = -0.15$, SE = 0.38, 95% CI [−0.90, 0.61], $t(157) = -0.38$, $p = .703$. Therefore H4 is not supported.

To conclude, the results do not support either hypotheses. Although the simple effects plot showed a trend that could visually support the direction H2 and H4, this effect was not statistically significant. Although mental effort tended to be higher for AI voices compared to human voices, this effect was not statistically significant, failing to support H2. The simple main effect analysis did not provide sufficient evidence to support H4. Specifically, within the AI voice condition, participants who were pre-exposed to the AI voice (M = 3.24) did not report significantly lower mental effort compared to those who were not pre-exposed (M = 3.38). Although the mean mental effort was slightly lower for pre-exposed participants, this difference was not statistically significant. Furthermore, the lack of a significant interaction effect indicates that pre-exposure does not moderate the influence of voice type on mental effort.

**Mediation by mental effort**

A mediation analysis was conducted to test whether mental effort mediated the relationship between pre-exposure and trust in the advertisement (H5a). Non-parametric bootstrapping with 5000 samples and a 95% confidence level was used. This approach

ensures that you get more accurate and robust estimates of the indirect effect (mediation) without relying on restrictive assumptions about the data distribution.

The results showed no significant indirect effect, b = 0.00, SE = 0.02, 95% CI [−0.03, 0.03], z = 0.03, p = 0.977. Additionally, the direct effect of pre-exposure on trust in the advertisement was not significant, b=0.05, SE=0.14, z=0.38, p=.703. These findings indicate that mental effort does not mediate the relationship between pre-exposure and trust in the advertisement and therefore failed to support H5a.

A mediation analysis tested whether mental effort mediated the effect of pre-exposure on voice effectiveness (H5b). While mental effort was a significant predictor of perceived voice effectiveness (b=-0.09, p=.043), the indirect effect of pre-exposure on voice effectiveness through mental effort was not significant (b=0.00, 95% CI [-0.04, 0.04], p=.977). The direct effect of pre-exposure on voice effectiveness approached significance (b=0.26, p=.064). These results suggest that mental effort does not mediate the relationship between pre-exposure and perceived voice effectiveness in this sample and therefore fail to support H5b.

**Table 7**

*Tested hypotheses*

| | | |
|---|---|---|
| H1a | Participants exposed to a human voice, will show higher levels of trust in the advertisement compared to participants exposed to an AI voice | Supported |
| H1b | Participants exposed to a human voice, will show higher levels of perceived voice effectiveness compared to participants exposed to an AI voice | Supported |
| H2 | Participants exposed to a human voice, will show lower levels of mental effort compared to participants exposed to an AI voice | Unsupported |
| H3a | Participants pre-exposed to an AI voice, will show higher levels of trust in the advertisement, compared to participants who did not get pre-exposed with the AI voice | Unsupported |

| | | |
|---|---|---|
| H3b | Participants pre-exposed to an AI voice, will show higher levels perceived voice effectiveness, compared to participants who did not get pre-exposed with the AI voice | Unsupported |
| H4 | Participants pre-exposed to an AI voice, will show lower levels of mental effort, compared to participants who did not get pre-exposed with the AI voice | Unsupported |
| H5a | Mental effort mediates the effect of pre-exposure on trust in the advertisement | Unsupported |
| H5b | Mental effort mediates the effect of pre-exposure on perceived voice effectiveness | Unsupported |

## Discussion

This study investigated the effects of voice type and pre-exposure on trust in the advertisement, perceived voice effectiveness and mental effort in the context of an audio advertisement. Lastly, it looked at the mediating role of mental effort in this relationship. An overview of the tested hypotheses including outcomes can be found in table 7.

### Trust in the advertisement

Fostering consumer trust in AI practices, such as AI-generated speech in advertisements, is essential, as it promotes consumer engagement, and enhances brand loyalty (Chen et al., 2022). As AI speech technologies are increasingly popular in advertising contexts, the first hypothesis in this study examined the difference in levels of trust in the advertisement between human and an AI voiceover. The results showed that human voices elicited significantly higher trust in the advertisement than AI voices, supporting H1a. This finding is consistent with prior research emphasizing that human voices are generally more trusted than artificial ones. Robotic-sounding AI voices can undermine credibility and reduce trust in the information provided (Torre & White, 2020b; Schreibelmayr & Mara, 2022b ; Audacy Inc., 2024).

This outcome aligns with Processing Fluency Theory (Reber, 2012), which suggests that people find familiar stimuli easier to process. Participants likely perceived the human voice as more trustworthy due to its familiarity. In contrast, artificial voices that deviate from this familiarity are often seen as less credible, undermining trust in the information they convey. This distrust can diminish confidence in the advertisement and the company, ultimately challenging brand loyalty and reducing consumer engagement.

It is evident that voice quality plays a crucial role in influencing trust, as highlighted in both this study and Chiou et al.’s (2020) study. Chiou's study demonstrated that higher-quality artificial voices are more trusted than lower-quality artificial voices, underscoring the

importance of enhancing artificial voice quality. Our study extends this understanding by showing that human voices are generally more trusted than artificial voices, regardless of their quality. These complementary findings suggest that while improving artificial voice quality is essential for building trust, the inherent trust associated with human voices cannot be overlooked. Therefore, when implementing AI-generated speech in advertisements focusing on the quality of artificial voices is a critical consideration for maximizing trust.

**Perceived voice effectiveness**

Given the fact that the effectiveness and quality of a voice influences listener's feelings and perceptions and consequently ad efficacy (Wiener & Chartrand, 2014), this study examined the difference in levels of perceived voice effectiveness between human and an AI voiceover. The findings revealed that participants perceived human voices as significantly more effective than AI voices, supporting H1b. This builds on previous studies that saw that voices in audio advertising are a crucial aspect in delivering a message, as they influence how people respond to advertisements (Gelinas-Chebat & Chebat, 1992).

The study's findings align with previous research indicating that the quality of a voice is crucial in delivering effective audio advertisements (Lu et al., 2024). A clear and engaging human voice can improve recall of the content and clarity of the message, as listeners are more likely to pay attention and retain information presented in a natural and expressive manner. AI voices, despite advancements, may still fall short in this area due to their relatively limited ability to convey emotion and emphasis. Favourable evaluations from listeners are more likely when the voiceover is of high quality (Dubey et al., 2018 ; Wiener & Chartrand, 2014). This is true for both human and AI voices; however, the inherent naturalness and expressiveness (rhythm, intonation, stress) of human voices are superior.

Mori's (2023) Uncanny Valley Theory provides further insights into this perceptual gap. AI voices that sound almost human, but not quite, can fall into the uncanny valley, especially when their prosodic features seem slightly off. Such deviations can make the voice sound robotic or awkward, contributing to a sense of unease and reducing perceived effectiveness (Staff, 2024). Research on the audio uncanny valley has shown that manipulating prosodic factors can improve the naturalness of artificial voices, making them more relatable and effective (Wang et al., 2024; Hirose, 2019; Rodero, 2017).

However, despite these advancements, participants in the study likely found human voices more natural and relatable, leading to higher perceived effectiveness. This outcome highlights the importance of naturalness in voice effectiveness (Duville et al., 2022; Rodero & Lucas, 2021). While high-quality AI voices can still be effective (Chiou et al., 2020), there remains a perceptual gap that they need to overcome to match the effectiveness of human voices.

*Effects of pre-exposure*

The lack of a significant effects of pre-exposure on trust in the advertisement (H3a unsupported) and perceived voice effectiveness (H3b unsupported) indicates that merely introducing participants to an AI voice prior to the main advertisement does not substantially enhance voice evaluations. These findings align with the Processing Fluency Theory (Reber, 2012), which suggests that the ease of processing information influences preferences and judgments. Human voices, being inherently more familiar, require fewer cognitive resources to process compared to artificial voices (Ingvalson & Stoimenoff, 2021; McKenzie et al., 2020), leading to higher trust and perceived effectiveness. Despite the Mere Exposure Effect (Zajonc, 1968), which posits that repeated exposure should enhance positive evaluations, this study's results indicate that familiarity alone may not be sufficient.

While studies like Nickerson (2023) suggest that repeated exposure can increase liking and acceptance of AI voices, the brief pre-exposure used in this study was insufficient to achieve this outcome. Moreover, the lack of a significant interaction effect between voice type and pre-exposure suggests that the perceived effectiveness of AI voices remains lower than that of human voices, regardless of prior exposure. This aligns with findings by Lu et al. (2024), who observed that AI voices consistently lag behind human voices in terms of effectiveness.

Prior research (Weisman & Peña, 2021; Holmes et al., 2021) has demonstrated that even brief exposure to new stimuli can improve familiarity and reduce discomfort. Although 65% of participants indicated that they did not feel it was necessary to extend the duration of the pre-exposure to become accustomed to the voice, the lack of a significant effect of pre-exposure in this study suggests that the duration and context of exposure may be critical factors. The brief pre-exposure in this study may not have been sufficient to overcome the inherent challenges of synthetic voices.

Also, it was found that participants reported relatively high familiarity with AI (M=3.08) and openness to technology (M=3.29), referring to already formed opinions about artificial voices which diminishes the impact of the pre-exposure. Moreover, the context of pre-exposure (listening to an introduction) might not have aligned with the context of the subsequent radio commercial evaluation. If the voice was perceived differently in these contexts, the effect of pre-exposure may have been mitigated.

Additionally, the cognitive demands of processing an artificial voice may still limit trust and perceived effectiveness, highlighting the importance of both familiarity and processing fluency. This focus on artificial voices addresses a critical gap in the literature, reflecting the increasing use of AI-generated voices in advertising contexts (Weitzman, 2023;

Amit, 2024), and contributes to a deeper understanding of how familiarity influences perceptions of AI-generated speech.

**Mental effort**

The hypothesis that human voices would result in lower mental effort compared to AI voices (H2) was not supported. While the trend suggested that AI voices required slightly more cognitive effort, the difference was not statistically significant. This partially aligns with research suggesting that synthetic voices demand greater cognitive resources due to their robotic intonation and occasional irregularities (Rodero, 2017; Schiller et al., 2023). However, advancements in AI voice technology may have mitigated these effects, resulting in comparable mental effort ratings for both human and AI voices. This finding contrasts with Cognitive Load Theory (Sweller, 1988), which posits that stimuli requiring additional cognitive processing should result in higher mental effort.

Similarly, pre-exposure did not significantly reduce mental effort for participants exposed to AI voices (H4 unsupported). While studies suggest that familiarity can ease cognitive load (Weisman & Peña, 2021), the brief pre-exposure used in this study may not have provided sufficient time for participants to adapt to the AI voice. The lack of significant effects of pre-exposure on any of the dependent variables indicates that brief familiarization with synthetic voices is insufficient to overcome their inherent limitations. While pre-exposure has been shown to enhance familiarity and reduce cognitive barriers in other contexts, its impact may be constrained in the context of AI voices. A possible explanation for this might be that if participants were not actively engaging with the content during the pre-exposure phase, the Mere Exposure effect might not have been triggered effectively. Prior studies have shown that a lack of attention to the stimulus can diminish the impact of the Mere Exposure effect (Yagi et al., 2009 ; Huang & Hsieh, 2013).

Moreover, the absence of a significant interaction effect between voice type and pre-exposure suggests that intrinsic qualities of the voice, such as clarity and naturalness, remain more influential than familiarity in determining cognitive effort. These findings underscore the importance of voice quality in reducing cognitive demands. Wang et al. (2024) highlighted that factors such as subtitles may play a greater role in easing mental effort than pre-exposure alone. This suggests that advertisers aiming to use AI voices should focus on enhancing the intrinsic naturalness and comprehensibility of the voice to minimize cognitive barriers, potentially by incorporating contextual aids like subtitles.

**Mediation by mental effort**

The mediation analysis revealed no significant indirect effects of mental effort on trust in advertisements (H5a unsupported) or perceived voice effectiveness (H5b unsupported). These findings suggest that mental effort does not play a mediating role in the relationships between pre-exposure and the dependent variables. While mental effort was a significant predictor of perceived voice effectiveness, the indirect effect of pre-exposure on effectiveness through mental effort was not significant. This indicates that while mental effort influences voice evaluations, pre-exposure does not sufficiently impact mental effort to mediate these relationships.

These results challenge prior theoretical frameworks, such as Processing Fluency Theory (Reber, 2012), proposing easier-to-process stimuli are more likely to be trusted and positively evaluated. This implies that reduced cognitive load enhances evaluative judgments. Although processing fluency may contribute to trust and effectiveness ratings, the brief pre-exposure in this study did not significantly change mental effort levels to facilitate such effects. A possible explanation for this outcome is that pre-exposure might not have sufficiently increased processing fluency to influence subsequent evaluations. Additionally,

the participants may not have perceived the AI voice as cognitively demanding even without pre-exposure, making it challenging to observe significant difference.

**Implications**

Firstly, the study breaks through the limitations of existing research on AI voices in the advertising context. Prior literature has predominantly focused on advertisement evaluations rather than assessing the AI voice specifically, lacking crucial insights of consumer's perceptions on artificial voices (Wang et al., 2024). Also limited attention has been given on the influence of familiarization of a voice and whether this has an impact on voice evaluations (Lauretta et al., 1990 ; Gong, 2023). Moreover, this study contributes to the growing body of literature on the use of AI voices in advertising, addressing gaps in understanding how consumers perceive and process artificial voices compared to human voices. The findings suggest that the Mere Exposure Effect and Processing Fluency Theory may not universally apply to AI voices in advertising, challenging the assumption that repeated exposure will automatically increase trust and perceived effectiveness. This calls for a refinement of the theoretical model, incorporating more nuanced factors like voice quality or gender and context or duration of exposure.

From a practical standpoint, the findings suggest that human voices remain more effective to use in advertising. Advertisers considering AI voiceovers should focus on the quality of AI voices, by enhancing naturalness and human-likeness, to bridge the gap in trust and perceived effectiveness. The study's findings on the limited impact of pre-exposure suggest that brief exposure may not be enough to reduce cognitive load for AI voices. Researchers could explore longer or repeated exposure strategies to familiarize audiences with AI voices, potentially improving their effectiveness over time.

**Limitations and future studies**

A limitation of this study that needs consideration is the fact that mental effort was only measured subjectively. Self-report may cause participants responses being biased due to misinterpretation, self-over- or self-under-estimation. This affects the validity of the results. In the future, researchers should minimise these biases by using additional objective measurements to test pre-exposure effects on mental effort, like performed in the Wang et al. (2024) study, using eye-tracking and dual-task performance.

Another limitation of this study is that it did not directly examine the quality of the AI voice used, despite evidence suggesting that higher-quality AI voices, which are clearer, more natural, and human-like, significantly enhance trust and perceived effectiveness (Chiou et al., 2020 ; Lu et al., 2024). The chosen AI voice may have lacked essential features such as natural intonation, emotional expressiveness, and seamless pronunciation, which are known to make synthetic voices more relatable and engaging to listeners (Wang et al., 2024; Hirose, 2019 ; Kühne et al., 2020 ; Rodero, 2017, 2021). This potentially limited the generalizability of the findings to more advanced synthetic speech technologies. Future research should incorporate quality assessments of AI voices to better understand their impact on voice evaluations. This would provide a more nuanced view of how voice quality interacts with consumer perceptions and cognitive effort.

While the study's sample size was determined through a power analysis (statistical power of 0.80, effect size of 0.25, and $\alpha = 0.05$), ensuring sufficient power for detecting medium-sized effects, it is important to note that subtler effects might not have been detected. For instance, the hypothesized effect of pre-exposure on mental effort, which showed a trend in the expected direction but did not reach statistical significance, might benefit from a larger sample size to clarify these relationships. Future research could increase the sample size to improve sensitivity to subtle effects, especially in exploratory analyses. Nonetheless, the

current sample size was adequate for the primary research aims, supporting reliable conclusions about the main effects.

Additionally, this study did not directly assess participants' emotional reactions to AI voices, leaving the role of uncanny valley effects speculative. Future studies should include measures for affective responses to AI voices, such as eeriness or discomfort, to directly evaluate uncanny valley effects.

This study focused on Dutch-speaking participants, which may limit the generalizability of findings across different languages and cultures. Notably, there has been limited research specifically investigating how cultural differences influence perceptions of AI and human voices, particularly in the context of AI voice processing. Future research could address this gap by exploring how these perceptions vary across different languages and cultural settings.

Lastly, since the majority of respondents in the current study (43 out of 67) recognized that they were listening to an AI voice, it would be wise for companies to be transparent about this from the beginning. Research has shown that consumer trust can increase by implementing AI disclosure (Kirkby et al., 2023, Yahoo Inc., 2024). This study did not disclose the AI nature of the voice until after participants completed the survey, to avoid biasing participant’s responses. Future research could explore whether prior disclosure of the AI voice impacts voice evaluations, providing valuable insights into the role of transparency.

## Conclusion

Overall, the results imply that the preference for human voices in advertising remains superior, even when attempts are made to familiarize consumers to the artificial voice. Human voices consistently outperformed AI voices in terms of trust and effectiveness, underscoring the continued preference for natural speech in advertising contexts. This study advances our

understanding of consumer perceptions of AI-generated voices in advertising, emphasizing the critical role of voice quality and familiarity. While there is still much to learn, the findings of this study contribute to a deeper understanding of AI voice perception in the advertising context, and open exciting opportunities for future research and innovation in this rapidly evolving area.

## References

Abdulrahman, A., & Richards, D. (2022). Is Natural Necessary? Human Voice versus Synthetic Voice for Intelligent Virtual Agents. *Multimodal Technologies And Interaction*, *6*(7), 51. https://doi.org/10.3390/mti6070051

Ameen, M. S., Heib, D. P., Blume, C., & Schabus, M. (2022). The Brain Selectively Tunes to Unfamiliar Voices during Sleep. *Journal Of Neuroscience*, *42*(9), 1791–1803. https://doi.org/10.1523/jneurosci.2524-20.2021

Amit, J. (2024, September 26). *Speech AI: A Guide to the Technology's Applications, Challenges, and Trends*. aiOla. https://aiola.com/blog/speech-ai/

Audacy Inc,. (2024, July 22). *Audio: A Beacon of Trust in the Age of AI*. Audacy Inc. https://audacyinc.com/insights/audio-a-beacon-of-trust-in-the-age-of-ai/

Balaratnam, R. (2024, June 18). Turning potential to profit: building consumer trust in AI. *www.cognizant.com*. https://www.cognizant.com/us/en/insights/insights-blog/building-consumer-trust-in-ai-wf2729750

Campbell C., Plangger K., Sands S., Kietzmann J., Bates K. (2022). How deepfakes and artificial intelligence could reshape the advertising industry: The coming reality of AI fakes and their potential impact on consumer behavior. *Journal of Advertising Research*, 62(3), 241–251.

Chen, H., Chan-Olmsted, S., Kim, J., & Sanabria, I. M. (2021). Consumers' perception on artificial intelligence applications in marketing communication. *Qualitative Market Research An International Journal*, *25*(1), 125–142. https://doi.org/10.1108/qmr-03-2021-0040

Chen, Y., Prentice, C., Weaven, S., & Hisao, A. (2022). The influence of customer trust and artificial intelligence on customer engagement and loyalty – The case of the home-

sharing industry. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.912339

Chiou, E. K., Schroeder, N. L., & Craig, S. D. (2020). How we trust, perceive, and learn from virtual humans: The influence of voice quality. *Computers and Education*, *146*, Article 103756. https://doi.org/10.1016/j.compedu.2019.103756

Colomé, J. P. (2023, April 5). Another job lost to AI: 'The human factor may soon disappear'. *EL PAÍS English*. https://english.elpais.com/science-tech/2023-04-05/another-job-lost-to-ai-the-human-factor-may-soon-disappear.html

Craig, L. (2023, November 8). AI Voice Over & Its Impact on Voice Over Industry. *Acadecraft*. https://www.acadecraft.com/blog/impact-of-ai-on-voice-over-industry/

Davenport, T., Guha, A., Grewal, D., & Bressgott, T. (2019). How artificial intelligence will change the future of marketing. *Journal Of The Academy Of Marketing Science*, *48*(1), 24–42. https://doi.org/10.1007/s11747-019-00696-0

Diel, A., & Lewis, M. (2023). The vocal uncanny valley: Deviation from typical organic voices best explains uncanniness. *Research Square (Research Square)*. https://doi.org/10.21203/rs.3.rs-2784067/v1

Dubey, M., Farrell, J., & Ang, L. (2018). How Accent and Pitch Affect Persuasiveness in Radio Advertising. In *Springer eBooks* (pp. 117–130). https://doi.org/10.1007/978-3-658-22681-7_9

Duville, M. M., Alonso-Valerdi, L. M., & Ibarra-Zarate, D. I. (2022). Neuronal and behavioral affective perceptions of human and naturalness-reduced emotional prosodies. *Frontiers in Computational Neuroscience*, *16*. https://doi.org/10.3389/fncom.2022.1022787

Gao, B., Wang, Y., Xie, H., Hu, Y., & Hu, Y. (2023). Artificial Intelligence in Advertising: Advancements, Challenges, and Ethical Considerations in Targeting, Personalization,

Content Creation, and Ad Optimization. *SAGE Open*, *13*(4). https://doi.org/10.1177/21582440231210759

Gelinas-Chebat, C., & Chebat, J. (1992). Effects of Two Voice Characteristics on the Attitudes toward Advertising Messages. *The Journal Of Social Psychology*, *132*(4), 447–459. https://doi.org/10.1080/00224545.1992.9924724

Gélinas-Chebat, C., Chebat, J., & Vaninsky, A. (1996). Voice and Advertising: Effects of Intonation and Intensity of Voice on Source Credibility, Attitudes toward the Advertised Service and the Intent to Buy. *Perceptual And Motor Skills*, *83*(1), 243–262. https://doi.org/10.2466/pms.1996.83.1.243

Gong, C. (2023). AI voices reduce cognitive activity? A psychophysiological study of the media effect of AI and human newscasts in Chinese journalism. *Frontiers in Psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1243078

Heald, S. L. M., & Nusbaum, H. C. (2014). Speech perception as an active cognitive process. *Frontiers in Systems Neuroscience*, *8*. https://doi.org/10.3389/fnsys.2014.00035

Herbert, C. & Pisoni, D. (2023, February 1). *Information Processing and Cognitive Audiology*. The Hearing Review. https://hearingreview.com/inside-hearing/research/information-processing-and-cognitive-audiology

Huang, Y., & Hsieh, P. (2013). The mere exposure effect is modulated by selective attention but not visual awareness. *Vision Research*, *91*, 56–61. https://doi.org/10.1016/j.visres.2013.07.017

Ingvalson, E. & Stoimenoff, T. (2021). Florida State University, & Northwestern University. Greater Benefit For Familiar Talkers Under Cognitive Load [Journal-article]. https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0177.pdf

Kaat, J. (2024, April 12). *Verdwijnt het werk van stemacteurs door AI?* [Radio broadcast]. NPO Radio 1. Verdwijnt het werk van stemacteurs door AI? | NPO Radio 1

Keim, K. (2024, May 4). *What is text to speech (TTS) and how it works with AI? - EITCA Academy*. EITCA Academy. https://eitca.org/artificial-intelligence/eitc-ai-gcml-google-cloud-machine-learning/introduction/what-is-machine-learning/what-is-text-to-speech-tts-and-how-it-works-with-ai/

Kendall, E. (2024, September 30). *Uncanny valley | Definition, Origin, & Facts*. Encyclopedia Britannica. https://www.britannica.com/topic/uncanny-valley

Kirkby, A., Baumgarth, C., & Henseler, J. (2023). To disclose or not disclose, is no longer the question – effect of AI-disclosed brand voice on brand authenticity and attitude. *Journal Of Product & Brand Management*, *32*(7), 1108–1122. https://doi.org/10.1108/jpbm-02-2022-3864

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*. https://doi.org/10.3389/fnbot.2020.593732

Kumar, V., Rajan, B., Venkatesan, R., & Lecinski, J. (2019). Understanding the Role of Artificial Intelligence in Personalized Engagement Marketing. *California Management Review*, *61*(4), 135–155. https://doi.org/10.1177/0008125619859317

Landwehr, J. R., Golla, B., & Reber, R. (2017). Processing fluency: An inevitable side effect of evaluative conditioning. *Journal of Experimental Social Psychology, 70,* 124–128. https://doi.org/10.1016/j.jesp.2017.01.004

Lang, A., Kurita, S., Gao, Y., & Rubenking, B. (2013). Measuring Television Message Complexity as Available Processing Resources: Dimensions of Information and

Cognitive Load. *Media Psychology*, *16*(2), 129–153. https://doi.org/10.1080/15213269.2013.764707

Lauretta, D. J., Rodman, R. D., & Antin, J. F. (1990). The Effects of Familiarization on the Comprehension of Synthetic Speech in Telephone Communication. *Proceedings Of The Human Factors Society Annual Meeting*, *34*(3), 189–193. https://doi.org/10.1177/154193129003400302

Leif, E. (2024, March 12). *AI vs Human Voiceovers – Pros and Cons*. Video Production Washington DC - MiniMatters. https://www.minimatters.com/ai-vs-human-voiceovers-pros-and-cons/#:~:text=Human%20Voiceover%20%E2%80%93%20Cons&text=There%27s%20a%20potential%20for%20extra,need%20for%20more%20administrative%20time.

Leong, W., Yuin-Y, C., Zulkifli, B. & Nie Suet, K. (2023). Perception of university students towards the use of artificial intelligence-generated voice in explainer videos. In *SEARCH Journal Of Media And Communication Research* (pp. 77–89) [Journal-article]. https://fslmjournals.taylors.edu.my/wp-content/uploads/SEARCH/SEARCH-2023-Special-Issue-SEARCH-Conf2022/SEARCH-2023-P6-15-SEARCHConf2022.pdf

Lovo.ai. (2024, July 18). *Using AI Generated Voices: Top 5 benefits*. https://lovo.ai/post/the-key-benefits-of-using-ai-generated-voices

Lu, S., Tran, H. T. T., & Ngo, T. (2024). Are we ready for artificial intelligence voice advertising? Comparing human and artificial intelligence voices in audio advertising in a multitasking context. *Quality & Quantity*. https://doi.org/10.1007/s11135-024-01967-x

Mabbott, J., Van Vlissingen, R. F., & Gillespie, N. (2023, February 22). Trust in Artificial Intelligence: Global Insights 2023. *KPMG*. https://kpmg.com/au/en/home/insights/2023/02/trust-in-ai-global-insights-2023.html

MarketResearch. (2023, August). *Text To Speech Market*. Text To Speech Market Size, Share, Growth and Forecast To 2032 (marketresearch.biz)

Martin, E. (2024, February, 9) *2024 State of AI in the Speech Technology Industry: AI Is Enabling Audiovisual Enhancements*. Speech Technology Magazine. https://www.speechtechmag.com/Articles/ReadArticle.aspx?ArticleID=162531

McKenzie, C., Hodgetts, W. E., Ostevik, A. V., & Cummine, J. (2020). Listen before you drive: the effect of voice familiarity on listening comprehension and driving performance. *International Journal Of Audiology*, *60*(8), 621–628. https://doi.org/10.1080/14992027.2020.1842522

Minczeles, C., Kasteleijn, N., Minczeles, C., & Kasteleijn, N. (2024, April 12). *Aldi vervangt stemacteur door AI in reclames, zullen meer bedrijven volgen?* NOS. https://nos.nl/artikel/2516494-aldi-vervangt-stemacteur-door-ai-in-reclames-zullen-meer-bedrijven-volgen

Mitchell, W. J., Szerszen, K. A., Lu, A. S., Schermerhorn, P. W., Scheutz, M., & MacDorman, K. F. (2011). A Mismatch in the Human Realism of Face and Voice Produces an Uncanny Valley. *i-Perception*, 2(1), 10–12. https://doi.org/10.1068/i0415

NOS. (2023, January 11). *Vliegbelasting omhoog, maar gaan we ook echt minder vliegen?* https://nos.nl/artikel/2459489-vliegbelasting-omhoog-maar-gaan-we-ook-echt-minder-vliegen

Ouwehand, K., Van Der Kroef, A., Wong, J., & Paas, F. (2021). Measuring Cognitive Load: Are There More Valid Alternatives to Likert Rating Scales? *Frontiers in Education*, *6*. https://doi.org/10.3389/feduc.2021.702616

Pols, G. (2023, October 27). Onrust onder Nederlandse stemacteurs: bedrijven vervangen mensen door AI. NOS. https://nos.nl/artikel/2495616-onrust-onder-nederlandse-stemacteurs-bedrijven-vervangen-mensen-door-ai

Rodero, E., Larrea, O., & Vázquez, M. (2012). Male and Female Voices in Commercials: Analysis of Effectiveness, Adequacy for the Product, Attention and Recall. *Sex Roles*, *68*(5–6), 349–362. https://doi.org/10.1007/s11199-012-0247-y

Rodero, E. (2017). Effectiveness, attention, and recall of human and artificial voices in an advertising story. Prosody influence and functions of voices. *Computers in Human Behavior*, *77*, 336–346. https://doi.org/10.1016/j.chb.2017.08.044

Rodero, E., & Lucas, I. (2021). Synthetic versus human voices in audiobooks: The human emotional intimacy effect. *New Media & Society*, *25*(7), 1746–1764. https://doi.org/10.1177/14614448211024142

Romportl, J. (2014). Speech Synthesis and Uncanny Valley. In: Sojka, P., Horák, A., Kopeček, I., Pala, K. (eds) Text, Speech and Dialogue. TSD 2014. Lecture Notes in Computer Science(), vol 8655. Springer, Cham. https://doi.org/10.1007/978-3-319-10816-2_72

Saumya. (2024, April 2). *Ethical concerns related to voice cloning & misuse of AI-generated Voiceovers*. Appy Pie. https://www.appypie.com/blog/ethical-concerns-ai-generated-voiceovers

Schiller, I. S., Aspöck, L., & Schlittmeier, S. J. (2023). The impact of a speaker's voice quality on auditory perception and cognition: a behavioral and subjective approach. *Frontiers in Psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1243249

Schirmer, A., & Gunter, T. C. (2017). Temporal signatures of processing voiceness and emotion in sound. *Social Cognitive And Affective Neuroscience*, *12*(6), 902–909. https://doi.org/10.1093/scan/nsx020

Schreibelmayr, S., & Mara, M. (2022b). Robot Voices in Daily Life: Vocal Human-Likeness and Application Context as Determinants of User Acceptance. *Frontiers in Psychology*, *13*. https://doi.org/10.3389/fpsyg.2022.787499

Sedivy, J. (2020). Mapping the Healthy Human Brain. *Language in Mind: An introduction to Psycholinguistics* (2nd ed., pp. 164-195). Oxford University Press

Shafiq Obeidat, M., Nazmy, P., & Rizvi. S. (2024). "The voice of trust and engagement: examining the impact of ai-generated content's tone on advertising success". *Academy of Marketing Studies Journal, 28*(5), 1-8.

Soh, H., Reid, L. N., & King, K. W. (2009). Measuring trust in advertising. *Journal Of Advertising*, *38*(2), 83–104. https://doi.org/10.2753/joa0091-3367380206

Staff, S. (2024, January 19). *The Uncanny Valley: AI's Struggle with Naturalness in Voices - SWAGGER Magazine*. SWAGGER Magazine. https://www.swaggermagazine.com/technology/the-uncanny-valley-ais-struggle-with-naturalness-in-voices/

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. Cognitive Science, 12(2), 257–285.

Sweller, J. (2011). Cognitive Load Theory. In Psychology of Learning and Motivation (Vol. 55, pp. 37–76). Elsevier. https://doi.org/10.1016/B978-0-12-387691-1.00002-8

Natural Reader (2024). *Free Text to Speech Online with Realistic AI Voices.* Free Text to Speech Online with Realistic AI Voices

Statista. (2023, January 6). *Global AI in marketing revenue 2028.* Statista. https://www.statista.com/statistics/1293758/ai-marketing-revenue-worldwide/

Taylor, J. (2024, June 29). Cheap AI voice clones may wipe out jobs of 5,000 Australian actors. *The Guardian*.

https://www.theguardian.com/technology/article/2024/jun/30/ai-clones-voice-acting-industry-impact-australia

Torre, I., & White, L. (2020b). Trust in Vocal Human–Robot Interaction: Implications for Robot Voice Design. In *Prosody, phonology and phonetics* (pp. 299–316). https://doi.org/10.1007/978-981-15-6627-1_16

Voicetune.ai. *(*2024, September 2). *The impact of Artificial Intelligence voices on marketing and advertising technologies*. https://www.voicetune.ai/post/the-impact-of-artificial-intelligence-voices-on-marketing-and-advertising-technologies

Wang, X., Zhang, Z., & Jiang, Q. (2024). The effectiveness of human vs. AI voice-over in short video advertisements: A cognitive load theory perspective. *Journal Of Retailing And Consumer Services*, *81*, 104005. https://doi.org/10.1016/j.jretconser.2024.104005

Weitzman, C. (2023, July 6). How AI voice overs are revolutionizing ad campaigns and boosting ROI. *Speechify.* https://speechify.com/blog/ai-voice-overs-are-boosting-roi-ad-campaigns/

Wiener, H. J. D., & Chartrand, T. L. (2014). The Effect of Voice Quality on Ad Efficacy. *Psychology And Marketing*, *31*(7), 509–517. https://doi.org/10.1002/mar.20712

Yagi, Y., Ikoma, S., & Kikuchi, T. (2009). Attentional modulation of the mere exposure effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*(6), 1403–1410. https://doi.org/10.1037/a0017396

Yahoo Inc. (2024, April 9). *Yahoo & Publicis Media Survey Reveals AI Ad Disclosure Increases Consumer Trust*. https://www.yahooinc.com/press/yahoo-publicis-media-survey-reveals-ai-ad-disclosure-increases-consumer-trust

# Appendix A

## Scripts

Dutch script of the introductory audio:

*"Hoi. Ik ben jouw voiceover vandaag. Ik spreek regelmatig commercials in voor op de radio of tv. Dit doe ik voor verschillende bedrijven, maar vandaag zul je mijn commercial voor luchtvaartmaatschappij SkyWings luisteren. Veel luisterplezier!"*

Dutch script of the radio advertisement:

*"Bij SkyWings draait reizen om meer dan alleen je bestemming… het start al bij de ervaring onderweg. Of je nu reist voor werk of ontspanning, wij bieden je comfort, service en een warm welkom aan boord. Boek nu jouw droombestemming voor 2025 bij SkyWings en ontvang onze vroeg boek-korting! SkyWings. Vlieg verder, beleef meer."*

## Appendix B

## Pre-test survey

Hallo!

Je bent uitgenodigd om deel te nemen aan mijn onderzoek. Mijn naam is Koosje Bogers en ik studeer momenteel de Master Communication & Information Sciences aan de Universiteit van Tilburg.

Voor mijn scriptie doe ik onderzoek naar voice-overs. Hiervoor heb ik jouw hulp nodig! Ik ben benieuwd naar jouw mening na het beluisteren van een korte radio commercial van een vliegtuigmaatschappij. Na het beluisteren van de commercial duurt het invullen van de enquête ongeveer 3 minuten. De enquête bevat **GELUID**, dus zet je koptelefoon op of zet het volume harder.

Jouw deelname is vrijwillig en blijft geheel anoniem. De data die via deze enquête wordt verzameld, zal alleen worden gebruikt voor dit onderzoek en zal naderhand worden verwijderd. Je zal me enorm helpen met je bijdrage, ontzettend bedankt!

Groetjes,

Koosje

Als je akkoord gaat met deelname aan dit onderzoek, klik dan op ‘doorgaan’

---

*Prior survey questions are identical to the final survey questions described in appendix B.*

---

Ten slotte ontvang ik graag jouw feedback op deze enquete.

Waren de instructies duidelijk?

Ja
Nee, … *

Ervaarde je enige problemen met het afspelen van de audio?

Ja, … *
Nee

Was de audio goed verstaanbaar?

Ja
Nee, … *

Je hoorde een vrouwelijke stem in de audio. In hoeverre ben je het eens met de volgende stelling:

"Als er een mannelijke stem was gebruikt, had dit mijn mening over de radio commercial veranderd"

Zeer oneens
Oneens
Neutraal
Eens
Zeer eens

Was er een begrip of term wat niet goed genoeg was uitgelegd waardoor je de enquête niet goed hebt kunnen invullen?

Ja, … *
Nee

De audio die je hoorde was een AI-stem. Had jij dit meteen door?

Ja
Nee
Na een tijdje

*Note:* only applicable for condition groups 2 and 4

De stem die je hoorde was gecreëerd met artificial intelligence (AI). In hoeverre ben je het eens met de volgende stelling:

"Als er een menselijke stem was gebruikt, had dit mijn mening over de radio commercial veranderd."

Zeer oneens
Oneens
Neutraal
Eens
Zeer eens

*Note*: only applicable for condition groups 2 and 4

De introductie audio die je op het begin te horen kreeg duurde ongeveer 20 seconden. In hoeverre ben je het eens met de volgende stelling:

"Ik had langer de stem willen horen om eraan gewend te raken"

Zeer oneens
Oneens
Neutraal
Eens
Zeer eens

*Note:* only applicable for condition groups 3 and 4

## Appendix C

## Survey

Hallo!

Je bent uitgenodigd om deel te nemen aan mijn onderzoek. Mijn naam is Koosje Bogers en ik studeer momenteel de Master Communication & Information Sciences aan de Universiteit van Tilburg.

Voor mijn scriptie doe ik onderzoek naar voice-overs. Hiervoor heb ik jouw hulp nodig! Ik ben benieuwd naar jouw mening na het beluisteren van een korte radio commercial van een vliegtuigmaatschappij. Na het beluisteren van de commercial duurt het invullen van de enquête ongeveer 3 minuten. De enquête bevat **GELUID**, dus zet je koptelefoon op of zet het volume harder.

Jouw deelname is vrijwillig en blijft geheel anoniem. De data die via deze enquête wordt verzameld, zal alleen worden gebruikt voor dit onderzoek en zal naderhand worden verwijderd. Je zal me enorm helpen met je bijdrage, ontzettend bedankt!

Groetjes,

Koosje

Als je akkoord gaat met deelname aan dit onderzoek, klik dan op ‘doorgaan’

---

Geef aan in hoeverre je het eens bent met de volgende stellingen

| | |
|---|---|
| Ik sta open voor nieuwe technologieën | Zeer oneens<br>Oneens<br>Neutraal<br>Eens<br>Zeer eens |
| Ik ben bekend met artificial intelligence | |
| Ik heb ervaring met artificial intelligence | |

---

*Additional part for condition groups 2 and 4:*

Om te starten ga je eerst een korte introductie van jouw voice-over beluisteren. Als je klaar bent met luisteren, klik op doorgaan.

*All condition groups:*

Je zult nu een radio commercial horen over een vliegtuigmaatschappij. Als je de audio heb geluisterd, klik op doorgaan.

De volgende vragen zullen gaan over de radio commercial die je zojuist hebt geluisterd.

Informatie in de radio commercial is…

| | Zeer oneens | Oneens | Enigszins oneens | Neutraal | Enigszins eens | Eens | Zeer eens |
|---|---|---|---|---|---|---|---|
| Eerlijk | | | | | | | |
| Waarheidsgetrouw | | | | | | | |
| Geloofwaardig | | | | | | | |
| Betrouwbaar | | | | | | | |
| Nauwkeurig | | | | | | | |
| Feitelijk | | | | | | | |
| Volledig | | | | | | | |
| Duidelijk | | | | | | | |
| Sympathiek | | | | | | | |
| Aangenaam | | | | | | | |
| Positief | | | | | | | |

Ik vond de stem in de radio commercial…

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
|---|---|---|---|---|---|---|---|---|
| Onduidelijk | | | | | | | | Duidelijk |
| Incorrect | | | | | | | | Correct |
| Onaangenaam | | | | | | | | Aangenaam |
| Niet geloofwaardig | | | | | | | | Geloofwaardig |
| Niet overtuigend | | | | | | | | Overtuigend |
| Onbegrijpelijk | | | | | | | | Begrijpelijk |

In hoeverre ben je het eens met de volgende stelling:

"Ik vond de stem in de radio commercial aangenaam om naar te luisteren"

Zeer oneens
Oneens
Neutraal
Eens
Zeer eens

Hoeveel mentale inspanning heb je gedaan om naar de advertentie te luisteren?

Mentale inspanning verwijst naar de hoeveelheid aandacht, concentratie en energie die je moet uitvoeren om naar de commercial te luisteren

Zeer, zeer lage mentale inspanning

Zeer lage mentale inspanning
Lage mentale inspanning
Vrij lage mentale inspanning
Noch lage noch hoge mentale inspanning
Vrij hoge mentale inspanning
Hoge mentale inspanning
Zeer hoge mentale inspanning
Zeer, zeer hoge mentale inspanning

*Additional questions for condition groups 2 and 4 (AI voice):*

De audio die je hoorde was een AI-stem. Had jij dit meteen door?

Ja
Nee
Na een tijdje

De stem die je hoorde was gecreëerd met artificial intelligence (AI). In hoeverre ben je het eens met de volgende stelling:

"Als er een menselijke stem was gebruikt, had dit mijn mening over de radio commercial positief veranderd."

Zeer oneens
Oneens
Neutraal
Eens
Zeer eens

*Additional question for condition group 4 (AI, with pre-exposure):*

Je hoorde vooraf een korte introductie van de AI stem. Vervolgens hoorde je deze zelfde AI stem in de radio commercial. In hoeverre ben je het eens met de volgende stelling:

"Ik had de AI stem in de introductie langer willen horen om er beter gewend aan te raken"

Zeer oneens
Oneens
Neutraal
Eens
Zeer eens

---

Wat is je leeftijd?

18-24
25-34
35-44

45-54
55-64
65+

Wat is je geslacht?

Man
Vrouw
Non-binair
Zeg ik liever niet

Wat is je hoogst afgeronde opleidingsniveau?

Geen opleiding afgerond
Basisonderwijs
Vmbo
Havo
Vwo
Mbo
Hbo-bachelor
Wo-bachelor
Hbo-master
Wo-master
Doctorale studie (Phd)

---

**Bedankt voor uw deelname!**
In dit onderzoek hebben deelnemers 2 verschillende voice-overs beluisterd (AI en een menselijke stem). Daarbij kreeg de helft de stem ook nog geïntroduceerd voorafgaand aan de commercial om aan deze "nieuwe" stem te kunnen wennen.

Dit experiment is bedoeld om te onderzoeken in hoeverre het type stem van een voice-over in advertenties invloed heeft op de mening van de luisteraar en in hoeverre een introductie van de AI stem voor gewenning zorgt en dit de mening van de luisteraar kan beïnvloeden.

Mocht u vragen hebben over mijn onderzoek, kunt u contact opnemen via k.v.bogers@tilburguniversity.edu

Delen van deze enquête in uw omgeving wordt ontzettend gewaardeerd :)