

AI-Narration Acceptance Among Frequent Audiobook Users: A Quantitative Study on Demographics and Personal Characteristics of Consumers

Bachelor's thesis in Retail Management
Academic year 2021-2023
Zulaika Molin
50579

Abstract

This comprehensive survey study examines consumer acceptance of AI-narration, along with demographic and individual characteristics that may influence the general acceptance of synthetic speech, including age, gender, importance of the narrator, previous technological experience, and frequency of audiobook use. The study also includes an experiment to evaluate general acceptance of AI-narration, which did not reveal any negative attitudes towards AI-narration, contrary to the initial hypotheses. Drawing insights from the analysis of open-ended and other responses, the research identifies a remarkably high level of general acceptance of AI speech. Factors contributing to this acceptance include the perceived importance of the narrator, male gender, and previous technological experience. This research provides valuable insights into consumer attitudes towards AI, aiding audiobook publishers and industry professionals in effectively implementing AI-narration and understanding the preferences of different consumer segments.

Table of content

1. Introduction	2
1.1 Background	2
1.2 Problem Statement	6
1.3 Purpose of The Research	8
1.4 Contributions	8
2. Theory on AI-narration	9
2.1 Consumer acceptance of AI	9
2.3 Humans interacting with robots	10
2.4 The role of the narrator	13
2.5 Demographics of AI-acceptance	14
2.6 Prior technological experience	14
2.7 Frequency of use	15
2.8 Areas of application	15
2.8 Summary of the hypothesis and research questions with respective literature	16
Section 3: Methodology	17
3.1 Research Design	17
3.2 Questionnaire Design	18
3.3 Experiment Design	18
3.4 Data collection	19
3.6 Reliability and validity of measurements	19
4. Analysis and Results Findings	22
4.1 Experiment: Attitudes, Likeness	22
4.2 Importance of the narrator	22
4.3 Demographic factors: age	23
4.4 Demographic factors: gender	24
4.5 Personal characteristics: previous technological experience	24
4.6 Personal characteristics: frequency of use	25
4.7 Analysis of the open-ended responses	26
4.8 Summary of the results	29
5. Discussion	30
5.1 Consumers acceptance of AI-narration	32
5.2 The role of demographic and personal characteristics in adopting AI-narration	33
5.3 Consumers perception of AI-narration: beliefs, fears and suitable fields of application	35
5.4 Managerial implications	35
5.5 Limitations	37
6. Conclusion	39
Appendix	43

1. Introduction

1.1 Background

According to De Carolis et al. (2021), audiobooks have evolved from tangible goods people purchase in stores to digital codes that people download to their smartphones. Historically, the audiobooks industry has not been as successful as the printed one, due to high production costs and misconceptions that for an audiobook to succeed it must generate the same sales as a print book and provide listeners with more entertainment than a print book could (Snelling, 2021). Targeted marketing is essential to a book's success because individuals frequently hesitate to change their habits, especially their reading manner (Dakic, 2019). Therefore, marketers in the audiobook's industry are actively targeting potential consumers in other audio-centered platforms, such as podcasts and music listeners. As a result, audiobooks are becoming more common, yet despite their expanding acceptance and profitable success, they are still not thought to be on balance with print novels (Snelling, 2022). Similarly to other industries, audiobooks' publishers turn to technological advancements to explore new ways of reducing production costs, effectivization of operations and attracting more customers. As high production costs of audiobooks could be tied to the high wages paid to narrators and long hours of production (Wright, 2020), publishers feel lured by an opportunity to substitute human narrators with a synthetic voice or AI-narration, as a way to improve profitability of audiobooks. However, little is known on the effects that this technology might have on consumers, due to the lack of knowledge on the audiobook's listeners' attitudes towards AI and synthetic speech (Zhang et al., 2021).

What do consumers of audiobooks think about non-human narrators, and how does the acceptance and propensity to adopt synthetic speech vary between different demographic groups of listeners? This paper explores consumers' attitudes and beliefs about non-human narrators and their susceptibility to synthetic speech in the audiobook format.

The audiobook market has been steadily expanding in recent years and is forecasted to grow by a compound annual growth rate of 24.4% between 2021-2027. This growth could be explained by changing consumer behavior, affordable subscription terms, and rapid technological advancements (Hanner and O'Connor, 2019). In addition, consumers want books in different formats to fit their needs. Thus, companies like Amazon, with its giant

audiobook service Audible, Storytel, and others, strive to match their offerings to these needs. The growth in the number of published audiobooks is also explained by the enormous amount of already published reading hard print and electronic books, which are required to be transferred into an audiobook format, creating the so-called "gap" between the number of audiobooks that get to the market and the number that could potentially be recorded (Thad McIlroy, 2021). For example, Google books have identified that 95% of their books in electronic reading format miss an accompanying audiobook. To tackle this issue and feed the growing demand, they have implemented a sophisticated auto-narrating program, which could narrate a whole book in less than 2 hours in 20 different male and female voices in English and other languages in progress (Abramov, 2021). Others like Storytel have launched their own AI machine Ivan the Robot, which is currently successfully operating in Russia (Erik Wikberg, interview 2021). Moreover, several other independent and highly technologically advanced providers and vendors of AI-narration are emerging in the market. All of this signals that the industry is responding to this unprecedented increase of demand in the market of audiobooks, and one of the ways to fill in this reading vs. listening formats gap, is to turn to auto-narration.

AI narration in the audiobook market is one of the hottest topics in the industry, with many weighing up the pros and cons and trying to reach a golden mean (Kuta, 2023). The advocates of AI-narrated books argue that the costs of production of an audiobook item are very high and time-consuming, with an average production time being about a couple of months long. In contrast, an AI program can narrate the same book in a few hours, costing about ten times cheaper than a human-narrated audiobook (Cecco, 2022). Moreover, technological advancement has enabled the so-called customization of a synthesized voice, where one can choose between over 250 varying tones of voice in 72 languages (Dubiel et al., 2020). Also, the outdated business model of audiobook production cannot catch up with the demands and needs of the market. CEO of one of the leading vendors of AI synthetic speech, Speechki, says that relying heavily on human narrators is the reason that 95% of the books do not get to the market: "This process is slow, expensive, and highly complex. Publishers' unit economy does not work with traditional methods of audiobook production." (Abramov, 2022).

Some in the publishing industry are also skeptical about the popularity of AI-narrated audiobooks. They fear structural changes, the loss of jobs for narrators and producers, and

that audiobooks will lose their authenticity and appeal to customers if non-humans tell them. As a representative of the Screen Actors Guild - American Federation of Television and Radio Artists put it, "audiobooks are human stories, and the incredible professionals who tell them often have a great sense of human storytelling." Amazon's audio streaming platform Audible has long avoided offering books with automated voices, and it is Amazon's policy not to allow artificial voice reproduction on its platform (although there is a universal voice synthesis feature on the Kindle). As Amazon has 50% of the audiobook market, producers and providers of synthesized voices see this policy as a huge barrier to the widespread adoption of human narration. However, some believe it is only a matter of time before Amazon follows suit and dozens of audio content providers, such as Apple, Google, and Storytel, adopt and actively promote AI-based narration. (Thad McIlroy, 2021).

While the shift towards AI-narrated speech, in the otherwise traditional book circles and industry, is very heated and segregated (Simonite, 2022), AI implementation in retail and business, in general, has met way less resistance (Mulqueen, 2017) and is welcomed as a useful complementation tool and solution to many problems. For a long time, there have been talks about the power of Big Data and the necessity to collect and protect this data for understanding customers' demands and preferences and as a driver of innovation (Eriksson and Reponen, 2021). Intel, a microprocessor giant, argues that collecting Big Data is herculean labor in itself for humans, which could otherwise be easily obtained by AI systems while processing this very data and integrating it to provide reasonable solutions is the real challenge that humans cannot fathom, and therefore leading retailers have unanimously implemented the AI in everyday tasks such as inventory management, planning and providing customers with personalized offerings: all of which reduces the waste and increases the efficiency of the businesses (Intel, 2022). However, retailers' use of AI has become more sensitive regarding direct human and robot interaction in retailing. Many studies have focused on human-nonhuman interaction, as chatbots in general, in fashion retail, brick-and-mortar stores, and medical care. The majority of these research papers are very fresh. Their multitude, focusing solely on the human-nonhuman interaction, signals a growing interest and need for retailers to understand how consumers perceive and react towards the novelty of AI substituting humans. Even as there are several papers delving more in-depth into the area of synthetic speech acceptance (a core subject of this paper), 1) none of them is based upon a recently updated AI synthetic speech, which differs greatly from the earlier versions; 2) and does not represent the attitudes of the consumers of the audiobooks

industry: which is a whole another group of consumers, differing from any other retail industry, such as fashion, brick, and mortar or medical care; and as previously stated is substantially growing worldwide. Moreover, NPD BookScan (a data-analytics service of retail book sales in the US) announced record-breaking book sales the previous year and argued that retail sales and the number of customers in the industry are expected to grow even further. Still, there would have to be a shift towards more digital formats, as the supply chain of physical books is limited by production capacity and disruptions. Therefore, this paper is unique in its ambition to explore the overlooked and growing market of audiobook customers and the specifics of their interaction with an updated AI-narrating program, which is likely to substitute human narrators fully or partially in the future.

1.2 Problem Statement

Whether we like it or not, AI-narrated books are rapidly entering the market, tempting publishers with various economic and time-saving advantages (Simonite, 2022). The question is no longer whether you need an AI narrative but where it fits best. Using AI narratives in text-rich non-fiction, such as education and student literature, has long been popular. Recent technological advances in speech synthesis have enabled AI voices to express a range of emotions, allowing for increased use of AI narratives in the non-fiction and fiction genres (Green et al., 2004). However, publishers are wary of publishing books with AI narratives because little is known about how consumers perceive and accept non-human narrators. Storytel, for example, has published dozens of fictional works with AI narratives. Still, the target audience for these narratives is limited to young teens to reduce the associated risks and uncertainty of audience response. Zhang et al. (2021) stated that this slow acceptance and promotion of AI narratives might be due to a lack of understanding of audience attitudes and expectations towards AI narratives, raising the question of which segments or audiences should be targeted by AI narrative content.

The mechanism that most AI-enabled audiobooks are built on takes their roots from more generic Text to Speech programs. A modern example is Apple's Siri and Google's Alexa synthesized voices that sound more natural and can take up various tones and emotions. The sophisticated text-to-speech synthesis systems described in this report and intended for use in the audiobook market have surpassed that of Siri and Alexa. Bradley Metrock, CEO of Project Voice, and Digital Book World, says, "With the current crop of high-end synthetic

voices, 95% of people would not recognize that they're artificially generated. In 12–24 months it will reach human levels.” In line with Masahiro Mori's term "Uncanny Valley," stating that humans might experience a feeling of disgust towards their non-human counterpart displaying highly human-like features; there is also a great deal of uncertainty if consumers of the audiobooks could potentially experience "eeriness" or other unpleasant feelings when listening to a perfectly human-like synthetic speech (Habib et al., 2022). Moreover, the fear of AI has long been a part of mainstream media reporting, with people like Stephen Hawking, Bill Gates, and Elon Musk warning consumers that the future misapplication and evolution of AI technology is detrimental to the development of human race, and fueling customer's minds with concerns, prejudices, and suspicions towards any form of non-human novelty (Schmelzer, 2019). If this was the case with audiobooks consumers, who would hypothetically at large resist or dislike the widespread implementation of AI, there is still a need for research on the reasons why consumers form negative attitudes towards AI-narration to overcome them; and which groups of people are more susceptible to synthetic speech than others.

1.3 Purpose of The Research

Although the debate about human versus non-human storytelling is widely discussed and considered in publishing circles, audiobook listeners are mainly unaware of the changes (McIlroy, 2021). Consumer perceptions and opinions on AI storytelling could help publishers allocate their human and non-human resources accordingly and target the right audience with the right (most appropriate) narrative format: a task that has been frustrating audio marketers for some time (Maughan, 2022). In other words, by increasing publishers' knowledge of listeners' beliefs and underlying characteristics when interacting with non-human narrators, audiobook content providers can make the right decisions to implement AI narratives and target audiences effectively.

Currently, consumer perception and acceptance of synthetic speech is a blind spot for which there is a great need for research and segmentation (Maughan, 2022). Previous research on synthetic voice preference and acceptance has been based on less intelligent and outdated narrators, such as the Alexa and IBM synthetic voice systems (Ilves and Surakka, 2013) mentioned above. Furthermore, there is no research on the average consumer of the audiobook market and their acceptance of non-human narrators. Therefore, this paper aims

to fill this gap in knowledge about human and non-human interactions in the audiobook context and provide managers with answers to questions, such as 1) whether listeners are receptive to the idea of AI-narration, 2) which groups are more likely to consume AI-narration, and 4) consumers perceived benefits, fears and areas of application for AI-narration.

1.4 Contributions

As the scope of this research is vast and encompasses various themes and aspects that are relevant today, it is intended that this paper will contribute to several theoretical and practical fields. Firstly, this paper might be a minor, but still a contribution to the broad field of research on the general human vs. non-human interaction. More specifically, the experiment conducted in this paper strives to confirm or reject Masahiro Mori's theory of the "Uncanny Valley." Secondly, no research has been conducted on synthetic speech acceptance using newer AI systems, which have already bypassed the more renowned and researched Alexa and Siri, and are still developing to excel in the upcoming year (Hoyer et al., 2020). Therefore, this paper might yield different or complementing results on synthetic speech acceptance. Thirdly, the main objective of this research paper is to deliver practical contributions to the audiobooks industry and provide retailers with concrete answers as to the overall susceptibility of audiobooks consumers to AI-narration, as well as genres and demographics, where non-human narrators could successfully substitute human narrators.

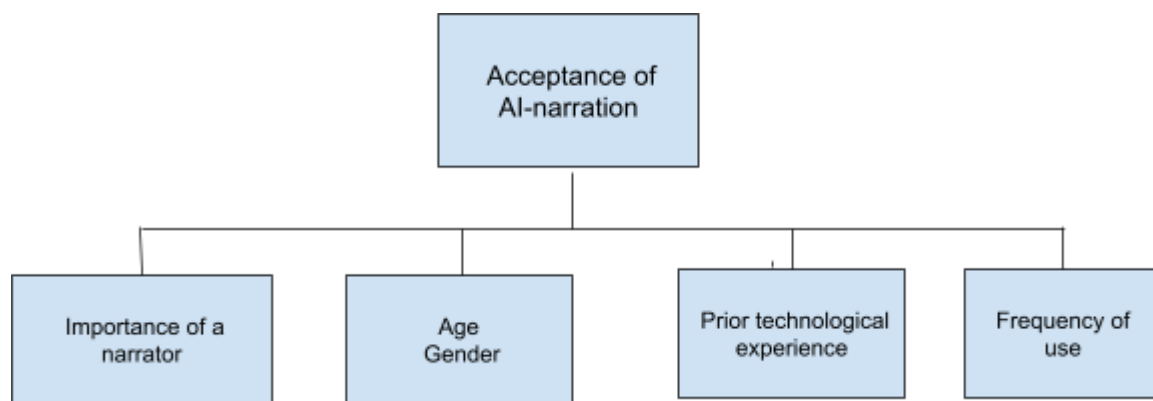
2. Theory on AI-narration

2.1 Consumer acceptance of AI

With the rapid technological advancements, AI takes an ever-growing place in modern humans lives, and thus is hard to avoid. Machine learning has developed beyond suggesting relevant ads and products to consumers: today AI can produce those products, create and innovate in the fields where humans have felt superior to machines. Retailers strive to outperform their competitors by creating a unique customer experience (Bäckström and Johansson, 2006) where AI may provide a more immersive experience, as in the examples of deep simulator rooms and interactive art exhibitions in museums. For instance, consumer interest in the AI-infused art has been growing in the past years, with reputable museums all across Europe exhibiting paintings made by robots with huge success (ArtMagazine, 2021). AI-infused collections are described by some as the future leaders of modern-art and consumers take it as seriously as a human-made art, with a portrait of Edmond de Belamy by an AI-artist Obvious been sold for staggering €432,000 in 2018. The creativity of AI machines is not bound only to painting, as robots can create poetry, music and even write fictional romances (Bidshahri, 2019). AI-artists in these fields are hugely popular and have their own broad fan-bases, with some consumers of such art describing their interest as “obsession” (Samuel, 2022). AI is not just creating art, but is also inspiring and enabling human artists to produce art in new ways, as the renowned artist Mario Klingemann says that AI allows him “to explore the creative potential of materials in ways that would be impossible without it” (Spratt, 2018). A Hungarian study of Nagy and Hajdu (2022), on the consumer’s acceptance of AI in online retail shopping, has found that trust is a crucial factor influencing the attitude towards non-humans. Museum workers share that there is interest in AI-work by consumers, who are interested in arts and technology, but to promote trust and acceptance of AI in broader audiences, there is a need for more informational events on the nature of AI and how it works (Canterbury AI, 2022). It is evident that today AI can be used not only for functional purposes, but for entertainment purposes too, with consumers' ever-growing interest and knowledge on the capabilities of AI. Narrating a book requires artistry, creativity and emotional connection to the content: if consumers' acceptance of AI in other entertainment and media sources is growing, could listeners of audiobooks accept non-human artist narrators?

2.2 Synthetic speech acceptance

The main objective of this study is to investigate the relationship between mainstream audiobook consumers and AI narratives and their propensity to consume such products. The main hypothesis is that most audiobook consumers have a negative attitude towards synthetic speech due to deep-seated psychological constraints and negative beliefs about AI, which prevent them from engaging with non-human counterparts who closely resemble humans (Jönsson, 2021). A theoretical framework below summarizes the factors that theoretically influence the adoption and acceptance of synthetic speech.



Below is a summary of various theories used to create this framework and hypothesize customers' attitudes toward AI narration.

2.3 Humans interacting with robots

As the previous section shows, we are dealing with a synthesized voice that is nearly perfect for humans and can become even more so a year from now. The main obstacle to its widespread adoption within the audiobook industry comes not from the publishers, who are admittedly ready to implement such a system and see its monetary and time-saving benefits, but from the consumers (Rodero and Lucas, 2021). Some theories suggest that humans are reluctant to accept any non-human counterpart with high or indistinguishable human traits. One of these theories is the theory of Speciesism, "a prejudice or attitude of bias in favor of the interests of members of one's species and against those of members of other species"

(Singer, 1990), widely used in biology and social sciences to explain the origins of racial, sexist and other prejudices. Bernd Schmitt, in his article: "Speciesism: an obstacle to AI and robot adoption", takes this notion further and illustrates that it's highly applicable in human-to-robot interaction context as artificial intelligence of robots becomes more and more advanced and, in some cases, their performance is indistinguishable or even better than that of a human. However, humans are afraid and reluctant to accept such an advanced robot, even if there's no difference between the level of service they receive from the AI and a human (Schmitt, 2020).

Another famous theory of the "uncanny valley" states that a robot's human likeness will increase. Still, at some point, the more human-like the robot is, the more likely humans will experience revulsion and eeriness towards this robot (Mori, 1970). Furthermore, most research suggests that humans prefer real humans over their non-human counterparts (Söderlund and Oikarinen, 2019) (Kuhne et al., 2020). Therefore, if we apply the previous research in human acceptance of human-like robots, we may hypothesize that listeners will prefer a human narrator over the AI-narration program, implying that humans might have negative attitudes towards overly human-looking or sounding robots.

When it comes to the area of research on synthetic speech acceptance, there have been several studies about human vs. non-human narrators: some of them are in line with the "uncanny valley" and "speciesism" theories, while others are yielding contrary results. For example, Rodero (2017) compared human vs. synthetic speech in their study on effectiveness, recall, and attention to the information presented in the advertisements, with concrete results that humans perceive real human speech better than artificial one on all tested aspects. However, even as humans perceive human speech better than synthetic one, the central area of application of synthesized voices is still bound to task-oriented activities, such as checking the weather, playing music, and controlling other devices (Lopatovska et al., 2018), rather than for social activities and companionship. This may lead us to think that consumers of audiobooks who use audiobooks for companionship might prefer real human narrators over synthetic ones. However, this may be because most voices available for general consumption are still artificial sounding ones.

Several studies researched exclusively synthetic speech, the results of which contradict Mori's and Schmitt's theories that humans would abstain from non-human species that resemble them too much and experience the so-called "eeriness." The research by

Schreibelmayer and Mara (2022) and Kuhne et al. (2020) looked at the levels of "humanism" that are acceptable by human listeners in the synthetic speech and came to the same conclusion that the more human-likeness of the synthetic speech led to more user-acceptance, and what is even more fascinating: the synthetic voices that were the most human-like scored highest on the pleasantness and least on the eeriness aspects, contrary to the "uncanny valley" theory. In line with Lopatovska et al. (2018), Schreibelmayer and Mara (2022) have concluded that synthetic speech is less accepted in social interactions (for companionships and entertainment) and more for information and navigation. However, the more human-like the voice sounded, the more likely users would accept it in social domains. Another study by Ilves and Surakka (2011) came to similar conclusions when comparing various synthetic voices. The most human and emotional-sounding ones were rated statistically significantly more natural, pleasant, and clear than the less human-sounding ones. These studies highlight that synthetic voices can be accepted by users without any feeling of eeriness, even in social domains: the more human sounding, the higher the acceptance rate. However, the studies that have yielded positive results have compared only various tones of synthetic speech (less human-sounding vs. more human-sounding ones) rather than real humans vs. non-humans. Another recent study has compared real human vs. non-human speech with results that when an actual human is in the picture, the synthetic counterpart is evaluated as "eery" and slightly less preferable than the human one. A study of DECtalk Express vs. real humans has yielded concrete results that humans prefer real humans over synthesized ones (Stern et al., 2000).

To conclude, the research on synthetic speech acceptance is rather incongruent in its results. However, applying it to the audiobook context is still promising, as there is a potential to overcome the "eeriness" aspect and expand the consumption of synthetic speech beyond the task-oriented applications to more social and companionship ones. The study by Kumar et al. (2020) comparing only synthetic speech highlights that humans perceive, process, and store human-sounding narrated speech better than artificial ones. While the studies comparing real human vs. synthetic speech ones suggest that humans prefer humans, however, all of those studies were based on older generations of speech-based systems, while the sophisticated, cutting-edge speech systems intended for use in the audiobooks market have long surpassed those of, for example, DECtalk Express, Alexa and Siri. Therefore, based on previous positive results of advanced synthetic speech acceptance, there is a huge potential that studies using more human-resembling synthetic speech, like that of ZenDeep and Speechki, might yield promising results for AI-narration acceptance in audiobook format. However, most of

the studies comparing real humans vs. non-humans concluded that real humans are preferred. Due to the specifics of audiobook consumers being "traditionally oriented" and seeking "human companionship," there is a huge risk that, specifically in the audiobook context, AI-narration might meet more resistance than in any other field of application.

Therefore, the first hypothesis is:

H1: Consumers of audiobooks, who are aware that the narrator is non-human, would score less on the attitudinal measures, compared to those, who perceive to listen to a real human.

2.4 The role of the narrator

This study would have been redundant if the consumers were to be indifferent to the role of the narrator in their consumption experience of audiobooks. However, several studies suggest speakers play a huge role in enabling listeners to create, imagine and perceive the communicated information (Dahmani et al., 2021). Another study on American audiobook consumers has concluded that consumers consider the narrator's voice before purchasing the title to listen to and has even provided a list of popular narrators (Kunzel et al., 2011). One could compare the role of the narrator in delivering the story's content to the actor's role in the movie.

Looking beyond the auditory studies, the theory of "parasocial interactions" (Horton and Wohl, 1956) states that humans feel a special connection and relationship with a performer or a celebrity persona, indicating that audiobook listeners might be attached to a particular narrator, and it could determine their listening behavior. This research delves deeper into the listener-narrator relationship and, based on previous research, hypothesizes that listeners have established relationships and the characteristics of the narrator are important. If consumers think that human narrators and personal connection with them are important, they would be less inclined to adopt synthetic non-human narrators.

H2: The narrator's importance negatively correlates with the probability of AI-narration adoption.

2.5 Demographics of AI-acceptance

Publishers have already accepted the unavoidable implementation of AI-narration (McIlroy, 2021). The main question on everyone's minds is where to use it and what customer segments to provide with AI-narrated content (Maughan, 2022). The research in the field of audiobook acceptance argues that young and middle-aged individuals are more likely to adopt technological innovations such as audiobook streaming platforms (Höglund and Wahlström, 2018). Broadening the research area beyond solely audiobooks towards technology and innovations acceptance in general, the studies suggest that younger individuals, especially men and those with previous technological experiences, are more likely to adopt technological innovations. Combining all this knowledge and applying it to the acceptance of AI-narration by consumers, we may hypothesize that:

H3: age is negatively correlated with the propensity to listen to AI-narration

H4: male listeners are more likely to adopt AI-narration

2.6 Prior technological experience

Several studies by Craig and Schroeder (2019); Pajupuu et al. (2019); Rodero (2017) on user acceptance models have concluded that prior technological experience, not necessarily in the same field, is a strong predictor of the individual's intention to use or adopt some other technological novelty. Moreover, having a positive attitude towards technology in general also means that an individual is more susceptible to trying or adopting a tech-novelty (Jönsson, 2021). Therefore, we assume that those with previous technological experience and a positive attitude towards technology, in general, are more likely to adopt AI narration.

H5: previous technological experience and attitudes towards technology positively correlate with the likelihood of adopting AI-narration

2.7 Frequency of use

The research on the relationship between the frequency of use of audiobooks and attitudes towards technology in general and synthetic speech is non-existent at the current moment. However, the researcher believes that there is a link between the frequency of use of audiobooks and the likelihood of adopting AI narration. O'Connor and Hedda (2019) argue that the audiobook streaming platform is an innovative technological novelty, and those using it are initially more predisposed to accepting new technologies. Thus, we may suppose that

the more frequently an individual listens to or uses audiobook services, the more he or she is prone to accepting AI narrators.

H6: frequency of use of audiobooks is positively correlated with the likelihood of adopting AI-narration

2.8 Areas of application

As touched upon in the introduction, the central area of application of synthetic speech is non-fictional books heavy with texts, where performance, emotionality, and improvisation of the narrator do not play such a significant role as in the cases of fictional books. The founder of one of the most prominent AI-conversational companies, VUX World, Kane Simms, says: "To create a synthetic voice that is capable of holding human attention for long periods—that's quite a challenge. The way that an audiobook narrator will read an audiobook, for example, is different from how one would answer a simple question such as 'What time is it?' In a book, you might need a news reader voice, an excited voice, a sad voice, a slow voice, a fast voice, a high-pitched voice, and all of this on a single page." However, recently two of the most developed vendors of AI speech in the market, ZenDeep and Scribe, have claimed that their technology is capable of narrating a fictional book as well as a human, with several dozens of successful examples, such as *Alice in the Wonderland* and more, narrated by AI. One of the studies, conducted by Pajupuu et al. (2019), has outlined three speaking styles that are appreciated by the listeners of the fictional books and calls for these styles to be embedded into the speech synthesis of the fictional book. Several other studies have successfully created synthetic speech capable of detecting emotions in the text and expressing them accordingly (Marcela & Ingmar, 2013; Dahmani et al., 2021; Pengfei et al., 2019). If manufacturers know what voices are required for the fictional books and how to reproduce them, then the only unanswered question remains the reaction of the consumers of the audiobooks to such shifts. Therefore, we want to explore how accepting consumers are of AI narration, what fears and challenges they perceive with non-human narrators?

What do listeners of audiobooks think of the non-human narrator? What are the perceived benefits, fears and areas of application for AI-narrators?

2.8 Summary of the hypothesis and research questions with respective literature

H1: Consumers of AI-narrated audiobooks score less on attitudinal measures, such as liking, engagement and naturalness, than consumers of human-narrated audiobooks.	Lopatovska et. al, 2018; Mori , 1970; Kuhne et al., 2020; Söderlund and Oikarinen, 2019; Schmitt 2020; Schreibelmayr and Mara 2022
H2: Importance of the narrator negatively correlates with the probability of AI-narration adoption.	Horton and Wohl, 1956; Kunze et al., 2011; Melanie C., 2004; Rodero, 2017
H3: Age is negatively correlated with the propensity to listen to AI-narration.	Höglund and Wahlström, 2018; Bergström et. al, 2017;
H4: Male listeners are more likely to adopt AI-narration.	Bruce A. Weinberg, 2004
H5: Previous technological experience and attitudes towards technology positively correlate with the likelihood of adopting AI-narration.	Bruce A. Weinberg, 2004; Tracy, 2000; Jönsson, 2021
H6: Frequency of use of audiobooks is positively correlated with the likelihood of adopting AI-narration	O'Connor and Hedda, 2019
What do listeners of audiobooks think of the non-human narrator? What are the perceived benefits, fears and areas of application for AI-narrators?	Pajupuu et al., 2019

3. Methodology

The methodologies and approaches utilized to analyze and respond to the research questions are described in this chapter. The chapter on methods is divided into several sections, such as the methodological role of research, the specific research methods used to address the research questions, the methods used to collect and analyze data, and a section on the validity and reliability of data.

3.1 Research Design

The research questions stated previously aimed at 1) assessing the aggregate level of synthetic speech acceptance by audiobook consumers, 2) the susceptibility of specific groups to AI-narration, and 3) identifying appropriate areas for AI-narration. A broad survey study has been conducted to answer the research questions and test the above-stated hypotheses. Firstly, a classical A/B experiment was conducted to measure if synthetic speech affects the listener's attitude toward the narrator. Secondly, the A/B experiment with two groups was part of an extensive survey study, with 400 consumers of audiobooks responding to multitudes of quantitative 5-point Likert-scaled questions, aiming to explore the relationship of propensity to consume AI-narration with participants' demographics (age, gender) and personal characteristics (frequency of use, previous technological experience, the degree of the importance of the narrator). Thirdly, several open-ended questions were asked to understand why listeners have particular attitudes towards synthetic speech and their personal opinions on the appropriate areas of AI speech application. While the answers to quantitative questions give clear answers as to the level of acceptance of AI-narration by listeners, open-ended questions add an important dimension, explaining the findings and capturing ideas and beliefs of consumers that are not included and predicted by the researcher in the initial questionnaire. Therefore, this questionnaire design enables the reader to understand the consumer's viewpoint on AI-narration and its application.

3.2 Questionnaire Design

The initial questionnaire comprised 30 questions in total, where 5 of those questions were dedicated to the experimental part of the research. The first question is the screening question, where respondents are asked to choose activities they've done in the past month.

Among five different answers, only those who choose "listened to an audiobook" can answer the questionnaire. The subsequent five questions are aimed at measuring respondents' listening habits, motivation to use audiobooks, preferred genres, and the importance of the narrator, after which an experiment is introduced. When the experiment is complete, respondents are asked more directly related to AI quantitative (Likert 5-points) questions assessing the level of their previous technological experience, attitudes towards technology in general, propensity to choose AI-narrator vs. human; as well as some open-ended questions, requesting respondents to motivate their previous answers and share their thoughts, fears, expectations in regards to AI-narration and areas of its application. Besides questions directly concerning the survey, respondents are also asked to provide their age, gender, ethnicity, country of residence, and spoken languages.

3.3 Experiment Design

To assess if listeners have any prejudice against AI narration, a classical A/B experiment has been conducted. Magnus Söderlund (2018) argues that a true social experiment with two randomly allocated groups can help researchers assess whether the independent variable affects the dependent variable. In this experiment, we want to measure listeners' reactions to the treatment of synthetic speech. The design of the experiment is such that the whole population of respondents is randomly allocated to 2 different test and control groups. Both groups were exposed to an audiobook excerpt narrated by an AI; the only difference between the groups is that the test group was provided with the description that the narrator is not a human, while the control group was unaware of the fact. A synthesized voice used to record the excerpt is an AI-narrator Nicholas Smith, available at Storytel and produced by an AI-speech company Speechki. The recording is considered to sound human enough, and the content of it is very plain (describing in a minute the concept of human happiness), thus reducing the risks that respondents' likeness of the text and content might interfere with the evaluation of the recording and narrator. After listening to the recording, respondents are asked to evaluate the narrator only and answer questions measuring behavioral and psychological reactions (Söderlund, 2018), such as how much they liked the recording, how engaging the narrator was, and the probability of listening to the same or similar narrator. If there is any difference between the groups on how much they liked and enjoyed the recording and how natural they found the narrator to be, then we would be able to claim that knowing

that the narrator is a non-human has an effect. In other words, we could claim that listeners have a negative prejudice against AI, affecting their evaluation of the speaker and content.

3.4 Data collection

A surveying platform, Pollfish.com, was chosen as a place for data collection. The main reason for using a surveying website is accessibility to the global market of millions of respondents, allowing this research results to be universally applicable and not pertaining only to Sweden. Also, it allowed for specific quotas to be enabled so that different demographics, groups, and even ethnicities are fairly represented, ensuring a truly random allocation (Tattersall, 2022). Polling websites are sometimes criticized for the quality of the responses due to respondents hastily completing the surveys for monetary rewards. However, the pollfish market specialists assured that they use a method of "organic sampling," and this survey would be spread on the various relevant platforms, where the audience is not indifferent to the subject of the survey. The survey was launched in mid-April and has collected roughly 400 responses in just 24 hours.

3.6 Reliability and validity of measurements

Likert (1932) has originally developed a 5-point Likert scale to measure specifically attitude and it's been proven to be both valid and reliable in various sectors of social sciences and humanities, such as psychology, sociology, education and business administration (F.Hair et al., 2019). An attitude can be defined as a preferred way of acting/reacting in each situation, based on a relatively stable organization of beliefs and thoughts around an object, topic, or concept acquired through social interaction. The main purpose of this paper is to shed light on the beliefs and thoughts of audiobooks' consumers in regards to AI-narration. Thus, using scales developed specifically for measurement of attitudes, and which has proven to be universally applicable in different fields, has been deemed to be an appropriate method to measure the attitudes of subjects of this study.

Attitude, being a complex concept, has been measured as a 'construct' (instead of a single variable) consisting of 4 items aimed at understanding the true nature of listener's feelings and beliefs about AI-narration. These items include questions (Q7-Q10, appendix), on the participant's perception of how 'likable', 'engaging', and 'natural' they found AI-narrative and likelihood they would listen to it again. All these items were chosen for their general

application in the research field of attitude; while measuring ‘engagement’ and ‘naturalness’ are recommended when assessing characteristics of AI performance in human interaction (Ref). The answers follow a five-point Likert scale ranging from strongly negative (liked a lot, extremely engaging, extremely natural) to strongly positive attitudes (disliked a lot, extremely unengaging, extremely unnatural). While it is generally recommended to use a seven-point Likert scale, the characteristics of this study allow for the 5-points. Chen, Yu and Yu (2015) found in their study that five-point Likert scale requires the least cognitive effort and described by the respondents as easiest to answer, therefore allowing for a better communication of the information. The survey of this study is extensive and requires a lot of cognitive attention to process complex descriptions, therefore a five-point scale would provide more reliable results. To ensure the reliability of the responses, measuring attitudes of the AI speech, three different items were asked, measuring attitude construct. As a result, Cronbach's alpha for these items is 0.839, which is above the recommended limit of 0.7 Söderlund et al. (2018), indicating that the measures are reliable.

The previous technological experience is considered as an essential variable that is hypothesized to have an effect on attitude towards AI narration. The variable consists of two items, asking participants 1) to choose their level of technological experience out of five different options, ranging from “I hate technological novelties” to “I am fish in the water when it comes to tech news” (Q12) and 2) their attitude towards advanced synthetic speeches such as Google’s Alexa and Apple’s Siri. While the scale itself follows Likert’s five-points, the wording and expression are adapted from Edison and Geissler’s (2003) study, conducted similarly to this study, with an aim to highlight personal characteristic factors for the adoption of technology. These two items have a Chronbach’s alpha of 0.741 which is above the recommended limit of 0.7 Söderlund et al. (2018), indicating that the measures are reliable.

Importance of the narrator is a factor that is predicted to negatively affect the attitude of the respondents. Following the recommendations of Michel and Siegrist (2019)— in their study on the development of scales for measurement of importance of naturalness, trust and authenticity— the shortest and the most efficient scales were adapted: Likert’s five-point scale with answers ranging from ‘extremely unimportant’ to ‘extremely important’.

Variables such as age and gender of the respondents are automatically collected and provided by the polling website, prior to respondents taking a survey, and are measured in ratio and

nominal scales respectively. “Frequency of use” variable was originally intended to serve as an easy opening question, with a purpose to immerse and involve respondents in the topic of the survey (McNeeley, 2012) , thus no special constructs are made. “Frequency of use” is measured on ordinal Likert scales (consistent with the rest of the survey), by asking Q2: “How often do you listen to audiobooks”, with answers ranging from “everyday” to “once a month” (note, the survey is aimed at frequent users of audiobooks, thus “once a month” is the rarest option allowed).

Propensity to consume is another crucial construct variable against which other variables, such as “importance of the narrator”, “frequency of use”, “age” and “previous technological experience”, are correlated. “Propensity to consume” is a construct variable based on three questions Q20, Q21 and Q22 (appendix), aimed at measuring the likelihood and intention of audiobooks listeners to consume AI-narrated content. The questions are adapted from Ajzen’s (2006) article on measurements of intention in the Theory of Planned behavior, and follow clear guidelines. Firstly: Target, Action, Context and Time of the intended action are clearly stated, as in the example of Q20: “Consider a situation: next time you search for an audiobook title to listen to, there is a broad selection of your favorite books narrated by an AI. How likely is it that you would listen to an AI-narrated audiobook?”, with answers consistent to Likert’s 5-points scales used throughout the survey, and ranging from “extremely unlikely” to “extremely likely”. And secondly: the construct provides a high reliable internal consistency of variables, with a Cronbach’s alpha of 0,817, above the recommended threshold of 0,7 (Söderlund et al., 2018).

The collected data has been meticulously cleared of all the respondents who did not pay attention to the recording and answered wrong on the control question and those whose open-ended answers suggested that they did not understand the topic of the survey. The final cleared data contains 322 valid respondents and approximately 15 different variables measuring crucial parts of research and five open-ended qualitative answers.

4. Analysis and Findings

4.1 Experiment: Attitudes, Likeness

The purpose of the experiment is to see if there is any difference in attitudes towards the narrator between the test group, AI, and the control group, Human. Due to the ordinal scales used in the measurement of the variables, a non-parametric Mann-Whitney U test is deemed the only appropriate way to analyze the differences.

Mann-Whitney Test: Attitudes					
Group	Mean	SD	U	P-value	Alpha
AI	3.6605	0.93799	11828	0.172	> 0.05
Human	3.5404	0.86998			

Assessing the group statistics, it becomes clear that the mean score for the AI group (3.6605) is higher than that of the Human group (3.5404): intuitively suggesting that AI listeners have more positive attitudes towards the narrator, rather than the Human group. However, Mann-Whitney U test concludes that this identified difference is insignificant ($U=11828$, $p=0.172$). Thus, we cannot support the hypothesis that there is a difference in attitudes between the AI and the Human groups.

H1: Consumers of AI-narrated audiobooks score less on attitudinal measures (liking, engagement and naturalness) than consumers of human-narrated audiobooks.

Rejected

4.2 Importance of the narrator

It is assumed that consumers of audiobooks place significant importance on the role of the narrator. Additionally, it is assumed that the more important the role of the narrator, the less likely consumers are to accept an AI-narrator, instead of a real human. Analysis of this assumption is done by correlating the “importance of the narrator” variable with the “propensity to consume” construct variable, with Spearman’s correlation deemed as an appropriate test, due to the ordinal scales.

Spearman's Correlation Test: Importance of the narrator		
Correlation Coefficient	P-value	Alpha

0.216	0.001	< 0.05
-------	-------	--------

The results of the Spearman's correlation test suggest that there exists a significant weak positive association between propensity to consume and importance of the voice of the narrator ($r=0.216$, $p=0.001$). It indicates that listeners, who think that the voice of the narrator is important for their listening experience, are more likely to adopt AI-narration, in comparison to the listeners for whom this factor is of less importance. This result is contrary to the initial hypothesis that importance of the narrator negatively impacts the chances of adopting AI-narration.

H2: Importance of the narrator negatively correlates with the probability of AI-narration adoption.

Rejected: weak positive correlation is identified.

4.3 Demographic factors: age

It has been previously hypothesized that younger listeners of audiobooks are more prone to adopting AI-narration, compared to older listeners of audiobooks. As the variables are in ordinal scales, Spearman's correlation is used to test the hypothesis.

Spearman's Correlation Test: Age		
Correlation Coefficient	P-value	Alpha
0.11	0.846	>0.05

The results of the Spearman's test suggest that there is an insignificant weak correlation between the age and propensity to consume AI-narrated audiobooks ($r=0.11$, $p=0.846$). As the p-value of 0.846 is above the accepted threshold of 0.05 (Söderlund, 2017), we cannot reject the possibility that there is no correlation between the age and likelihood to adopt AI-narration.

H3: Age is negatively correlated with the propensity to consume AI-narration.

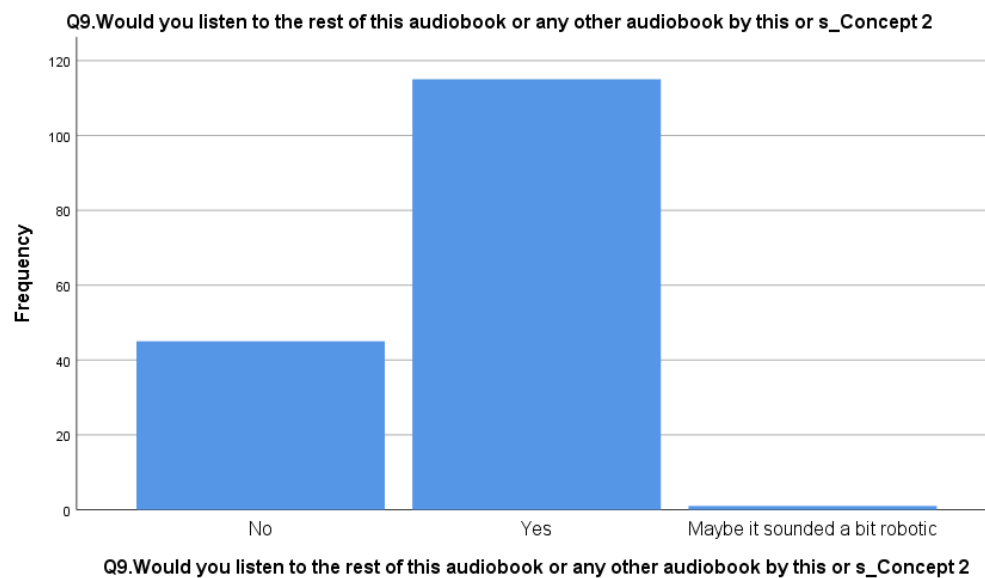
Rejected: no significant correlation is found

4.4 Demographic factors: gender

Initially, it was hypothesized that male consumers of audiobooks are more likely to adopt AI-narration, compared to women consumers, and the results of the Mann-Whitney test ($U=11122$, $p=0.03$) reinforce this belief.

Mann-Whitney Test: Attitudes					
Group	Mean	SD	U	P-value	Alpha
Female	2.4212	0.64778	11122	0.03	< 0.05
Male	2.5625	0.63387			

Moreover, assessing the results of descriptive statistics of Q9 (likelihood to listen to a similar AI-narrator): most of the respondents were positively installed towards using AI-narration in the future. For example, when asked if respondents would listen to another AI-narrator, 73% of the respondents answered: Yes. However, it's worth noting that significantly more women (32%) than men (19%) answered No. A similar scenario is evident from other resembling questions (Q20-Q26), where roughly 70% of the total population is positively installed towards AI-narration.



H4: Male listeners are more likely to adopt AI-narration than females.

Supported

4.5 Personal characteristics: previous technological experience

It is hypothesized that consumers of audiobooks, who have had positive attitudes and experience with technological devices and applications, are more prone to adapting AI-narration, compared to those, who have had limited prior technological experience or negative attitudes towards technology. Spearman's correlation test is used to test the hypothesis.

Spearman's Correlation Test: Previous technological experience		
Correlation Coefficient	P-value	Alpha
0.316	< 0.001	< 0.05

The results of the Spearman's test suggest that there is a significant moderate positive association between the previous technological experience and propensity to consume AI-narration ($r=0.316$, $p<0.001$). This indicates that listeners who have positive attitudes towards technology and previous experience, are indeed more likely to adopt AI-narration

H5: Previous technological experience and attitudes towards technology positively correlate with the likelihood of adopting AI-narration.

Supported

4.6 Personal characteristics: frequency of use

It is hypothesized that the more frequently consumers listen to audiobooks, the more likely they are to adopt AI-narration. The hypothesis is tested by using Spearman's correlation, due to the ordinal scales of variables.

Spearman's Correlation Test: frequency of use		
Correlation Coefficient	P-value	Alpha
- 0.178	0.002	< 0.05

The results of the Spearman's test suggest that there exists a significant weak negative association between the frequency of use of audiobooks and the likelihood to adapt

AI-narration ($r=-0.178$, $p=0.002$). As the p-value is below the threshold of 0.05 (Söderlund, 2018), we may conclude that the negative correlation identified is significant.

H6: Frequency of use of audiobooks is positively correlated with the likelihood of adopting AI-narration.

Rejected: significant negative correlation is identified

4.7 Analysis of the open-ended responses

Analysis of the open-ended responses reinforces the findings of the quantitative research. Majority of audiobooks listeners have positive attitudes towards AI narration, due to audiobooks perceived as being more user-friendly and efficient. Some consumers consider this new technology as an innovation in the field of audio-recording and are confident that it will become more common and better in the future, thus sharing views of some of the publishers. According to listeners, the main significance of using AI narration is that it would provide error free work with more accurate narrations, compared to the real human narrator: "I like AI-narration because it does not have an accent.", and "I think AI doesn't make mistakes, very accurate compared to people." Therefore, some of the suggestions for the use of synthetic speech were in the 'business books', 'news related', 'scientific fiction' and even 'bible'. In addition, some audiobooks' consumers would highly appreciate the possibility to choose and tailor narrative-voice according to their preferences. When asked what their preferred style of narrator-voice was, most recurring answers indicated that audiobook's listeners favored 'female' 'energetic' and 'seductive' voices. Therefore, it could give a clue to the publishers and synthetic speech providers on the types of voices to develop for the audiobooks market.

When assessing some of the respondents and asking them why they are positively inclined towards AI-narration, participants say that they value practical and economic advantages that come with AI, for example, access to more audiobooks than there currently are: "Something is better than nothing. When a human voice is not available, you may go for AI-narration." As the participants of the study are international, it was interesting to note that some hispanic respondent complained on the lack of spanish audiobooks compared to the english titles, and voiced a hope that AI-narration would bring more equality and variety to other narrated languages. Respondents who were extremely positively inclined towards AI-narration happened to be young males with strong previous technological experience: "yes, it (AI-narration) should be introduced, i am a tech savvy and ai is my favorite field".

However, some think AI-narration should not be implemented and encouraged, but those were surprisingly few. The most recurring explanation to negative attitudes towards synthetic

speech is that listeners fear that technique cannot express the same emotions as real humans and audiobooks would become 'weird' and 'boring like reading an encyclopedia': "the lack of emotion leaves everything bland and mundane, no excitement, creates anxiety as if I'm at work meeting". Also, the presence of the human-narrator is important to some listeners for the whole audiobook consuming experience and visualization of the narrator: "people have the habit of listening and the same time they can imagine the person image in their mind narrating the story, but now with AI its not". When assessing how appropriate AI-narraion is for different genres, some listeners are not ready to consume AI-narrated fiction, due to these two factors: the lack of emotions and connection to the real person, whilst others thought that it could and should be implemented in all genres. This difference could be described by the differing knowledge and beliefs that listeners might have about the capabilities of AI. There are also those who try to find a middle ground and say that AI should be complementary to human narration and co-exist: "In the future, it will be a luxury to have a choice of AI or a human, and they both should be used for everyone's good and different demands."

4.8 Summary of the results

<i>Assumptions</i>	<i>Findings</i>	<i>Conclusions</i>
H1: Consumers of AI-narrated audiobooks score less on attitudinal measures, such as liking, engagement and naturalness, than consumers of human-narrated audiobooks.	Means: AI (3.6605) Human (3.5404) Mann-Whitney U-test: U-value (11828) P-value (0.172) Alpha (>0.05)	Hypothesis is rejected: no difference in attitudinal measures is found between the consumers of AI-narrated and human-narrated groups.
H2: the importance of the narrator negatively correlates with the probability of AI-narration adoption.	Spearman's correlation test: Correlation coefficient (0.216) P-value (0.001) Alpha (<0.05)	The hypothesis is rejected: a weak positive correlation is identified between the importance of the narrator and the probability of adopting AI-narration
H3: age is negatively correlated with the propensity to listen to AI-narration.	Spearman's correlation test: Correlation coefficient (0.11) P-value (0.846) Alpha (>0.05)	The hypothesis is rejected, as no significant correlation is found between the age and the likelihood to consume AI-narration
H4: male listeners are more likely to adopt AI-narration.	Means: Female (2.4212) Male (2.5625) Mann-Whitney U-test: U-value (11122) P-value (0.03) Alpha (<0.05)	The hypothesis is supported, as males are significantly more likely to adopt AI-narration, compared to females
H5: previous technological experience and attitudes towards technology positively correlate with the likelihood of adopting AI-narration.	Spearman's correlation test: Correlation coefficient (0.316) P-value (<0.001) Alpha (<0.05)	The hypothesis is supported, as consumers with previous technological experience are more likely to adopt AI-narration
H6: frequency of use of audiobooks is positively correlated with the likelihood of adopting AI-narration.	Spearman's correlation test: Correlation coefficient (-0.178) P-value (0.002) Alpha (<0.05)	The hypothesis is rejected, instead it is identified that the more frequently listeners consume

		audiobooks, the less likely they are to adapt to AI-narration
What do listeners of audiobooks think of the non-human narrator? What are the perceived benefits, fears and areas of application for AI-narrators?	AI-narration is seen by consumers as “user-friendly”, “error-free” and “effective”. However, many are skeptical that AI-narrator “lacks emotions”	AI-narration would allow for greater access to audiobooks, especially for non-english speaking consumers of audiobooks. But due to the perception that AI is robotic sounding and emotionless, its use in fiction literature is still a highly arguable topic for consumers, leaving non-fiction genres as a widely accepted area for AI.

5. Discussion

The purpose of this research is to contribute to the growing field of knowledge in the area of synthetic speech acceptance. It was clearly identified that the research in the area of synthetic speech acceptance is gravely outdated and highly inconclusive. Moreover, with the introduction of new synthetic speech technology, powered by Artificial Intelligence, the previous research, based on old technology, could not provide a reliable compass for the understanding of the mechanisms of new synthetic speech acceptance; nevertheless, it served as a starting point for formation of assumptions and hypotheses of this study. These assumptions and hypotheses were quantitatively tested, and this section intends to discuss how the results of these tests correspond to the various theories and literatures used in this study. However, it is important to remember that the main purpose of this study is not to merely contribute to the theory on synthetic speech acceptance, but to apply this knowledge to a practical field and industry of audiobooks production and consumption: that is to say, to provide audiobooks publishers and distributors with clear answers to three main questions: 1) the degree of consumer acceptance of AI-narration, 2) the role of demographic and personal characteristics in adopting AI-narration, and 3) consumers perceptions of AI-narration: fears, beliefs and suitable fields of application.

5.1 Consumers acceptance of AI-narration

In an attempt to predict consumer behavior of audiobooks users in relation to AI-narration, this research heavily relied on the theories of “uncanny valley” of Masahiro Mori (1970) and “speciesism” of Bernd Schmitt (2020). The theory of the "uncanny valley" suggests that as a machine's resemblance to a human increases, humans' emotional response becomes more negative and eerie (Mori, 1970). In the context of audiobooks, this theory implies that the more human-like the AI-generated narration sounds, the less appealing it would be to human listeners. The theory of "speciesism" proposed by Bernd Schmitt in 2020 focuses on the preference for human interactions over interactions with non-human entities. In the context of audiobooks, this theory suggests that consumers would prefer human narrators over synthetic speech, regardless of how human-like the AI-generated narration may sound. Together these theories set the foundation for the hypothesis (H1) that consumers of AI-narrated audiobooks

would score less on attitudinal measures such as liking, engagement, and naturalness compared to consumers of human-narrated audiobooks.

To test the hypothesis, an experiment was conducted involving two groups. These groups were compared based on attitudinal measures, such as liking, engagement, and naturalness, depending on whether they perceived to listen to an AI narrator or a human narrator. The aim of this experiment was to investigate whether consumers of audiobooks indeed prefer human narrators over synthetic speech, even if the AI narration is almost indistinguishable from humans. The experiment itself did not find any significant difference in attitudinal measures between the groups that listened to AI-narrated audiobooks and those that listened to human-narrated audiobooks. This finding contradicts the theories of the uncanny valley and speciesism, which predicted a preference for human narrators over synthetic speech.

However, the analysis of other survey questions and open-ended responses suggests that consumers have a positive inclination towards the adoption of AI-narration. The responses indicate that participants are open to the idea of AI-narration and believe that synthetic speech could be valuable in providing accessible audiobooks materials. This finding suggests a potential acceptance and willingness to embrace AI technology in the context of audiobooks. On the question of preference between human-narrators and AI, the results of Questions 17 (Appendix) indicate that a majority of consumers (77.3%) would choose a human-narrator over AI-narration in general; so there is still a preference for human narrators when given a choice.

In summary, while the results challenge the uncanny valley theory (Mori, 1970) by demonstrating acceptance and positive inclination towards AI-narration, the preference for human narrators over the machines aligns with the concept of speciesism (Schmitt, 2020), emphasizing the continued appreciation for human involvement in the audiobook narration process.

5.2 The role of demographic and personal characteristics in adopting AI-narration

This research aims to investigate how demographic factors and individual personal characteristics contribute to the acceptance and adoption of AI-narration in the audiobooks market. By examining variables such as importance of the narrator, age, gender, technological familiarity, and frequency of use, this study seeks to understand the factors that shape consumers' acceptance and attitudes towards AI-narrated audiobooks. The findings provide valuable insights into the drivers and potential barriers to the adoption of AI-narration, allowing for targeted strategies to enhance user experience and promote wider acceptance of AI-narration.

Consistent with previous research by Dahmani et al. (2021) and Kunzel et al. (2011), the survey results reveal that a significant proportion of respondents (78.88%) consider the role of the narrator to be either "important" or "extremely important" in their audiobook consumption experience. Drawing on the theory of "parasocial interactions" proposed by Horton and Wohl (1956), it was hypothesized that the greater the perceived importance of the narrator, the less likely consumers would be to accept any form of substitution of human narrators with AI-narrators. However, the findings of this study present a contrasting result. The analysis demonstrates a positive relationship between the importance of the narrator and the likelihood of accepting AI-narration. In other words, when consumers place greater importance on the role of narrators, they are more inclined to accept AI-narration.

The acceptance of AI-narration is predicted to be influenced by two demographic factors: age and gender. When it comes to the age, no significant correlation has been identified between the age of the listeners and their probability to adopt AI-narration; contrary to the study of Höglund and Wahlström (2018) which suggested that age influences the probability of adapting to new technology. On the other hand, gender indeed influences the probability of adopting AI-narration, as male listeners are significantly more likely to adopt AI-narration than female listeners (H4). This result confirms Weinberg's (2004) study of males being more receptive and accepting of new technologies than females.

Some other personal characteristics that are predicted to impact the probability of adopting AI-narration are previous technological experience and frequency of use of audiobooks.

Several studies on user acceptance models (Rodero, 2017; Craig and Schroeder, 2019; Pajupuu et al., 2019) suggest that extensive technological experience in any field positively influences the likelihood of adopting new technologies. In the context of this research, a hypothesis (H5) was formulated that posits a positive relationship between consumers' previous technological experience, their positive attitudes towards technology in general, and their propensity to adopt a technological novelty like AI-narration. The quantitative analysis conducted in this study confirms this hypothesis (H5), revealing a significant positive correlation between participants' prior technological experience and their likelihood of adopting AI-narration. Similarly, it was assumed that the frequency of audiobook use would also impact the probability of adopting AI-narration. However, contrary to the initial hypothesis (H6), the analysis demonstrates a negative relationship between frequency of audiobook consumption and the likelihood of adoption. This implies that the more frequently listeners consume audiobooks, the less inclined they are to adopt AI-narration.

In summary, the research findings indicate that certain groups exhibit a higher tendency to adopt AI-narration compared to others. Specifically, factors such as the perceived importance of the narrator and previous technological experience show a positive correlation with the probability of adopting AI-narration. On the other hand, the frequency of use negatively affects the likelihood of adopting AI-narration. Moreover, the findings suggest that males are more inclined to adopt AI-narration than females, whereas age does not appear to have a significant relationship with the likelihood of adopting AI-narration.

5.3 Consumers perception of AI-narration: beliefs, fears and suitable fields of application

Consumers hold positive beliefs regarding AI-narration, perceiving it as user-friendly and efficient. They value its potential to provide error-free and accurate narrations, making it suitable for various audio-genres such as business, news, science fiction, and even biblical texts. Moreover, non-English listening consumers hold a belief that AI-narration would allow for more audiobooks content in their native languages, than there currently is.

Some consumers express fears related to the lack of emotions and human connection in AI-narration, perceiving it as potentially "weird" and "boring." Consumers' belief that AI is emotionally-deprived is a significant hurdle to implement AI-narration in fiction literature. However, they believe that AI-narration can be complementary to human narration and coexist in the future. Young males with previous technological experience display particularly positive beliefs towards AI-narration, voicing their readiness to consume even fiction books narrated by AI. Overall, consumers appreciate the practical advantages AI-narration offers, including increased access to audiobooks in other languages, while recognizing the importance of maintaining emotional connection in fiction genres.

5.4 Managerial implications

The introduction of intelligent synthetic speech systems has started debates within the audiobook publishing industry (Kuta, 2023). Producers advocate for the implementation of these efficient and cost-effective systems, while human narrators oppose the change, asserting that AI-narration cannot replicate the entertainment value provided by human narrators (Mellroy, 2019). The scarcity of human and other resources further exacerbates the situation, with 95% of potential audiobooks failing to reach the market (Abramov, 2022). As a result, AI-narration has become a necessity to meet the growing demand for recorded titles. In this context, the opinions of audiobook consumers play a pivotal role in resolving this ongoing debate. The aim of this research is to provide audiobook publishing industry managers with evidence-based arguments and findings, shedding light on consumers' acceptance of AI-narration, vulnerable target groups, and potential fields of application. The following recommendations are intended to guide managers in navigating this complex landscape:

1. **Implementation of AI-Narration:** The research findings strongly support the recommendation for managers to implement AI-narration in the audiobook publishing industry. Both consumers and publishers recognize the necessity for AI-narration due to resource constraints and the increasing demand for recorded titles. Managers should emphasize the cost-effectiveness and efficiency of AI-narration as key benefits to justify its adoption.
2. **Targeting Specific Consumer Groups:** Managers should consider a holistic approach when targeting specific consumer groups, taking into account both tailored marketing

strategies and promotions, as well as the allocation of AI-narration for these groups. The research indicates that males who highly value the role of the narrator and possess technical expertise are more likely to embrace AI-narration. To effectively capture their interest and engagement, managers can strategically allocate AI-narration for genres that particularly appeal to this target group, such as science fiction and fantasy for teenagers and young adults. By combining tailored marketing efforts and genre-specific allocations, managers can maximize the acceptance and impact of AI-narration in the audiobook publishing industry.

3. **Addressing Consumer Emotional Concerns:** While AI-narration is widely accepted in non-fiction genres, consumers express concerns about its perceived lack of emotional depth in fiction genres. Overcoming this hurdle requires managers to focus on enhancing the expression of emotions through AI-narration. This may involve investing in advanced AI technologies capable of replicating human-like emotional nuances, especially in genres where emotional engagement is crucial, such as romance or suspense. By prioritizing the development of AI systems that can deliver compelling emotional performances, managers can increase the acceptability and effectiveness of AI-narration in fiction genres. This strategic approach will ultimately enhance the overall consumer experience, ensuring that AI-narration becomes a viable and engaging option across various literary genres.
4. **Human vs. Non-Human Narrators:** The research findings indicate that consumers of audiobooks perceive AI-narrators as acceptable; however, they also value the role of human narrators. Consumers are not yet ready to completely abolish human narrators and instead believe that a coexistence of both human and non-human narrators is ideal. These views align with those of publishers, highlighting that consumers understand the need for AI-narration and do not object to its implementation. In light of this understanding, managers should carefully consider the preferences and expectations of consumers and strike a balance between utilizing AI-narration for its efficiency and cost-effectiveness while maintaining the presence of human narrators for their unique qualities and emotional connection. By finding the right equilibrium between human and non-human narrators, managers can ensure a satisfactory audiobook experience that meets the desires of consumers and aligns with the evolving landscape of the audiobook publishing industry.
5. **Consumer and AI-narrator relationship:** Based on the theory of parasocial interactions (Horton and Wohl, 1956), consumers of audiobooks value the role of the narrator,

considering it crucial for their satisfactory listening experience. Therefore, it is recommended that AI-narrators possess distinct personalities and identities, akin to human narrators. For instance, Storytel, an audiobook platform, assigns unique names to AI-narrators to provide them with individuality (Speechki, 2021). Another example, a producer of synthetic speech, ZenDeep works with recreating celebrities' voices, such as Samuel L. Jackson, and has even licensed actor Edward Herman's voice (McIlroy, 2021). By creating AI-narrators with their own traits and characteristics, consumers can develop relationships with them, resulting in stronger connections and deeper engagement. Implementing this recommendation can contribute to the acceptance and success of AI-narration in the audiobook publishing industry, acknowledging the significance consumers place on the narrator's role in enhancing their listening experience.

5.5 Limitations

The present study aimed to explore consumer acceptance of AI-narration within the context of audiobooks. To investigate this as part of a bigger survey study, a carefully designed experiment was conducted, utilizing the latest AI-narration systems. The experiment involved 320 participants who were divided into two groups. One group listened to an excerpt narrated by an AI-narrator named Nicholas Smith, while the other group listened to the same excerpt but were led to believe that it was narrated by a human. The primary objective was to compare the level of liking and acceptance between participants who were aware of the AI-narration and those who were not. However, despite careful planning and execution, the experiment did not yield statistically significant results. Several limitations in the study design and methodology should be considered when interpreting these findings. These limitations encompass factors such as the use of a single synthetic voice, the rating scales employed, and the scale type utilized. Addressing these limitations in future research endeavors would provide a more comprehensive understanding of consumer beliefs, fears, and suitable fields of application for AI-narration in audiobooks.

One limitation of the experiment was the use of the same synthetic voice for both the AI-narrator and human-narrator conditions. Ideally, it would have been preferable to have a distinct human narrator and an AI narrator for a more comprehensive comparison. This

limitation may have hindered the ability to capture the full range of differences between human and AI narration, potentially leading to less distinguishable outcomes between the two conditions. Another limitation lies in the scales used to measure participant responses. The study utilized 5-point Likert scales, which may have limited the sensitivity of participants' ratings. Implementing a more nuanced 7-point Likert scale could have potentially generated more significant results and provided a finer understanding of participants' attitudes and preferences towards AI-narration. Furthermore, the study employed ordinal scales instead of interval scales, which restricted the type of statistical analyses that could be performed. By utilizing interval scales, more sensitive results could have been obtained, enabling the use of regression analysis, such as ANOVA. Incorporating regression analysis would have allowed for a better understanding of the relative impact of various factors on the acceptance of AI-narration, enabling resource prioritization and identification of consumer segments more likely to adopt AI-narration.

These limitations highlight areas for improvement in future research. Conducting experiments with multiple human and AI narrators, utilizing more sensitive rating scales, and employing interval scales for robust statistical analyses would enhance the validity and depth of insights into consumer acceptance of AI-narration.

6. Conclusion

Consumer attitudes, reactions, and thoughts towards AI narration in the industry of audiobooks have been relatively unexplored until now, leaving retailers and producers uncertain about the extent to which synthetic speech should be implemented and its potential impact on consumer behavior and intentions. However, this research has shed light on this important topic, revealing that audiobook listeners across different regions and demographics are receptive and positive towards AI narration. This finding provides valuable insights for publishers seeking to bridge the gap between reading and listening formats in the audiobook industry. It is evident that synthetic speech can be used to address the challenges identified in earlier stages of this research, expanding the availability and accessibility of audiobooks to a wider audience. However, it is important to recognize that certain groups display a higher propensity towards accepting AI narration than others. Currently, young males with technological experience demonstrate the greatest openness to synthetic speech, indicating a specific target audience for publishers and vendors. To capitalize on the potential of AI narration, publishers should collaborate with vendors of synthetic speech and work together on the emotional expression of machines. Enhancing the emotional aspects of synthetic speech can contribute to a more immersive and engaging audiobook experience, helping to overcome concerns raised by listeners regarding the perceived lack of emotions and connection. Overall, this research signifies that audiobook consumers worldwide are ready and receptive to AI narration. Publishers and vendors can benefit from this positive inclination towards synthetic speech to bridge the gap between reading and listening formats and deliver a more diverse and accessible range of audiobooks. By addressing the needs and preferences of different consumer segments and continually improving AI technology, the audiobook industry can effectively meet the growing demand for audio listening products.

References

- Bidshahri, R. (2019, June 17). The Rise of AI Art and What It Means for Human Creativity. *Singularity Hub*. Available at: <https://singularityhub.com/2019/06/17/the-rise-of-ai-art-and-what-it-means-for-human-creativity/>
- Bäckström, K., & Johansson, U. (2006). Creating and consuming experiences in retail store environments: Comparing retailer and consumer perspectives. *Journal of Retailing and Consumer Services*, 13(6), 417-431. doi: 10.1016/j.jretconser.2006.05.005
- Catalan, M.D. and Marcelo, D.N., 2022. Development and validation of an audiobook in Afro-Asian literature. *Luke: An Online Journal of Pedagogy*, 2(1), pp.74-94.
- Cecco, L. (2022, January 4). Apple unveils new artificial intelligence technology for audiobooks. *The Guardian*. Available at: <https://www.theguardian.com/technology/2023/jan/04/apple-artificial-intelligence-ai-audiobooks>
- Craig, S.D. and Schroeder, N.L., 2019. Text-to-speech software and learning: Investigating the relevancy of the voice effect. *Journal of Educational Computing Research*, 57(6), pp.1534-1548.
- Dahmani, S., Colotte, V., Girard, V. and Ouni, S., 2021. Learning emotions latent representation with CVAE for text-driven expressive audiovisual speech synthesis. *Neural Networks*, 141, pp.315-329.
- Dakic, M., 2019. Preferences and attitudes of audiobook users in Sweden: Surveying Swedish audiobook groups on Facebook.
- De Carolis, B., D'Errico, F. and Rossano, V., 2021, August. Pepper as a Storyteller: Exploring the Effect of Human vs. Robot Voice on Children's Emotional Experience. In *IFIP Conference on Human-Computer Interaction* (pp. 471-480). Springer, Cham.
- Dubiel, M., Halvey, M., Gallegos, P.O. and King, S., 2020, July. Persuasive synthetic speech: Voice perception and user behavior. In *Proceedings of the 2nd Conference on Conversational User Interfaces* (pp. 1-9).
- Endrstova, B., Macik, M. and Treml, L., 2018, August. Reprobooktor: A concept of audiobook player for visually impaired older adults. In *2018 9th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)* (pp. 000063-000068). IEEE.

- Eriksson, M. and Reponen, V., 2021. Designing a Shared Listening Experience for an Audiobook Streaming Service: A UX and UI Project.
- Festerling, J., Siraj, I. and Malmberg, L.E., 2022. Exploring children's exposure to voice assistants and their ontological conceptualizations of life and technology. *AI & Society*, pp.1-28.
- Green, M.C., Brock, T.C. and Kaufman, G.F., 2004. Understanding media enjoyment: The role of transportation into narrative worlds. *Communication theory*, 14(4), pp.311-327.
- Habib, F., Fatema, T., Khan, M., Khan, N.I. and Islam, M.N., 2022, April. Exploring Design Attributes and Development of an Acoustic VR Game to Improve Ethical Values of Visually Impaired People. In *2022 IEEE 7th International Conference for Convergence in Technology (I2CT)* (pp. 1-8). IEEE.
- Hoyer, W.D., Kroschke, M., Schmitt, B., Kraume, K. and Shankar, V., 2020. Transforming the customer experience through new technologies. *Journal of Interactive Marketing*, 51(1), pp.57-71.
- Ilves, M. and Surakka, V., 2013. Subjective responses to synthesised speech with lexical emotional content: the effect of the naturalness of the synthetic voice. *Behaviour & Information Technology*, 32(2), pp.117-131.
- Irani, T., 2000. Prior experience, perceived usefulness and the web: Factors influencing agricultural audiences' adoption of internet communication tools. *Journal of Applied Communications*, 84(2), pp.49-63.
- Jönsson, J., 2021. AI acceptance and attitudes: People's perception of healthcare and commercial AI applications.
- Kozlowski, M. (2020, June 20). Audiobook Trends and Statistics for 2020. *Good e-Reader*. Available at: <https://goodereader.com/blog/audiobooks/audiobook-trends-and-statistics-for-2020>
- Kumar, S., Gahalawat, M., Roy, P.P., Dogra, D.P. and Kim, B.G., 2020. Exploring impact of age and gender on sentiment analysis using machine learning. *Electronics*, 9(2), p.374.
- Kuta, M. (2023, Month Day). You Can Now Listen to Audiobooks Narrated by Artificial Intelligence—But Is That a Good Thing? *Smithsonian Magazine*. Available at: <https://www.smithsonianmag.com/smart-news/you-can-now-listen-to-audiobooks-narrated-by-artificial-intelligencebut-is-that-a-good-thing-180981402/>

- McIlroy, T. (2019, November 1). AI Comes to Audiobooks. *Publishers Weekly*. Available at: <https://www.publishersweekly.com/pw/print/20111101/87762-ai-comes-to-audiobooks.html>
- McNeeley, S. (2012). Sensitive Issues in Surveys: Reducing Refusals While Increasing Reliability and Quality of Responses to Sensitive Survey Items. *Research in Higher Education Journal*, 17, 1-16.
- Mulqueen, T. (2017, October 17). Most Consumers Welcome AI: Here's What It Means for Retailers. *Forbes*. Available at: <https://www.forbes.com/sites/tinamulqueen/2017/10/17/most-consumers-welcome-ai-heres-what-it-means-for-retailers/?sh=2b184aba43ff>
- Pajupuu, H., Altrov, R. and Pajupuu, J., 2019. Towards a vividness in synthesized speech for audiobooks. Eesti ja soome-ugri keeleteaduse ajakiri. *Journal of Estonian and Finno-Ugric Linguistics*, 10(1), pp.167-190.
- Rodero, E. and Lucas, I., 2021. Synthetic versus human voices in audiobooks: The human emotional intimacy effect. *New Media & Society*, p.146.
- 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, pp.336-346.
- Simonite, T. (2022). The rise of synthetic voices: Audiobooks' new chapter. *Wired*. Available at: from <https://www.wired.com/story/audiobooks-synthetic-voices/>
- Singh, A. and Alexander, P.A., 2022. Audiobooks, print, and comprehension: What we know and need to know. *Educational Psychology Review*, pp.1-39.
- Snelling, D. (2021). The Value of Art and Creativity in the Age of AI. *Academic Questions*, 34(3), 331-345. doi: 10.1007/s12109-021-09838-1
- Söderlund, M., 2018. Experiments in marketing. *Studentlitteratur*.
- Spjeldnæs, K. and Karlsen, F., 2022. How digital devices transform literary reading: The impact of e-books, audiobooks and online life on reading habits. *new media & society*, p.14614448221126168.
- Spratt, E. (2018). Virtual Reality: Exploring the Impact on Society and the Ethical Considerations. *XRDS: Crossroads, The ACM Magazine for Students*, 25(4), 20-25. doi: 10.1145/3186677
- Stern, S.E., Mullennix, J.W., Dyson, C.L. and Wilson, S.J., 1999. The persuasiveness of synthetic speech versus human speech. *Human Factors*, 41(4), pp.588-595.

- Tattersall Wallin, E. and Nolin, J., 2020. Time to read: Exploring the timespaces of subscription-based audiobooks. *New media & society*, 22(3), pp.470-488.
- Tattersall Wallin, E., 2022. SOUND READING: Exploring and conceptualising audiobook practices among young adults (Doctoral dissertation, Höskolan i Borås).
- Wright, K. (2020). Why Are Audiobooks So Expensive? *Chirpbooks*. Available at: <https://www.chirpbooks.com/audiobook-lists-and-spotlights/why-are-audiobooks-so-expensive>
- Zhang, L., Jiang, L., Washington, N., Liu, A.A., Shao, J., Fourney, A., Morris, M.R. and Findlater, L., 2021. Social Media through Voice: Synthesized Voice Qualities and Self-presentation. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1), pp.1-21.

Appendix

Questionnaire

Survey name:

Copy of AI-narration 1

Q1 Please choose activities that you have done in the past two months (MultipleSelection) (Screening question)

- Listened to classical music
- Watched Youtube videos on personal growth
- Listened to audiobooks
- Listened to podcasts on Spotify
- None of the above

Q2 How often do you listen to audiobooks? (SingleSelection)

- Everyday
- A couple of times a week
- Once a week
- Every other week
- Once a month

Q3 What is the main reason/motivation as to why you listen to audiobooks? (SingleSelection)

- Easier than reading a book
- Convenience while doing something else/multitasking
- Helps to focus
- Experience the content better when listening than reading
- Have favourite narrators
- Other

Q4 How important is the voice of the narrator(reader) of the book when you choose an audiobook to listen to? (SingleSelection)

- Extremely unimportant
- Unimportant
- Neutral
- Important
- Extremely important

Q5 What kind of voice do you prefer the most in audiobooks? For example, male, female, energetic, pitchy and more. In other words, what characteristics of the narrator are important to you? (open-ended)

-

Q6 What genre of books do you prefer and listen to the most? (SingleSelection)

- Fiction
- Non-fiction
- Other

Q7 How did you like the recording? (SingleSelection)

- Disliked a lot
- Disliked
- Neither liked nor disliked
- Liked
- Liked a lot

Q8 How engaging did the narrator sound? (SingleSelection)

- Extremely not engaging
- Not engaging
- Neutral
- Engaging
- Extremely engaging

Q9 How natural did the reader's voice sound? (SingleSelection)

- Extremely Artificial
- Artificial
- Neutral
- Natural
- Extremely natural

Q10 Would you listen to the rest of this audiobook or any other audiobook by this or a similar narrator? (SingleSelection)

- No
- Yes
- Other

Q11 What was the audio recording about? (SingleSelection)

- Climate issues
- Psychology of a human being
- Culinary achievements
- Swimming activities

Q12 You just listened to an audiobook extract read by a sophisticated AI system that can take up to 260 varying voices and is described by some as indistinguishable from human speech, and is expected to excel a year from now. Adopting AI-narration would enable publishers to record many more books than there currently are and would cost ten times cheaper and faster to record per book. (Description)

-

Q13 How would you describe your previous technological experience: (SingleSelection)

- I hate technological novelties and try to stay away from them
- I'm not so good with technology but I use it if I must to
- I am a moderate tech user and adopt something that I find useful and interesting
- I like technological appliances and apps and would consider trying something new
- I'm a fish in the water when it comes to tech news and love to try something new

Q14 Some highly advanced text-to-speech AI systems, like Apple's Siri, and Google's Alexa, are already widely used. What's your attitude towards them and technology in general? (SingleSelection)

- I definitely will NOT use anything like that
- I'd rather not use them
- I'm unsure
- I'm open to using them someday
- I would strongly like to use them / already use them
- Other

Q15 How important is it for you to be notified that the book's narrator is not a real human? (SingleSelection)

- Not at all important
- Slightly important
- Moderately important
- Very important
- Extremely important

Q16 Consider a situation: you choose an audiobook title to listen to, and there is a broad choice of your favourite books narrated by an AI system and a human. When you choose the title, would you consider AI-read books along with human narrated? (SingleSelection)

- Absolutely would not consider
- Would not consider
- Neutral
- Would consider
- Definitely would consider

Q17 If you want to listen to a particular audiobook and it's available only in AI-narrative, would you still listen to it? (SingleSelection)

- Absolutely would listen
- Would not listen
- Neutral
- Would listen
- Definitely would listen

Q18 If there was a choice of the same book narrated by a human and AI, which one would you choose? (SingleSelection)

- Human
- AI

Q19 Please, motivate your answer in the previous question.{{refId:f097fd59-31b9-49ad-9f6b-8fd33d3945f8}} (OpenEnded)

-

Q20 Consider a situation: next time you search for an audiobook title to listen to, there is a broad selection of your favourite books narrated by an AI. How likely is it that you would listen to an AI-narrated audiobook? (SingleSelection)

- Extremely unlikely
- Unlikely
- Neutral
- Likely
- Extremely likely

Q21 Consider a situation: next time you search for an audiobook title to listen to, there is a broad selection of your favourite books narrated by an AI. You would plan to listen to an AI-narrated audiobook: (SingleSelection)

- Extremely false
- Probably false
- Neither true nor false
- Probably true
- Extremely true

Q22 Consider a situation: next time you search for an audiobook title to listen to, there is a broad selection of your favourite books narrated by an AI. Would you intend to listen to AI-narrated audiobooks? (SingleSelection)

- Absolutely not
- Probably not
- Might or might not
- Probably yes
- Definitely yes

Q23 Consider a situation: next time you listen to your favourite/interesting book narrated by an AI system. How pleasant would you expect your listening experience to be? (SingleSelection)

- Extremely unpleasant
- Unpleasant
- Neutral
- Pleasant
- Extremely pleased

Q24 Consider a situation: next time you listen to your favourite/interesting book narrated by an AI system. How enjoyable would you find it? (SingleSelection)

- Extremely unenjoyable
- Unenjoyable
- Neutral
- Enjoyable
- Extremely enjoyable

Q25 Consider a situation: next time you listen to your favourite/interesting book narrated by an AI system. How efficient would you find it? (SingleSelection)

- Not efficient at all
- Slightly efficient
- Moderately efficient
- Very efficient
- Extremely efficient

Q26 Consider a situation: next time you listen to your favourite/interesting book narrated by an AI system. How emotional would you expect it to be? (SingleSelection)

- Extremely unemotional
- Unemotional
- Neutral
- Emotional
- Extremely emotional

Q27 Consider a situation: people who are important to you, such as friends and family, would listen to an AI-narrated audiobook. How likely is it that he or she would like to do it? (SingleSelection)

- Extremely unlikely
- Unlikely
- Neutral
- Likely
- Extremely likely

Q28 Consider a situation: next time you listen to your favourite/interesting book narrated by an AI system. Would the people important to you, such as friends and family, approve of your choice? (SingleSelection)

- They would NOT approve
- They would NOT approve
- Neutral
- They would approve
- They would approve

Q29 Please share your positive/negative views as a consumer of the widespread introduction of AI-narration. Do you think it should be introduced and which genres/types of books, in your opinion, would fit the best? Thank you, your answers are extremely valuable for the industry. (open-ended)