Bachelor thesis

Simone van Horen

s1080614

simone.vanhoren@ru.nl

Deepfake Audio Detection on YouTube: Influences of Source Platform, Speaker, Language, and Age

Wordcount: 6823

06-06-2024

International Business Communication

Faculty of Arts, Radboud University

Supervised by Cristian Tejedor-Garcia

Abstract

Previous studies have highlighted deepfake as one of the most dangerous forms of deception. However, studies investigating the human ability to detect deepfake is scarce, especially in music audios. Therefore, this experiment investigates the human ability to detect deepfake audios on YouTube. Twelve audios, specifically songs, were displayed to n = 68 individuals and they were asked if they thought the audio was real. The experiment was carried out in English to determine whether native language affects the detection accuracy. In addition, it was tested whether age, presence of source and speaker familiarity affect detection accuracy of real and fake audios, as these variables are important in the deepfake field. The experiment was carried out in an online questionnaire, as a within-subject design. The whole sample showed an effect of speaker familiarity and age, but only for the fake audios. The more voices were recognized, the higher the detection accuracy. Regarding age, the detection accuracy decreased as the age increased. No effect was found for native language and presence of source. After analysing the bar chart of each audio to see the individual responses on the scale, it became clear that people are not really good at detecting deepfake audios. The implications are explored in the discussion section.

Deepfake, audio, YouTube, detection ability

<u>Table of contents</u>

Introduction

Nowadays, the Internet is playing an important role in society. It is involved in people's lives and it influences the way people think and perceive things (Gambín et al., 2024). According to the report on the website Statista (2024), there were 5.35 billion internet users worldwide as of January 2024. This represents 66.2 percent of the global population. Of this 66.2 percent, 62.3 percent were social media users, which are 3.33 billion people. With social media platforms, it is easy to distribute information easy and quickly (Bengtsson & Johansson, 2022). For instance, on YouTube, people are able to watch, upload and share videos, and interact with each other (Liikkanen & Salovaara, 2015).

The information on the Internet does not always necessarily need to be true, there is a possibility for the information to be fake (Borges-Tiago et al., 2020). In recent years, the spread of fake news has become an issue (Figueira & Oliveira, 2017). People are even able to create fake content themselves and share this online. Because of the growing importance of social media, fake news can be distributed at a fast rate (Figueira & Oliveira, 2017). In 2017, the term deepfake appeared on the Internet (Westerlund, 2019). It is a form of artificial intelligence and it is possible to create fake images, videos, speech and texts. On YouTube, it is easy to disseminate this fake information. In addition, nowadays, a deepfake audio is created with only a few real audio samples (Choi et al., 2020). It is expected that in 2026, 90% of online content will be synthetically generated (Schick, 2020). Therefore, it is important to know how people interact on social media, with the presence of deepfakes (Hancock & Bailenson, 2021). It is even more important to understand whether people are able to detect whether information they encounter is real or fake (Shu et al., 2017). Previous research mainly focusses on the human ability to detect video or image deepfakes, only a few studies focus on audio deepfakes (Kimberly et al., 2023; Müller et al., 2021; Watson et al., 2021). In these studies, participants are specifically presented with real and fake audios. After listening to these audios, the participants are asked to identify the audios.

In this thesis, the problem of detecting deepfake audios on YouTube will be investigated, as YouTube is the platform with the largest amount of deepfake content (Poon, 2021). At the end of this thesis, an understanding of the relationship between detecting deepfake audios on YouTube and use of social media source, speaker familiarity, age and native language will be reached. This study will contribute to the scarce research on audio deepfake detection.

Theoretical framework

Social media

In the evolving digital world, social media platforms have become increasingly popular. Social media is defined as a digital technology that allows the sharing of ideas and information, including text and visuals, through virtual networks and communities (Kaplan & Haenlein, 2010). Social media has been studied in relation to a wide range of topics, including their role in politics, social relationships, activism, identity construction, and youth cultures.

Especially young adults are using social media (Shannon et al., 2022). In 2016, 97.5% of the adolescents in the US noted using at least one social media platform frequently (Villanti et al., 2017). In the Netherlands, over 15 million of the almost 18 million people are active social media users and 55% of this 15 million use social media for media sharing on platforms as Snapchat, Instagram and YouTube (Statista, 2024). This means that people obtain a lot of information from social media. The main power of social media platforms is giving people the chance to become content distributors (Weeks & Holbert, 2013). Social media use (SMU) is now seen as one of the most popular leisure activities (Lenhart et al., 2010). Therefore, SMU contributes to the dissemination of deepfake content (Westerlund, 2019).

YouTube

YouTube was founded in 2005 and it enables people to watch, upload and share videos (*YouTube*, n.d.). It is a popular social media platform to listen to music, by means of audiovisual content and YouTube is a common platform for celebrities to post their songs with a video (Liikkanen & Salovaara, 2015). Music videos hold top positions in different YouTube charts.

After Google, YouTube is the most visited website in the world (Dean, 2024). Around 2.49 billion people use YouTube, which is almost a third of the world population, and 300,000 new videos are posted every day. In addition, the platform is relevant in the Netherlands as Statista (2024) reports that YouTube became popular in 2015 and in 2020, over 9 million people were using this platform in the Netherlands.

As mentioned before in the introduction, YouTube is known for hosting the largest amount of deepfake content on the surface web, besides closed discussion boards and forum-based platforms, such as Reddit (Poon, 2021). Therefore, Google has come up with a policy regarding deepfake videos. Digwatch (2023) states that ‘The policy focuses on creator disclosures and labelling for AI-generated content, with plans for disclaimers on YouTube for video descriptions and within videos’. Lee et al. (2021) investigated the impact of the 10 most watched deepfake videos on YouTube. In the study they report on when YouTube deepfake videos are shown with commentaries, people think the deepfake is realistic, but it leads to a less positive attitude. Even with the knowledge of a video being fake, it led to fear or distrust in the technology.

Deepfake

Deepfakes, Explained | MIT Sloan (2020) explains that ‘A deepfake refers to a specific kind of synthetic media where a person in an image or video is swapped with another person's likeness’ and in 2017 this term appeared on the internet for the first time. This technology developed and more recently, deepfake technology can create audios that sound like somebody else. The term consists out of ‘deep learning’ and ‘fake’ and it is a form of artificial intelligence (Westerlund, 2019). The deep learning system will study photographs or videos of a certain person and mimic their speech and behaviour. Nowadays, there are deepfakes in the forms of images, videos, audios and text.

Before the appearance of deepfake technology, fake news has been an issue (Borges-Tiago et al., 2020). According to Aldwairi and Alwahedi (2018, p.1) ‘fake news refers to fictitious news style content that is fabricated to deceive the public’. It can range from simple hoaxes to sophisticated disinformation campaigns designed to influence public opinion or sow discord within communities (How to Spot “fake News” Online, n.d.). As of 2024, a lot of information is shared via social networks, fake news can spread through society very quickly (Figueira & Oliveira, 2017). Due to this quick distribution, it is hard to know what is true and what to trust online (Borges-Tiago et al., 2020).

In summary, fake news involves the dissemination of false or misleading information through conventional media channels, whereas deepfake involves the creation of manipulated media using advanced artificial intelligence techniques. Both fake news and deepfake contribute to the broader landscape of misinformation and disinformation, posing challenges to media literacy, trust in information sources, and the integrity of digital content (*A Family Guide to Talking About Misinformation and Deepfakes*, n.d.).

Deepfake technology and development

In the early 2010s, deep learning techniques appeared such as convolutional neural networks (CNNs) and generative adversarial networks (GANs) (Gambín et al., 2024). Researchers explored the potential of neural networks for tasks such as image recognition, natural language processing, and computer vision (Sarker, 2021). In the field of speech and audio processing, these advance techniques are applied into medical or speech recognition applications (Gómez-Zaragozá et al., 2023; Parikh et al., 2023). In 2017, the term deepfake appeared from a Reddit user who used deep learning algorithms to superimpose celebrities' faces onto pornographic videos (Deepfakes, Explained | MIT Sloan, 2020). Deepfake technology quickly garnered attention for its ability to create highly realistic but fake videos, raising concerns about its potential for misuse and the spread of misinformation (Naffi et al., 2023).

Researchers and developers continued to refine deep learning models, improving their ability to generate and manipulate media content. Techniques such as autoencoders, variational autoencoders (VAEs), and recurrent neural networks (RNNs) were adapted for generating realistic images, videos, and audio (Sarker, 2021). Commercial applications and services for creating deepfake content emerged, offering user-friendly interfaces and pre-trained models for generating manipulated media. The democratization of deepfake technology made it accessible to a wider audience, including individuals with limited technical expertise (Gambín et al., 2024).

As concerns grew over the potential misuse of deepfake technology, policymakers, technological companies, and researchers began to explore regulatory measures and detection methods (Westerlund, 2019). Efforts focused on developing algorithms and tools for detecting and mitigating deepfake content, as well as raising awareness about its implications for privacy, security, and trust in media (Godulla et al., 2021). The development of deepfake technology continues to evolve, with ongoing research into more sophisticated algorithms, techniques, and applications. Challenges persist in distinguishing between authentic and manipulated media, addressing ethical concerns, and mitigating the negative consequences of deepfake technology on society (Hancock & Bailenson, 2021).

Deepfake technology alone is harmless and could be used with good intentions. Westerlund (2019) stated that deepfake technology is used ‘in fields as education, entertainment, online social media, healthcare, fashion, and marketing’. It is also already used in movies and TV shows to make the appearance of deceased actors. Deepfakes appear on social networks for entertainment or to make viral videos (Gambín et al., 2024). However, the harmful use of deepfake is greater than the harmless use. According to Nguyen et al. (2019), deepfakes can be used to create fake speeches of world leaders. It can cause political and religious tensions and influence the results of elections. Dobber et al. (2021) tested the attitude towards a politician after seeing a deepfake, meant to discredit the politician. Results show that people had a worse attitude towards the politician after seeing the deepfake. It can also create chaos in the financial world or generate fake satellite pictures.

Mirsky and Lee (2021) state that the evolution of deep neural networks and the availability of large amounts of data have ensured that the deepfake images and videos are hard for humans and even computer algorithms to spot. The creation of deepfake now is simple, and it does not need a lot of real material to create the deepfake. This endangers the privacy and identity also of individuals. Furthermore, the extensive use of social media networks makes it easy to share the deepfakes (Yazdinejad et al. 2020). In addition, a study done by Chadwick and Vaccari (2020) about deepfake in news and the effect on trust shows that after seeing a deepfake, the uncertainty about news media was higher. As deepfake posit threats to society, it is important that people are able to detect when they are being misled.

Human ability to detect deepfake

It is no surprise that the accessibility and dissemination of deepfakes have consequences on society (Hancock and Bailenson, 2021). They gained insight that deepfakes are deception (Levine, 2019). It is found that people are not really good at detecting deception and it is easy to believe something that is not true (Bond & DePaulo, 2006). This level of accuracy did not depend on the medium, it was relatively the same for messages conveyed through text, audio or a video (Hancock et al., 2009). This might be surprising as video holds more details, and the impact of deepfake video might be worse. Humans tend to be primed more by visual communication for perception (Posner et al., 1976).

Deepfake technology, especially deep voice, which has been derived from artificial intelligence in recent years, is potentially harmful, and the public is not yet aware (Chang, 2023). Not only humans but also machines struggle to identify deepfakes. Current speaker and facial recognition systems might be easily fooled by carefully prepared synthetic media – deepfakes (Firc et al., 2023). There are two main methods for creating deepfake speech: text-to-speech synthesis (TTS) and voice conversion (VC) (Gómez-Zaragozá et al., 2023; Parikh et al., 2023). Currently, for both methods, no specific detection method exists. However, there are general detection methods for speech. The problem with this is, as soon as the known database changes, the reliability and accuracy decrease, as the deepfake detectors do not know this new database. Therefore, it is important that humans are able to detect deepfake audios.

First, Müller et al. (2021) investigates the difference between humans and machine audio deepfake detection. ‘Audio deepfake detection, also termed audio spoofing detection, denotes the capability of identifying generated audio data’. They note that human deepfake detection mostly have been investigated with image or video deepfake. In these experiments, machines detect deepfake with a higher accuracy, but both humans and machines are easy to fool. They also report that people under 35 years old are better at detecting deepfake audio than people older than 35 years.

Watson et al. (2021) specifically investigates the human perception towards audio deepfake. The focus was on people at college, as they are exposed to social media the most. The language used was English, which was a limitation in the study of Müller et al. (2021), as that study was done in German. Participants had to listen to eight audio files and state whether they think the audios are real or fake. The audio files differed in grammar complexity and length. The length ranged from 3-10 seconds. Results show that accuracy did not differ between grade level, but complex audio files were easier to detect. Furthermore, short clips were detected more easily. A limitation of this study is that all participants were English, in the future it might be interesting to see how non-native English speakers react to deepfake audio.

Kimberly et al. (2023) investigated the human ability to detect deepfake speech, with accounting for different languages in audio, namely English and Mandarin. Thus, the experiment was carried out in both English and Mandarin. However, the participants carried the experiment out in their native language. The performance was similar between the languages and the detection abilities were limited for participants.

The current study

The aim of this study is to investigate whether people are able to detect deepfake audios on the platform YouTube and therefore are able to notice whether they are deceived. To the best of my knowledge, there is not a lot of research available on human abilities to detect deepfake audio, as this is a relatively new field. This is necessary to study in the evolving digital world and with the dangers of deepfake. The effect of the two chosen independent variables, use of source and speaker familiarity, is also not researched in the audio deepfake domain, but is relevant to investigate as they impact perception and detection. Based on the studies of Müller et al. (2021) and Watson et al. (2021), it will be investigated whether age and native language affect detection ability. The outcome of this study contributes to the effect of deepfake on society and the importance of detecting deepfake.

Source credibility

The source credibility theory by Hovland and Weiss (1951) explains how perceived trustworthiness and expertise of a message source affect the persuasiveness of a communication. They found that messages from credible sources were more effective in changing opinions than messages from less credible sources. For example, Jin et al. (2023) investigates the influence of heuristic cues on perceived credibility. Results show that 'the number of the source's followers and the video's popularity are positively associated with perceived credibility'. As YouTube is such a popular source in the world, it holds a certain power. It would make information more credible. In this thesis this will be investigated, by using screenshots of the YouTube platform to make the audio possibly more credible.

Ahmed (2021) carried out an investigation regarding informative cues used in deepfake videos. The deepfake videos were either deceptive or educational, depending on having informative cues. The deceptive deepfake videos were perceived as more accurate, which contained no informative cues. This raises a concern regarding well-made deepfake videos.

As mentioned before, humans are more primed by visual information than other sensory information when it comes to perception (Posner et al., 1976). This is called the Colavita visual dominance effect (Koppen & Spence, 2007). In the Colavita paradigm, 'participants have to make speeded responses to a random series of auditory, visual, or audiovisual stimuli' (Hancock and Bailenson, 2021). There is no problem when responding to audio or video separately, but when this is presented together, participants have problems with responding to the audio. It is also known that people recall visual messages more than verbal messages (Graber, 1990), and misleading visual information tends to provoke false perceptions more than misleading verbal content (Sundar, 2008). This is because of the realism heuristic, which means that people tend to trust audiovisual content more as it is more similar to the real world. This will be taken into account when presenting the audio files with a screenshot of the source.

Speaker familiarity

People are able to identify someone by listening to their voice (Hansen & Hasan, 2015). Interestingly, humans recognize individuals by their voices with a high accuracy, especially when they are familiar with the speaker. Many times, even a short non-linguistic queue, such as a laugh, is enough for us to recognize a familiar person (Eriksson, 2003).

Another phenomenon is the truthiness effect (Newman et al., 2015). This means that people are more likely to accept information as accurate when it is perceived as familiar. This would posit when people are familiar with the person speaking in the deepfake audio, they are able to detect less accurately if the audio is real or fake.

Native language

In the domain of audio deepfake detection, language has not been investigated extensively, as mentioned by Watson et al. (2021) and Kimberley et al. (2023). In this study, the audio files will be presented in English. According to Kilman et al. (2014), speech comprehension in a non-native language requires the full attention of the listener. In their study, when the speech is in the non-native language of the participant, they encountered more difficulties. This effect will be further investigated in this study, as the native language of the participants may differ.

Based on the previous literature, the following research questions and hypotheses are made (see Appendix 1):

RQ1. To what extent do the presence of the social platform (source) and speaker familiarity influence human accuracy in detecting deepfake audio on YouTube?

RQ2: Does age influence the ability to detect deepfake audio on YouTube?

RQ3: Does the native language influence the ability to detect English deepfake audio on YouTube?

H0: There is no effect of presence of source, speaker familiarity, age and native language on audio deepfake detection accuracy.

H1: If there is a source present, the accuracy will be lower.

H2: If the participant is familiar with the speaker, the accuracy will be lower.

H3: The younger the participants are, the higher the accuracy will be

H4: Native English speakers are better at detecting English deepfake audio on YouTube

Methodology

Materials

The current study has four independent variables (see Appendix 1), presence of source, speaker familiarity, native language and age. The materials used are 12 audio files (8-15 seconds long), 6 deepfake and 6 authentic audio files. Each group of 6 audios has 3 audios accompanied by a source and 3 audios not accompanied by a source, thus there are 4 conditions. The audio files are famous people singing a part of a song. Every audio contains the voice of a different person and the audio is in English. All audio files were taken from YouTube.

Presence of source

The audio files were presented with a YouTube picture or without a YouTube picture. It was a screenshot of the template you see when clicking on a video (see Appendix 2). Only the video screen was presented with the play button, play bar, next button, time of the video, subtitles button and screen option buttons. To keep the conditions similar, every picture contained the mark of Vevo and the normal template of YouTube containing the logo, next video options, search bar, YouTube account and comments were left out. This was done to test whether the detection accuracy was different when the source is present.

Speaker familiarity

The audio files contain the voices of famous singers. To keep the experiment equal, 6 female voices and 6 male voices were used. The voices used are from Adam Levine, Kanye West, Dua Lipa, Beyoncé, Pink, Ariana Grande, Billie Eilish, Justin Timberlake, Lil Uzi Vert, Taylor Swift and Jason Derulo. The participants did either know the voice or not and they had to state this after listening to the audio files with answering a yes/no question. This was done to check whether people who know the speaker were better at determining whether the audio was fake or not.

Subjects

In total, 99 participants started the questionnaire, 68 participants (68.7%) finished the questionnaire and were used for this study. Since the experiment took place in English, the native language of the participants was important. The study aimed to investigate a general effect of presence of source and speaker familiarity on detection accuracy, therefore there were no requisites regarding educational level and gender. People were not asked for their educational level, as this was not important for this study. The minimum age was 16 years old, there was no maximum.

Table 1 shows the distribution of gender, age and native language of the participants. 24 participants were male (35.3%), 42 were female (61.8%), 1 was non-binary (1.5%) and 1 preferred not to say their gender (1.5%). The age varied from 16 to 80 years old (M= 31.2, SD= 17.7). Lastly, there were 8 different native languages. 52 participants were Dutch (76.5%), 4 were German (5.9%), 3 were English (4.4%), 3 were Spanish (4.4%), 3 were Arabic (4.4%), 1 was French (1.5%), 1 was Hungarian (1.5%) and 1 was Portuguese (1.5%).

To test whether detection accuracy was influenced by gender, a Chi-square test was carried out. A Chi-square test did not show a significant relation between gender and detection accuracy ($\chi2$ (72) = 59.54, p = .853).

Table 1

Demographic Characteristics of Participants

Characteristic	*n*	Percentage
Gender		
Male	24	35.3%
Female	42	61.8%
Non-binary	1	1.5%
Prefer not to say	1	1.5%
Native language		
Dutch	52	76.5%
German	4	5.9%
English	3	4.4%
Spanish	3	4.4%
Arabic	3	4.4%
French	1	1.5%
Hungarian	1	1.5%
Portuguese	1	1.5%

Note. Total *n*= 68. Participants were on average 31 years old (*SD* = 17.7)

Design

The present study had a within-subjects design, every participant listened to the 12 same audio files that are real or fake. They answered the same questions afterwards. The order was randomized.

Instruments

The study was carried out with a questionnaire on Qualtrics. The dependent variable was detection accuracy. To test whether the participants think the audio file is real or not, they had to answer a question after listening, on a 7-point Likert scale. The question was: ‘How do you feel about the following statement? Rate this on a scale from 1 (strongly disagree) to 7 (strongly agree)'. The statement was 'this audio is real'. The score for detection accuracy ranged from 1-7, where 1 is very low and 7 is very high.

For the fake audio files, the polarity of the scale was reversed. The reliability of 'detection accuracy of fake audios' comprising six items was medium: $\alpha = .389$. The reliability of 'detection accuracy of real audios' comprising six items was medium: $\alpha = .384$. The reliability would not significantly increase when one item was deleted. After this, two new variables were constructed.

For the variable speaker familiarity, it was also necessary to compose two new variables. The reliability of 'speaker familiarity of fake audios' comprising six items was high: $\alpha = .827$. The reliability of 'speaker familiarity of real audios' comprising six items was high: $\alpha = .845$. Therefore, two new variables were constructed.

Procedure

The online questionnaire was distributed mainly via WhatsApp, but also via Instagram stories. The questionnaire was online from 22-04-2024 until 01-05-2024. The participants were sent a hyperlink to a Qualtrics survey with an informed consent. The questionnaire was filled out individually. The link was sent out with a small text saying 'Hello everyone! I am currently conducting my Bachelor's thesis study about the quality of audios on YouTube. Would you please fill in this quick survey (5 minutes)? Thank you in advance!'. The real purpose of the study was not mentioned beforehand, they were told that the study was about the quality of audios. They were asked to share the Qualtrics survey with other people to ensure a high response rate. There was no reward or incentive for the subjects who participated in the survey. Before starting the experiment, the participants were asked to either use headphones or go to a quiet setting where they can listen to the audio files carefully. It was possible to listen to every audio multiple times.

In the questionnaire, participants were asked to fill in their age, gender and native language before starting the real experiment. They were informed beforehand that answering was completely anonymous. If they wished to participate, they could continue the study. The procedure was the same for every participant, except that the order of audios and speakers was randomized. After these questions, the real purpose of the experiment was revealed and participants were asked if they would have answered differently if they knew the purpose before.

Statistical treatment

Before deciding which tests to carry out, tests of normality and homogeneity were carried out. Based on this, the following treatments were chosen.

A one-way MANOVA with native language as independent variable was carried out. The two dependent variables were detection accuracy of fake audios and detection accuracy of real audios.

For the detection accuracy of real audios, a paired samples t-test was carried out. For the detection accuracy of fake audios, a Wilcoxon signed rank test was carried out.

Lastly, four Pearson's correlation analysis were carried out individually for age and speaker familiarity, again differentiating between real and fake audios.

Results

Before starting the statistical tests, the data needs to be checked for a normal distribution and on homogeneity. This was done for the variables native language and presence of source. The difference was made between detection accuracy for real and fake audios. First, for native language, a Shapiro-Wilk test showed that the normality for both fake and real audios higher was than $p = .063$ for Dutch, English, German and Spanish. However, only Arabic showed $p = .000$ for real audios and $p = .726$ for fake audios. Therefore, Arabic was omitted for further testing. Hungarian, Portuguese and French were not tested because only 1 participant belonged to these groups. In addition, a Levene's test for native language showed that the groups are homogenous for both fake audios ($L = 0.50$, $p = .733$) and real audios ($L = .67$, $p = .614$).

Furthermore, presence of source was tested for normal distribution. A Shapiro-Wilk test for the fake audios showed that the data was not normally distributed for audios with source ($p = .025$) and without source ($p = .049$). The data of real audios was normally distributed for both audios with source ($p = .322$) and audios without source ($p = .120$). The values of the pre-tests are shown in Table 2.

Table 2

Pre-tests normality

Variable	Shapiro-Wilk (fake)	Shapiro-Wilk (real)
NL Arabic	$p = .726$	$p < .001$
NL Dutch	$p = .063$	$p = .173$
NL English	$p = .253$	$p = .702$

NL German	$p = .900$	$p = .233$
NL Spanish	$p = .537$	$p = .144$
With source	$p = .025$	$p = .322$
Without source	$p = .049$	$p = .120$

Note. Total $n = 68$. NL= Native Language.

First, to check whether the detection accuracy was different for the fake audios than the real audios, a paired samples t-test was carried out. A paired samples t-test showed a statistically significant difference between the real audios and the fake audios. This difference, -0.9, 95% CI [-1.22, -0.58] was significant, $t\,(67) = -5.63$, $p < .001$ and represented a moderate-sized effect, $d = -0.68$. Detection accuracy for the real audios ($M = 4.72$, $SD = 0.86$) was shown to be higher than detection accuracy for the fake audios ($M = 3.83$, $SD = 0.85$). The mean and standard deviation are presented in table 3.

A significant negative correlation was found between age of the subject and detection accuracy of fake audios ($r = -.307$, $p = .011$, $N = 68$). The detection accuracy of fake audios decreased with the age of the participants. This means the older the people are, the lower the detection accuracy.

A significant negative correlation was not found between age of the subject and detection accuracy of real audios ($r = -.082$, $p = .508$, $N = 68$). The detection accuracy of real audios did not change with the age of the participants.

A significant positive correlation was found between speaker familiarity and detection accuracy of fake audios ($r = .390$, $p = .001$, $N = 68$). The detection accuracy increased with the speaker familiarity. This means the more voices people knew, the better they were at recognizing the deepfake audio.

No significant positive correlation was found between speaker familiarity and detection accuracy of real audios ($r = .076$, $p = .540$, $N = 68$). The detection accuracy did not increase with the speaker familiarity. This means the more voices people knew, the detection accuracy did not change. A summary of the values for the correlations can be found in table 5.

A paired samples t-tests did not show a difference between detection accuracy with picture and detection accuracy without picture for real audios. This difference, -0.04, 95% CI [-0.41, 0.32] was not significant, $t\,(67) = -.242$, $p = .810$ for real audios.

A Wilcoxon signed rank test did not show a difference between detection accuracy with picture and detection accuracy without picture for fake audios. This difference was not significant $T = 956$, $p = .762$. The means and standard deviations are also presented in table 3.

A one-way multivariate analysis for detection accuracy of fake audios and detection accuracy for real audios, with native language as factor, did not find a significant multivariate effect of native language (F (2, 12) = 1.026, p = .430). The univariate analyses did not show an effect of native language on detection accuracy of fake audios (F (6, 58) = 1.132, p = .355) and on detection accuracy of real audios (F (1, 18) = 0.700, p = .651). The means and standard deviations of native language are presented in table 4.

Table 3

Descriptive statistics for detection accuracy

Detection accuracy	*M (SD)*	*M (SD)*	*p*-value
Fake audios	3.83 (0.85)		<.001
Real audios	4.72 (0.86)		<.001
Fake audios*	3.78 (1.02)	3.88 (1.14)	0.762
Real audios*	4.70 (1.13)	4.75 (1.15)	0.810

Note. Total n = 68. *= first is with picture, second is without picture.

Table 4

Descriptive statistics for native language

Native language	*M (SD)* (fake)	*M (SD)* (real)
(Arabic)		
Dutch	3.82 (0.87)	4.75 (0.87)
English	4.11 (0.63)	4.33 (0.93)
French	3.33 (-)	3.33 (-)
German	3.46 (0.55)	4.54 (0.93)
Hungarian	4.67 (-)	5.50 (-)
Portuguese	3.50 (-)	5.00 (-)
Spanish	4.83 (0.60)	4.89 (1.11)
p-value	0.355	0.651

Note. Total n = 65. French, Hungarian and Portuguese had 1 participant, so no *SD*.

Table 5

Correlation for detection accuracy

Factor	Pearson (p)
Age (fake)	-.307 (0.011)

Age (real)	-.082 (0.508)
Speaker familiarity (fake)	.390 (0.001)
Speaker familiarity (real)	.076 (0.540)

Note. Total *n* = 68.

Discussion

It was found that humans are able to detect audio deepfakes, but not persistently. The first hypothesis presented that when a source is present, the detection accuracy will be lower. On the contrary, the present study does not show a difference in detection accuracy of audios presented with a source and audios without a source. Therefore, H1 cannot be confirmed. This does not correspond with the findings of Hovland and Weiss (1951) and Jin et al. (2023). Hovland and Weiss (1951) explain that messages from credible sources were more effective in changing opinions than messages from less credible sources. In addition, Jin et al. (2023) showed that the number of followers and video's popularity affect the credibility. A possible explanation could be that people did not recognize the picture as a source from YouTube. Appendix 3 shows the bar charts of answers to every separate audio. Rows 1 and 3 represent the rows with source, row 2 and 4 without source. For the fake audios (row 1), the charts show a tendency to the right, which means that the participants thought the audio was real. For the other 3 fake audios (row 2), there was no tendency to each of the sides. Furthermore, for the real audios (row 3), the tendency was also more to the right and people thought it was real, which was the correct answer. The 3 other real audios (row 4) also showed no tendency to either the left or right side. Thus, the materials could be an explanation for the occurred results, as there was no clear tendency.

Secondly, it was hypothesized that when participants are familiar with the speaker's voice, the accuracy will be lower. The opposite is true for this study, people were better at detecting deepfake audios when they recognized more voices and therefore H2 cannot be confirmed. This corresponds with the findings of earlier studies, namely Hansen and Hasan (2015) showed that people are able to recognize voices of individuals, especially when they know the speaker and Eriksson (2003) even showed a short audio is enough to recognize a familiar person. On the contrary, the truthiness effect explained by Newman et al. (2015) does not occur in the present study. They could detect deepfake audios even when the voice was familiar. This only occurs for the deepfake audios, no effect was found for the real audios. A possible explanation for the occurred effect could be that when people recognized the voice, it was easier for them to realize the audio was fake. This is especially true for songs, as this was used in this study and not a random audio. When they have heard the voice in another song or recognized the specific song, it is more likely that they will have the voice in their head during the experiment and therefore could better detect the fake audio.

Thirdly, it was expected that younger people were better at detecting deepfake audios. Müller et al. (2021) showed that people under 35 years old are better at detecting deepfake audios than people over 35 years old. The present study showed a similar effect, as detection accuracy decreases as the age increases. H3 can be confirmed. An explanation for this could be that older people are less familiar with social media and online interaction (Shannon et al., 2022).

Lastly, native speakers were expected to be better at detecting deepfake audios, in this case that were native English speakers as the audios were in English. However, this study did not show an effect of native language on detection accuracy. Therefore, H4 cannot be confirmed. This coincides with the results of Kimberley et al. (2023) but not with the study of Kilman et al. (2014). According to the second study, it should be more difficult for non-native speakers to detect deepfake, which is not the case. A possible explanation for this could be that the present study only had 3 English native speakers, the majority were Dutch. However, Dutch people are an exception, as their English proficiency is the highest in the world (EF EPI, 2022). Therefore, it would have been expected that Dutch people are also better at detecting deepfake audios than the other native languages, but this is not the case.

Limitations

Every audio was in English and contained a voice of different people. As discussed before, the images used to present YouTube as a source could have been not clear to the participants. The sample size of the experiment was also relatively small, compared to the study of Müller et al. (2021) for example, which had more than 200 participants.

In addition, the study has a within-subject design. Possible downsides of this design are that participants might try to guess the purpose of the study and alter their behaviour to match their perceptions, leading to biased results or participants might perform differently based on the order in which they experience the conditions or treatments. For example, they might get better at a task simply because they've done it before (practice effect) or get tired or bored as the experiment progresses (fatigue effect).

As mentioned before, the reliability of the scales for detection accuracy was relatively low. This means that the items do not consistently measure the same underlying construct or idea. A possible explanation could be that the items in the scale are measuring very similar aspects of the construct, without capturing its full variability. The real purpose of the study was not mentioned beforehand and in a few cases, participants found the statement 'this audio is real' confusing and did not understand the meaning confounding with the quality of audios, which was the fake purpose. In addition, almost a third of the participants stated that if they had known the real purpose, they would have taken this into account and listened more carefully. Eight participants found the experiment difficult.

A possible explanation and limitation for this study are the materials, as are the questions and scales. Appendix 3 shows the bar charts of every separate audio. The first six are of the fake audios, the other six of the real audios. For the fake audios, the desired answer should be on the left, however, almost no audio shows a tendency to the left. On the other hand, for the real audios, with the desired answer on the right, it can be noticed that some charts have a tendency to the right, but not all. Therefore, it can be concluded that the materials are a problem of this study. It might be that it was too difficult for the participants or that the audios did not have anything in common.

Lastly, the majority of the participants knew every singer, except for 2 singers (Adam Levine and Lil Uzi Vert). Therefore, it can be concluded that most singers were famous, and no distinction was really made when the majority of the people did not know the voice.

Future research

Based on the previous findings, suggestions for future research are to carry out a similar experiment, but then changing the language of the experiment to the expected participants' language. A similar study could also be done with speakers in another field than music to investigate speaker familiarity. Examples could be marketing or politics. Or it would be interesting to see what results will be yielded when infamous speakers are used. It would also be interesting to investigate presence of source in other fields, as advertising, marketing or politics. Another suggestion is to carry out a similar study but then with a between-subjects design or a bigger sample. My hope from this study is also to highlight the importance of educating people about deepfake and making sure that they are aware of this on the Internet, and maybe they will be able to identify audio deepfakes.

Implications

The findings have implications for the music industry, as it is easy to clone a voice and that people are not really good at noticing the difference between real songs and deepfake songs. In technology policy, YouTube and other social media platforms could focus on alerting users to potentially manipulated content, to also protect the original content. In governance, policymakers can draft guidelines to protect the more vulnerable section of society (e.g., older people) against misinformation by including informational cues on

YouTube posts to buffer against misinformation. In addition, education efforts could be tailored to different age groups, emphasizing the development of skills to recognize deepfake content, especially targeting older adults. Improving AI algorithms for detecting deepfakes could take into account the influence of speaker familiarity, possibly enhancing accuracy when familiar voices are involved.

Conclusion

Summarized, this study is one of the few that investigates the human ability to detect deepfake audios on YouTube. It is explored what role native language, age, speaker familiarity and presence of source play in the deepfake detection field. Our study shows that fake audios are more difficult to detect than real audios. To answer the three research questions, statistical analyses were carried out. A significant effect was found of age and speaker familiarity, regarding the deepfake audios. The older the people were, the more they were deceived by deepfake audios. In addition, the more voices the participants recognized, the better the detection accuracy was. Future research could focus on carrying out a similar experiment but in a different language, using infamous singers, a between-subject design or a bigger sample size. In addition, it would be a suggestion to carry out an experiment with deepfake audios in a different field, as advertising or marketing. Lastly, this study has implications for the music industry, technology policy, governance and education.

Bibliography

A family guide to talking about misinformation and deepfakes - Hwb. (n.d.). https://hwb.gov.wales/keeping-safe-online/resources/a-family-guide-to-talking-about-misinformation-and-deepfakes

Ahmed, S. (2021). Fooled by the fakes: Cognitive differences in perceived claim accuracy and sharing intention of non-political deepfakes. *Personality and Individual Differences*, *182*, 111074. https://doi.org/10.1016/j.paid.2021.111074

Ahmed, S., & Wen, C. H. (2023). Perception and deception: Exploring individual responses to deepfakes across different modalities. *Heliyon*, *9*(10), e20383. https://doi.org/10.1016/j.heliyon.2023.e20383

Aldwairi, M., & Alwahedi, A. (2018). Detecting fake news in social media networks. *Procedia Computer Science*, *141*, 215–222. https://doi.org/10.1016/j.procs.2018.10.171

Bengtsson, S., & Johansson, S. (2022). The meanings of social media use in everyday life: filling empty slots, everyday transformations, and mood management. *Social Media + Society*, *8*(4), 205630512211302. https://doi.org/10.1177/20563051221130292

Bond, C. F., & DePaulo, B. M. (2006). Accuracy of deception judgments. *Personality and Social Psychology Review*, *10*(3), 214–234. https://doi.org/10.1207/s15327957pspr1003_2

Borges-Tiago, T., Tiago, F., Silva, O., Martínez, J. M. G., & Botella-Carrubi, D. (2020). Online users' attitudes toward fake news: Implications for brand management. *Psychology & Marketing*, *37*(9), 1171–1184. https://doi.org/10.1002/mar.21349

Chang, H. (2023). Deepfake Speech Recognition and Detection. *International Journal of Pattern Recognition and Artificial Intelligence*, *37*(09). https://doi.org/10.1142/s0218001423500155

Choi, S., Han, S., Kim, D., & Ha, S. (2020). Attentron: Few-Shot Text-to-Speech Utilizing

Attention-Based Variable-Length Embedding. Interspeech 2020, 21st Annual Conference of the International Speech Communication Association, Virtual Event, Shanghai, China, 25-29 October 2020. ISCA. https://doi.org/10.21437/interspeech.2020-2096

Dean, B. (2024, February 8). *How many people use YouTube? [New data]*. Backlinko. https://backlinko.com/youtube-users

Deepfakes, explained | MIT Sloan. (2020, July 21). MIT Sloan. https://mitsloan.mit.edu/ideas-made-to-matter/deepfakes-explained

Digwatch (2023, December 11). Google takes steps to regulate deepfake content on YouTube |Digital Watch Observatory. *Digital Watch Observatory*. https://dig.watch/updates/google-takes-steps-to-regulate-deepfake-content-on-youtube

Dobber, T., Metoui, N., Trilling, D., Helberger, N., & De Vreese, C. (2021). Do (Microtargeted) deepfakes have real effects on political attitudes. *The International Journal of Press/Politics*, *26*(1), 69–91. https://doi.org/10.25384/sage.c.5073707.v1

EF EPI. (2022). EF English Proficiency Index. In EF EPI [Report]. https://www.ef.com/assetscdn/WIBIwq6RdJvcD9bc8RMd/cefcom-epi-site/reports/2022/ef-epi-2022-english.pdf

Eriksson, A. (2003). Tutorial on Forensic Speech Science Part I: Forensic Phonetics. *Proc. European Conf. Speech Communication and Technology*. https://www.york.ac.uk/media/languageandlinguistics/documents/currentstudents/Eriksson_tutorial_paper.pdf

Figueira, Á., & Oliveira, L. (2017). The current state of fake news: challenges and opportunities. *Procedia Computer Science, 121*, 817–825. https://doi.org/10.1016/j.procs.2017.11.106

Firc, A., Malinka, K., & Hanáček, P. (2023). Deepfakes as a threat to a speaker and facial recognition: An overview of tools and attack vectors. *Heliyon*, *9*(4), e15090. https://doi.org/10.1016/j.heliyon.2023.e15090

Gambín, Á. F., Yazidi, A., Vasilakos, A. V., Haugerud, H., & Djenouri, Y. (2024). Deepfakes: current and future trends. *Artificial Intelligence Review*, *57*(3). https://doi.org/10.1007/s10462-023-10679-x

Godulla, A., Hoffmann, C., & Seibert, D. (2021). Dealing with deepfakes – an interdisciplinary examination of the state of research and implications for communication studies. Studies in Communication, Media, 10(1), 72–96. https://doi.org/10.5771/2192-4007-2021-1-72

Gómez-Zaragozá, L., Wills, S., Tejedor-Garcia, C., Marín-Morales, J., Alcañiz, M., & Strik, H. (2023). Alzheimer Disease Classification through ASR-based Transcriptions: Exploring the Impact of Punctuation and Pauses. arXiv preprint arXiv:2306.03443.

Graber, D. A. (1990). Seeing Is Remembering: How Visuals Contribute to Learning from Television News. Journal of Communication, 40(3), 134–156. https://doi.org/10.1111/j.1460-2466.1990.tb02275.x

Hancock, J. T., & Bailenson, J. N. (2021). The social impact of Deepfakes. *Cyberpsychology, Behavior, and Social Networking*, *24*(3), 149–152. https://doi.org/10.1089/cyber.2021.29208.jth

Hancock, J. T., Woodworth, M., & Goorha, S. (2009). See No Evil: The Effect of Communication Medium and Motivation on Deception Detection. *Group Decision and Negotiation*, *19*(4), 327–343. https://doi.org/10.1007/s10726-009-9169-7

Hansen, J. H. L., & Hasan, T. (2015). Speaker Recognition by Machines and Humans: A tutorial review. *IEEE Signal Processing Magazine*, *32*(6), 74–99. https://doi.org/10.1109/msp.2015.2462851

Hovland, C. I., & Weiss, W. (1951). The influence of source credibility on communication effectiveness. *Public Opinion Quarterly, 15*(4), 635. https://doi.org/10.1086/266350

How to spot "fake news" online. (n.d.). Beyond Fake News. https://www.bbc.co.uk/beyondfakenews/how-to-spot-fake-news-online/

Jin, X., Zhang, Z., Gao, B., Gao, S., Zhou, W., Yu, N., & Wang, G. (2023). Assessing the perceived credibility of deepfakes: The impact of system-generated cues and video characteristics. *New Media & Society*. https://doi.org/10.1177/14614448231199664

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! The challenges and opportunities of Social Media. Business horizons, 53(1), 59-68.

Kimberly, T., Bray, S. D., Davies, T., & Griffin, L. D. (2023). Warning: Humans cannot reliably detect speech deepfakes. *PLOS ONE*, *18*(8), e0285333. https://doi.org/10.1371/journal.pone.0285333

Koppen, C., & Spence, C. (2007). Seeing the light: exploring the Colavita visual dominance effect. Experimental Brain Research, 180(4), 737–754. https://doi.org/10.1007/s00221-007-0894-3

Lee, Y., Huang, K., Blom, R., Schriner, R., & Ciccarelli, C. (2021). To believe or not to believe: Framing analysis of content and audience response of top 10 deepfake videos on YouTube. *Cyberpsychology, Behavior, and Social Networking, 24*(3), 153–158. https://doi.org/10.1089/cyber.2020.0176

Lenhart, A., Purcell, K., Smith, A., & Zickuhr, K. (2010). Social Media & Mobile Internet Use among Teens and Young Adults. Millennials. *Pew Internet & American Life Project.* http://files.eric.ed.gov/fulltext/ED525056.pdf

Levine, T. R. (2019). *DUPED: Truth-Default Theory and the Social Science of Lying and Deception*. https://muse.jhu.edu/book/70119/

Liikkanen, L. A., & Salovaara, A. (2015). Music on YouTube: User engagement with traditional, user-appropriated and derivative videos. *Computers in Human Behavior, 50*, 108–124. https://doi.org/10.1016/j.chb.2015.01.067

Mirsky, Y., & Lee, W. (2021). The creation and detection of deepfakes. *ACM Computing Surveys*, *54*(1), 1–41. https://doi.org/10.1145/3425780

Müller, N. M., Markert, K., & Williams, J. L. (2021). Human perception of audio deepfakes. *arXiv (Cornell University)*. https://doi.org/10.48550/arxiv.2107.09667

Naffi, N., Charest, M., Danis, S., Pique, L., Davidson, A., Brault, N., Bernard, M., & Barma, S. (2023). Empowering Youth to combat malicious deepfakes and disinformation: an experiential and reflective learning experience informed by personal construct theory. *Journal of Constructivist Psychology*, 1–22. https://doi.org/10.1080/10720537.2023.2294314

Newman, E. J., Garry, M., Unkelbach, C., Bernstein, D. M., Lindsay, D. S., & Nash, R. A. (2015). Truthiness and falsiness of trivia claims depend on judgmental contexts. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *41*(5), 1337–1348. https://doi.org/10.1037/xlm0000099

Nguyen, T. T., Nguyen, Q. V. H., Nguyen, D. T., Nguyen, D. T., Huynh-The, T., Nahavandi, S., Nguyên, T. T., Pham, Q., & Nguyen, C. M. (2022). Deep learning for deepfakes creation and detection: A survey. *Computer Vision and Image Understanding*, *223*, 103525. https://doi.org/10.1016/j.cviu.2022.103525

Parikh, A., ten Bosch, L., van den Heuvel, H., & Tejedor-García, C. (2023, December). Comparing Modular and End-To-End Approaches in ASR for Well-Resourced and Low-Resourced Languages. In Proceedings of the 6th International Conference on Natural Language and Speech Processing (ICNLSP 2023) (pp. 266-273).

Poon, J. (2021). GANs gone wild: Public perceptions of Deepfake technologies on YouTube.

Posner, M. I., Nissen, M. J., & Klein, R. M. (1976). Visual dominance: An information-processing account of its origins and significance. *Psychological Review*, *83*(2), 157–171. https://doi.org/10.1037/0033-295x.83.2.157

Sarker, I. H. (2021). Deep Learning: a comprehensive overview on techniques, taxonomy, applications and research directions. *SN Computer Science*, *2*(6). https://doi.org/10.1007/s42979-021-00815-1

Schick, N. (2020). Deep fakes and the infocalypse: What you urgently need to know. Hachette UK.

Shannon, H., Bush, K., Villeneuve, P. J., Hellemans, K. G. C., & Guimond, S. (2022). Problematic social media use in adolescents and young Adults: Systematic review and Meta-analysis. *JMIR Mental Health*, *9*(4), e33450. https://doi.org/10.2196/33450

Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. *arXiv (Cornell University)*. http://export.arxiv.org/pdf/1708.01967

Statista. (2024, January 31). *Worldwide digital population 2024*. https://www.statista.com/statistics/617136/digital-population-worldwide/

Sundar, S. S. (2008). The MAIN model: A heuristic approach to understanding technology effects on credibility (pp. 73-100). Cambridge, MA: MacArthur Foundation Digital Media and Learning Initiative.

Vaccari, C., & Chadwick, A. (2020). Deepfakes and Disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social Media + Society*, *6*(1), 205630512090340. https://doi.org/10.1177/2056305120903408

Villanti, A. C., Johnson, A. L., Ilakkuvan, V., Jacobs, M. A., Graham, A. L., & Rath, J. M. (2017). Social media use and access to digital technology in US young adults in 2016. Journal of Medical Internet Research, 19(6), e196. https://doi.org/10.2196/jmir.7303

Watson, G., Khanjani, Z., & Janeja, V. P. (2021). Audio deepfake perceptions in college going populations. *arXiv (Cornell University)*. http://arxiv.org/abs/2112.03351

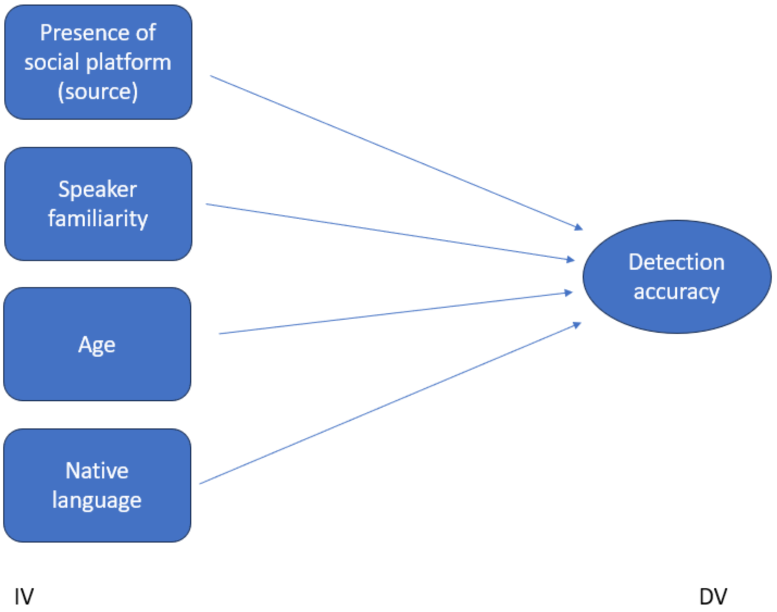
Weeks, B. E., & Holbert, R. L. (2013). Predicting dissemination of news content in social media. *Journalism & Mass Communication Quarterly*, *90*(2), 212–232. https://doi.org/10.1177/1077699013482906

Westerlund, M. (2019). The Emergence of Deepfake Technology: A review. *Technology Innovation Management Review*, *9*(11), 39–52. https://doi.org/10.22215/timreview/1282

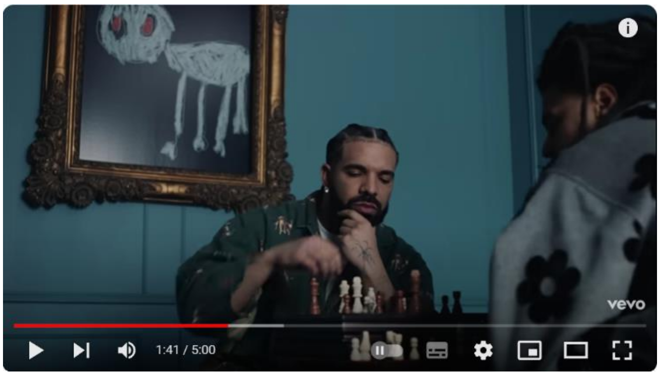
Yazdinejad, A., HaddadPajouh, H., Dehghantanha, A., Parizi, R. M., Srivastava, G., & Chen, M. (2020). Cryptocurrency malware hunting: A deep Recurrent Neural Network approach. *Applied Soft Computing*, *96*, 106630. https://doi.org/10.1016/j.asoc.2020.106630

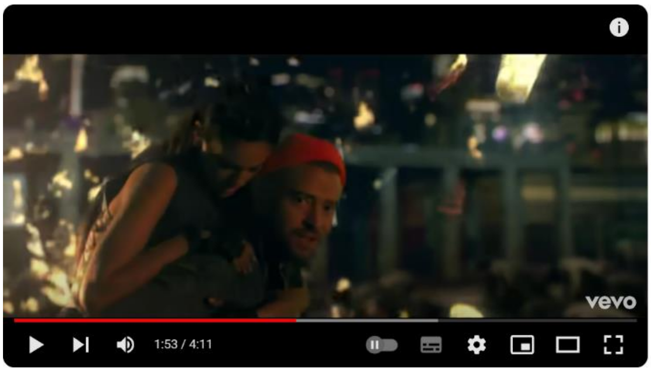
YouTube. (n.d.). https://www.youtube.com/?app

Appendix



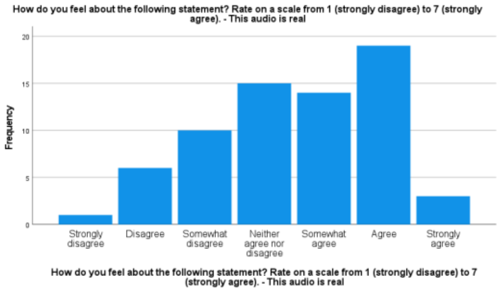
Appendix 1: analytical model



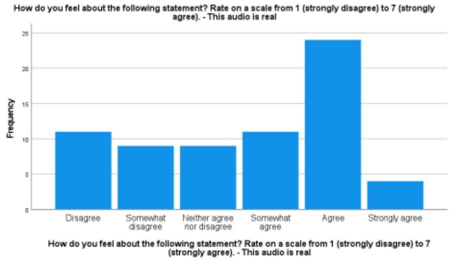


Appendix 2: materials

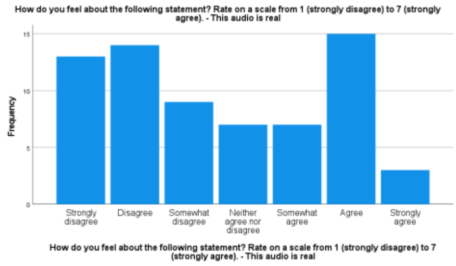
Adam Levine



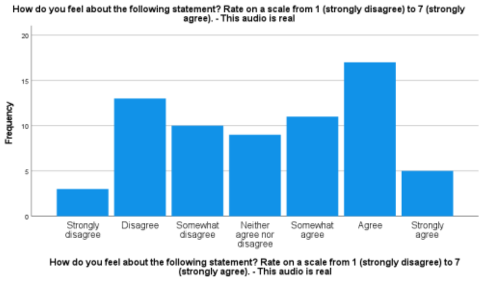
Drake



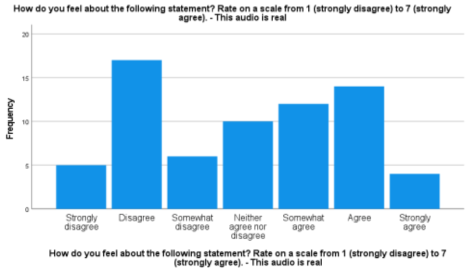
Taylor Swift



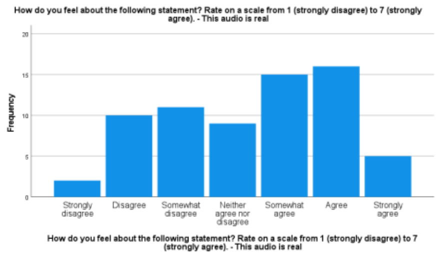
Dua Lipa



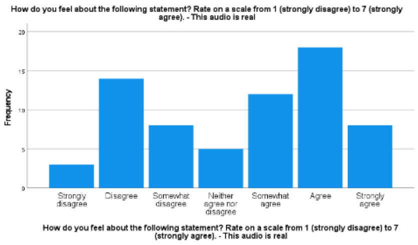
Kanye West



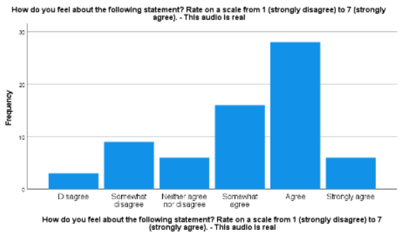
Lil Uzi Vert



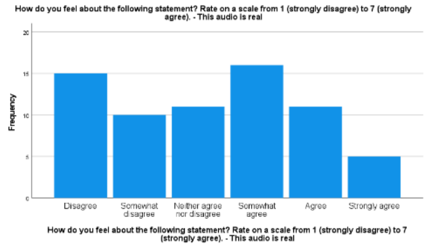
Ariana Grande



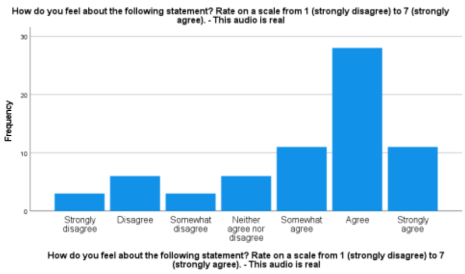
Beyoncé



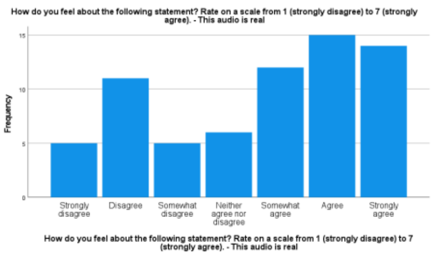
Justin Timberlake



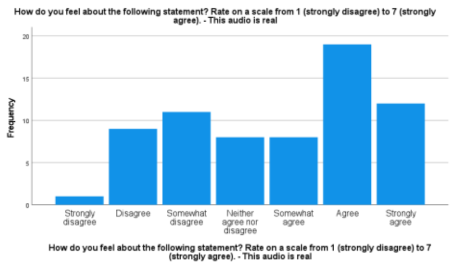
Billie Eilish



Jason Derulo



Pink



Appendix 3: bar charts for separate audios (row 1 is fake with source, row 2 is fake without source, row 3 is real with source and row 4 is real with source). Row is horizontal.

Statement of own work

Statement of own work – BA thesis

By signing this declaration, the undersigned

Simone van Horen, s1080614

Bachelor's student at the Radboud University Faculty of Arts,

declares that the submitted BA thesis is entirely original and was written exclusively by himself/herself, and without the use of AI tools, such as ChatGPT. The undersigned has indicated explicitly and in detail where all the information and ideas derived from other sources can be found by referencing all sources used, both in the text and in the bibliography.

By signing this declaration, the undersigned also declares that the research data presented in this thesis were collected by the undersigned himself/herself using the methods described in this thesis.

Place and date:

Roermond, 06-06-2024

Signature:

Checklist EACH (version 1.11, September 2023)

Simone van Horen s1080614

1. Will you be collecting data from social media platforms?

☐ Yes → consult the guidelines. If in doubt, contact the EACH to see if assessment is necessary

☒ Nee → continue with questionnaire

2. Will you use an existing dataset?

☐ Yes → continue with questionnaire

☒ No → go to question 4

3. When using an existing dataset, do you comply with the EACH guidelines**?

☐ Yes → continue with questionnaire

☐ No or in doubt → contact the EACH to see if assessment is necessary

** Guidelines: ethics approval is obtained for the original data collection, participants have consented to the reuse of the research data, or the reuse fits within the original research purpose. Or: dataset is completely anonymous

4. Will you be collecting data from participants?

☒ Yes → continue with questionnaire

☐ No → end of questionnaire

5. Is a health care institution involved in the research?

Explanation: A health care institution is involved if one of the following (A/B/C) is the case:

A. One or more employees of a health care institution is/are involved in the research as principle or in the carrying out or execution of the research.
B. The research takes place within the walls of the health care institution and should, following the nature of the research, generally not be carried out outside the institution.
C. Patients / clients of the health care institution participate in the research (in the form of treatment).

☒ No → continue with questionnaire

☐ Yes → Did a Dutch Medical Institutional Review Board (MIRB) decide that the Wet Medisch Onderzoek (Medical Research Involving Human Subjects Act) is not applicable or does the research comply with one of the standard research methods described by the EACH

☐ Yes → continue with questionnaire

☐ No → This application should be reviewed by a Medical Institutional Review Board, for example, the Dutch METC Oost-Nederland . If review by an MIRB has already taken place

→ continue with questionnaire. If this review has not yet taken place → end of questionnaire

6. Does the research include medical-scientific research that might carry risks for the participant?

☐ Yes → This application should be reviewed by a Medical Institutional Review Board, for example, the Dutch METC Oost-Nederland → end of questionnaire
☒ No → continue with questionnaire

Standard research method

7. Does this research fall under one of the stated standard research methods of the Faculty of Arts or the Faculty of Philosophy, Theology and Religious Studies?

☒ Yes → 12 standard questionnaire method→ continue with questionnaire

☐ No → assessment necessary, end of checklist

Participants

8. Is the participant population a healthy one?

☒ Yes → continue with questionnaire

☐ No → assessment necessary**, end of questionnaire → go to assessment procedure

**Exception for studies with patients participating in one of the described standard studies in the field of language and speech pathology

9. Will the research be conducted amongst minors (<16 years of age) or amongst (legally) incapable persons?

☐ Yes → assessment necessary, end of questionnaire → go to assessment procedure

☒ No → continue with questionnaire

Method

10. Will the study heavily burden participants?

☐ Yes → assessment necessary, end of questionnaire → go to assessment procedure

☒ No → continue with questionnaire

11. Are the estimated risks connected to the research minimal?

☒ Yes → continue with questionnaire

☐ No → assessment necessary, end of questionnaire → go to assessment procedure

12. Are the participants offered a higher compensation than the usual one?

☐ Yes → assessment necessary, end of questionnaire → go to assessment procedure

☒ No → continue with questionnaire

13. Should deception take place, does the procedure meet the standard requirements?

☐ Yes → continue with questionnaire

☐ No → assessment necessary, end of questionnaire → go to assessment procedure

☒ deception is not applicable

14. Are the standard regulations regarding anonymity and privacy met?

☒ Yes → continue with questionnaire

☐ No → assessment necessary, end of questionnaire → go to assessment procedure

Conducting the research

15. Are participants recruited via the Radboud Research Participation System (SONA) and/or is the research conducted in the CLS Lab?

☐ Yes → assessment necessary, end of questionnaire → go to assessment procedure

☒ No → continue with questionnaire

16. Will the research be carried out at an external location (such as a school)?

☐ Yes→ Do you have/will you receive written permission from this institution?

☒ No → assessment necessary, end of questionnaire → go to assessment procedure

☐ Yes → continue with questionnaire

☒ No → continue with questionnaire

17. Is there a contact person to whom participants can turn to with questions regarding the research and are they informed of this?

☒ Yes → continue with questionnaire

☐ No → assessment necessary, end of questionnaire → go to assessment procedure

18. Is it clear for participants where they can file complaints with regard to participating in the research and how these complaints will be dealt with?

☒ Yes → continue with questionnaire

☐ No→ assessment necessary, end of questionnaire → go to assessment procedure

19. Are the participants free to participate in the research, and to stop at any given point, whenever and for whatever reason they should wish to do so?

☒ Yes → continue with questionnaire

☐ No→ assessment necessary, end of questionnaire → go to assessment procedure

20. Before participating, are participants informed by means of an information document about the aim, nature and risks and objections of the study? (see explanation on informed consent and sample documents).

☒ Yes → continue with questionnaire

☐ No→ assessment necessary, end of questionnaire → go to assessment procedure

21. Do participants and/or their representatives sign a consent form? (see explanation on informed consent and sample documents).

☒ Yes → questionnaire finished

☐ No→ assessment necessary, end of questionnaire → go to assessment procedure

If you want to record the results of this checklist, please save the completed file.

If you need approval from the EACH due to the requirement of a publisher or research grant provider, you will have to follow the formal assessment procedure of the EACH.