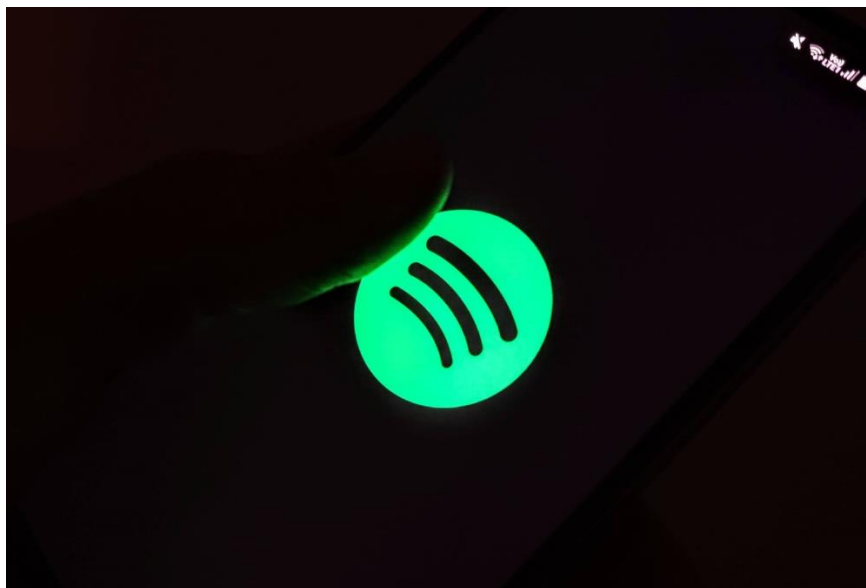


Roskilde University 2nd semester project

**Voice Data, Power Structures and Technological Enframing:
*A Critical Examination of Spotify as Irrational Regress***



By group S2140698517:

Anna Marion Mills

Anna Østerskov Gammelgaard

August Sebastians Bek

Marius Risbæk Thomsen

Supervisor:

Anne H. Fabricius

Number of Characters:

99.669

Table of content

Abstract	3
Introduction.....	4
Problem Area	5
Theory and Conceptual Framework.....	9
How Sound is Produced and a Working Definition	9
Speech Sounds.....	11
Speech Variations.....	13
Speech Variability based on Emotions.....	13
Speech Variability based on Depression and Mental Stress	14
Speech Recognition.....	15
Surveillance Capitalism.....	17
Dangers of Surveillance Capitalism	18
Digital citizenship.....	19
Voice as an individual property.....	20
GDPR	21
Human Rights.....	22
Methodology	23
General Introduction to the Methodology chapter	23
Digitalization as the latest expansion of modernity	24
Critical analysis	26
Analysis.....	27
Spotify, the Voice and personal data.....	27
Special Categories of Personal Data	28
Human Rights and Spotify	29
Terms of Service	29
Manipulation and Exploitation.....	32
Asymmetrical Power Relations	34
Discussion	36
Knowledge as a Resource for Change.....	37
Conclusion	40
Bibliography.....	42

Abstract

Spotify's patented technology is an example of how Adorno and Horkheimer's concept 'irrational regress' is inherent to surveillance capitalism. The data collected through Spotify's new patented voice recognition technology may be legal according to the GDPR and Human Rights Declaration, however, ethical concerns arise. The data collected through this technology has the potential for manipulation and exploitation, as well as producing and upholding existing asymmetrical power structures which turn individuals into data subjects. Knowledge of the issues this presents is crucial for collective action which is deemed necessary to change current social reality.

This paper presents a critical examination of Spotify's new patent titled 'Identification of taste attributes from an audio signal' and their product 'Car Thing'. It will provide a general introduction to the voice and its embodied characteristics which functions as background knowledge for further analysis of voice recognition technology, as well as introduce important concepts such as Surveillance Capitalism and Digital Citizenship. This paper adopts a Critical Theory approach when analyzing and discussing the problems and dangers of collecting data on users with voice recognition technology.

Introduction

In this paper, the voice as data, mass personalization and profiting of human behavior is central with a special focus on the company Spotify and their new patented technology which allows the company to listen to users' voices, and further using that data to suggest music and advertisements specifically for individual users.

We find it interesting to investigate the issues of companies harvesting voice data from their users. Technologies allowing companies to perform this form of data collection are advancing rapidly. We as researchers, but also as users of digital services such as Spotify, deem it highly relevant to raise questions regarding this advancement.

To analyze this topic, the following dimensions are used: Text and Sign and Science and Philosophy. The Text and Sign dimension provide insight and knowledge on what speech is, as well as what constitutes the voice. It will further improve our understanding of how voice characteristics are able to reveal personal information of an individual.

Applying the Science and Philosophy dimension allows us to raise questions about the ethics of Spotify having access to our voices so that they can recommend music directly to us. We question whether it's ethical for Spotify, and other platforms, to collect, analyze and sell our personal data in the search of a profit.

Problem Area

Citizens in the age of digitalization are used to technology which is constantly changing, evolving and improving. This technological progress serves as important factors in how we live and can very quickly transform certain aspects of our daily lives; an example of such technology is speech recognition software, also referred to as speech recognition technology or voice recognition technology.

Speech recognition technology enables technology, such as phones and computers, to understand human utterances and voices, and it is often promoted as much easier than 'older' methods of input such as typing on a keyboard (Summa Linguae, 2017). Speech recognition technology has taken great strides over the past 10 years, and today the technology is found in most phone devices; especially known is 'Siri' from Apple's iPhones as well as the popular Amazon 'Alexa'. However, contrary to what many might believe, this sort of technology is not new; in fact, speech recognition technology dates all the way back to the 1950s where the first example of modern speech recognition software can be found (Summa Linguae, 2017).

Bell Laboratories' 'Audrey' was groundbreaking technology in the 1950s. The device was originally designed to help secretaries and could recognize the numbers 1-9. The next piece of speech recognition technology took 12 years to develop. This device, IBM's 'Shoebbox', was able to pick up 16 different words. About nine years later came 'Harpy' which could recognize up to 1000 words. 'Harpy', had been invested in by the American Department of Defense who had begun to see the value in this kind of technology. However, so had the toy industry who in the 1970s and 1980s began to implement speech recognition technology in their products. In 1978 the doll 'Julie' was born. 'Julie' had the ability to not only recognize words and respond to the speaker, but she could also recognize different speaker's voices (Summa Linguae, 2017).

These early inventions serve as important tools for where voice recognition technology is today. When Google launched their 'Google Voice search App' in 2010, they had taken great inspiration from 'Julie' and her ability to recognize voices. A distinct feature of the Google app was that it was personalized to the user's voice as well as the user's speech patterns (Summa Linguae, 2017). This ability has paved the way for the technology we see today, and today artificial intelligence (also referred to as AI) is able to analyze user's age, gender and emotional state based on the voice data collected by voice recognition technologies (Justia Patents, 2021).

However, even though this sort of technology is often marketed as ‘easy’ and ‘convenient’, speech recognition technology also paves the way for certain issues in the form of surveillance capitalism. This has evolved by companies storing and capitalizing on users’ data on a mass scale. This data is being analyzed by artificial intelligence which can produce predictions of users’ behavior, interests and preferences. This is highly commercially valuable and has constituted the digital profiling of both individuals and groups which is conceptualized as ‘mass personalization’, which is the digital personalization of services to fit the individual’s preferences, interests and behavior. This personalization can be used for advertising, or to increase the time individuals spend using the service (Yeung, 2018). As a result, this process of companies profiting from human behavior can, for example, enable the potential for manipulation of individuals. In the case of the Facebook and Cambridge Analytica scandal, data had for example been used to influence individuals politically which affected the 2016 US election and the Brexit referendum (Yeung, 2018).

In this paper, we, the researchers, wish to engage in a critical discussion of the dangers of this new form of capitalism that is surveillance capitalism, and the ethics surrounding this issue, which will be an underlying topic throughout the paper. To accomplish this, we have examined Spotify’s new patent for a voice recognition technology, which was applied for in 2018 but obtained in January 2021 (Justia Patents, 2021). This patented technology represents an example of the ethical issues concerning companies profiting from human emotion and behavior, and the potential for exploitation and manipulation.

We have chosen to examine the Spotify patent as a case, which serves as the foundation of our analysis and critique. Spotify is just one example of a company which employs this sort of technology; however, a case-based analysis allows us to ground our critique within the specific example and provide a focused analysis. Considering this, it is necessary to briefly explain what Spotify is, what the company’s new patent concerns, and, lastly, touch upon the company’s new invention: ‘Car Thing’.

The company Spotify first launched in 2008, and it is today one of the world’s most popular audio streaming services with more than 356 million users across 178 different markets (Spotify, seen May 4th 2021). The streaming service provides its users with access to millions of songs, podcasts and videos, and can be used both as a free service or as a paid premium subscription service which costs \$9.99 a month (Pocket-lint, February 2021).

The company obtained its new patent named ‘Identification of taste attributes from an audio signal’ (Justia Patents, 2021) in January 2021. This patented technology makes it possible for the company to collect data on, and analyze, users’ emotional state, accent, gender, physical environment etc. based on voice data which they gather using this patented voice recognition technology. Spotify will then, using an algorithm, be able to recommend songs and advertisements to users based on the data highlighted above (Justia Patents, 2021).

Additionally, we, the researchers, are also including Spotify’s new product ‘Car Thing’ as a part of this paper. The ‘Car Thing’ was announced on April 12th, 2021, and is a controller which is mountable in cars, and features four microphones for easy voice control of the device (Pitchfork, seen April 28th 2021). We concern ourselves with this product because we believe that it presents an example of a physical use of the patented technology. Furthermore, this product is initially marketed as free to Spotify premium subscribers. The ‘Car Thing’ then also serves as an example of the business model of ‘free for services’ where users agree to further surveillance in exchange for a free service (Yeung, 2018).

We, the researchers, argue that the consequences of using these services which Spotify provide are highly unclear to users, and the constraints they present contradict the notion that technological advancement is good and equals freedom which is inherent to western modernity (Allen and Goddard, 2017). Adorno and Horkheimer (referenced in Allen and Goddard, 2017), the fathers of Critical Theory, which is the position this paper takes, argue that modernity’s idea of enlightenment creates a system of control that inhibits rational thought. Technological systems marketed as freedom are then in fact restraints to our freedom. Adorno and Horkheimer describe this as initially being rational progress which reverses and becomes the opposite: ‘irrational regress’ (Allen and Goddard, 2017). Withal, we arrive at this problem formulation:

Problem formulation

How does Spotify’s new voice recognition patent represent the dangers of irrational regress in the age of mass digitalization?

Research questions:

- What is the voice? What linguistic data shows individuality in the voice? How is this data used to make speech recognition technology work?
- How is voice data exploited in a digitalized world?
- How can citizens challenge the companies profiting from human behavior?

Theory and Conceptual Framework

In the following chapter, phonetic theory is explored in order to understand the voice and the characteristics which it embodies. First, we will present how sound is produced followed by a working definition. Then, speech sounds and speech variations will be touched upon followed by an introduction to speech recognition. Furthermore, this chapter includes the conceptual framework which serves as the basis for our analysis. This second part of the chapter includes concepts such as Surveillance Capitalism, Digital Citizenship and Voice as an Individual Property, as well as legal frameworks such as the GDPR and the Human Rights Declaration.

How Sound is Produced and a Working Definition

Before beginning, we, the researchers, find it important to mention that although we find it necessary to provide some information on how sound is produced, it should not be seen as the main focus of this chapter. The reason for this, is because knowing how sound is produced is crucial to understanding the voice and seeing as “the main obstacle in studying speech is that the object of study is for the most part invisible” (Zsiga, 2013, p. 3), we hope that studying the bodily functions of speech will help diminish that obstacle. However, we will not go into great detail in this part of the chapter as the important part, for our paper, is simply understanding the general functions of the body which are necessary for producing sound and, therefore, speech and voice.

The vocal tract consists of the structures of the body which are used in the creation of sound, and as a result speech. The overall structure of the vocal tract can be divided into sub-groups with the larynx (also called the laryngeal system) as the dividing point. Everything below the larynx will be referred to as the sub-laryngeal system, and everything above is named the supra-laryngeal system (Zsiga, 2013).

The sub-laryngeal system is comprised of the trachea (windpipe) and the lungs. According to Zsiga (2013), speech is the “act of making the movement of air out of the mouth and nose audible” (Zsiga, 2013, p. 5), and speech then begins in the sub-laryngeal system when air is “drawn into the lungs, and then forced out” (Zsiga, 2013, p. 6). Most languages, English included, make use of the pulmonic airstream mechanism (meaning that the movement of air is powered by the lungs), and the airstream is generally pulmonic egressive in all languages when speaking (meaning that speech happens

on an exhalation of breath, and with a few exceptions, not on an inhalation). When speaking, voice is the result of air which is made audible, either by making it turbulent and fast, by building up pressure, or by making air molecules resonate (Zsiga, 2013).

The larynx is also commonly called the ‘Adam’s Apple’ and sits on top of the trachea which connects the mouth and lungs. In simple terms, the larynx is made up of the vocal folds (which consists of thickened tissue in the shape of folds/flaps), four cartilages which protect and manipulate the vocal folds, and the muscles which connect them all. The vocal folds can move and manipulate the flow of air, and if they are held in a specific position where they are touching but not closed, they will vibrate and create ‘voicing’ which is the basis of many sounds of speech. The human larynx is unique in the way that it is placed much lower in the throat compared to other animals. This allows for more space in the back of the mouth and greater freedom of movement for the tongue which is what makes the phonetic complexity of human language possible (Zsiga, 2013).

The supra-laryngeal system further shapes the air as it moves from the lungs; this is how we can distinguish individual sounds of speech. We can further divide the supra-laryngeal system into active and passive articulators. The active articulators are the lips and tongue which move in relation to the passive articulators (which consists of everything from the upper lip/teeth and all the way across the roof of the mouth to the back of the throat) to create a narrowing or widening of the vocal tract (Zsiga, 2013).

Having now explained how sound is produced within the vocal tract we can begin to arrive at a working definition of the voice which we, the researchers, will use as a point of reference for our paper. The voice is not easy to define, and this is perhaps due to the fact that it is for the most part invisible as mentioned earlier. The definition which we are choosing to work from was developed by Watt et. al. (2019) and is as follows:

“The voice is the output in the sound domain of the coordinated movements of an individual human talker’s speech organs, which is used principally but not exclusively by the talker for the acoustic encoding of linguistically meaningful utterances, and which after a period of exposure listeners may come to associate with that talker.” (Watt et. al., 2019, p. 144)

This definition of a voice takes into account the fact that not all sound produced by the vocal tract are considered linguistic (an example could be involuntary sounds of anger, surprise etc.), however,

sounds like these might be enough to reveal the identity of the person to the listener if the two have a certain degree of familiarity. Due to this, the importance of the listener is also allowed for in this definition. Lastly, it is important to note that this definition does not exclude sounds produced by a vocal tract which is not functioning normally, e.g., due to physical illness or psychological disorders (Watt et. al., 2019).

Speech Sounds

Contrary to popular belief, it has not been proven whether our voices are actually unique to us as individuals. Furthermore, the amount of detail needed to define a human's voice as unique is also unknown (Watt et. al., 2019).

Even so, as explained in the previous section, it becomes evident that the individuality of voice can be somewhat determined by anatomy. However, it should be mentioned that what is called 'vocal twins' do not have to resemble each other anatomically or be genetically related and identical twins might not use their vocal organs the same, making them sound different from each other (Watt et. al., 2019).

Nevertheless, a very basic example of what differences in anatomy can determine would be the difference in pitch between men and women. When young boys reach puberty, their vocal folds become 50% longer and thicker than women's. In general, this means that difference in the structure of the larynx creates different types of voices. So, in this example, larger vocal folds, mean slower vibration in the larynx, which results in a lower-pitched voice. Another example, still in the realm of difference between men and women, could be that for some women their vocal folds never close completely which creates what is called a 'permanent glottal chink'. This causes these women to have a breathier voice. (Zsiga, 2013)

So, speech sounds, and sounds in general, are created in the larynx. Sound is a pattern of pressure variation which propagates in waves from a voice source. To count as linguistic sound, the size and rapidity of these pressure variations must be within the ranges to which the human ear is able to hear. There are two types of sound 'periodic' and 'aperiodic':

- Periodic: a pressure wave of specific shape repeated e.g., the same musical note continuously played (vowels and nasal plosives)

- Aperiodic: a moment to moment/random variations e.g., leaves rustling (fricative consonants such as 's')

There is also a category called 'transient' sounds which is a part of the aperiodic sound. What differentiates transient sounds are that they are instantaneous e.g., slamming a door (comparable to the oralplosive sounds 'p, t, k') (Zsiga, 2013). Speech is a combination of all of these types of sound making speech sounds very complex.

As mentioned earlier, sound moves out from a source in waves. To better understand this, we should view a sound wave as a pendulum. The ball in a pendulum moves in a set pattern. The ball starts in the middle (A), then someone moves the ball or rather dislocates it (B), and then lets it go which starts the swinging motion of the pendulum. The ball then swings through its original placement (A) over to the other side (C). So, the pattern looks like this: **A – B – A – C – A**

This pattern repeats over and over, and the cycle takes the same amount of time (if we ignore energy loss). This is the same motion a tuning fork makes when struck, however, the 'swings', also called oscillations, of the tuning fork are within the range of frequencies and amplitudes which the human ear is able to hear. A young and healthy ear can pick up frequencies as low as about 20 cycles per second to as high as about 20.000 cycles per second (Zsiga, 2013).

Since larger vocal folds equal slower vibrations of the larynx and hereby a lower pitch, this is, in general, explained as frequency and pitch being proportional. The higher the frequency the higher the note and vice versa. Furthermore, for a given frequency, the amplitude of the vibration is proportional to volume (Zsiga, 2013). A soundwave can be described as a 'sinusoid'. Furthermore, Jean Baptiste Joseph Fourier proved in 1807 that every kind of vibration including speech sounds can be described as the sum of a set of simple sinusoids of various frequencies and amplitudes. The more complex sound, the more sinusoids. Complex sound waves are the sum of a set of sinusoids with different amplitudes and frequencies. The different sinusoids are components which create harmonics (Zsiga, 2013).

Lastly, to relate this to the individuality of voice we find it useful to also present source/filter theory. The source of the sound creates periodic sounds by the vocal folds opening/closing and vibrating. This creates harmonic frequencies, and these are filtered by the vocal tract. This is where articulators, lip and tongue shape, play a role. It is only frequencies that match the natural resonance of the vocal

tract which get amplified and propagate through the filter with high amplitude. Frequencies that do not match the natural resonance of the vocal tract are therefore filtered out (Zsiga, 2013).

Speech Variations

The Spotify Patent states that the voice recognition feature will be able to analyze personal voice data and reveal age, gender, accent and emotional state combined with the users' physical environment (Justia Patents, 2021).

Some speech variations are created by the physical features of the speakers, others are created by the influence of the people around them. Though factors like these are not the main topic of our project they are worth keeping in mind. There are multiple factors which make our voices individual: "we all have different dialects and other factors that makes the way we speak individual to us. Important influences on variation include situation, socioeconomic class, ethnicity, gender, age, and sexuality" (Zsiga, 2013 p. 427).

A human listener is quick to differentiate between the voices of men and women, as previously mentioned this is because of the traits we have been given biologically. Another speech variation which is conditional upon anatomy is the difference of voice at different ages. Younger children have more high-pitched voices since their bodies are smaller than for example adults, whose voice will increasingly become lower and creakier caused by stiffening of the vocal folds. (Zsiga, 2013)

Lastly, as stated in the patent, Spotify will be able to analyze the accent of its users. Everyone has a dialect and accent shaped by their regional environment. "A dialect is a variation of speech, and a regional dialect is a language variety spoken by a group of people who live in a particular place" (Zsiga, 2013 p. 428) e.g., Australian dialect of English. The term 'termed register' refers to the variation that occurs based on the social environment. Whether delivering a professional speech or conversing casually with friends, the same person will speak differently depending on the context of the situation. We use more formal registers with people we don't know well or with people who appear superior to us, and we use informal registers with people we know well (Zsiga, 2013).

Speech Variability based on Emotions

It is commonly known that the speech signal (or voice) which a speaker produces, contains many types of information about the speaker, the emotional state being one of them (Benzeguiba et. al.,

2007), and it is therefore relevant to consider how the voice of a speaker might change and reflect their emotional state.

Changes in the voice of a speaker may occur when the speaker experiences 'normal' emotions such as anger, happiness, sadness, and fear. These changes in the speaker's voice are based on the emotions which the speaker experiences at the time of speaking. Examples of acoustic changes of speech which can be noted in a speaker experiencing sadness is a slightly slower speech rate, a narrower pitch range, and a more slurred articulation. Likewise, speakers experiencing anger may appear to speak slightly faster, with a higher intensity and a much higher pitch average (Khanna and Sasikumar, 2011).

It is important to note, that not all emotions are as easily recognized by speech recognition software. For example, is 'anger' the most recognizable while 'happy' seems to be the least recognizable; however, this might be largely dependent on which feature is analyzed by the software and if it is the same for all emotions. Additionally, it seems to be easier to recognize emotions within male voices compared to female voices (Khanna and Sasikumar, 2011).

Speech Variability based on Depression and Mental Stress

In addition to this a more general overview of how speakers' emotions are reflected through their voice we, the researchers, would like to briefly introduce depression and stress as, perhaps, more extreme examples of how a voice can change. The reason for this is because they represent changes within the voice over a longer period seeing as e.g., depression is usually a long(er) struggle for the individual compared to momentary sadness.

Speakers who are diagnosed as depressed tend to speak in a more monotone manner than those who are not depressed. Their voices fluctuate less in pitch, are slower than average, and appear more slurred, or perhaps just less intense, which might sound like the person is mumbling. Their speaking pitch range is reduced as well as the intonation (the patterns of rise and fall of the voice) which combined with a lack of linguistic stress (the acoustic differences between syllables in a word) all help making the voice of the depressed speaker appear more monotone to the listener (Long et. al., 2017). Of course, the severeness of the depression also has a say in how obvious these differences in voice are perceived to be. Lastly, while the content of speech can be controlled, these sorts of prosodic characteristics of the voice are not under conscious control. This means that it is very easy to

detect these sorts of characteristics through speech recognition software since they cannot be hidden by the speaker (Yan, 2012).

Mental stress (which might for example be related to anxiety), has also been shown to influence the linguistic cognitive complexity of the speaker, both in cases of long-term and short-term stress. This means that the speaker might speak with less complexity when under stress (Saslow et. al., 2014). Concerning the voice more specifically, stress is shown to affect the quality of the voice because the muscles in the body, including those in and around the vocal tract, tighten and become tense. Speakers dealing with long-term, chronic stress might experience sudden pitch changes which sound as if the voice ‘cracks’, or they might experience their voice as ‘shakier’ and hoarser than usual (Folk, seen March 25th 2021). All in all, speakers dealing with mental stress, especially long term, seems to be in less control of their voice.

Speech Recognition

We, the researchers, have now established voice as an intangible thing which varies based on anatomy, regional location, social context and emotional state. This section will now examine how these features are analyzable by speech recognition technology, which allows companies to gather data that identifies and characterizes individuals, based on the above explained characteristics and features. A technology that identifies individuals is called biometric software; it measures and analyses data to determine people’s behavioral and physical traits such as facial features or voice features. This software often serves as authentication, as it has the potential to identify people based on patterns that algorithms are trained to extract from datasets that for example measure facial features (Crampton, 2019).

Facial recognition software is often used as authentication for different purposes on a smartphone. However, further, than just identifying specific individuals, algorithms are also trained to identify factors such as ethnicity, gender, age and emotions. An example of this is MS Face which has the purpose of analyzing people’s emotions based on their face. The MS Face algorithm works by mapping facial expressions to a certain value which is one of seven categorized emotions (Crampton, 2019).

Speech recognition technology can also identify features such as emotion and gender. This software has the overall function of translating spoken language into text. As with facial recognition, speech recognition technology works by training an algorithm to map certain acoustic features to certain

meanings. By analyzing the soundwave of a person's speech, it's possible to identify certain phonemes based on the shape of the soundwave. Phonemes are the smallest contrastive units in language that distinguishes words and thereby meaning, and speech recognition works (very simplified) by training algorithms to create words by combining these phonemes (Cambridge Dictionary, "Phoneme", 2021 and Robertson, 2016).

However, artificial intelligence is capable of and do much more than transcribing speech samples to text. As noted above, the acoustics of how people project their voice, change based on their emotional state. For example, within speech affected by depression, there is a difference in the acoustic space variability. So, by teaching an algorithm to map certain data on acoustic space variability onto the characteristics of depressive speech, it is made possible to identify if a person is depressed or not based on their recorded speech (Cummins et. al., 2015).

As mentioned, biometric technology is also used with voice features and can be used for speaker recognition. This technology can generate 'voiceprints' that distinguish speakers from each other. The technology works by analyzing multiple voice data samples, and if enough data points align between different recordings of speech, the technology will determine that it is the same speaker presented in multiple samples. There will always be similarities in the data between different speakers, but the idea is that speech samples recorded by the same speaker will be more similar to each other than they are to samples of other speakers (Watt et. al., 2019).

Companies using this technology as authentication often advertise these 'voiceprints' as completely unique and as reliable security against hackers. This is the case for HSBC giving assurance to their customers that "Fraudsters and hackers may be able to steal or guess your security number, but they can't replicate your voice" (Watt et. al., 2019, p. 147). However, even without complex technology, the HSBC biometric security system has been spoofed. For example, Joe Simmons the twin brother of British Journalist Dan Simmons had gained access to Dan's bank account by vocally impersonating him. There is also the risk of these forms of gatekeeping technology being subject to spoofing by 'voice cloning' software that artificially mimics biometric data. Voice cloning software such as Adobe's VoCo can produce highly realistic emulations of speaker's voices. This is popularly termed as deepfake technology (Watt et. al., 2019).

Companies gather voice data on a large scale through technologies such as Amazon 'Alexa', Google Home, 'Siri' and Microsoft's 'Cortana' which are designed to learn differences between individual voices. This indicates that person-specific voice data is being gathered on a mass scale. If these

samples of speaker's voices recorded by 'Alexa' or Google home were illicitly obtained by hackers, they could be used with voice cloning software as a tool for identity theft. One way this type of identity theft can take place is through 'phishing' which is when scammers try to obtain personal information from people by pretending to be someone else. So, biometric voice data being gathered by these companies has created potential dangers that users are generally not informed about (Watt et. al., 2019).

Surveillance Capitalism

A concept central to our analysis is surveillance capitalism. Because of this, we will now define surveillance capitalism and introduce possible dangers inherent to this concept.

Shoshana Zuboff (referenced in Kulwin, 2019) argues that capitalism evolves by taking things that exist outside the market dynamic and claiming them for the market dynamic. The term surveillance capitalism describes how this process has occurred in the commodification of the private human experience. This commodification has developed through companies gathering and selling behavioral data and combining these data with machine learning or artificial intelligence which then produce predictions of human behavior. This has created a market where these predictions of individuals' behavior are sold to business customers that have an interest in knowing how we will act now and in the future. (Kulwin, 2019).

The commercial value in these predictions exists in multiple markets where there is a benefit in tweaking services based on people's behavior. These predictions are used to optimize personalized advertisements via social media platforms and websites and are used to design all kinds of apps in ways that make people engage with them as much as possible i.e., making them as addictive as possible. So, through knowing our behavior these predictions are also used to control and manipulate our behavior by designing digital platforms in certain ways that impact and change our behavior to create an economic profit. (Kulwin, 2019)

Data that show how we act is valuable, and gathering, selling and using these data has become the business model of big tech companies such as Facebook and Google. However, we can also observe how the principles of this business model are spreading into other sectors. For example, in late 2018 Ford's CEO described how Ford would improve their price-to-earnings ratio by 'becoming a data company'. It is valuable for Ford to gather all kinds of information, ranging from how people are

driving to what they are talking about when they drive as this information can be sold to many different business customers (Kulwin, 2019).

Dangers of Surveillance Capitalism

Predictions of human behavior are used to personalize online services for users. Yeung (2018) calls this phenomenon ‘mass personalization’. Personalization of services leads to asymmetry of power as the mass surveillance of individuals changes the relationship between the ones in possession of this data and the surveilled population. This has various potential implications for the surveilled population. Yeung (2018) mentions consumer exploitation as an essential problematic implication of surveillance capitalism.

AI technology can identify who and when individuals are more likely to make impulse purchases, and the kinds of offers specific individuals are more susceptible to. This analysis happens based on users’ click-through behavior and creates profiling of individuals including characteristics such as ‘maximum willingness to pay’. This analysis allows retailers to offer personalized products, but also to personalize products’ prices to optimize the retailer’s revenue. This means that different individuals can buy the same product at the exact same time but at different prices. Yeung (2018) argues that data-profiling of individuals, which allows price discriminations, indicates that there are groups of individuals that are routinely deprived of equal price offers. She mentions these groups as being:

“[...] those consumers who are algorithmically identified as those likely to have difficulty making good decisions, either due to lack of knowledge, poor digital literacy or due to consumer disengagement and who therefore do not actively shop around for better offers, nor switch to alternative providers who are willing to offer them better” (Yeung, 2018, p. 5).

In general, the asymmetry of power between data subjects and the ones gathering data is increased by the surveilled subjects’ unawareness of companies’ profiling practices and their implications. Undoubtedly, people’s lack of knowledge is fundamentally correlated to the opacity of these processes. However, according to Yeung (2018) the lack of knowledge can also be constituted to how “this creeping development as a whole, and its impact on human rights and society, has received little attention and there has been scarcely any fundamental political and public debate so far” (Yeung, 2018, p. 4). So, there is also a problem with a lack of public debate regarding the development of surveillance capitalism (Yeung, 2018).

The practices of mass predictive personalization lack transparency in nature, but this lack of transparency is also enhanced by companies' use of 'nudging' techniques. Nudging techniques rely on and exploit individuals' cognitive weakness, rather than conscious reflection, in making decisions. Nudging, in the context of mass personalization, works in powerful and subtle ways by continuously refining the users' choice environment in response to the user's behavior. Nudging further increases the potential for users to receive increasingly worse offers as their choice environment is nudged based on their behavior. Nudging techniques can generally be understood as manipulative but combined with the subtle power of the unclear practice of mass personalization their manipulative power is enhanced. Because of this, Yeung (2018) describes this process as 'hypernudging' (Yeung, 2018).

An essential component of surveillance capitalism is mass predictive personalization. This dynamic has, amplified by its murkiness, evident implications concerning what it means to be a digital citizen (Yeung, 2018).

Digital citizenship

In this section, it will be discussed what surveillance capitalism means for the concept of digital citizenship. We, the researchers, do not speak of digital citizenship as the formal status of being a country's citizen, but about the practice of active citizenship online.

So, what does it mean to be a good digital citizen? Citizenship is people creating and performing their role in society. In this way, citizenship is understood as empowerment, as it describes citizens taking an active part in society, for example by creating social change within their communities. Digital citizens take an active part in society through, for example, social media which has been described as a 'liberating technology' as it lets people take an active part in political discussions and makes citizen journalism more feasible. In this way, citizens have been empowered by the digitalization of society. This is one side of the coin; Citizens' digital, performative enactment of rights and construction of their position in society (Gangneux, 2017).

Digital citizens are also data subjects being monitored by tech companies. This means that being a digital citizen also entails a compromise regarding the right to privacy. Human Rights protect citizens from arbitrary attacks against their privacy; however, companies' extensive data collection on citizens' behavior gives unprecedented insight into these citizens' lives to the people in possession of this data. Furthermore, tech companies' data collection on citizens does not coincide with the aim of

digital citizenship, with citizens reasserting their position in society by claiming individual rights, because data collection actually limits their privacy. Over and above the invasion of privacy, data collection can potentially allow the state and companies to “compartmentalize us according to consumption habits, political preferences, or the likelihood of committing a crime” (Gangneux, 2017, p. 734). The profiling of citizens as data subjects, therefore, holds potential for discrimination. This is also a clear challenge for citizenship (Gangneux, 2017).

So, digitalization has both empowered citizens but has also increased surveillance, which compromises the human right to privacy. The citizen is left with the choice of agreeing to the Terms of Service on digital platforms, allowing surveillance of their behavior, or deciding to not use these platforms, but then withholding themselves from the empowerment and utility of these platforms. Again, the issue is also that users have to make this decision regarding highly opaque processes and consequences inherent to surveillance capitalism, which leads to their decision not being fully informed.

Voice as an individual property

Since voice data is collected from digital citizens, it is relevant to introduce the ownership rights of voice. So, to whom does a voice belong? The concept of ‘property’ is what prevents others from doing or essentially using our ‘things’. The law has history of protecting property and determining exactly what property is, and what is or isn’t allowed to be done with our property: “We have got used to the idea of property rights in tangible items (things we can touch)” (Watt et. al., 2019, p. 150).

However, in this case we are looking at ownership of voice which is an intangible thing.

As an intangible ‘thing’ is not something we can touch, it cannot be physically locked away or be protected like other ‘touchable’ things can. Therefore, it is necessary to determine a way in which we can define the legal limits of protection of an intangible item, this way people can get a clearer picture of what is owned by us as consumers (Watt et. al., 2019).

However, since technological advancement is rather rapid, it makes it difficult for the legal system to keep up and further makes it harder to determine laws which protect intangible items, such as the voice (Thrasher, 2019). The goal of intellectual property laws is to protect a wide range of intangible concepts (for example, through patents, trademarks or copyright). In each case, however, the law must distinguish between intangible and similar existing items. In order to make these determinations, the law must consider the reason for its existence (Watt et. al., 2019).

Essentially, the issue boils down to this; in order to give protection patents, trademarks and copyrights have very specific requirements, such as context, profit and brand value. Evidently, it is not possible to define an individual person's voice within these very specific requirements (Watt et. al., 2019).

GDPR

The European Union's General Data Protection Regulation, also known as the GDPR, was put into effect on May 25, 2018, and was proclaimed as the toughest privacy and security law in the world. Even though the GDPR was passed by the European Union it affects companies and organizations all over the world, if they collect data related to people living in the EU (GDPR Archives, seen 3rd of May 2021). Admittedly, The GDPR is a very extensive piece of legislation, hence within the framework of this paper, we will, mostly, focus on the three listed paragraphs below, as we deem those especially important to the project:

1. Article 4 of the GDPR consists of a list of definitions. The first paragraph is what constitutes 'personal data':

"1. 'personal data' means any information relating to an identified or identifiable natural person ('data subject'); [...], in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person;" (GDPR.eu, 2018, Art. 4.1)

2. The GDPR does not mention voice data specifically, however it mentions 'biometric data', which is Paragraph 14 of Article 4:

"14. 'biometric data' means personal data resulting from specific technical processing relating to the physical, physiological or behavioural characteristics of a natural person, which allow or confirm the unique identification of that natural person, such as facial images or dactyloscopic data;" (GDPR.eu, 2018, Art. 4.14)

3. Article 9 revolves around ‘Processing of special categories of personal data’:

“4. Processing of personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person’s sex life or sexual orientation shall be prohibited.” (GDPR.eu, 2018, Art. 9.1)

What these three paragraphs determine and how they affect, or should affect, the new Spotify patent will be examined further in later sections of this paper in form of a critical analysis.

Human Rights

The Universal Declaration of Human Rights was signed in 1948 by representatives from all over the world with different cultural and legal backgrounds. It was a result of the experiences of the Second World War and was seen as a tool to ensure “the rights of every individual everywhere” (United Nations, “History of the Declaration”, seen April 20th 2021).

More specifically, in this paper we, the researchers, concern ourselves with Article 12 of the Human Right Declaration which states:

“No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honor and reputation. Everyone has the right to the protection of the law against such interference or attacks.” (United Nations, “Universal Declaration of Human Rights”, seen April 20th 2021).

We are especially concerned with the very first sentence as it defines privacy as a fundamental human right, and we believe that this right is particularly challenged by the new Spotify patented technology. This will be further elaborated during our analysis of the case.

Methodology

General Introduction to the Methodology chapter

It is perhaps no surprise by now, that the critique of the rapid technological developments concerning voice recognition, the recording of voice data, and the companies which capitalize on this is central to our research paper. As a result, this paper takes on a Critical Theory approach which originated from the Frankfurt School seven decades ago and was heavily inspired by Marx's critique of capitalism (Budd, 2008). Critical Theory is generally concerned with critiquing and changing society and is heavily focused on issues of power relations and inequality. According to Horkheimer, Critical Theory must:

1. Explain what is wrong with current social reality.
2. Identify the actions needed to change it.
3. Provide clear norms for both criticism and transformation.

(Asghar, 2013)

Because we, the researchers, work from the perspective of Critical Theory we take the ontological position that reality is historically shaped by different factors such as political, cultural, and economic factors. Reality is then created over time and accepted as the norm although this political, cultural and economic reality might not be beneficial to the entire population equally. Epistemologically, we believe that knowledge is subjectivist and transactional – it is influenced and mediated by the values which people hold. (Plack, 2005 and Guba, 1994).

Before elaborating on why we still find the Critical Theory perspective to be relevant today, we would like to make clear why we did not choose to work from the perspective of Critical Realism which is undoubtedly familiar to Critical Theory. We did not find the Critical Realism perspective relevant to our paper because we are not concerned with studying an underlying, objective reality which is independent of individuals – in fact, we do not believe that this exists. Additionally, because we believe that knowledge is subjective, we do not see how it is possible to avoid imposing existing preconceptions and ideologies when conducting research, which is important within Critical Realism. Lastly, the focus within Critical Realism seems to be on explaining and understanding outcomes and reality.

We do not see this as being sufficient to this paper; our focus should be on critiquing and changing social reality (Clark, 2008).

Additionally, we would like to mention that our paper examines digitalization as the latest development of western modernity, meaning that our point of view is western-based. However, even if the phenomenon of surveillance capitalism has occurred as a result of western society's belief in technological advancement, every person with a smartphone can be subject to tech-companies capitalizing on their data. So, because the developing world has become increasingly digitalized, surveillance capitalism is happening all over the world.

As researchers, growing up and living in the western world we acknowledge that our perspective has been shaped in the context of western modernity and its arguably blind trust in technological advancement. Essentially, we are citizens in the western world critiquing the latest form of western modernity. A perspective from outside the western world could potentially have different critiques of digitalization as their understanding of humans' and technology's place in the world is formed in a different context.

Digitalization as the latest expansion of modernity

During the seventeenth century, modernity emerged because a new thought about the human condition developed. It began to be believed that humanity could be able to control nature through reason and, as a result, create a better world. It is also true that modernity has resulted in major benefits and improvements to human society e.g., economic and political systems (supposedly) guaranteeing civil rights, freedom and the knowledge to navigate these systems as individuals, again through the use of reason (Allen and Goddard, 2017).

However, it can be questioned whether modernity has met all these promises. The expansion of capitalism, which has developed alongside modernity, has been criticized for corrupting humanity and created "a world of illusory desires" (Allen and Goddard, 2017, p. 102) which uphold the injustices inherent to capitalism.

In this paper, we offer the idea that digitalization is the latest form of modernity and surveillance capitalism the latest expansion of the capitalist system. The rapid advancement of technology today is a consequence of the original belief in reason, rationality and scientific progress (Allen and Goddard, 2017). Obviously, the advancement of technology has brought some major benefits to human

society. In the context of the COVID-19 pandemic, one excellent example of the benefits of technology is the way in which education and workplaces have been able to continue virtually during periods of lockdown. The usual face-to-face processes of teaching and collaboration have been moved en masse to online communication platforms such as Microsoft Teams or Google Meet. In general terms, then, this paper should not be perceived as fundamentally hostile towards technology and digitalization; its purpose is simply to raise critical questions.

Martin Heidegger, a German philosopher, who was critical towards technological modernity, was also not hostile towards modern technology from that time such as electric kettles or telephones. However, he was concerned about the effects that technology as an ‘enframing’ had on humans. He describes enframing as systems of how humans understand and perceive their place in the world that impede other ways of understanding the world. Therefore, he argued that technological enframing could lead to an end to the development of new understandings and possibilities of what it means to be human. This would ultimately mean the destruction of the very essence of being human (Allen and Goddard, 2017). Evidently, this critique appears to be of utmost importance, and maybe even more prominent today than it was then. It is critique such as this that will act as a basis for our critical analysis.

In previous times, people worked with an acceptance of the natural properties of things. According to Heidegger, modernity challenges nature “treating it with a violent disregard for their essential being i.e., exploiting natural resources to maximize its utility to human beings” (Allen and Goddard, 2017, pp. 110-111). So, this technological enframing has made humans come to see themselves as economic units to be utilized for the achievement of material ambitions (Allen and Goddard, 2017).

As previously introduced in the ‘Problem Area’, Adorno and Horkheimer argue that modernity’s idea of enlightenment, which is grounded in the notion of reason and rationality, creates an established system of control that inhibits thought and thereby becomes a mythology. We can observe this mythology within the issue of digitalization as these technological systems marketed as freedom is a restraint to our freedom. Adorno and Horkheimer describe this as rational progress reversing and becoming ‘irrational regress’ (Allen and Goddard, 2017).

Adorno and Horkheimer argue that the notion of reason should not be an instrumental and computational science that lacks the consciousness of its own history but one that includes a “reflective consideration of its own goal” (Allen and Goddard 2017, p. 118).

Critical analysis

To sum up, in this paper we offer the idea that digitalization is the latest overarching form of modernity and surveillance capitalism is the latest expansion of the capitalist system. Our case analysis of Spotify provides an example of how digitalization can be harmful to society although this process is commonly believed to be good. This constitutes a case of critical analysis as it is a critique of the system of thought that is prevalent in Western modernity. In order to try and challenge the status of scientific progress in the twenty-first century, it is necessary to emphasize the continuing relevance of critical reflectiveness on how humans are formed by our attempts to understand who we are, in relation to the non-human otherness around us (Allen and Goddard, 2017). This critical reflectiveness is the purpose of this paper.

In the following chapter, we will engage in a critical analysis of Spotify, the company's new patent and their product 'Car Thing'. We decided that it would be most fruitful to approach this topic with a case-based analysis rather than looking more generally at the issue, as we will be able to ground our critique within the specific case and hopefully provide a much clearer and focused analysis. To begin, we will engage in a careful comparison between the Spotify patent and the aforementioned paragraphs from the GDPR and the declaration of Human Rights. This section will be followed by segments elaborating further on the potential dangers of the patented technology and 'Car Thing', including the risk of manipulation and unequal power relations which could have implications for digital citizenship.

Analysis

Spotify, the Voice and personal data

The new Spotify patent is titled ‘Identification of taste attributes from an audio signal’ (Justia Patents, 2021) and the patent claims to do exactly what the title entails. More specifically, the patent provides a series of claims describing how the patented technology will be able to identify these ‘taste attributes’ from an audio signal. In the context of this paper, Claim 6 of the patent is especially significant.

Claim 6 states the following: “The method of claim 1, further comprising: determining at least one of the emotional state, gender, age, or accent.” (Justia Patents, 2021, Claim 6).

To not cause any confusion, claim 1 is mentioned because it refers to the method of processing an audio signal (Justia Patents, 2021, Claim 1). Furthermore, Claim 4 of the Spotify patent states that speech content will be gathered from a provided audio signal through speech recognition (Justia Patents, 2021, Claim 4).

However, continuing from Claim 6, it can be established that Spotify is gathering personal data if the definition of personal data from the GDPR Article 4, Paragraph 1 (listed previously in the ‘GDPR’ section) is applied. Spotify’s patent states that the technology will determine “emotional state, gender, age, or accent” (Justia Patents, 2021 Claim 6) of a listener, which would appear to be within the GDPR’s definition. Paragraph 1 of Article 4 of the GDPR reads that any information related to an “identifiable natural person” and “factors specific to the [...], mental, [...], cultural or social identity” (GDPR.eu, 2018, Art 4.1) constitutes to personal data.

It is uncertain whether the voice itself counts as an ‘identifier’ (and hereby personal data), e.g., the same way an identification number or a name does (GDPR.eu, 2018, Art 4.1). As mentioned, these identifiers could be specific to physical, physiological, cultural and social identity. Based on previous sections, we know these four are embodied in the voice. However, we also know that the voice has yet to be proven as uniquely individual (Watt et al., 2019).

Although the voice is perhaps too intangible to act as an identifier, the voice could very likely be a part of what is called biometrics, or biometric data. The GDPR states examples of specific biometrics, such as facial images and dactyloscopic data, also known as fingerprints (GDPR.eu, 2018, Art.

4.14). This implies that biometrics could be more than just these two examples. Here it is important to mention that in the GDPR's definition of 'processing' it says:

"[...] set of operations which is performed on personal data or on sets of personal data, [...], such as collection, recording, organisation, structuring, storage [...]" (GDPR.eu, 2018, Art 4.2).

The essential part here is the word "recording" which in theory could be a voice recording.

Since the voice is not mentioned explicitly, a degree of interpretation is needed to determine how the voice is protected. Nevertheless, what is certain, is that Spotify will be able to collect personal data through the voice of its users. Depending on the type of personal data which is processed there exists implicit protection of the voice, which will be elaborated further in the following section.

Special Categories of Personal Data

The processing of personal data is not illegal per se, but could potentially be, because of the type of personal data Spotify is planning to process. The Spotify patent states it will determine "at least one of the emotional state, gender, age, or accent" (Justia Patents, 2021, Claim 6) which are strikingly reminiscent of the type of data processing Article 9 in the GDPR prohibits. Article 9 in the GDPR revolves around the 'Processing of special categories of personal data' and it declares that the processing data, including biometric data, revealing ethnicity, religion, sexuality etc. alongside the processing of biometric data with the purpose of identifying a person shall be prohibited (GDPR.eu, 2018, Art. 9.1).

Although, it would appear Claim 6 of Spotify's patent clashes with Article 9 in the GDPR if read very exactly Claim 6 does not assert to determine the specific categories of Article 9. Spotify mentions emotional state, gender, age, and accent, the GDPR mentions "racial or ethnic origin, political opinions, religious or philosophical beliefs" and "a natural person's sex life or sexual orientation" (GDPR.eu, 2018, Art. 9.1). It appears that the Spotify patent has been worded carefully because the patent would not be affected by Article 9 unless 'accent' is interpreted as a way of revealing ethnicity and 'gender' is a way of revealing sex life and sexual orientation. Continuing with 'accent', it seems evident that determining accent will reveal ethnicity, however, a person with Chinese ethnic origin born and raised in Australia could very likely just speak Australian English. In this hypothetical example, accent does not reveal ethnic origin.

Nevertheless, even if the patent would be affected by Article 9, it must be mentioned that Paragraph 2 of Article 9 consists of a list of points which nullify the first paragraph; the first point of Paragraph

9.2 is that Paragraph 9.1 shall not apply if the user has given explicit consent to the collection of their data (GDPR.eu, 2018, Art. 9.2). Evidently, this seems like an easy workaround for a company as dominant as Spotify, a music streaming service that already integrated into millions of people's daily lives.

Human Rights and Spotify

Approaching Article 12 of the Human Rights regarding privacy, it presents itself as very vague compared to the GDPR. The first part of Article 12 states:

“No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence...” (United Nation, Universal Declaration of Human Rights”, seen April 20th 2021)

From this, it is mostly up to interpretation whether Spotify, through its new patented technology, exercises ‘arbitrary interference’ since the user consents to the Terms of Service when using Spotify’s services, and is therefore presented with reasons for this seemingly invasive technology. The reasons for implementing this new patented technology, which are stated in the patent, include that features such as generated personalized experience already exist in on-demand media streaming services and these processes sometimes require tedious input answers (Justia Patents, 2021). As Article 12 continues it shifts its focus to ‘attacks’ on reputation but yet again nothing in the Spotify patent implies this to be the case. However, the next section will present an example of how the Spotify patent could create problems within Article 12 of the Human Rights Declaration.

Terms of Service

As it has been previously touched upon, Spotify’s new patented technology appears to be legal, but possibly problematic, when observed through the lenses of Human rights and the GDPR data protection law. It has also been mentioned how, according to the GDPR, the processing of data revealing personal information about race, sexuality etc. can indeed be legal according to the GDPR law if “the data subject has given explicit consent to the processing of those personal data for one or more specified purposes” (GDPR.eu, 2018, Art. 9.2.a). However, according to the patent, Spotify’s new technology is not only able to collect data on one individual; it is also able to collect so-called environmental meta-data, which could reveal the physical environment in which the data subject appears, such as sounds of birds, printers etc. However, environmental meta-data could also be voices

of other people which could potentially reveal who and how many people are in the background (Justia Patents, 2021). This raises a series of new questions surrounding the issue of whether simply giving explicit consent for personal data to be used and collected on one's own behalf is enough.

We, the researchers, will attempt to highlight this issue by the use of a hypothetical example. The core of this example will be Spotify's new 'Car Thing' which is a physical example of how Spotify's patented technology could be used. The 'Car Thing' is specifically designed to allow users to play music hands-free on Spotify while driving on the basis of voice commands. This device will bring four microphones into the car, and although disabling the microphones should be possible there is no clear 'privacy' button (Pitchfork, seen April 28th 2021). This means that users will have four microphones picking up every conversation happening inside their car unless they actively choose to disable the microphones which would diminish the purpose of the device.

Imagine a scenario where a woman has this 'Car Thing' in her car. She has agreed to the Terms of Service (whether she has properly understood the conditions she has agreed to is another discussion which will not be engaged in this example) and has therefore given explicit consent to the processing of her voice data by Spotify. The woman owns her car; however, she drives her boyfriend to and from work five days a week since they work at the same company. This means that the woman's boyfriend spends almost the same amount of time in the car as the woman. The boyfriend has not given explicit consent to the processing of his data by Spotify, however, since the technology is able to pick up the previously mentioned environmental meta-data, we will have to assume that the boyfriend's voice data is being picked up, analyzed and stored in the same, or at least similar, way to the data of the woman.

The first question raised is then whose responsibility it is to make sure that all passengers of the car are aware of the possible recording of voice data when getting into the car. Is it Spotify who, as the company providing the service and therefore benefitting from the collected data, has the responsibility of ensuring that all data collected by them has happened legally in the sense that all subjects have given their explicit consent to the company? Or is it the woman who, because of the ownership she has of the car and the consent she has given, has the responsibility of passing on the information to her passengers and make sure that they acknowledge and agree to the fact that they subject themselves to this surveillance by getting into her car? This is an example of technological progress moving faster than the regulations and laws which are meant to protect the users of said technology.

The above is just one aspect of this problem, and other issues might potentially arise with time. To highlight these problems, we will return shortly to our hypothetical example. After having driven in the woman's car (mostly as a passenger) for more than a year now it is very plausible that Spotify has collected a large amount of data on the boyfriend. The boyfriend is now being investigated for an alleged felony and in relation to the case, police has asked Spotify to hand over any information they have on the man which has been collected through the 'Car Thing'. According to the company's own privacy policy they:

“share [...] personal data when [they] in good faith believe it is necessary for [them] to do so in order to comply with a legal obligation under applicable law, or respond to valid legal process, such as a search warrant, a court order, or a subpoena” (Spotify, seen April 29th 2021, Law enforcement and data protection authorities).

This means that the boyfriend's data would be handed over to the authorities in such a case. Whether the company records and saves whole conversations or more specific data points regarding mood etc. is unclear. However, the issue here is that there is reason to believe that the boyfriend has not been aware that any kind of voice data which could be used to reveal his identity and, therefore, could be used against him even exists since he has not consented to it.

Previously, it has been mentioned how Spotify's collection of voice data through their patented technology might be considered legal if viewed through the Human Rights article 12 since it can be argued that the user is not subjected to “arbitrary interference with his privacy [...]” (United Nations, “Universal Declaration of Human Rights”, seen April 20th 2021) as a result of the user agreeing to the Terms of Service in which the reasons for this interference are presented. However, the above example could present an issue in terms of the Human Rights Article 12. Because the boyfriend has not agreed to the terms of service the collection of his voice data could be understood as an arbitrary invasion of his privacy. It is arbitrary because he has never been presented with reasons for, or agreed to, this interference.

Another issue connected to this example is the potential for the boyfriend's data to be manipulated. As previously mentioned, we do not know whether whole conversations are recorded or if it is more along the lines of data points. Nevertheless, it is possible that the data could be cut and manipulated in such a way that would make the boyfriend seem nervous, stressed etc., which could potentially portray him in a bad light in front of a judge.

What we, the researchers, have attempted to highlight with this hypothetical example are some of the different issues and questions surrounding the collection of this environmental meta-data. Although the collection of this data at first glance might seem innocent, and as if its only purpose is to help target ads and music recommendations to Spotify's users, it is surrounded by several ethical concerns such as the lack of consent surrounding the storing of this data, it is an invasion of privacy, which is against the Human Rights Article 12, and the possibility of manipulation concerning this type of data.

Manipulation and Exploitation

As touched upon previously, an essential danger of surveillance capitalism is its inherent potential for manipulation (see 'Dangers of Surveillance Capitalism' section).

This often occurs as a result of digital services that are designed to impact or manipulate the users' behavior with the intention of increasing profit. Another factor enabling manipulation is the lack of transparency inherent to the exploitative processes of surveillance capitalism. Companies can manipulate people's actions because these individuals lack awareness of how they are being exploited (Yeung, 2018). The hypothetical of the boyfriend in the car is an example of this, as he would be unaware of the existence of profiled voice data that could be used to reveal his identity. Currently, without the use of the patented technology, Spotify recommends music based on queries that identify the user's preferred content. This also includes identifying characteristics such as gender and age. The stated use of Spotify's patent is to eliminate the time that these queries require for the user. Letting Spotify identify personal taste based on continuously recorded voice samples is time-efficient for the user as the input is passive (Justia Patent, 2021).

However, the implications of Spotify gathering and analyzing this data on a mass scale go beyond increasing the ease of use of the platform. This patent allows for a form of mass predictive personalization that has implications on users' privacy as well as the potential for manipulation (Yeung, 2018).

The use of AI to categorize individuals' emotions can potentially be a manipulative endeavor. Emotion-reading AI has generally been shown to contain biases regarding, for example, race and gender (Rhue, 2018). AI technologies are often regarded as an objective way of analyzing emotion, however, these technologies seem to reflect the same biases as the people who trained the technology. These biases often include a pattern of labelling black people as angry (Rhue,

2018). So, hypothetically the result of AI's biases could be that Spotify will recommend music based on false stereotypes and that Spotify will falsely label people's emotional state based on these stereotypes.

Furthermore, the idea of the interpretation of emotions potentially being objective can be criticized seeing as people perceive voice and facial expressions differently depending on their cultural background. So, even though an AI's interpretation of emotion is similar to people's average interpretation in one regional area, it cannot necessarily be universalized as a successful technology for everybody because the perception of emotion is subjective to culture (Keshtiari and Kuhlmann, 2016).

Concerning the issue of privacy, the safety of the gathered data used to personalize Spotify's platform is also relevant. In the case of the Facebook and Cambridge Analytica scandal, personal data was unlawfully harvested and stored (CNN Business, 2018, seen May 14th 2021). This data was used to influence individuals politically which affected the 2016 US election and the Brexit referendum (Yeung, 2018).

If we imagine a similar breach of data happening in the case of Spotify, or if Spotify starts selling these data to third parties, it could have powerful consequences. As Spotify will implement their product called 'Car Thing', Spotify will be able to gather plenty of data on people's emotional state when they drive (Pitchfork, seen April 28th). Hypothetically, if insurance companies come in possession of this data, we can imagine that it could cause the occurrence of higher premiums for costumers characterized as angry by Spotify's AI. This could potentially happen with, for example, car insurance if insurance companies have data showing that angry drivers are more likely to damage their car.

So, depending on the AI's biases it could implicate a tendency for certain groups, that are more likely to be identified as angry, because of these biases, to receive inferior offers by insurance companies. This form of discrimination could be a general issue if these data will be used in other contexts. Potentially, companies in possession of data on people's emotional state could, for example, use it in their consideration of a person's employability.

So, the use of AI to identify users' emotional state can be misleading. This can have manipulative consequences if people act based on the assumption that these algorithms' identifications are accurate. The unawareness of potential problems and the blind trust in this emotion-reading technology

can increase the power of this manipulation as people act based on a lack of knowledge regarding these processes. This lack of knowledge can, according to the Rathenau Institute, partly be constituted to the ‘creeping development’ of surveillance by tech-companies (Yeung, 2018). Spotify’s new product ‘Car Thing’ is seemingly an example of a subtle, opaque implementation of further surveillance as its marketing completely avoids addressing the fact that the product’s four microphones can be used to implement the technology of the company’s patent on a large scale (Pitchfork, seen April 28th 2021).

Another factor that is fundamental to this potential for manipulation is Spotify’s commercial interests. What drives this development in increased surveillance is its inherent profitability (Yeung, 2018). It is not necessarily in Spotify’s interest to avoid potential manipulation and exploitation as long as the technology increases their profits. So, in the case of Spotify, we can identify an ethical issue regarding the company profiting from human emotions as there is a potential for exploitation and manipulation caused by a possibly systematized misinterpretation of people’s voice data by AI. Furthermore, the case of Cambridge Analytica shows how profiled data also impacts political power relations as it can be used to interfere with people’s political behavior (Yeung, 2018).

Asymmetrical Power Relations

Evidently, in the age of digitalization services are moved to online platforms, which means if citizens want to access these digital platforms, such as Spotify, they must give the platforms permission to collect their personal data. Essentially, users have to agree to everything included in the Terms of Service if they want to access these different platforms.

Supposedly, digital citizens are expected to be able to exercise their societal role on digital platforms and further empower their agency when engaging with ‘liberating technology’ (Gangneux, 2017). But how much power do citizens have when companies essentially are creating the rules? The following statement is from the Spotify patent:

“What is needed is an entirely different approach to collecting taste attributes of a user, particularly one that is rooted in technology so that the above-described human activity (e.g., requiring a user to provide input) is at least partially eliminated and performed more efficiently.” (Justia Patents, 2021, Description of Related Art).

This statement clearly shows how the patent is presented as liberating users from time-consuming queries, although, the queries still allowed users some agency to make choices for themselves.

This transition is indeed opaque, but Spotify also states that the purpose of the patented technology is to eliminate human activity. This is a further expansion of technological enframing and shows a clear lack of “reflective consideration of its own goal” (Allen and Goddard 2017, p. 118). Because of the opacity of the transition created by the nature of enframing, it keeps citizens from realizing the injustices and hereby retaliating. This prevents a common understanding of the injustices concerning this issue from forming, which prevents change, as this is an important factor for political organizing to take place (Bek, et. al., 2020).

Additionally, when companies, such as Spotify, collect users' personal data it creates an asymmetrical power structure where the companies gain more power over the individual users (Freedman, 2020). This is problematic as it permits exploitation of the users, but the asymmetrical power structure also prevents change by reducing the users to merely being data subjects. Companies implementing similar technologies as Spotify have access to the subjects' behavioral data and individual preferences. Not only does this allow companies to operate manipulatively but companies will be able to sell and share the data collected through these implemented technologies, thus further enhancing the asymmetrical power structure (Kulwin, 2019).

This patent puts Spotify in a new position of power. It makes them more dominant than users as they have access and the rights to user's personal data, as soon as they consent to Spotify's Terms of Services. Because of the number of users who use Spotify, it is a significant group of citizens who will fall victim to this further expansion of surveillance capitalism.

Discussion

As stated in the earlier section about ‘Speech Sounds’ it is yet to be proven that our voices are uniquely our own and it is unknown how much information is needed to prove this (Watt et. al. 2019). However, throughout this paper, we, the researchers, have been able to establish that, even though the voice may not count as a personal identifier, the voice embodies characteristics which are able to reveal personal data and emotions. So, if a company, in this case Spotify, has access to its users’ voices the company will be able to collect excessive amounts of personal data and behavioral information about its users. Spotify does not necessarily own individual person’s voices (in a literary sense), but they own the data collected from the voice. As also established personal data is a powerful manipulative tool and users are rapidly being treated merely as data subjects. After a critical analysis of the new Spotify patent ‘Identification of taste attributes from an audio signal’ (Justia Patents, 2021) we must accept that the implementation of speech recognition technologies appears to be legal, in the context of the privacy and data protection laws examined in this paper. However, even though this invasive data collection via speech recognition software is legal, we have identified ethical problems concerning Spotify’s users’ privacy and digital citizenship.

We can in the case of surveillance capitalism conclude that legality does not imply that something is ethical or that it is beneficial to the users. As Spotify is just one example of the exploitative processes inherent to surveillance capitalism, we argue that the processes of surveillance capitalism are generally unethical even if they are legal.

The acceptance of the state of society, which in this case is surveillance capitalism, as necessarily good or ‘ethical’ lacks the rational consideration of the purpose of surveillance capitalism. As touched upon in the previous section, Adorno and Horkheimer explain irrational regress as when a system of thought lacks a “reflective consideration of its own goal” (Allen and Goddard 2017, p. 118). Because of this, we argue that the lack of ethical consideration in the development of new technology constitutes irrational regress.

The idea that legality and ethics are not the same might be presumed as obvious. However, the technological enframing of surveillance capitalism prevents people from considering whether this mass collecting of personal data is beneficial to society’s citizens. So, surveillance capitalism presents a phenomenon in which there is a blind trust in something problematic, even though these problematic practices are not presently illegal. We argue that Spotify is an example of this larger issue.

It is unethical for a company, or anyone, to have access to profiled data, as it is an extreme invasion of privacy. However, because this collecting of data is so 'invisible' this is not always obvious. Imagine if an unknown person would be allowed to enter your private property, or sit in your car, and listen to your private conversation. This person would gather information, analyze your emotions and behavior in order to supposedly improve a product and present it as a favorable change for you, the user. Furthermore, you would have no say in how this person interprets your data and you would not be allowed to access the information yourself. Most would agree that this seems strange and undesirable and would probably not give consent to this. This shows the murkiness of these practices which are already commonplace.

So, technological enframing prevents people from reflecting on the problems of surveillance capitalism which also means that there is less potential for collective protesting of the legal system that allows this. As explained digital citizens are losing agency, but in the case of Spotify, digital citizens also lose ownership of their voice because of Spotify's continued access to their voice data. Apart from collective action, the speed of technological advancement also prevents the actuality of legislation that protect citizens from cases such as Spotify's technology (Bennett Moses, 2007).

Knowledge as a Resource for Change

After having now been presented with an extensive critique of Spotify's patent, voice recognition technology, and technological progress without ethical reflection in general, we, the researchers, find it valuable to present a possible solution to the identified issues. The solution we propose, within the already existing institutions of western modernity, is knowledge, or rather, knowledge through education. School is, among many other things, a very important place for the reproduction of society's institutions and cultural values, as well as for learning how to be a citizen (Durkheim, 1956). It is arguable, that due to digitalization it is today equally as valuable to teach students how to be digital citizens as it is teaching students how to be national citizens.

According to Yeung (2018), people who lack knowledge or have poor digital literacy are more likely to be taken advantage of for example when shopping online. This is due to the asymmetrical power structure created by surveillance capitalism and mass personalization where the companies collecting data will have greater and more detailed knowledge of the users. Individuals do not, in the same way, have access to knowledge of how the companies work which makes them more likely to be

cheated or taken advantage of (Yeung, 2018). By teaching humans what it means to be a digital citizen, what kind of data might be collected on them when they use the internet or accept terms of services on digital platforms, and, most importantly, what kinds of complications exist as a result of the collecting of those data. Making sure that people are knowledgeable and informed about what they agree to when using online platforms may make the power structure between the companies who are collecting the data and individuals less asymmetrical and create greater equality amongst the two.

When knowledge is gathered and implemented it leads to action (CEBI, seen May 15th 2021). The hope is that through education people will become more knowledgeable on the topic of digital citizenship and begin demanding new laws which take into consideration the ethical concerns of implementing voice recognition technology and gathering data on consumers on a mass scale. One example of this could be that individuals become more aware of the importance of the Terms of Service and, as a result, organize in groups to actively change the Terms of Service, with the intention of making the contracts shorter and easier for the average person to read and understand. Today, it takes almost 36 minutes to read Spotify's Terms of Service; this is longer than the U.S Bill of Rights which takes about 31,5 minutes. However, this is nothing compared to Microsoft's terms of service which takes 1 hour 3 minutes and 30 seconds. In fact, the average American would need to spend about 250 hours if they were to read every contract they agree to while using online services (Social Media Today, seen May 15th 2021). Not only is this very time consuming, but even if a person was to sit down and read everything there is almost no possibility of that person being able to change the conditions on his/her own (Berreby, 2017). Therefore, it is crucial that increased knowledge leads to organization and action for things to really change.

It is important to recognize that the idea of change through the education system could be subject to critique. A neo-Marxist perspective on the sociology of education would be that the education system functions as a reproductive mechanism of society's power structures (Carney and Madsen, 2009). From this perspective, in order to achieve change, this system needs to be revolutionized. So, regarding surveillance capitalism, education would only uphold existing power structures, which means one could argue that a digital revolution is needed. Change through a digital revolution, whether by boycotting or other means, would entail legislation which protects individuals from the exploitation of surveillance capitalism hereby creating a more symmetrical power structure. A digital revolution would require digital citizens to have a politicized collective identity which is created by

the common understanding that their position is unjust. For a collective identity to be politicized the creation of a common enemy is necessary (Bek et. al. 2020). In this case, the common enemy provides a service, which the 356 million users enjoy. This results in an ethical ambivalence within the collective identity.

Conclusion

The voice propagates as sound waves, and a human ear will pick up the frequencies it is able to hear. To the human ear, the voice consists of words and utterances. The way these words and utterances are used and articulated reveals characteristics about the speaker. A speaker's way of talking can be influenced by anatomy, social environment and emotions. Voice recognition technology can then pick up audio signals and determine these characteristics because it is programmed to map certain data-values to e.g., certain categorized emotions. However, interpretation of emotion is not objective, and it can have manipulative consequences if an AI systematically misinterprets emotion.

Companies, such as Spotify, are able to collect, capitalize and exploit the voice data gathered by technology since it reveals identifiers, personal information and preferences concerning users of the platform. Continuous surveillance and data harvesting allow companies to mass personalize services, operate manipulatively, and hereby uphold control over their users, or rather data subjects in this context.

Due to digitalization, it would be valuable to teach the proficiency and knowledge necessary for digital citizenship as an essential part of public education. However, even if individuals are aware of the exploitative struggles of surveillance capitalism, individuals cannot change the system on their own. Collective action is needed to cause legislative change and protect individuals from the exploitation of surveillance capitalism. Nevertheless, knowledge and awareness are requisites to challenge the technological enframing of surveillance capitalism, which we argue is necessary for collective action leading to long-term change.

Spotify plans to implement features which are not necessarily beneficial for the user. However, these features have the potential of providing the company with a larger profit and control over the users. Although the technology appears legal with the current privacy and data protection legislation examined in this paper, Spotify's patented technology comes with a series of ethical concerns. This patented technology is an example of how the asymmetrical power structure between users and companies, who gather data on their users, is upheld.

It should be alarming when a music streaming service introduces a technology that serves the purpose of eliminating human activity (Justia Patents, 2021, Description of Related Art) because it is supposedly more efficient. Further technological enframing deprives digital citizens of the opportunity to develop their own (human) understanding of something as basic as their own music preferences. As

a result of this, Spotify's new patent 'Identification of taste attributes from an audio signal' and 'Car Thing' is an excellent example of the dangers of irrational regress as it lacks serious "reflective consideration of its own goal" (Allen and Goddard 2017, p. 118). Since this enframing is so opaque, Spotify further solidifies current unequal structures within the social reality.

Bibliography

- Allen, A. and Goddard, R. (2017) “Modernity and it’s problems.” *I: Education and Philosophy: An Introduction*. London, UNITED KINGDOM: SAGE Publications. <http://ebookcentral.proquest.com/lib/kbdk/detail.action?docID=5942892>.
- Asghar, J. (2013) “Critical Paradigm: A Preamble for Novice Researchers”. *Life science* 10 (December 2013): 3121–27.
- Bek, A. S., Müller M, Middelkamp M., and Thomsen, M. R. (2020) “Acting on Sustainability: A Contestable Concept”. Roskilde University.
- Bennett Moses, L. (2007) “Recurring Dilemmas: The Law’s Race to Keep Up With Technological Change.” *SSRN Scholarly Paper*. Rochester, NY: Social Science Research Network, April. <https://doi.org/10.2139/ssrn.979861>.
- Benzeguiba, M. et. al. (2007) “Automatic Speech Recognition and Speech Variability: a Review”. *Speech Communication* (2007), doi: 10.1016/j.specom.2007.02.006
- Berreby, D. (2017). “Click to Agree with What? No One Reads Terms of Service, Studies Confirm”. *The Guardian*. <http://www.theguardian.com/technology/2017/mar/03/terms-of-service-online-contracts-fine-print>. (Accessed May 15th, 2021)
- Budd, J. M. (2008) “Critical Theory” in Lisa Given (ed) (2008): *Encyclopedia of Qualitative Research Methods*, LA: Sage, 174-178
- Cambridge Dictionary. “Phoneme.” <https://dictionary.cambridge.org/dictionary/english/phoneme> (Accessed March 31st, 2021)
- Carney, S. and Madsen, U. A. (2009) “A place of one’s own: schooling and the formation of identities in Modern Nepal”. In: *Nation-Building, Identity and Citizenship Education: cross-cultural perspectives (Globalisation, Comparative Education and Policy Research, Volume 3)*. Springer, pp. 171-187.
- CEBI. “Knowledge-to-Action Framework”. <https://www.hhscebi.ca/resources/knowledgetranslationframework/>. (Accessed May 15th, 2021)

- Clark, A. (2008) "Critical Realism" in Lisa Given (ed) (2008): *Encyclopedia of Qualitative Research Methods*, LA: Sage, 167-170
- CNN Business. (2018) "What you need to know about Facebook's data debacle". <https://money.cnn.com/2018/03/19/technology/facebook-data-scandal-explainer/index.html>. (Accessed May 14th, 2021)
- Crampton, J. (2019) "Platform Biometrics." *Surveillance & Society* 17 (March 31st): 54–62. <https://doi.org/10.24908/ss.v17i1/2.13111>.
- Cummins, N., Sethu, V., Epps, J., Schnider, S. and Krejewski, J. (2015) "Analysis of Acoustic Space Variability in Speech Affected by Depression." *Speech Communication* 75 (December 1st): 27–49. <https://doi.org/10.1016/j.specom.2015.09.003>.
- Durkheim, E. (1956). "The Nature of Education, Education and Sociology". The Free Press, 57-61.
- Folk, J. "Anxiety and Voice Changes". anxietycentre.com. <https://www.anxietycentre.com/anxiety-symptoms/voice-changes.shtml>. (Accessed March 25th, 2021)
- Freedman, M. (2020) "Business News Daily". [online] *Businessnewsdaily.com*: <https://www.businessnewsdaily.com/10625-businesses-collecting-data.html>. (Accessed May 12th, 2021)
- Gangneux, J. (2019) "Digital Citizenship in a Datafied Society". By Arne Hintz, Lina Dencik and Karin Wahl-Jorgensen. Cambridge, UK. Polity Press. 2019. ISBN 9781509527168." *Information, Communication & Society* 22 (June): 1-3. <https://doi.org/10.1080/1369118X.2019.1635186>.
- GDPR.eu (2018) "Art. 4 GDPR – Definitions". (November). <https://gdpr.eu/article-4-definitions/> (Accessed May 3rd, 2021)
- GDPR.eu. (2018) "Art. 9 GDPR - Processing of Special Categories of Personal Data". <https://gdpr.eu/article-9-processing-special-categories-of-personal-data-prohibited/>. (Accessed April 28th, 2021)
- GDPR.eu. "GDPR Archives". <https://gdpr.eu/tag/gdpr/> (Accessed May 3rd, 2021)
- Guba, E. G., and Lincoln, Y. S. (1994). "Competing paradigms in qualitative research". In N. K. Denzin and Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 105-117). Thousand Oaks, CA: Sage

- Justia Patents. (2021) “US Patent for Identification of taste attributes from an audio signal Patent (Patent # 10,891,948 issued January 12, 2021) - Justia Patents Search”.
- <https://patents.justia.com/patent/10891948>. (Accessed March 24th, 2021)
- Keshtiari, N. and Kuhlmann, M. (2016) “The Effects of Culture and Gender on the Recognition of Emotional Speech: Evidence from Persian Speakers Living in a Collectivist Society.” *International Journal of Society, Culture & Language* 4, no. 2: 71–86.
- Khanna, P. and Sasikumar, M. (2011) “Recognizing Emotions from Human Speech”. In *Thinkquest 2010*. Edited by S. J. Pise, 219–23. New Delhi: Springer India. https://doi.org/10.1007/978-81-8489-989-4_40.
- Kulwin, N. (2019) “Shoshana Zuboff on Surveillance Capitalism’s Threat to Democracy”. <https://nymag.com/intelligencer/2019/02/shoshana-zuboff-q-and-a-the-age-of-surveillance-capital.html>
- Lieber, C. (2018). “Psychologists are speaking out against tech companies that use psychology to lure kids in”. [online] *Vox*: <https://www.vox.com/2018/8/8/17664580/persuasive-technology-psychology> (Accessed May 13th, 2021).
- Long, H., Wu, X., Guo, Z., Liu, J. and Hu, B. (2017) “Detecting Depression in Speech: A Multi-Classifer System with Ensemble Pruning on Kappa-Error Diagram”. *Journal of Health & Medical Informatics* 08, nr. 05. <https://doi.org/10.4172/2157-7420.1000293>.
- Murray, I.R. and Arnott, J.L. (1993) “Towards a simulation of emotion in synthetic speech: A Review of the literature on human vocal emotion”. *JASA* 93 (2), 1097-1108
- Nordquist, R. (2019) “Do You Know What a Dialect Is?”. [online] *ThoughtCo*: <https://www.thoughtco.com/dialect-language-term-1690446> (Accessed March 28th, 2021)
- Pitchfork. “What the Hell Is Spotify’s Tragically Named ‘Car Thing’?”. https://pitchfork.com/thepitch/what-the-hell-is-spotifys-tragically-named-car-thing/amp/?_twitter_impression=true&fbclid=IwAR1TN9J-SeHfJhY5V5DkhgkyGgAGyyF9TvpwrG-YxZNaTDfDVMXfqMWjPk8. (Accessed April 28th, 2021)

- Plack, M. (2005) “Human Nature and Research Paradigms: Theory meets practice”. *The Qualitative Report 10*: 223–45. <https://doi.org/10.46743/2160-3715/2005.1847>.
- Pocket-lint (2021). “What Is Spotify and How Does It Work?”. <https://www.pocket-lint.com/apps/news/spotify/139236-what-is-spotify-and-how-does-it-work>. (Accessed May 14th, 2021)
- Rhue, L. (2018) “Racial Influence on Automated Perceptions of Emotions.” *SSRN Scholarly Paper*. Rochester, NY: Social Science Research Network, November 9. <https://doi.org/10.2139/ssrn.3281765>.
- Robertson, B. (2016) “Q: How Does Speech-Recognition Software Work?” *Science and Children 54*, no. 3 (November): 64–68.
- Robinson, J. (2019) “Received Pronunciation”. [online] *The British Library*: <https://www.bl.uk/british-accent-and-dialects/articles/received-pronunciation#> (Accessed March 27th, 2021)
- Saputra, M. and Siddiq, I. (2020) “Social Media and Digital Citizenship: The Urgency of Digital Literacy in The Middle of A Disrupted Society Era”. [online] *Online-journals.org*: <https://online-journals.org/index.php/i-jet/article/view/13239/6843>. (Accessed May 13th, 2021).
- Saslow, L. et. al. (2014) “Speaking under pressure: Low linguistic complexity is linked to high physiological and emotional stress reactivity”. *Psychophysiology 51*, nr. 3: 257–66. <https://doi.org/10.1111/psyp.12171>.
- Social Media Today. “How Long Does It Take to Read the Terms of Service for Each App? [Infographic]”. <https://www.socialmediatoday.com/news/how-long-does-it-take-to-read-the-terms-of-service-for-each-app-infograph/577235/>. (Accessed May 15th, 2021)
- Spotify. “Spotify - Company Info”. <https://newsroom.spotify.com/company-info/>. (Accessed May 4th, 2021)
- Spotify. “Spotify Privacy Policy ”. <https://www.spotify.com/us/legal/privacy-policy/> (Accessed April 29th, 2021)

- Summa Linguae (2017). “Speech Recognition Software: History, Present & Future”, <https://summa-linguae.com/language-technology/speech-recognition-software-history-future/>. (Accessed May 14th, 2021)
- Thrasher, J. (2019) “Self-Ownership as Personal Sovereignty”; Capman University. https://digital-commons.chapman.edu/cgi/viewcontent.cgi?article=1044&context=philosophy_articles
- United Nations. “History of the Declaration”. *United Nations*. <https://www.un.org/en/about-us/udhr/history-of-the-declaration> (Accessed April 20th, 2021)
- United Nations. “Universal Declaration of Human Rights”. *United Nations*. <https://www.un.org/en/about-us/universal-declaration-of-human-rights>. (Accessed April 20th, 2021)
- Watson, S. (2019) “The Unheard Female Voice”. [online] *The ASHA Leader*: <https://leader.pubs.asha.org/doi/10.1044/leader.FTR1.24022019.44> (Accessed March 27th, 2021)
- Watt, D., Harrison, P. S., Cabot-King, L. (2019) “Who owns your voice? Linguistic and legal perspectives on the relationship between vocal distinctiveness and the rights of the individual speaker”. *The International Journal of Speech, Language and the Law*, vol 26.2 2019 137–1: <https://rucdk.sharepoint.com/sites/GlobalHumanities2ndSemester-semesterstartandprojectgroupform/Shared%20Documents/Anne%20Fabricius/Watt%20et%20al%20IJSLL%202020%20Who%20owns%20your%20voice.pdf>
- Yan, J. (2012) “Listen Closely to Patient’s Voice—You May Hear Depression Signals”. *Psychiatric News*. https://doi.org/10.1176/pn.47.20.psychnews_47_20_17-a.
- Yeung, K. (2018) “Five Fears about Mass Predictive Personalisation in an Age of Surveillance Capitalism.” *International Data Privacy Law* 8, no. 3 (November 22, 2018): 258–69. <https://doi.org/10.1093/idpl/ipy020>.
- Zsiga, E. C. (2013). “The Sounds of Language: An Introduction to Phonetics and Phonology”. John Wiley & Sons, Incorporated. ProQuest Ebook Central. <https://ebookcentral.proquest.com/lib/kbdk/detail.action?docID=1120543>.