Rathenau Instituut

Look who's talking

Tools for the responsible use of speech technology

Report

Authors
Hamer, J., S. Doesborgh and L. Kool

Editor
Rathenau Instituut

Illustrations and photographs
L. Marienus

Cover photograph
ANP

Please cite as:
Rathenau Instituut (2020). *Look who's talking – Tools for the responsible use of speech technology*. The Hague (authors: Hamer, J., S. Doesborgh and L. Kool)

Foreword

In recent years, speech technology has become commonplace. Many drivers give verbal instructions to their cars and some people even wake up in the morning with the voice of their digital voice assistant, to wish them good morning and to provide the weather forecast. We are increasingly talking to computers – and that has consequences.

Digital technology has had a profound impact on our culture and social norms over the past few decades, such as the etiquette for video calls and debates on social media. Speech technology will continue to influence our social relationships and norms even further. After all, nothing is more human than our speech. We express ourselves and develop social interactions in our conversations. It is therefore important to manage speech technology effectively.

Moreover, speech technology gets very close to us: we install speech systems in our living rooms and offices. In the wrong hands, a voice-activated computer is a surveillance tool that can unlock our secrets. You can even clone voices and put words in someone's mouth. Our autonomy is also at stake. Speech technology increasingly functions as a guide that helps the user navigate the digital world. But this guide is created by companies that pursue their own interests, which do not necessarily coincide with the interests and wishes of individuals.

The Rathenau Instituut has therefore devoted this study to speech technology. Based on desk research and interviews, we show how the technology works, what it is used for and what ethical questions it raises. We are looking for ways in which the government, businesses and individuals can help develop speech technology that will enrich our society and social relationships, rather than erode them.

The study calls for the development of ethical speech technology that, among other things, is inclusive, respects our private lives and is marketed in a fair way. The study also calls for societal dialogue and political debate. The rise of speech technology brings with it questions that we need to answer together. For example, do we want to be disciplined by a voice assistant? In the past, this question would have sounded fanciful, but today it is real. Computers have started talking: time for a serious talk.

Dr Melanie Peters
Director, Rathenau Instituut

Summary

Introduction

Computers are getting better at recognising, interpreting and producing human speech. Thanks to improvements in speech technology, it is possible to talk to computers, and users can control the digital world with their voice. Speech technology is already widely used in cars and homes, and companies and organisations are experimenting with it in many other fields, including healthcare and the security sector. The increasing use of speech technology has important consequences for society. Our speech is an essential part of who we are as human beings, and of our social relationships. Our conversations also contain highly sensitive information – about our identity, the type of conversations we have, and even about our health and mood. Our speech therefore has to be protected. This study examines ways in which society can shape this protection.

The study is based on desk research and interviews. Given that the widespread application of speech technology is a relatively recent phenomenon, the desk research consisted of studying a combination of academic literature and grey literature, The interviews were intended to get a better picture of the technical possibilities of speech technology and are exploratory in nature.

Speech technology is getting better all the time

In this study, we first analyse the technical situation: how does speech technology work, and how good is it? Speech technology consists of three key processes: recognising speech, interpreting speech, and producing speech, which is referred to as speech synthesis (see Figure 1). Progress has been made in all three areas, mainly thanks to richer and larger datasets, advanced machine learning technology and faster processing power of computers. However, despite the improvement in quality, it is a mixed picture.

Figure 1 Elements of speech technology

Speech recognition already works pretty well. In ideal conditions, speech computers achieve an error rate of around 5%. But conditions matter a lot: the error rate increases sharply in a noisy setting, when technical words are used or when the system listens to voices of groups that are less strongly represented in the training data, such as those of children. Nevertheless, speech recognition is sufficiently accurate to provide many useful services, for example when it comes to remotely controlling music or transcribing an interview. But there are plenty of applications, such as in healthcare or heavy industry, where such an error rate is not acceptable.

Progress is less unambiguous in the area of **speech interpretation**. When performing tasks, help is needed from the environment and the user: he or she must give simple commands and formulate and answer questions correctly. Although it was promised that computers would learn our language, a human being still has to adapt to speech technology if it is to be interpreted properly.

Speech synthesis, on the other hand, has become much better. In short, speech systems can already make themselves clearly understood. Developers have raised the bar: speech synthesis has to be so good that people are no longer aware that they are talking to a computer. This is not yet the case for the vast majority of applications, but things are developing fast. Some speech systems, such as Google Duplex, come very close to producing human speech, including "ums" and "ahs".

Speech technology is our guide in the digital world

This study also reviewed the application of speech technology. Speech technology is already widely used in cars and in the home. Technology providers and

companies are also experimenting with speech. The applications can be divided into two groups: applications that **control devices** and applications that **support or provide services**. In the first category, we are familiar with speech technology in the car (hands-free calling) and in the home (voice assistants such as Google Assistant or Amazon's Alexa). But machines can be voice-controlled in industry as well. The second category includes voice assistants who book trips for us, assist us at the office, and check our identity, for example when we want to do our banking.

Figure 2 Ethical aspects of speech technology

Impact on social relationships and norms

This wide range of applications raises various societal and ethical issues (see Figure 2). First, speech technology interferes with people's social lives. This raises questions about the desired relationship between people and computers: do we want to, and should we always know, that we are talking to a computer instead of a human being? Do we actually hear who, or what, is saying something? Is it problematic if users consider their voice assistant to be their best friend? And how do we ensure that speech technology respects existing social norms, for example

with regard to equal treatment and disciplining? We have to make sure that speech technology does not compromise our dignity as human beings.

The voice as a new data source

Moreover, all these applications collect data, by means of both call logs and audio recordings. Our study shows this means that the voice acts as a new data source. The data is used by developers to personalise speech systems, and forms the basis of analyses in the field of emotion recognition and the diagnosis of diseases. These analyses are often not scientifically proven, but various companies are expecting a lot from the future possibilities of audio recordings. Speech data contains very sensitive information: more than anywhere else, people reveal themselves in conversations at home, in the car and at work. This requires extra attention from developers and regulators to ensure that our private and family life remains respected.

Our autonomy is at stake

The use of speech technology also affects our autonomy. This technology helps users in many domains to perform tasks, take decisions and have a pleasant experience. This offers opportunities but also raises concerns. Does the use of speech technology lead to the loss of skills, and does it exert a malign influence and mislead? Take, as an example, deep fake videos, in which someone's appearance and voice are faked ("cloned"), and which can fool people and undermine public debate. In addition, speech technology offers fewer possibilities for nuance and questioning compared with screens. Who controls and decides on the answer the voice assistant gives? Finally, an empathetic and comfortable voice assistant can be so useful that people overuse it and become addicted to it.

The importance of safe and healthy use

Speech technology can also compromise people's security. Speech data can be stolen and misused, for example to commit identity fraud. And, despite the improvements made, speech technology is not perfect and accidents can happen. Before speech technology is used in critical applications in healthcare, defence or manufacturing industry, the reliability of the technology will have to be beyond doubt and money will have to be invested in technologies to combat misuse.

Tech giants' growing market power

Finally, the study shows that the power of several large technology companies is growing even faster thanks to speech technology. The objective of several technology giants such as Google and Amazon is to create a broad platform of speech applications and link them to a voice assistant, such as Alexa and Google Assistant, which can perform a multitude of tasks. In this way, these assistants will take on the role of a guide who helps us navigate through the digital world, while

keeping us as deep as possible within the environment of a particular platform. To achieve this, technology giants are buying up start-ups and making significant investments. Although other players are also active in the speech technology market, such as the Houndify platform, and companies sometimes develop their own voice assistants, the question is how these players will hold their own against the tech giants' increasingly dominant position.

Recommendations for government and industry

Our voices and conversations are an essential part of who we are as human beings and the relationships we enter into with others. Speech technology gives us someone to talk to at all times – at home, in the car, at work and when shopping – and this will affect our speech and our relationships – both with each other and with computers. In addition, speech technology creates a new source of data, containing highly sensitive information. Our speech is at stake.

Speech technology adds a new dimension to the general task of managing digital technology effectively and shows that government and industry are once again taking the lead. After all, speech technology not only influences the way individuals use computers, it also affects the behaviours we develop together. It changes not only the way in which individuals acquire knowledge, but also the knowledge on which public debate is based. And it has an impact not only on the relationship between customers and companies but also on the platform economy as a whole.

The Rathenau Instituut is therefore making six recommendations to government and industry to protect human speech and to manage the speech technology applications effectively:

Recommendations	Government	Industry
1. Ensure effective privacy protection	Introduce permit requirement for voice analysis. Monitor use of speech technology by law enforcement agencies.	Implement privacy principles rigorously.
2. Promote inclusive speech technology	Invest in a Dutch speech database. Call for industry to accept its responsibilities.	Beware of stereotyping in use of voice and promote recognition of diverse language use.
3. Create a fair market	Tighten up competition law. Provide opportunities for alternative suppliers.	Prioritise the rights of consumers.
	Government and industry	
4. Protect human dignity	Initiate an ethical dialogue on the use of speech technology and make agreements.	
5. Make sure speech technology is reliable	Address disinformation. Reduce error rates of speech systems.	
6. Invest in technological citizenship	Educate individuals to deal responsibly with speech technology, and boost research.	

1. Ensure effective privacy protection

Speech technology makes it possible to collect sensitive voice data from people and use it to influence them. This includes biometric and health data. This means the processing of voice data poses risks to people and their fundamental rights. Existing privacy rules must be enforced more vigorously. The Rathenau Instituut is therefore calling on the government to introduce a permit system for biometric voice analysis and to develop strategies to regulate emotion recognition and health analysis. It is also important to monitor the use of speech analysis by law enforcement agencies: is it desirable for the police to scrape voice data from social media? Finally, it is incumbent on industry not to follow the minimum of privacy rules in their product development and service provision, but to implement them vigorously – for example by investing in technologies that minimise the data use.

2. Promote inclusive speech technology

Speech technology provides opportunities to make information more easily accessible. But speech systems can also exclude groups of users, confirm biases, and encourage discrimination. It is very important to ensure that everyone can use speech technology. To this end, government can invest in a Dutch speech database on which numerous players can base their speech technology. Industry also has responsibilities in this regard. In particular, the Rathenau Instituut calls on

industry to combat stereotyping, for example by offering a diverse range of voice assistants.

3. Create a fair market
Concerns have been raised in the data economy with regard to the dominance of a few large technology companies. Speech technology gives these companies an opportunity to expand this dominant position even further. In order to make the market accessible and fair to all players, government can tighten up competition law – steps are being taken to this end at European level. It is also important to provide opportunities for alternative providers, and not just to work with the tech giants. Industry is recommended to apply consumer rights, such as the right to request information, effectively and generously.

4. Protect human dignity
The Rathenau Instituut calls on government and industry to initiate an ethical dialogue on speech technology. Particular attention should be paid to protecting human dignity: guaranteeing the right to human contact and preventing situations in which users confuse computers with people. Government and industry should reach agreements in this regard.

5. Make sure speech technology is reliable
Speech technology has a lot to offer society, provided that it is reliable. It is up to both government and industry to take the following steps: act decisively to combat disinformation and voice cloning, reduce the error rate of speech technology, invest in technology that prevents misuse and develop security standards.

6. Invest in technological citizenship
Responsible and effective use of speech technology also requires knowledge and skills, for example in terms of searching for knowledge and setting up routines, and the information the devices collect. It is therefore necessary to assist people to deal with speech technology, which requires investment in education and training in media literacy. In addition, government, knowledge institutes and industry must invest in research to analyse the impact of speech technology on our physical and mental health. Finally, individuals also have an important role to play. They can make their voices heard and put speech technology on the agenda for public debate. Our speech is a vulnerable and meaningful commodity – and worthy of debate.

Contents

Introduction

1.1 Computers that talk

Sometimes science fiction becomes reality. Decades ago, writers and film-makers such as Isaac Asimov – I, Robot – and Stanley Kubrick – 2001: A Space Odyssey – fantasised about computers that can talk like people. This fantasy has become more and more real in recent years.

With the advent of artificial intelligence and the increased computing power of processors, computers are increasingly able to recognise, interpret and produce speech. Human language is no longer the exclusive domain of humans – in more and more conversations, speech systems are joining us. Like Google Duplex, which calls restaurants for you and makes a reservation on your behalf. Or Microsoft Cortana, which allows you to manage your diary and send emails. Or Athena, a voice assistant for industry, which enables machines to be controlled by voice.

The market for speech technology is currently growing rapidly. For example, 6% of Dutch households purchased a voice-operated speaker in 2018, which grew to 19% in 2019 (Multiscope, 2020). And things are developing even faster in America and China (Kimmich, 2019). According to some analyses, the rise of smart speakers seems to be even faster there than the rise of mobile phones previously – the mobile phone that you can increasingly control with your voice nowadays as well (Kinsella & Mutchler, 2018).

Because you can interact quickly and intuitively with a speech system, interest in speech technology is growing. Quickly, because people can say an average of 150 words per minute but can only type forty words per minute. Intuitively, because for the first time we do not have to learn computer language, as computers use our language.

The advent of speech technology will have a major impact on society. Our relationship with the digital world changes when we can talk to computers – and all kinds of digital devices. Where we used to search a screen by means of search engines and click through, one answer now comes back. Where we used to call customer service to be helped by a human being, we can now get a digital voice assistant on the line, which sounds more and more like a human being.

Speech technology is also changing our relationship with each other. With family, it is not necessary for everyone to look at their own screen, as they can all have a conversation with the voice assistant. And maybe the way we talk to computers also influences the way we talk to each other.

Finally, our relationship with companies and institutions is changing, for example because we no longer come into contact with a human employee, but speak to a company's digital voice assistant. Or because our voice is used as a means of verifying our identity in transactions with companies. Companies are going to see both *what* we say to computers and *how* we say it: our voice, conversations and emotions form a new, interesting source of data that companies can analyse, for example to detect diseases and identify people.

Our speech, and our conversations, are precious. We use our language to express who we are and enter into many valuable relationships. Nothing is more intimate, and more personal, than our voice. Therefore, given the rise of speech technology, it is now important to investigate the societal and ethical questions associated with speech computers. What does the use of speech technology mean for our privacy? What happens when we learn from childhood to give commands to digital devices? Can we still keep speech computers and people apart? What are the consequences if speech computers guide us through the digital world? And how will speech technology affect the platform economy and the competition between technology giants like Google and Amazon?

This study will address these and other questions. We will review technological developments, take stock of where speech technology is used and identify the societal and ethical aspects. The study will show why a societal and political debate on speech technology is necessary, and what that debate should be about. The study suggests actions for government and industry, and identifies central themes for public debate.

Isaac Asimov foresaw that intelligent robots would need rules and boundaries. He therefore devised three rules, the first of which was the most important: "A robot may not injure a human being or, through inaction, allow a human being to come to harm" (Asimov, 2008). Now that a new piece of science fiction is becoming reality, our society will also have to think about the public values and norms that can manage the rise of speech technology.

1.2 Delimitation

This study is about speech technology. By this we mean technology that enables computers to recognise and interpret human speech and to talk by themselves – summarised as speech recognition, speech interpretation and speech synthesis (see Chapter 2).

A computer system that can perform one or more of these three tasks is known as a **speech system**. Different types of speech systems are available on the market. The main one is the **voice assistant**, a speech system that can usually perform a wide range of tasks. Amazon's Alexa and Google's voice assistant are well-known examples. These assistants can be installed on all kinds of digital devices, including a mobile phone, a desktop PC or a **smart speaker.** The study also looks at other speech systems, such as transcription software and navigation systems.

Voice assistants are also known as **cognitive** or **virtual assistants**. These digital systems can also perform tasks and are usually capable of interpreting text. They do not have to be based on speech technology. In this study we focus on systems equipped with speech technology. We will therefore use the term "voice assistant".

1.3 Research questions

Speech technology is making rapid advances. The key research question of this study is therefore:

How can the rise of speech technology be managed effectively?

In order to address this question, this study performed three analyses. First of all, we analysed the **technical potential** of speech technology. How has speech technology developed in recent years? What is possible now, what is not, and what is speech technology expected to look like in future?

We then analysed **which applications** speech technology already has, and which are emerging. What is speech technology used for? And by whom? What happens to the data that speech technology collects? What can you tell from a voice? And what does the market for speech technology look like?

Finally, we analyse the **societal** and **ethical aspects** associated with the rise of speech technology. How is speech technology changing the interaction between people and devices and between individual people? What questions does that raise? Which public values are compromised by speech technology and which public values can be promoted by speech technology?

Based on these three analyses, we present a number of actions in the conclusion and identify themes for public debate.

1.4 Research method

This study is based on desk research and interviews.

The desk research consisted of a combination of academic literature and grey literature, given that the widespread application of speech technology is a relatively recent phenomenon. In this case, "grey literature" refers to non-academic sources such as government documentation and reports, journalistic articles and publications from the private sector. Exploratory in nature, the interviews were intended to get a better picture of the technical possibilities of speech technology. They were held with experts in speech technology: Roeland Ordelman of TU Twente, Olya Kudina of TU Delft, Jeroen Vonk of Novum, Piet van Dosselaar of Statistics Netherlands, Voice Consultant Maarten Lens-FitzGerald, Vanessa Hendriks of Google Netherlands, Jeff Adams of Cobalt Speech and Clive Summerfield of Auraya Systems.

2 Speech technology: a technical analysis

If we are to believe technology companies, speech technology promises a lot. Technology companies view speech as a central means of controlling the digital world, with all its possibilities, from the comfort of your armchair. See, for example, Microsoft's *Voice Report* (Olson & Kemery, 2019). Using your voice to write text, retrieve data, turn on music, open doors, drive cars, control construction machinery – speech , they say, is the interface of the future.

Whether this dream will come true depends, among other things, on how well the technology works.[1] This chapter charts the technological development of speech technology. We explain exactly what speech technology is and investigate how speech technology makes different applications possible. We also investigate the future promises of speech technology – which applications are currently out of reach, but could be realised in the near future?

This chapter is mainly based on the work of computer scientists who conduct research into speech technology.

2.1 Speech technology in a nutshell

Speech technology allows the user to control applications by using his or her voice. For example, if you are curious about today's weather, do not type anything in, just ask a device such as a smartphone or smart speaker out loud what the weather is going to be like. The advantage of speech technology is readily apparent: for most people, talking is faster and easier than typing. This is also shown by scientific research on human-machine interaction (Nass & Brave, 2005).

In this study we look at three components of speech technology. First, speech has to be **recognised**. The system has to convert the captured sounds into text – that is why this technology is also known as ″speech to text″ (Liu et al., 2019). Anything can go wrong in this conversion: differences between words, such as "lunch" and "punch", are sometimes minimal. And words can be lost through noise.

1 How a technology is ultimately used depends on many factors, including how users will appropriate the technology, and how players deal with various societal and ethical issues. These factors are discussed in greater detail in the following chapters.

The system then has to **interpret** the captured text (Mittal, 2019). If someone asks "how old is the Utrecht Dom tower?", it is crucial that the software interprets these words in the right way and in the right context. For example, the user does not want to know how old the city of Utrecht is – and the system needs to understand that.

Then the system has to do what was asked. This can be anything from looking up and playing a song, to calling people and turning up the central heating. Sometimes the device ends the interaction with a confirmation such as: "All right, I'll turn up the heating". But for other tasks, one question is not enough and the user enters **into conversation** with the device. The device must be able to speak in both cases – this is referred to as **speech synthesis** or text to speech (Kuligowska et al., 2018). During such a conversation, the system alternates between speech recognition, speech interpretation and speech synthesis.

In summary, speech technology consists of **speech recognition, speech interpretation** and **speech synthesis** (see Figure 1). Much progress has been made in each area in recent years (see box 1). We discuss these three elements in more detail below.

Figure 3 Elements of speech technology

Box 1 Milestones in the history of speech technology

1922 – **Radio Rex**. Rex was a toy dog that came out of his kennel when you called his name. The dog responded to his name because the kennel had a built-in electromagnet that was sensitive to frequencies of about 500 Hertz, the sound of the vowel "E" in REX.

1939 – **VODER**. Homer Dudley presented VODER (Voice Operating Demonstrator) at the New York World's Fair. It was the first electronic version of previously developed mechanical speech machines. A trained employee was able to produce sounds of speech by pressing key combinations.

1962 – **Shoebox**. IBM developed Shoebox, a computer that could recognise sixteen spoken words, including the digits zero to nine. The Shoebox was linked to a calculator and, as the name suggests, was the size of a shoebox.

1979 – **MITalk**. Researchers at MIT developed MITalk, a system that could produce speech by merging sound waves. One of its users was Stephen Hawking.

1982 – **Dragon Systems**. Dragon Systems produced speech recognition software that used statistical models and was able to convert spoken words into text that appeared on a screen. These systems were the forerunners of today's dictation software.

2001 – **Natural Voices**. AT&T introduced Natural Voices in 2001, a natural-sounding speech synthesiser that pasted small pieces of speech together. The system was widely used in online applications, such as websites that could read emails aloud.

2011 – **Siri**. Apple launched Siri in 2011, as a software component of the iPhone 4S. Siri heralded the breakthrough of voice assistants that could recognise, interpret and synthesise speech well enough to perform a wide range of tasks.

2.2 Speech recognition

2.2.1 How does speech recognition work?

The purpose of speech recognition is to convert a speech recording into a clearly arranged text (Huang, 2014). Speech systems can do this in two ways: by performing a **hybrid analysis** or an **end-to-end analysis**.

Hybrid analysis
A hybrid analysis is the most commonly used method for converting audio into text and consists of **four elements** that work together to produce the best result (Jyothi, 2018).

1. An **acoustic analysis** takes place when the computer divides the recorded sound into chunks of a few milliseconds and analyses it (O'Shaughnessy, 2013). Some sound chunks are louder than others and have a different pitch. The chunks are therefore graded according to their volume – in **decibels**, and their frequency – in **Hertz.**
2. The computer continues to analyse the chunks using an **acoustic model** (Li et al., 2017). This model detects distinguishing features of sound – known as **phonemes**. These include the **b** and the **d** that distinguish the words **beer** and **deer**. They are the smallest units of language from which you can then build sentences. Based on the acoustic analysis, the computer calculates the probability of certain phonemes belonging to certain audio clips, and then makes a proposal, for example that the sound consists of **l**, **ay**, and **k**. Computers can perform this task better if they are trained with lots of audio clips (Jyothi, 2017).
3. The **pronunciation model** makes words out of the proposed combinations of phonemes: **l, ay,** and **k,** for example, form the word **like** (Jyothi, 2017). The computer usually performs this task on the basis of a dictionary created by language experts that links combinations of phonemes to words.
4. Finally, the **language model** turns the words into sentences (Sundermeyer et al, 2012). Here, too, the computer calculates the probability that certain words will follow one another. And here, too, the software is trained on the basis of datasets consisting of many existing texts containing countless sentences. The model actually asks: what is the probability that the word "dog", for example, is followed by "runs", "suns" or "tuns"? The sentence construction with the greatest probability then emerges. In this case, "runs" is the most probable, followed by "suns" – but definitely not "tuns".

The different elements are compared and repeated by a **decoder** until the computer produces the most probable sentence (Jyothi, 2017). The probability calculation is especially complex in this step, which therefore requires the most processing power.

Over the past ten years, hybrid systems have increasingly made use of **deep learning** (Huang, 2014). This is a type of artificial intelligence that uses neural networks and large amounts of training data. A neural network is best understood as a complex calculator modelled on the human brain, which can calculate and compare many different outcomes and scenarios using probability calculation. As a result, a neural network can learn for itself how best to use training data to categorise data. The neural networks improve the aforementioned elements based on probability calculation: the acoustic analysis and the language model.[2]

End-to-end analysis
The most recent development in the field of speech recognition is end-to-end analysis. In this analysis, the entire conversion from audio to words is performed by a neural network (Chiu et al., 2018). This means that certain results are not checked with a language model, but the model calculates which letters, words and sentences are most likely to fit a fragment of audio. The advantage of this is that these systems are easy to train and require less computing power than the hybrid systems (Synnaeve et al., 2020). This ensures that these systems can be installed more easily on devices themselves – and are not dependent on external computers that store files in the cloud or online (Li et al., 2020; Meyer, 2019; Mwiti, 2019). The disadvantage of end-to-end systems is that they can also record incomprehensible text and non-existent words, because results are not checked with a language model created by experts. It is therefore not yet clear whether end-to-end systems will replace hybrid systems.

2.2.2 How well does speech recognition work?

Errors can occur in all elements of speech recognition, which can lead to peculiar examples, for example when the sentence "Recognise Speech Using Common Sense" is recognised as "Wreck a Nice Beach You Sing Calm Incense"(Lieberman et al., 2005). In this case, the words were all recognised incorrectly and are in the wrong order. Sometimes entire sentence fragments can be lost because the microphone fails to pick up the sounds properly or because fragments are lost when the various chunks of audio are cut up. In 2019, for example, the White House's

2 All major commercial speech recognition systems use neural networks (Kepuska & Bohouta, 2018; Jyothi, 2018).

speech recognition system was found to be making mistakes, resulting in an incomplete transcript of a telephone call being sent to Congress (Barnes, 2019).

Nevertheless, speech recognition has progressed by leaps and bounds in recent years (Huang, 2014; Kodish-Wachs et al., 2018; Meeker, 2018; Sokol et al., 2017; Protalinski, 2019). This is mainly due to the increased **availability of data**, **increased computing power** and the use of **deep learning technology** (Hoy, 2018). Scientists, as well as commercial speech applications already on the market, have been collecting more and more speech data in recent decades. More and more language areas are being opened up so that the speech systems can recognise not only English and Spanish, but also other languages. As a result, Google Assistant can now recognise Dutch. Faster computing capacity makes it possible to process and analyse these large amounts of data, and, finally, deep learning is making the analysis more and more accurate. As a result, speech recognition is showing rapid improvement and can be used successfully for a growing number of applications (see Chapter 3 below).

The progress of speech recognition can be measured by the error rate: the percentage of words that were mistranslated, omitted or added by mistake. The best speech systems, the voice assistants of major technology companies such as Alexa and Google Assistant, now achieve an error rate of around 5%. This is a significant achievement because the error rate of human speech recognition is about the same (Jawad, 2017; Meeker, 2018; Chiu et al., 2018).

However, this 5% error rate is misleading. The high success rates only occur in speech systems trained with very specific datasets and operating under optimal conditions. Adams (personal communication, 15 June 2020), project leader of the team that developed Alexa, agrees that these figures are misleading. For example, if speech is more accented or if a sound recording contains more ambient noise, most systems will have higher error rates.
Speech systems can also have difficulty with the specific speech of children, older people or certain dialects because insufficient data is available from these groups. Companies focus on markets and languages for which a lot of data is available, such as US English and Mandarin Chinese. Moreover, speech systems have difficulty with the specific terminology of disciplines such as law or health care.

Speech recognition works best when certain conditions are met: (1) the system has been trained with fragments from all speakers using the system, (2) the system knows all the words used and (3) the system has been trained in all possible recording conditions (O'Shaughnessy, 2008). It is a huge challenge to satisfy these conditions because of the enormous variation in language and locations where language is used. Often voice assistants perform especially well when the user is

forced to use a limited set of simple commands – Condition 2. But this does of course mean that the user has to adapt to the speech system.

The error rate in various speech systems and applications is well above 5%. The error rate is around 10% to 12% for transcribing telephone calls, between 10% and 20% for dictating texts and subtitling and around 40% to 50% for transcribing a meeting between a group of people (3PlayMedia, 2019; J. Adams, personal communication, 15 June 2020). Moreover, these figures refer to systems that recognise accentless English – the systems make more mistakes with other languages and accents. Finally, the error rate can be even higher than 50% in specific usage contexts, such as a hospital (Kodish-Wachs et al., 2018).

Although speech recognition is clearly improving, the challenge of increasing accuracy in difficult conditions remains and a lot of work will have to be done to recognise accents and less common languages as well (O'Shaughnessy, 2008; J. Adams, personal communication, 15 June 2020). For use in critical applications in particular, such as in a medical procedure or controlling a lorry, speech recognition is not yet good enough.

2.2.3 How does speech interpretation work?

If the previous step, speech recognition, has gone well, a device will have the correct text. Sometimes this is also the end point, for example when you dictate and the computer writes down the speech. But in many cases the application is more complicated and the text gives an instruction, for example to buy a scarf. Then it is important to interpret the text correctly. We refer to this as **text analysis**, a kind of analysis that is closely related to **natural language processing** (NLP). In addition, the audio file can also be analysed in detail, for example to detect an emotion or to diagnose an illness. This is known as **voice analysis.**[3] We'll discuss them one after the other.

Text analysis
Text analysis is based on NLP: the automated interpretation of text. NLP is interdisciplinary, combining artificial intelligence with linguistics, computer science and informatics (Chowdhury, 2003). NLP's applications are many, and more varied than just speech technology. They include the operation of spellcheckers, spam filters, search engines and the monitoring of social media.

3 Voice analysis is sometimes regarded as part of speech recognition. Recognition and interpretation are closely interlinked. Because interpretation becomes richer when voice analysis is added, we discuss the technology as part of speech interpretation.

Understanding text is one of the most difficult elements of speech technology – and that is not surprising. After all, human language is rarely precise and tends to be unstructured, ambiguous and context-dependent. For example, a sentence like "I'll pick it up" can refer to picking up a physical object, but also to learning to perform a certain task. These subtle variations are a challenge for speech systems because it can be difficult to determine which possible interpretation is the right one (Manning, 2017). This is often not clear even to people – we regularly ask each other to repeat or clarify a request.

This problem can be solved in two ways: by requiring users to use simple, unambiguous commands or by improving the speech system itself. The second solution is, of course, the most attractive – if a speech system understands you more easily, it will be more efficient and convenient to use.

NLP focuses on the latter and tries to understand language by performing various analyses. These can be classified according to three levels: a **syntactic analysis**, a **semantic analysis** and a **pragmatic analysis**. We discuss them briefly below:

1. A **syntactic analysis** looks at the grammar of a sentence. It breaks the sentence down into words and punctuation and traces words back to their root. It also looks at prefixes and suffixes such as "un" and "ly". Next, the words are classified grammatically, for example by indicating which word is an article and which is a noun, thereby creating a grammatical overview of the text.
2. A **semantic analysis** looks at the meaning of the words: do they describe people or locations or products? In doing so, the system labels the words, paying close attention to the context – a task that is difficult to train. The automated comprehension of language is known as natural language understanding (NLU).
3. **Pragmatic analysis** figures out the intention of a text. What's the message? Roughly speaking, there are two common intentions in speech technology: to look something up and to perform an action. This intention is usually reflected in the words used, such as "what is" when looking up or "play" when performing an action. The user often knows which specific words to choose in order to convey the right intention. The more of these keywords the digital application understands, the better the system will be able to figure out the intention.

Contemporary speech technology often uses a combination of these processes to determine the meaning of a text.

Voice analysis

Speech data makes it possible to perform all kinds of voice analyses. These are based on "vocal biomarkers", biological characteristics that can be measured on the basis of a person's voice. This kind of analysis has aroused the interest of both scientists and industry. In Chapter 3 we look in more detail at various applications used by industry.

Voice analysis can be used, among other things, to try to determine someone's age, identity, gender and height (Chaudhari & Kagalkar 2012; Müller, 2006; Pisanski et al., 2014). Work is also being done on detecting diseases such as alcoholism, sleep apnoea, lung disease, depression, PTSD, emotional stress, anxiety disorders, autism, Parkinson's, Alzheimer's and even brain tumours (Place et al., 2017; Sondhi et al., 2015; Senseable Intelligence Group, 2019; Marmar, 2019). Let's be clear: it is valuable information that voice analysis is trying to retrieve. Sometimes companies try to leverage that value right away, for example by providing feedback to an employee in a call centre based on voice analysis, or providing advice if the system detects that someone is getting upset (Simonite, 2018).

Adding voice analysis to text analysis enriches the interpretation of someone's speech. After all, people also understand each other by paying attention not only to words, but also to the intonation and the way in which someone says something. Expectations are high, especially for some commercial stakeholders. For example, health technology company Vocalis Health claims that voice analysis can be used to draw up a personality profile (Chen, 2019).

2.2.4 How well does speech interpretation work?

Are today's computers equal to the richness and complexity of human language? Speech technology is also improving in this area. Again, we make a distinction between text analysis and voice analysis.

Text analysis
Once again, the increase in data, computing power and the use of deep learning systems is making text analysis faster and more accurate (Young et al., 2018). However, although speech interpretation is generally improving, some developments are faster than others.

For example, while software is increasingly able to dissect sentences grammatically, label words correctly and identify spam text, it still finds it difficult to

properly interpret more complex questions, understand ambiguous words, conduct dialogues and summarise the content of texts (MacCartney, 2011; Young, 2018).[4]

Speech interpretation works particularly well where the environment and the user cooperate – just like speech recognition. The environment can help by, for example, setting up websites "voice first", using certain key words and incorporating short chunks of information that can be recognised by the voice assistant. This makes it easy for a speech application to find information (Koksal, 2018). The user can help by using specific, recognisable words and articulating them properly. If users start speaking in a more natural and varied way, for example using a metaphor, speech applications quickly lose track.

This means that speech technology can be used successfully in some applications, such as a voice assistant that looks up the weather forecast and plays music, while other applications are still a long way off, such as a voice assistant that you can have an entertaining conversation with (see Chapter 3).

Voice analysis

And how good is voice analysis? This question mainly concerns the quality of the vocal biomarkers: do certain diseases and certain emotions really have distinctive vocal markers? In order to find out, you need to compare two groups and, say, investigate whether a group of Alzheimer's patients has a certain vocal marker that is lacking in a group of non-Alzheimer's patients. These groups must not differ in other respects, for example the age of the two groups must be the same.

This example neatly illustrates the challenge faced by voice analysis. It requires a high quality dataset to train the artificial intelligence. Many researchers maintain that recordings of tens of thousands of correctly selected subjects are needed before sufficient proof can be provided that the algorithms work (French, 2019). Moreover, a correlation differs from a causal relationship: it is often unclear why voice markers correlate with a certain disease or characteristic. Several scientists are therefore sceptical about the ability of voice analysis to discover characteristics and create profiles (Chen, 2019). Caution is advised.

Speech technology does seem to be able to reliably establish certain characteristics, such as identity, gender or age group (Sedaaghi, 2009; Gupta et al., 2019; Abdulsatar et al., 2019). Moreover, not every application requires the same level of reliability. Diagnosing a disease will have to meet completely different

4 https://nlpprogress.com/ is a website that developers use to track the development of NLP topics.

requirements than determining a certain music or film preference. For the time being, speech technology is unable to provide sufficiently reliable voice analyses for most characteristics (Wenderow, 2019).

2.3 Speech synthesis

2.3.1 How does speech synthesis work?

Finally, there are also digital speech applications that not only recognise and interpret text, but can also speak for themselves. For example, information about a known person can be read out, or the digital speech system confirms a certain command: "Here's John Mayer's Spotify playlist". Also, the application may ask for further information in order to answer more complex questions. We briefly explain how speech synthesis works.

Roughly speaking, there are two types of speech synthesis: **concatenative synthesis** and **model-based synthesis** (Jothilakshmi & Gudivada, 2016):

1. **Concatenative synthesis** (also known as sample-based synthesis) uses a database of recorded segments of speech and chops them into small fragments (Kishore & Black, 2003). These fragments are then used to form other words. For example, the word "impressive" can be synthesised from previous recordings of "imp" as in impossible, "pres" as in president, and "ive" as in detective. In this type of speech synthesis, you always have to strike a balance between larger speech fragments that sound more natural and smaller, flexible speech fragments, which can come across as faltering, but can form more words.
2. **Model-based synthesis** (also known as parametric synthesis) mimics speech by extracting acoustic properties such as frequency spectrum, intonation, and emphasis from datasets of speech fragments (Tokuday & Zen, 2015). In this type of speech synthesis, a digital application is trained on the basis of speech recordings with the corresponding text. This creates a model that can reproduce these acoustic properties. New text is then entered to convert the trained model with the acoustic properties to speech. The advantage of this method is that the voice is very flexible. Based on a limited number of recorded fragments of text, the model can generalise the voice and produce sentences that have not been recorded. The accuracy of the voice improves as more speech data is recorded. The use of deep learning has boosted the development of this type of synthesis (Zen, 2018).

2.3.2 How well does speech synthesis work?

Can computers talk like people? Or do we recognise a robot's voice right away? For the developers of speech technology, these questions are crucial. As Sundar Pichai, CEO of Google LLC, said about their digital voice assistant: "We want the assistant to be something that is natural and comfortable to talk to. And to do that we need to start with the foundation of the Google Assistant, the voice". (2018, 16:55)

According to scientists, a human computer voice reduces misunderstandings, makes it more pleasant to work with and even creates trust, says psycholinguist Wagner (Stinson, 2017). Humans naturally attribute human characteristics to computers, a phenomenon known as the "Eliza effect" (Weizenbaum, 1966). Voice assistants can enhance this effect by using a human voice. But has there also been success in creating a human computer voice? Are digital speech systems so good that they are indistinguishable from humans – and therefore pass the so-called Turing[5] test?

This question cannot be answered with a simple "yes" or "no". On the one hand, computer voices can usually be distinguished from human voices. On the other hand, computer speech is rapidly improving. For example, developers are successfully making better use of pauses and intonations in speech software, and making the systems speak more fluently (Oord et al., 2016). Google Duplex, a Google Assistant service that can make reservations for you, uses sounds like "um" and "hmm" and informal language ("oh yeah") to come across as even more human. The technology behind Duplex produces speech that is sometimes indistinguishable from human speech, as Sundar Pichai demonstrated during the Google developer conference in 2018.[6]

It is, however, not surprising that Duplex achieved this because the application was limited to a single task: making reservations. For this reason, a lot of attention can be devoted to improving a small set of answers. Another factor is the fact that Duplex has an English voice. More than enough data is available to produce speech in the English language. After all, progress will continue to depend on the amount of available data. At the same time, it is clear that high-level speech synthesis can be achieved once the data has been collected.

But speech synthesis is also improving outside more specific domains. Recent developments show that, regardless of the application, it is possible to generate a

5 "Turing" refers to Alan Turing, a British mathematician, computer pioneer and computer scientist.

6 Other Google systems are also achieving high-quality speech synthesis. The quality of voice synthesis is measured using the Mean Opinion Score, which is based on the assessments of a group of people (Streijl, Winkler & Hands, 2014). They then give a score between 1 and 5. For example, WaveNet, a Google speech system, scores a 4.21 for US English and 4.08 for Mandarin Chinese (Oord, 2016).

voice even with relatively small datasets (Arik et al., 2018). For example, the company Resemble AI needs a recording of at least three minutes to make a voice synthesis (Schwartz, 2019). In 2017 the startup Lyrebird demonstrated that they could make a voice synthesis with a sixty-second audio clip. They demonstrated this by presenting audio clips in which voice characteristics of Barack Obama, Donald Trump and Hillary Clinton could be heard (Vincent, 2017). That said, these clips were still clearly recognisable as computer-generated speech.

Speech applications can be personalised using "voice cloning". Instead of choosing a pre-programmed voice, users can set their own voice, or that of family or friends. The Waze navigation system already provides this service (Welch, 2017). To make their voice assistants even more attractive, major technology companies are now extending their voice offering to include celebrities such as Samuel L. Jackson (Amazon's Alexa) and John Legend (Google Assistant) (Faulkner, 2019).

Ultimately, the goal of developers is to create as diverse a range of voices as possible, allowing users as well as companies and organisations to choose the voice that best suits their identity or brand. This is known as "voice branding".

In summary, almost all speech systems produce speech that is easy to understand and some systems can also formulate complex sentences. Speech capacity is good enough for more and more practical applications, from navigation systems to reservation services. Work is also being done on cloning voices. Technological development is now focusing on making voices more human, an aspect where considerable progress can still be made.

2.4 The future of speech technology

This chapter has provided an overview of the current possibilities of speech technology. But what can we expect in future? Here is a list of the developers' ambitions.

When it comes to **speech recognition**, developers want to unlock more languages. Currently, speech recognition works particularly well for frequently spoken languages such as US English or Mandarin Chinese. At the moment, for example, there is only one voice assistant with an adequate command of Dutch: Google Assistant. This is mainly down to the challenge of collecting enough data from smaller language areas, and the willingness of developers to make products for a smaller market.

Developers are also keen to drive down the error rate. The only way to do so is to improve the digital models to ensure that the software also understands ambiguous, rare or vague words and expressions. In addition, it is important to reduce noise and develop speech technology that also works well in more difficult environments such as a busy railway station or even a dance party.

The future of **speech interpretation** seems to lie primarily in combining various meaningful elements, such as the words a person uses, the way a person speaks the words and the body language, facial expression and emotion with which the words are spoken (Jyothi, 2018; Henzi & Wright, 2019; Mattin, 2019) (see also Chapter 3). It is also intended that speech systems will understand the context of a question or remark much better and keep the purpose of a conversation clearly in mind (Hirschberg & Manning, 2015; Pundak et al., 2018; Zen, 2018).

When **synthesising speech**, developers want to focus even more on the user in future – the user should feel that he or she is being talked to in a personal and friendly way and that the information he or she receives, exactly matches his or her needs (Pichai, 2019). This will undoubtedly require personal data. The question is actually whether we *want* to talk to computers that are indistinguishable from humans, and whether we want to transfer all the necessary personal information – we will discuss this further in Chapter 4.

All in all, developers want to create speech technology that can be used in every conceivable setting, accurately provides numerous services and answers many questions, and works as the user's empathetic, personal assistant. This will require more data, more intimate personal data and smarter digital models. In addition, it will be necessary to keep the required computing power to a manageable level or even to reduce it, so as to make it easier to install speech technology in many devices (Pichai, 2019; Li et al., 2020).

2.5 Conclusion

In this chapter we investigated what speech technology is and how well it works. Impressive progress has been made over the past decade in every key element of speech technology – speech recognition, interpretation and synthesis. It is now possible to ask digital device questions and get an immediate and understandable answer. This progress is down to the enormous increase in data available for training the software, improved digital models and increased computing power. In particular, the use of deep learning has brought about a rapid improvement in speech technology.

At the same time, it is clear that the three elements of speech technology also have their limitations. Speech technology usually needs help from the environment and the user to enable it to work properly. In the area of speech recognition, the extent to which systems make mistakes is decreasing, but the error rates are still significant and even high in difficult situations. The technology is not yet good enough for critical applications. Developers are therefore not only improving their systems, but also want the digital environment to be designed "voice first", with recognisable keywords and short pieces of text that are easy for speech technology to find.

Speech interpretation also often needs the user's help to clarify the intention of given commands, especially if they are complicated. Although it can complete a simple task successfully, more complex conversations remain difficult. To improve interpretation, developers want to combine different types of data, including the spoken words and the emotion used to express them. Technology and service providers have high ambitions for voice analysis, but as yet scientists are critical of the reliability of this type of application.

Finally, the main challenge for speech synthesis is to further personalise the voice and make it even more human-sounding. Ultimately, developers want to have systems whose speech is indistinguishable from human speech, and which can synthesise a rich range of voices and personalities.

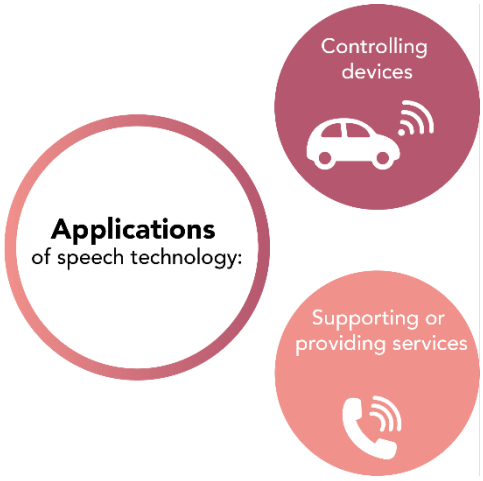
3 What is speech technology used for?

This chapter provides an overview of the current use of speech technology in society. It discusses **the applications** that speech technology can perform (3.1). We then look at **the analyses** that can be based on speech data (3.2) and sketch out **the market** for speech technology (3.3). In particular, the chapter makes use of the work of research agencies and scientists mapping the market for speech technology.

3.1 The applications that speech technology can perform

Speech technology can perform a variety of applications. We provide an overview of the domains into which speech technology is being introduced, based on two categories: speech technology that **controls devices** and speech technology that **supports** or **provides services** (see Figure 2). Speech technology already plays a much greater role in some domains than in others. For example, the market for speech technology in the home and in the car already has many users, while in other domains speech technology is still mainly at the experimental stage, and it is unclear which types of applications will ultimately predominate. We first describe the applications used in the home and in the car.

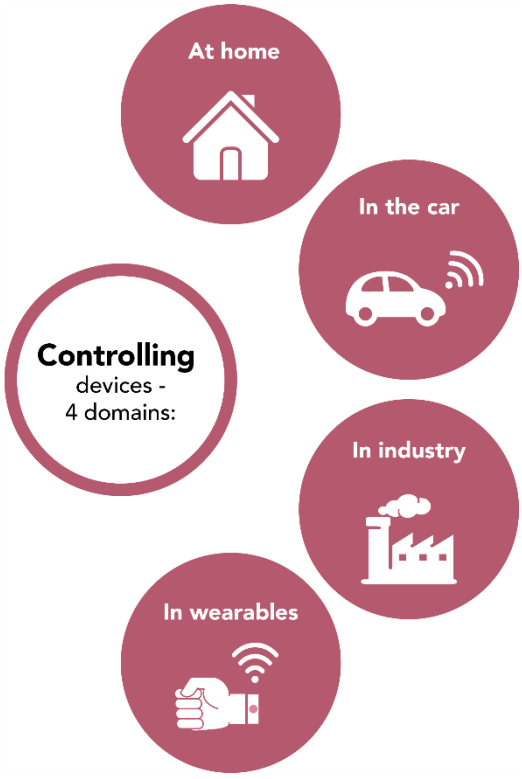
Figure 4 Applications for speech technology



3.1.1 Controlling devices

In recent years, many devices have been connected to the Internet, and an Internet of Things has emerged (Vermeend & Timmer, 2016). This network can increasingly be controlled by speech technology. We discuss four domains in which speech technology is involved in controlling devices: at **home**, in the **car**, in **industry** and in **wearables** (see Figure 3).

Figure 5 Control by speech technology



At home
Many of our homes are already "smart homes", in which household electronics are connected to each other. There's a "smart" version of almost every electronic device available that can be connected to the internet: TVs, lights, thermostats, video doorbells, cameras, domestic appliances and of course smart speakers. In the Netherlands, 3.3 million households have one or more smart home products, most of which fall into the categories of "lighting and switches" and "energy and heating". The proportion of households that have a smart speaker in the home is growing fastest: from 6% in 2018 to 19% in 2019 (Multiscope, 2020).

These devices can be linked to applications on a smartphone or tablet and the applications can then be controlled by a voice assistant. This means that, while sitting at your dining table, you can use your voice to dim the table lamp, turn up the heating or play a certain piece of music. You can even set up a certain routine and simply say "Hey Google, dinner time". The assistant can then combine different actions, simultaneously dimming the light and playing atmospheric music.

The use of speech in the smart home is not just a matter of comfort. Scientists, companies and (care) institutions have been looking for decades at how smart devices can enable elderly or disabled people to live more independently for longer (e.g. under the heading of home automation or Ambient Assisted Living). Speech also offers possibilities in this regard. ANBO, a Dutch elders' organisation, Achmea, an insurance company, SVB, the Dutch social insurance provider, and Google have therefore set up a trial in which elderly people were given a Google Home device to use in their homes (Lens-FitzGerald et al., 2020).

Finally, smart toys often contain speech technology. This has created a billion-euro market for automated toys, which can perform all kinds of tasks, from playing music to conducting conversations. For example, there's Golden Pup, a robotic dog, which barks in a specific way in response to a child's speech, or Hello Barbie, a Barbie doll that can talk back and play all kinds of games and even teaches children to say sorry (Vlahos, 2015). In addition to physical toys, there are also speech-based games, such as Wild Island, which you can play on your phone or on a smart speaker. They tell an interactive story in which the listener has to make choices.

In the car

Another environment where speech technology has been used for a long time is the car, especially to enable hands-free calling. More functions have recently been introduced to ensure that drivers can keep their hands on the wheel and focus their attention on the road (Muller, 2019). Other functions include navigation, texting and playing audio from a radio station or podcast, which are used particularly frequently in the United States.[7] As yet, less is known about this in the Netherlands. Alexa is not yet available in Dutch and Android Auto (Google Assistant) is not yet officially available either (Kamp, 2019).

Several car brands, including Mercedes and BMW, are developing their own assistant and therefore also a voice for use in the car. They are looking for the voice that can best support the brand. A young, female and intelligent voice may be chosen to try to appeal to a certain audience. In Chapter 4, we discuss various ethical aspects to be considered in the selection of voice profiles. Marketing choices

7 In the United States, there are 77 million monthly active users, more than the use of smart speakers in the home, which stands at 45.7 million monthly active users (Kinsella & Mutchler, 2019).

regarding voice profiles fall within the new domain of voice branding (Vernuccio et al., 2020).

Speech can also be used to control parts of the car itself, such as turning on the heating or opening the windows, either via the car or via smartphones (Carmen, 2018). Several patents from companies show that work is being done on controlling self-driving cars through speech technology(Szymkowski, 2020). They are also looking at using speech technology to control tractors or excavators. It seems that the technology is not yet suitable for this purpose because the error rate is too high (Kartalidis, 2018).

In industry
Speech technology provides opportunities to improve production processes. A well-known application in the logistics sector is Pick by Voice, where distribution centre employees are guided by speech. Messages received by ear tell the user the location of the item to be picked, the quantity to be picked and the most efficient way of navigating a distribution centre (Glynn, 2020).

New applications are voice assistants in the production hall and manufacturing industry. For example, IT Speex built the voice assistant Athena, specially developed to control machines such as drills and lathes on the factory floor (Leventon, 2019). The assistant can be used in tandem with a screen. The aim is for the voice assistant not only to provide support for the direct control of devices, but also for maintenance checks, the planning of production processes and the calculation of costs – it must be an assistant that is of use in virtually all brainwork and control operations. As in cars and excavators, maintaining a low error rate is essential. For example, the speech recognition scores of current voice assistants, which reach 5% in optimal conditions, are too low for industrial applications.

Speech developers are therefore looking for ways of reducing the error rate. Moreover, an assistant like Athena works subject to receiving explicit permission: if the assistant intends to carry out a task, it is announced first. The user then gives permission. In addition, in case of ambiguity, the assistant puts the ball back to the user. Where words are unclear, there is no guessing at what they might mean. Finally, a voice assistant can also increase security by working through checklists and asking the user questions when processes are risky.

In wearables
Speech technology is increasingly portable technology, installed on **wearables** such as smart watches, earphones and glasses (Williams, 2019). Watches and earpieces are available to consumers, while smart glasses focus on sectors such as construction, healthcare and logistics. Smart glasses can display information in

the field of vision and play sound through built-in speakers or sometimes even through bone conduction.[8] An example of this is the Google Glass which can be controlled using speech and a touchpad. Speech is used, for example, to navigate through the menu, start a video recording, and call friends or colleagues. Smart glasses are mainly used in the professional sector to improve work processes. The upcoming report of the Rathenau Instituut on augmented reality shows how speech technology is being used in construction, logistics and neurosurgery (see Snijders et al., 2020). Virtual reality headsets are not classified as wearables in this instance, because they isolate you from the physical environment. However, developers of these headsets are experimenting with speech control (Zhu et al., 2015).

3.1.2 Supporting or providing services

Speech technology can support, and sometimes even provide, services in many different ways. For example, like a human assistant, a voice assistant can help people keep track of their diary, schedule appointments, keep a to-do list and send messages. Microsoft's assistant, Cortana, focuses specifically on these kinds of general work processes. The assistant can speed up the work in basically all domains, therefore saving time. Speech technology is at the experimental stage in most domains and is not yet widely used.

Numerous speech systems are also being produced for more specific services. We provide a wide but not exhaustive overview below, and briefly discuss applications in the fields of **travel** and **tourism**, **shopping**, **healthcare**, **education**, **call centres**, **media broadcasts**, **language applications**, **identification** and **verification**, **online searches** and **advertisements** (see Figure 4).

8 Bone conduction works by transmitting sound by means of vibrations through the bone of the skull.

Figure 6 Services provided by speech technology

Companies are developing voice assistants in the field of **travel** and in the **tourism sector**, for example for booking restaurants, such as Google Duplex. Alexa for Hospitality, a version of Alexa that Amazon has developed specifically for hotels, helps users to order room service or extra towels, control the air conditioning and check out (Amazon, 2020). Likewise, voice assistants can already perform tasks in the aviation industry. For example, the "smart assistant" used by Dutch airline KLM can help you find a destination, search for a flight and draw up a packing list, and tell you what time to leave for the airport. Similar services are available in the banking sector. For example, you can ask Rabobank's smart assistant for your balance, request a payment or order a statement of your latest transactions.

Shopping is another emerging domain for speech technology. Its uses consist, in part, of direct online ordering of products and services. In the United States, for example, you can already ask the voice assistant Alexa to stock up on groceries: "Alexa, order more dog food." This is still in its infancy in the Netherlands – especially since the largest Dutch-speaking assistant, Google Assistant, is not yet

able to place orders. Google is, however, working on a payment function for Google Assistant (Hager, 2020).

Part of speech technology consists of supporting orders, for example by recommending choices and keeping shopping lists up to date. This is already more common in the Netherlands. Dutch supermarket chain Albert Heijn has created a speech application you can ask what time the delivery person will arrive, which products are on offer or what recipes are available for an oven dish with vegetables. Albert Heijn wants to make shopping and cooking easier (Van Bergen & Van Gelder, 2018). Dutch online stores Bol.com and Jumbo have also started working with speech technology that does not place orders but does help you do so.

Amazon regards speech as ushering in "a new era in retail" (Meridith, 2019). Critics point out, however, that only 2% of Alexa users have ever bought products using speech technology (Pardes, 2019). Only 10% of these users did so again afterwards (Anand, 2018). The main reason for this is that shopping is, in many cases, a visual activity. You want to see what you are buying and compare products by sight. Screens are more suitable for this than smart speakers – so perhaps the combination of speech and screen is the future of commercial transactions.

Another emerging field of application is **healthcare**. Speech technology is being developed to use voice assistants as a first point of contact: you can make an appointment via the assistant, request prescriptions and the assistant will attempt to perform a basic triage by asking questions, and then refer you to the right care provider. Suppliers of the Dutch Electronic Patient Record software are also collaborating with US companies to use speech technology to update patient records, write referral letters and prepare notes (Jacobs, 2018). These innovations are regarded as an opportunity to make healthcare more efficient and to reduce growing healthcare costs. In addition, the developers are trying to make the voice assistant as engaging and empathetic as possible (Kritzler et al., 2019). The voice assistant can spend time on patients and patiently guide them through the complex and busy healthcare domain.

As well as facilitating administrative processes, voice assistants are also to be used to support the work of doctors. The Dutch publishing company Wolters Kluwer (2020), for example, is working with speech recognition company Nuance to enable doctors to consult clinical databases by voice. Alexa offers an automated, medical transcription service for medical professionals. Scientists and industry are also focusing on diagnosing diseases on the basis of voice data (see chapter 3.3).

This also applies to **education**. For example, a consortium of research organisations and commercial players has developed iTalk2Learn, an open source

learning platform that teaches pupils mathematics (iTalk2Learn 2015). The software asks questions and recognises pupils' speech, and analyses their attitudes – is a particular assignment appreciated or not? Duolingo, a widely used application for learning foreign languages, also uses speech technology, particularly speech recognition (Sawers, 2019). It is developing bots that users can converse with.

Speech technology is also used in **call centres**. For example, Cogito uses speech recognition to support its agents. During calls, it detects whether the caller is enthusiastic or frustrated and then gives the agent real-time advice. An agent can also be notified to take a break or have a cup of coffee when his or her voice sounds tired (Dzieza, 2020). Also available are voice assistants such as Amelia, from developer IPsoft, that take calls themselves to answer callers' questions (Baraniuk, 2018). In addition, the aforementioned assistants as used by companies such as Albert Heijn and KLM also help to answer customers' general enquiries.

Speech technology is also used in **media broadcasts**. This is done in particular by asking voice assistants to play certain programmes, such as the news: they then play a recorded news bulletin prepared by a specific journalistic source, such as a newspaper or a daily news programme. The user can set the source. The Dutch public broadcaster (NPO) has collaborated extensively with Google to investigate which parts of its programming can best be offered via voice assistants and smart speakers (NPO, 2020). In China, they're already going a step further: Xinhua News Agency, a Chinese state press agency, has built a male and a female AI newsreader, based on human newsreaders. They have already been used during major events in China (Meiling, 2019).

Speech technology can also take on all kinds of **language applications**, for example translating speech. During a conversation between two people who do not speak each other's language, speech technology such as Google Translate and Skype translator can immediately recognise and translate sentences from one person (Kohn, 2019). Speech technology can also take minutes of meetings and transcribe recordings. It is also possible to use speech technology to automatically subtitle videos or live presentations. All these applications could have great potential: imagine a world where everyone with a smartphone has access to translation technology and can therefore communicate relatively easily with people who speak a completely different language. However, these applications require speech technology to master many more languages than is currently the case.

Speech technology is also seen as a promising development in the field of security as it can be used to **identify** and **verify** people. In identification you compare someone's voice with a wide dataset, thereby ascertaining their identity. This requires a comprehensive dataset. Verification works in a different way. In this

case, someone claims a certain identity and the software establishes – based on an audio recording – that a voice is authentic and belongs to the stored identity. This only requires one pre-saved recording per person. This choice is important for the impact of speech authentication on privacy and data protection (see Chapter 4).

Banks and telecom companies are exploring the possibilities of speech verification (Bharadwaj, 2019), for example to identify customers calling customer service or to make payments or request a balance. Various government agencies have also started to work on speech verification. New Zealand's tax authorities, for example, use this technology. Individuals and businesses could retrieve information and change data after their speech had been verified (Sankaranarayanan, 2017).

Speech verification can provide an additional security factor to combat online fraud or make buildings secure. In some cases, it can be used as a signature. In the United States, for example, audio recordings of the words "I agree" can be used to sign a digital contract.

A voice ID can also be used to track criminals. In which case, it is a matter of identification. For example, an audio clip can be recorded during a robbery, and the police can identify the perpetrators on the basis of a broad dataset of voice IDs. Interpol announced in 2018 that it had built an international speech identification database, with voice IDs from 192 different law enforcement agencies around the world (Kofman, 2018).[9]

Finally, speech technology can change the way we **search online** and, with it, the way we're exposed to **online advertising**. Speech technology allows us to search using a voice assistant, which usually does not give us a list of search results (like an online search engine), but tells us an answer – or asks a follow-up question. This search method has major consequences – first and foremost for players who want to be found on the internet. They need to make their websites "voice first", so that the voice assistant can find the information easily. In the Netherlands, various government organisations, such as the Dienst Uitvoering Onderwijs (DUO), the Kamer van Koophandel (KvK), the Uitvoeringsinstituut Werknemersverzekeringen (UWV) and the Sociale Verzekeringsbank (SVB), are already working on this.

In addition, the online advertising market will change as a result of voice search (Olson & Kemery, 2019). You will no longer be able to place an ad above an online search page or try to appear somewhere in a list of search results. Although new forms of advertising are likely to be experimented with, there is as yet less advertising space in the environment of voice assistants. Companies and

9 https://www.interpol.int/en/Who-we-are/Legal-framework/Information-communications-and-technology-ICT-law-projects/Speaker-Identification-Integrated-Project-SIIP.

organisations will therefore have to ensure that voice assistants can easily find their websites and that their products and services are so popular with users that the voice assistants will recommend them.

3.2 Data analysis by companies

Speech applications collect and process data while performing all the tasks and roles discussed. In this chapter, we review the ways in which companies analyse these new sources of data. Here, as in Chapter 2, we make a distinction between the analysis of **text** and the analysis of **voice**.

3.2.1 Text analysis

The call logs of speech applications show what has been said to a speech system, including a particular user's choice of words and sentence structure.
Major technology companies such as Google, Apple and Amazon are analysing the call logs to personalise their voice assistants Google Assistant, Siri and Alexa, so that the assistant can, for example, adapt to users' speech patterns, vocabulary and personal preferences, and therefore recommend certain services or tasks (also referred to as "skills" in the jargon of the industry (Dawar, 2018). In so doing, the tech companies are using emotion recognition to ascertain the user's emotional state of mind. After all, if Alexa knows what the user does and does not like, this helps enormously with the personalisation of the voice assistant and the user is more likely to keep talking to it. Emotion recognition is a key technology in the development of empathetic, intelligent voice assistants.

3.2.2 Voice analysis

The audio recordings are also used to allow the assistant to adapt to the user. This analysis could focus on recognising volume. For example, Alexa can whisper back if you ask it something quietly in the morning and do not want to wake up your partner (Gershgorn, 2019).
A number of patent applications and developments show that companies are also exploring other analyses based on audio, particularly emotion recognition (Fussel, 2018). Amazon, for example, has developed an application for recognising frustration so that they can engineer Alexa to generate as little frustration as possible (Gershgorn, 2019). Cogito uses voice analysis to monitor customer service

calls and gives employees suggestions such as "talk more calmly" or "apologise" (Simonite, 2018).

Companies are also interested in health analyses – for the recognition of diseases (Kinsella, 2018). Examples include the voice assistant Alexa, who suggests chicken soup when she hears the user coughing. The start-up Vocalis Health also uses voice analysis to conduct health-related research. Its CEO has described the voice as "the sensor of the body" (Wenderow, 2020). Vocalis Health has developed an application that lets people analyse their voices and monitor the progression of certain diseases (Staff, 2020b). During the COVID-19 pandemic, the company asked people to send in speech recordings, with the aim of identifying carriers of the virus (Staff, 2020a).

In collaboration with a major US insurance company, speech company Canary Speech analysed a file containing millions of phone calls collected over fifteen years. These calls were analysed with the aim of detecting Parkinson's disease and Alzheimer's disease. This data could be used to adjust insurance premiums. When asked about this, the insurance company's CEO indicated that such an application would have to be subject to regulation (Reynold, 2017).

Voicesense says that it can create a complete personality profile based on speech data (Chen, 2019). The company offers to assess existing health issues, including the likelihood that people will make an insurance claim and the likelihood that a patient can be discharged from hospital. Voicesense also analyses the probability of customers defaulting and profiles employees at risk of burn-out. Finally, Voicesense also offers to analyse the likelihood of a customer remaining loyal to a certain product.

In Chapter 2.2.3, we stated that voice analysis is still in its infancy – the promises made by the technology companies should therefore be taken with a pinch of salt. But their ambitions are great.

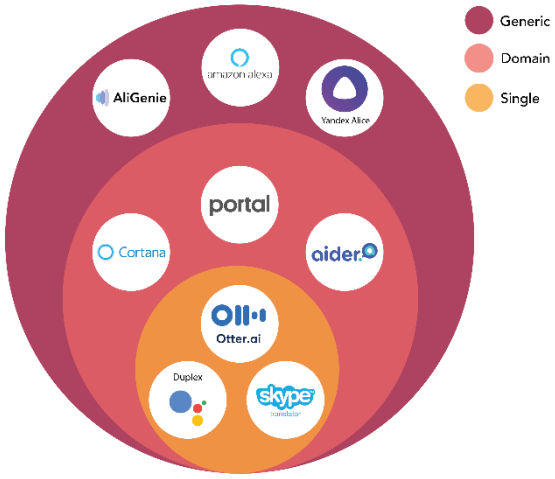
3.3 The market for speech technology

In the previous chapters we have seen increasing activity in the field of speech technology. This chapter looks at the bigger picture and gives an impression of the market for speech technology. What are the ambitions of the tech companies? Is most of the technology in the hands of a few major players or are there alternatives? When it comes to speech technology, we often think of the voice assistants of the major technology companies, Amazon's Alexa or Google

Assistant. But there are more providers. They differ in the way they **bundle tasks** and **develop applications**.

Bundling tasks
In the field of task bundling, we distinguish three types of speech applications: generic, domain-specific and single (see Figure 5).

Figure 7 Bundling tasks



Nowadays, almost every technology giant is offering a voice assistant. This includes Amazon Alexa, Google Assistant and AliGenie, the assistant produced by the Chinese company Alibaba. These assistants combine various tasks: they can play music, turn up the heating, etc. They are **generic speech applications.** These voice assistants are intended to serve us wherever we are and guide us through the digital world. Some smaller companies are also marketing generic voice assistants, such as Alice from the Russian company Yandex.

There are also **domain-specific** assistants. They focus on performing as many tasks as possible within a certain domain. Examples include Microsoft's office assistant Cortana; Portal from Facebook, which mostly focuses on video calls; Athena for machines in a factory; and Aider.ai's speech services which are mainly aimed at smaller businesses. The advantage of this type of assistant is that it can specialise in a specific domain and therefore provide better services. When Microsoft found Cortana was lagging behind the assistants of the other tech giants, the company decided to provide this more focused service (Warren, 2019).

The large technology companies are focusing on improving and expanding their services in specific domains. In doing so, they are competing with companies

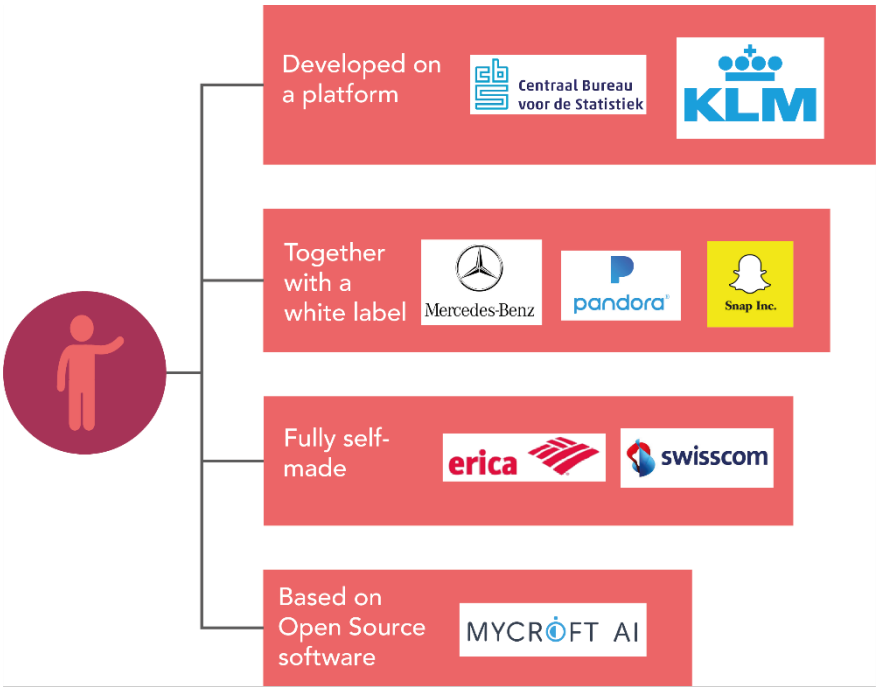
operating in these specific domains. Amazon, for example, is developing Alexa healthcare services, as a counterpart to Nuance.[10] In April 2020, Google teamed up with medical voice assistant Suki from the company of the same name.[11]

Some speech applications offer a **single** application, i.e. they can only perform one task. These include Google Duplex, which makes restaurant reservations. Because it focuses on this one task, Duplex is very good at it. The software's speech sounds very human and you can conveniently indicate your preferences during a chat with the software. But Duplex does not perform any tasks other than making reservations. In addition to Google Duplex, there is also Skype Translator and Otter.ai, which, for example, automatically generates transcriptions of speeches.

Product development and supply

Another difference is the way speech applications are developed and supplied. Providers can create their own speech applications, develop them on a platform or base them on white label or open source software (see Figure 6).

Figure 8 Product development and supply



10 https://voicebot.ai/2020/06/11/nuance-connects-wolters-kluwer-clinical-database-to-dragon-medical-virtual-assistant/

11 https://voicebot.ai/2020/04/13/how-clinical-voice-assistant-startup-suki-lowers-insurance-claim-rejection-by-19/

A number of providers are developing their own speech technology and supplying it themselves. Examples include Erica, Bank of America's voice assistant and Swisscom, the assistant produced by the telecommunication company of the same name. These companies developed the software themselves and supply it direct to customers.

However, many companies and organisations use the platforms of major technology companies such as Amazon and Google. Companies and organisations can supply "skills" and "actions", applications that allow users to interact with these companies and organisations, via the platform. The "information dialogue" launched by Centraal Bureau voor de Statistiek (CBS) is an example of this, as is KLM's speech application. In this case, the companies and organisations use a licence for the speech software of major technology companies.

It is not just the tech giants that offer this "platform variant". Car manufacturers Honda, Mercedes-Benz and Peugeot, as well as other companies such as Deutsche Telekom, Motorola, Snapchat and Pandora, for example, are using Soundhound's Houndify platform for their speech software.[12] This platform is a "white label" product, which means that the software is made by Soundhound but sold by other companies with the addition of their own branding and specifications. This gives companies more control over their data and the ability to position their brand more centrally. On the other hand, they do not benefit from the rich functionality provided by a major voice assistant. Houndify now offers more than fourteen languages, including English, Chinese, Spanish, German and Japanese. Small businesses that want to use Houndify pay with "credit points", where each enquiry received via Houndify costs a number of credit points. A business can also choose to have part of the functionality executed by Houndify, and part by another assistant.

Finally, there is also an open source alternative from Mycroft, the source code of which has been made publicly available and high privacy standards are observed.[13] Mycroft offers a generic voice assistant that can be freely reused and adapted and can be installed on Mycroft devices as well as on other hardware. At the moment, Mycroft is many times smaller than the other speech technology providers and depends on donations and crowdfunding. It cannot keep up with the investments of a company like Amazon. Moreover, it has less training data at its disposal and, because of tough privacy agreements, it also collects less data (Newman, 2018). The question is therefore whether Mycroft can compete with the large platform companies in terms of services, hardware and languages supported. At any rate, it is not a fully-fledged alternative at the present time.

12 https://www.houndify.com/
13 www.Mycroft.ai

Speech technology is our guide to the digital world

One market trend is clear: the technology giants are competing to see who can supply the biggest speech platform. This fierce competition is easy to understand: the more users and applications the platform has, the more attractive it is – for users, providers and advertisers. This is called the network effect. The platform owner can use all the new data to continue improving the voice assistant and make the platform even more functional.

The ambition of companies like Google and Amazon is to create such a widely accessible platform that users will want to use it at home, in the car, at work, and wherever else they may be. The speech platform is a seamless environment, a guide that opens up the digital world, which you as a user always want to take with you. Commentators describe the battle for this large, all-encompassing speech platform as the "voice war" (Yoffie et al., 2018; Kinsella, 2019a). The US technology giants, including Google and Apple, seem for the time being to be ahead of their Chinese counterparts, such as Baidu, in this war. For the time being, Chinese companies are focusing on the Chinese market, which is mainly down to the Chinese language (Harris, 2015).[14]

The technology giants are investing heavily in research and development (R&D) in order to build a widely accessible platform. Amazon, for example, has set up an Alexa fund which is investing USD 200 million in small businesses and has instituted an Alexa award aimed at sponsoring university research (Amazon, 2020). And Google has set up Google Assistants Investments, which mainly enters into partnerships in the healthcare sector and in tourism and hotels (CBinsights, 2019). We have already noted that the large technology companies are also becoming increasingly active in more domains, competing with companies that make domain-specific and single speech applications.

The question is what the competition between the big companies will mean for the market in the longer term and for the broad group of suppliers currently operating in it. Will smaller suppliers be bought up by the large technology companies or will alternatives continue to be available? In Chapter 4 we look in more detail at the conditions needed to create a fair, open and accessible market.

14 This market trend is well documented in the figures. Over 2.5 billion devices worldwide have a built-in voice assistant. Most of this consists of the mobile phones of Google and Apple, with their respective voice assistants, Google Assistant and Siri, pre-installed. For example, the Google Assistant is now installed on more than 1 billion devices and 500 million people use it every month. Apple Siri can be found on more than 500 million devices and 375 million people use this assistant every month. But other voice assistants can also be found on hundreds of millions of devices, such as Microsoft Cortana, Baidu DuerOS, Amazon Alexa and Samsung Bixby (Canalysis, 2020).

3.4 Conclusion

In this chapter we looked at the domains in which speech technology is being used. We saw that speech software already has a wide range of applications. Speech technology has developed furthest in the car and in the home. But companies, governments and organisations are also exploring possibilities in other domains, such as telecom, banking, shopping, aviation, the automotive industry, manufacturing, the office environment, healthcare and education. However, this technology is not yet widely used in these areas. The applications do, however, show a clear direction in how the technology has been used to date: speech technology is being sold to us as a guide, which opens up numerous services, talks to us, answers our questions and can even identify us.

The widespread use of speech technology requires the increasing collection and analysis of speech data: the voice is a new data source, enabling new services and applications. Call logs, and often audio recordings of human voices, form the basis of analyses that enable speech applications (such as disease detection) or improve them (such as personalisation based on a person's preferences). In some cases, speech analyses are sold for the purpose of analysing someone's health or assessing a person's work profile. Both major technology companies and small start-ups are keenly searching for the vocal markers for diseases and emotions. In the next chapter we will look in more detail at the societal and ethical questions that this raises.

The speech technology market is dominated by the major technology companies, which are investing huge sums and are competing to become the biggest speech platform. The already high market share of these companies is further enhanced by the network effect: the platform becomes more attractive to suppliers and users the more applications and users are added. The aim is to use speech technology to create not only a convenient gateway to the digital world but also a guide that stays with the user at all times. This can be while shopping or when someone is looking something up or listening to music. In this way, a user does not have to leave the platform in question.

Yet there are other platform environments, created by white label and open source providers, that enable companies to build and supply their own applications. The question is what the increasing competition between the major companies will mean for the market in the longer term: will alternatives continue to be available and, if so, how? We'll discuss this in the next chapter.

4 Societal and ethical aspects of speech technology

This chapter provides an overview of the societal and ethical aspects of speech technology. It explores what significance speech technology has, and can have, in society. In addition, it reviews the issues that speech technology raises. First, however, we discuss why our speech is so important to us.

This chapter is based on the work of ethicists, media researchers and journalists who have analysed the societal and ethical aspects of speech technology. We also make use of various studies by the Rathenau Instituut, in particular "Urgent Upgrade*" (2017)* and "Human Rights in the Robot Age" (2017*),* which drew attention to the societal and ethical aspects of digital technology, including robots and AI.

4.1 The importance of speech

The ability to speak is essential to people's lives in many ways. Our speech is an **expression of our personality**: you can speak patiently, or hastily. You can say something calmly or passionately. Speech is one of our most important forms of expression (Cox, 2019). For example, an accent says something about your background, as does the vocabulary you use and the way you form a sentence. Your speech, along with your appearance, is your business card.

Media researcher Sherry Turkle also points out the importance of conversation between people: it supports empathy, friendship, love, learning and productivity (Turkle, 2017). In the conversations we have with each other **we develop ourselves**. We exchange knowledge and discuss new ideas. A good conversation can change your mind or strengthen your conviction. In our conversations we describe new places where the other has not been yet, and pass on our experiences. A conversation can take us to a wider world. And of course we also learn about each other. We exchange our personalities and form an image of the other in our conversations.

Our aim is not just to acquire knowledge as our conversations are also the dominant way in which we **relate to each other**. In our conversations we get closer to each other or argue with each other, imitate each other or respect each other's

differences. So there's a lot at stake in our conversations. In a conversation you can make or lose a friend, stand up for your interests or harm them.

Finally, our conversations are therefore also a dominant context in which **to develop and apply social norms and values** (Habermas, 1997). The conversation is the cornerstone of a democracy, in which we jointly shape society. These norms are numerous and ubiquitous. When you are speaking or just listening, you have to behave yourself. You cannot just talk over each other, you are supposed to listen to each other with attention, and you do not have to tolerate speakers who express themselves in an unnecessarily rude or hurtful way. We have social norms about when to talk and when to remain silent, how to give compliments and how to criticise. All these norms are constantly evolving – consider the use of less formal language, for example. And they adapt to different contexts – from a pleasant conversation between friends to a debate between politicians.

Clearly, our conversations are important for our social interactions and personal development. People can develop and express themselves through their speech. **Almost nothing is more intimate, and more personal, than our voice**. The words and sentences we speak are the building blocks of our personality. Moreover, we use speech to shape complex and essential social processes. Our speech is a precious commodity.

Precisely because our speech is so important, speech technology can have great societal significance. After all, speech technology interferes with our conversations, and is increasingly becoming part of our social world. We talk to speech technology and speech technology listens to us. This raises questions, not only when the technology is working properly, but also when speech technology makes mistakes.

Companies rely on speech as an interface between people and the digital world. But what should this interaction look like? It is important to manage the development and implementation of speech technology effectively. We identify five key ethical aspects in this respect (see Figure 7). It is important to ensure that speech technology:

- strengthens **social relationships and norms** rather than blurring them,
- respects **privacy** rather than violating it,
- supports **autonomy** rather than harming it,
- can be used **safely** and **healthily** and
- leads to balanced **power relationships** between businesses and between businesses and individuals.

We will discuss the societal and ethical aspects one by one.

Figure 9 Ethical aspects of speech technology

4.2 Social relationships and norms

Speech technology can increasingly mimic human speech and participate in human conversations. This means that the difference between computers and people becomes blurred, changing our relationship with computers. This is not a neutral matter: speech technology can both damage and improve our social relationships and norms. It is therefore important to manage change and develop appropriate social norms. Five issues must be considered: confirming or fighting **bias**, **treating computers as human beings**, **educating and disciplining** users, **accepting or excluding** users and **taking over human work and contact**.

Bias

The Rathenau Instituut has already noted in its *'Urgent Upgrade'* report that biases, such as the idea that real women raise children or the idea that real men are heterosexual, are a persistent problem in the development of AI technology (Kool et

al., 2017). If such views creep into the technology, unintentionally or unnoticed, this can lead to discrimination or the exclusion of certain groups (Barocas & Selbst, 2016). Bias is also a factor in speech technology (Howard & Borenstein, 2017). For example, UNESCO is critical of the fact that the names of most major voice assistants, such as Cortana and Siri, are female and they speak with a female voice by default (UNESCO, 2019). In addition, voice assistants behave in a submissive and expressive way, pandering to the bias that women should always be submissive and spontaneous.[15]

It is therefore desirable to discourage bias at the design stage; however, it is not immediately clear *how* this can best be done. Gender-neutral voices are an option that does not sound like a man or a woman (Simon, 2019). Another option is an electronic voice, which we also associate with a computer. However, users find both of these options less appealing to listen to. Many companies therefore give users various options to set a voice of their choice, but this does not mean that biases will disappear. Another option is to alternate the assistant's voices, in order to interact more consciously with different types. The right answer won't be found any time soon.

Treating computers like people

It is well known that people attribute human characteristics to non-human objects in their use of language (Nass & Brave 2005; Ruane et al., 2019). For example, we call teddy bears "lovable", although they will never behave in a lovable way – they are only lovable in our imagination. This is known as anthropomorphism, where people project human traits on to the world around them. That is what people do with cars, computers and also with voice applications.[16] In fact, developers want to develop speech technology so that the software sounds as human as possible.

This can improve the user experience, so that we can perform tasks pleasurably and effectively using speech technology. Attributing human traits may also make us less lonely or sad. But it can also be a confusing and unpleasant experience if you can no longer tell computers and people apart.[17] Speech technology brings the blurring line closer (Shank et al., 2019). It is not inconceivable that vulnerable groups such as older people and small children will no longer be able to distinguish some of today's voice assistants, such as Google Duplex, from living people and

15 Although the software has now been updated, Siri used to reply to "Siri, you're a bitch" by answering "I'd blush if I could". The assistant now answers "I don't know how to respond to that".

16 Many books and films have been written about this blurring of boundaries, including the film *Ex Machina*, in which a programmer has to resist the temptations of a lifelike robot.

17 A debate on the issue of telling humans and robots apart has been going on for some time with regard to the development and deployment of social robots in healthcare and similar settings. In various reports on robots, the Rathenau Instituut has raised the question as to how far we want to stimulate and exploit the emotional bond between humans and machines (Royakkers et al., 2012; Van Est et al., 2017).

may even become emotionally attached to them.[18] A technology such as voice cloning could also lead to confusion and the blurring of boundaries, as well as abuse (we will discuss this in Chapter 4.5).

If you mimic human behaviour in a credible way, human reactions will automatically come to the surface. Especially since speech systems don't usually have robot bodies, enabling us to recognise that a computer is talking.[19] Like human radio voices, speech systems are voices produced by speakers and earphones. This raises ethical concerns. If people really think a computer is a human being, this deception seems to go against their dignity – making a fool of people is disrespectful. This raises questions: what degree of attribution of humanity do we find acceptable or even welcome? And at what point does the confusion clearly go too far? A minimum requirement is that computers should always indicate that they are a computer.

Undesirable language use and disciplining

Speech technology can capture our voices, understand them and respond to them. This also means that speech applications can respond to swear words, rudeness or grammatically incorrect sentences. This raises a number of questions. First of all, speech technology, just like chatbots, can learn bad habits and use swear words (Ruane et al., 2019). For example, users managed to get voice assistant Siri to say "motherfucker" (Leswing, 2018). Developers want to prevent that, of course, and look for ways of training speech systems so as to make this kind of abuse impossible.

Another debate concerns disciplining users' language use by means of speech software. A voice assistant can, for example, admonish users for shouting or reward them for politeness. Amazon has already implemented software in Echo smart speakers in which Alexa says "by the way, thanks for asking so nicely" (Baig, 2019). Google has also introduced "pretty please" software. The software may help combat certain habits or educate children. But it also raises the question of whether we are then crossing an ethical boundary: do we want robots to discipline and educate us? Is that not a task for people only? And do speech systems also teach us bad habits, such as giving each other commands with a degree of frustration?

In some human interactions, it is legitimate to ask whether robots indeed should play a role – such as in an escalating marital quarrel or in a parent's attempts to calm an angry child. There's already software that gives a signal when decibels are too high, a (possible) sign of an argument (Kool et al., 2014). In summary, respect

18 And what about a baby that does not call his or her parents "daddy" or "mummy", but "Alexa", as speculated by columnist Niels van Waarlo in de Volkskrant, a Dutch newspaper (Waarlo 2019)?

19 There are speech systems that are embodied, such as Tinybot Tessa. See tinybots.nl.

for human dignity must be an important principle to be observed in the development of speech technology (see also the study "Urgent Upgrade" by the Rathenau Instituut, Kool et al., 2017).

Including or excluding users

At the moment, voice technology can serve certain groups better than others for a variety of reasons, including the language the software has mastered. Speech applications recognise and speak one language better than another, depending on how they have been trained. Because some markets are more attractive than others and there is much more data available for some languages, speech systems are generally better able to speak English and Spanish than Dutch and Polish. And within certain languages, the system masters the majority's way of speaking better than the minority's accent. After all, the software works particularly well when people and their environment cooperate: there must be clear articulation, with familiar words, in a quiet environment and in a language that the system has a good command of (see Chapter 2).

This can lead to exclusion of certain groups and put pressure on minorities to conform to the language used by the majority. This dynamic can already be seen in the global language landscape, where many languages and dialects have already disappeared (Dujardin, 2017). Other forms of exclusion are also conceivable, such as speech systems that find women harder to understand than men, find children harder to understand than adults or that are simply too complicated for some groups of users (Tatman, 2017).

Developers, of course, are not aiming for exclusion – but it is attractive for them to focus on dominant language areas, and bias (i.e. an inbuilt bias or biases) all too readily occurs when training AI systems. Moreover, it is not always easy to detect or rectify these biases (Barocas & Selbst, 2016), for example because it is hard to build good datasets of children's speech for privacy reasons. In addition to increased awareness, a targeted effort to promote inclusion is therefore needed.

Taking over human work and contact

In the previous chapter, we saw that in addition to support in the personal domain, voice applications can also take over work with and contact with people, for example in largely automated call centres. In this way, speech technology touches on the societal and political debate on the future of work, in which the preservation and creation of high-quality work is a key issue. Instead of just looking at ways of replacing people's work with technology, this requires companies to take a broader view of how people and technology can complement and enrich each other wherever possible.

The use of speech technology, for example in customer contact centres or in healthcare, also means that we will have to deal with speech robots in our daily lives. Although this will sometimes be a solution in certain instances, it can also harm us. Is proper help really being given and can technology really "take care" of us? Is *care* not a prime example of something that people give or should give each other? The Rathenau Instituut previously advocated a right to human contact, so that people can choose whether they want to be spoken to by a bot or by a human being (Van Est et al., 2017).

4.3 Respect for privacy

Speech technology can be our guide in more and more places, but in the same way can also listen in to us, talk to us and collect data about us. In the United States, where speech technology is growing fastest, there are even users who have more than ten smart speakers installed in their homes, so they can voice-control their digital applications from any room (Kinsella, 2019b).

This raises questions about the protection of our privacy. It is precisely in those settings where we have traditionally imagined ourselves to be unseen and uninhibited, such as our house and car, where technology is now listening in. Speech technology affects our private lives in several ways: through the **collection** and **use of** personal data, through the **intrusion of** speech technology into our private lives and through the impact of speech technology on our **self-expression**.

Data collection
First of all, speech technology raises the question of whether our data is being collected properly. If someone is arguing with his or her lover, that argument is private. But what if there's a smart speaker in the bedroom? Who may be listening in? And when? And what information may be retrieved?

This question largely revolves around the "wake word". The idea is that voice assistant users can activate the technology themselves, for example by saying "Hey Siri" or "Hey Google". Nobody listens in if the user does not want them to.

But this system is not perfect. Sometimes a user may accidentally activate the technology by saying something that sounds like "Hey Siri" (Hern, 2019; Cox, 2020). This can happen several times a day. In addition, the system starts recording a few seconds before you say anything, because otherwise it cannot hear the wake word, and listens for a few seconds longer than strictly necessary, so as not to miss anything (Shulevitz, 2018). No matter how hard developers try to avoid it – speech technology does sometimes listen in when it is not supposed to.

One of the legal principles in this regard is the requirement to collect as little data as possible and to destroy collected data as quickly as possible (the data minimisation principle). This reduces the risk of abuse. In the case of voice assistants, this would mean, among other things, that only the converted text is collected and not the audio sound, so that a command can be executed correctly. But developers want to make sure the audio is being properly converted into text – and that means their employees listening to a selection of collected audio files (Day et al., 2019). These checks can quickly lead to privacy violations. For example, a whistleblower who worked for Apple reported listening to audio recordings of people having sex, listening in on drug deals and hearing people's medical details (Hern, 2019).

Data use

Privacy requires not only that our information is collected correctly, but also that it is used correctly. The General Data Protection Regulation (GDPR) sets out the important principles in this respect. In addition to data minimisation, these include consent, providing information and purpose limitation: data processing must be limited to a predetermined purpose.[20] This seems clear, but all three of them raise questions in practice, for example if law enforcement and security services want to seize the data. For example, the German authorities are considering requesting data in order to assist criminal investigations and this is already possible in the United States (Reuters, 2019; Statt, 2018; Reilly, 2016). We saw in Chapter 3 that Interpol has also created a voice database. When should a democratic society allow such requests and when should it not? There's hardly any public or political debate on this at the moment.

Users are not always aware of how their data is being analysed by providers and other players. What information and profiles are being recorded? And how are they being used? At the moment, it is mainly to learn how to improve personal interaction with a system. But the voice is increasingly acting as a new data source. Our voice is very valuable. It will be tempting for developers to stretch the purposes of data processing, given the emerging trade in voice analyses (see Chapter 3). All kinds of players, from credit rating agencies and advertisers to insurers, will be interested in these analyses and profiles. This requires an active regulator to enforce as effectively as possible the principles already laid down by law.

This includes a fundamental debate about how people can gain a better understanding of profiling. Although the GDPR offers protection, it does not prevent profiling. In its report "Human Rights in the Robot Age", the Rathenau Instituut

20 GDPR, Article 5.

therefore called for a right not to be measured, analysed and influenced (Van Est et al., 2017).

Providers could also step up their efforts to keep the use of voice to a minimum. Various solutions can be envisaged. For example, voice actors could be hired to train a speech system, or another option would be to adopt the use of "federated learning", where the data is not stored centrally but remains on users' devices.

Intrusion into our private and family life
The right to respect for our private and family life is enshrined in the Dutch Constitution and international treaties, including the European Convention on Human Rights (Dutch Constitution, Section 10; ECHR, Article 8). In addition to disclosing or leaking personal data, speech technology can also intrude into our private lives, for example if, due to an misunderstood wake word, the voice assistant interrupts a sensitive conversation or quarrel – or if a family member calls in the voice assistant uninvited. Our private lives are filled with social interactions that change when a computer is involved. This can also have positive aspects. With speech technology, for example, family members do not evade a conversation by watching TV or their mobile phone – the conversation is shared.

Self-expression
An important part of privacy is having the freedom to express your own identity (Koops, 2019; Amnesty International, 2019). If people know they are being watched, "chilling effects" can occur, where people do not feel free to express themselves. This could also occur in the context of speech technology. If speech technology is omnipresent, in people's homes but also in public spaces, chilling effects can occur everywhere. Imagine not daring to speak freely in your car, or even in your living room – a dystopian image. However, it is not known exactly which chilling effects may occur due to speech, and how they relate to other digital technology, such as cameras.

4.4 Autonomy

Speech technology promises to help users perform actions and make decisions. The name "assistant" says it all: a handy aid we can use to shop, make reservations, make calls and answer questions. We are guided easily and efficiently through the digital world. At least, that is the promise. But a guide also has another side: guides do not show everything and they make decisions for us. Speech technology can therefore affect our skills and autonomy.

Skills

A recurring debate when technology is being used is how it will affect our abilities. We learn certain tasks and can lose others or perform them less well, or "deskilling", as it is known (Danaher, 2018). If we are able to control the world with speech, this can provide a solution for illiterate or visually impaired people (Arcon et al., 2017). It can also mean that if we speak more, we will write less and become less good at it. We mentioned media researcher Sherry Turkle at the beginning of this chapter. In her years of research she has found evidence that interaction via screens – instead of directly with each other – reduces our empathy (Turkle, 2017). Turkle argues that we must reclaim conversation and talk to each other more. It is still too early to know how the rise of speech technology and conducting conversations will affect our relationships. These changes are not necessarily bad. The question is what exactly is the value of the skill we are losing and how we can make sure that we are better off in the new situation (Danaher, 2018).

Directing information

The way we retrieve information by speech is different than on a screen. A screen usually shows more information: multiple search results, with various measurement indicators and underlying sources already visible. A Wikipedia page or social media message also has additional information concerning its reliability or sender. With voice technology, these possibilities are usually fewer. Developers often try to get voice assistants to give a single answer to a question, for example how many Catholics there are in the Netherlands. The assistant will then mention a figure, without specifying whether they are Catholics who also go to church, or Catholics who are registered with the church and the source from which the information comes. Although it is useful to get an answer quickly, so that the user learns quickly, the question is: exactly what kind of answer is this? It could be a simplified version of reality, in which the user ultimately learns little (Vlahos, 2019). Another question raised is how the answer comes about and who decides that. Is it the first search result? Answers are more controlled and curated via speech than in our interaction via screens.

Another design choice is to let the voice assistant conduct a dialogue. For example, Statistics Netherlands (CBS) has used Google technology to create a speech application that can conduct an "information dialogue" with the user. A more nuanced approach is possible in this dialogue, which is necessary in view of the complex dataset managed by Statistics Netherlands. When searching for medical information, it is also important to highlight specifications and subtleties in the dialogue. However, this is a more complex task and it may require more user knowledge.

Directing decisions
With the information that speech technology gives us, voice assistants can also help us take decisions, for example which route to follow or what to put on the shopping list. Ideally, the technology aids the human thought process: someone is aiming for a certain goal and the voice assistant makes it possible to achieve that goal.

But sometimes the relationship between what someone wants and what speech technology offers is more complicated and influence may be exerted (Stucke & Ezrachi, 2016). This can occur if the voice assistant already thinks it knows what we want and makes suggestions. Take the example of an assistant that helps create shopping lists and suggests a certain menu. These suggestions influence a user's thought process and final decision – and this could make him or her less autonomous. As already noted in the previous paragraph, a respectful assistant does not unconsciously or improperly influence users.

Directing user experience
Speech technology is more than just a means of achieving something else, it also delivers an experience. This is most evident in voice assistants specially developed to serve as companions. The customer pays for a user experience. The digital world now has many attractive products and services that individuals use all the time. There's nothing wrong with this either. If a digital service is enjoyable and adds fun to our lives, that is a good thing. But attraction can turn into addiction. This is certainly conceivable with speech technology. It can be *too much* fun or *too* convenient to use voice assistants – after all, they are there for a user 24/7 (Kretzschmar et al., 2019). We can interact with speech technology more than is good for our health.

This addiction is aptly portrayed in the film *Her*, about a man who falls in love with his voice assistant, and in series such as *Black Mirror*. These are, of course, science fiction scenarios, but there are already specific developments that give cause for concern. For example, the chatbot Replika focuses solely on personal contact and can hold conversations on beauty (Harris, 2019). This kind of chatbot can be addictive.[21] The introduction of the social robot Jibo has also shown that users can form an emotional bond with a talking robot, which can encourage addiction (Van Camp, 2019). Finally, voice-controlled games are also being developed and speech technology can make addictive applications even more attractive, such as virtual reality games (Conner, 2020). So it is not inconceivable that we will become as attached to our speech technology as we are to our mobile phones.

21 For a sample conversation, see: https://www.youtube.com/watch?v=eOqeSC28EhU.

Directing news
These design choices are also important for the dissemination of news and the debate surrounding disinformation. The Rathenau Instituut's recent study "Digitalisation of the News" on ways of countering disinformation, emphasises the importance of having both a pluralistic media offering and media consumption (Van Keulen et al., 2018). In the Netherlands, most people consult multiple news sources every day. Voice assistants provide a new channel for bringing news to you, and "playing" the news on a regular basis. Users can choose the source themselves, for example NOS or nu.nl. A user will then hear the news highlights, say every hour in a three-minute bulletin. Just as when searching the web on a screen, there is less space available. And so, here too, choices are made about what the user gets to hear in those three minutes. This news is not personalised at present and users also consult other news sources. Currently, the "skill" of playing the news does not have a direct impact on the dissemination of disinformation.

But there are concerns in other areas. Speech technology offers new possibilities for creating fake messages, for example through voice cloning. In June 2020, Twitter was the first major social media platform to add a speech function: users can record an audio clip and place it on their timeline. Just like questions surrounding deep fakes, the question is how platforms can check the authenticity of these messages. With regard to disseminating disinformation, many measures on social media aim to provide additional information to accompany messages. In the case of speech, there will generally be less scope for this – how will providers and policy makers deal with this limitation?

4.5 Secure and healthy use

Ever since computers came on to the scene, cybersecurity has been a focus of attention. Any computer can be hacked and corrupted – and unfortunately, there will always be people who are tempted to take advantage of this. What's more, any computer can make mistakes and cause a major or minor accident. Speech technology can also cause damage to health. We therefore discuss three elements here: protection against **malicious actors**, **making** speech applications **reliable** and **preventing hearing loss**.

Malicious actors and security
Malicious actors, such as cybercriminals and hostile intelligence services, can abuse speech technology in various ways. First of all, speech technology can be hacked, allowing someone to access certain speech data or listen in via a speech application. For example, a criminal may eavesdrop on a password and steal money or blackmail individuals with sensitive information. In addition, stolen speech

data can be used to clone someone's voice, which in turn makes other abuses possible, such as identity fraud. Keep in mind that companies rely on biometric voice IDs precisely to improve the security of their services. But if these voice IDs are cloned, the security of these new identification systems will be compromised.

Another danger lies in "adversarial examples" (Alzantot et al., 2018). This is data that can be used to fool deep learning systems, such as a picture of a cat that has been manipulated in such a way that the computer still categorises it as a dog. Research is currently being conducted to ascertain whether speech technology can also be fooled, for example by adding certain background noises. Researchers claim they can control speech technology using sounds that are imperceptible to the human ear (Smith, 2018).

These risks call for action. For example, research is being conducted into training methods that protect AI systems against cloned voices and adversarial examples (Wu et al., 2020). But it is virtually impossible to make digital devices completely secure. Vulnerabilities are regularly discovered, even in the more reliable systems, such as Apple's IOS operating system, as shown in Rathenau Instituut's publication "Cyberspace without Conflict" (Hamer et al., 2019). However, security can be improved. Some devices, such as the domestic appliances that make up a large part of the Internet of Things, are known to leave a lot to be desired in terms of cybersecurity. If speech technology is installed on more of these appliances, the security problem could quickly escalate.

Reliable applications

Unsafe situations can also arise accidentally if the technology makes mistakes or distracts us too much. This is not a problem if it is a minor incident, for example if translation software leaves out a definite article. But if speech technology is performing important functions in a car or transmitting medical information, a technical error or a moment of inattention can have serious consequences (Strayer, 2015). As noted above, speech technology is rapidly improving, especially when it comes to speech recognition, but error rates of 5% are unacceptable in critical applications – and the error rate is often much higher.

The use of speech technology in specific domains therefore calls for a further reduction in error rates and for attention to be paid to the impact of speech technology on our concentration – before it is implemented. On the other hand, companies are using speech technology to try to make processes more secure, for example by having voice assistants submit a checklist of questions during certain processes, so that a care worker, or someone operating an industrial machine, does not forget anything (Softengi, 2020).

Prevention of hearing loss
Speech applications can also lead to hearing loss. This is typically important when the speech technology is in or listened to through earphones or headphones, particularly in the event of prolonged use and at excessive volumes, earphones and headphones can cause hearing loss (RTL News, 2019). This damage is no trivial matter: ear specialists warn that young people in particular suffer hearing loss that will trouble them for life. Hearing loss is usually incurable. It is not inconceivable that the rise of speech technology will further exacerbate this growing problem. After all, the philosophy of speech developers, and voice assistants in particular, is that the technology is in constant contact with the user and accompanies them everywhere. This could make it attractive for users to have their earphones in a lot of the time, thereby increasing the risk of hearing loss.

4.6 Power of technology companies

In Chapter 3 we saw how the speech technology market is dominated by the big technology companies. Almost every technology giant has developed its own assistant, which can be used on an increasing number of devices. The tech companies are mainly betting on developing assistants which, as multifunctional guides, are able to perform many tasks well, even in specific domains such as healthcare. They are investing heavily in research and in buying up start-ups (CBinsights, 2019). There is therefore a real chance that the speech technology market will fall into the hands of just a few companies. In this way, the technology companies are strengthening their already powerful position, which is increasingly being called into question by society and politicians.

The companies that sell speech technology have different ways of exercising power over the individuals who buy their products and services. For example, companies try to entice their users to buy all kinds of services within their commercial ecosystem, as evidenced by the Alexa voice assistant being linked to Amazon's online store. The tech companies aim to have as many interactions as possible (search, shop, travel, pay) taking place within their own domain. Consumers could, of course, still decide to sign up with another company. But by linking services and products, companies definitely exert influence on individuals, creating a lock-in effect (Van Dijck et al., 2016). This makes it more difficult for users to switch to another developer.

Policy makers in the Netherlands, Europe and elsewhere are trying in various ways to counter the technology giants' burgeoning market power. These efforts are aimed at more active enforcement of existing legislation and regulations, while

policymakers are looking to adapt various legal frameworks, including competition law, consumer law, and privacy law.

But more is needed. We have already stated that, thanks to their powerful position, the technology giants process a lot of user data – in a growing number of domains. Can smaller players still compete with that? And how do we ensure that languages, accents and target groups that do not interest the major companies are included? This seems to call not only for regulation, but also for active government investment in speech technology that makes speech systems more inclusive.

4.7 Conclusion

This chapter provided an overview of the societal and ethical aspects associated with the rise of speech technology. Based on this overview, we conclude that **people's speech is an essential attribute**. Through their speech, people develop and express themselves, relate to each other and maintain or develop important social norms. The use of speech technology will change the interaction among people, between people and devices, and between companies and institutions. Speech technology is a guide that helps people navigate the digital world. This chapter discussed various examples of this, both positive and negative, and illustrated the ethical and societal questions raised by such a guide. Not all these questions can be answered with a simple "right" or "wrong". We do not yet know exactly how human relationships are going to change, how far speech technology intrudes into private life and exactly how skills and decisions are influenced.

However, there are a number of minimum requirements that the use of speech technology must meet. Clearly, developers, companies and other institutions must be considerate of people's voices, conversations and social processes. Five ethical, societal and legal aspects are particularly important: respecting **social norms,** protecting **private and family life**, promoting **autonomy**, ensuring **secure and healthy** use and creating a **fair and accessible market** for speech technology.

5 Conclusion: time for a meaningful conversation

Speech technology has undergone major developments in recent years. Compared to ten years ago, computers are much better at recognising, interpreting and producing human speech – although the user often still needs to lend a helping hand. Some speech computers are even difficult to tell apart from people. Now, speech technology is all around us: in our living rooms, in our cars, in our laptops and in our phones.

Our speech is an essential part of who we are as human beings, and of our social relationships. Our conversations also contain highly sensitive information – about our identity, the type of conversations we have, and even about our health and mood. Our speech therefore has to be protected. Speech technology should strengthen our voices, not abuse them or discriminate against them. Speech technology should enrich our relationships – with other people and with computers – and not disrupt them. And speech technology should make our economy fairer, not increase the dominance of big technology giants.

What does it take to achieve this? In this final chapter we look ahead and make proposals. But we will first summarise the most important findings of the previous chapters.

5.1 The current situation

5.1.1 Speech technology has come of age

Chapter 2 showed that speech technology consists of three key processes: **recognising** speech, **interpreting** speech, and producing speech, which is referred to as **speech synthesis**.

Speech recognition is often quite good, with, in ideal circumstances, error rates of around 5%. For many applications, this level of speech recognition is sufficiently accurate to provide a useful service, such as remotely controlling music or transcribing an interview. However, this error rate is significant – there are plenty of

applications, such as in healthcare or heavy industry, where such an error rate is not acceptable. Moreover, the error rate increases in noisy environments.

Progress is more ambiguous in the area of **speech interpretation**. When performing tasks, help is often needed from the environment and the user: he or she must give the right commands and formulate and answer questions in the right way. Information on websites also has to be made accessible specifically to accommodate speech technology. Although it was often said that computers would learn our language, people still need to adapt to speech technology if it is to be interpreted properly.

Like speech recognition, **speech synthesis** is much improved. In short, speech systems can make themselves clearly understood. Developers have now set a higher goal: speech synthesis has to be so good that people are no longer aware that they are talking to a computer. This is not the case in the vast majority of applications, where people are aware that speech technology is being used. But developments are racing ahead – some specific speech systems, such as Google Duplex, come very close to producing human speech, including "ums" and "ahs".

Speech applications are mastering more and more languages, especially major languages such as English, Spanish and Chinese. At the moment, only Google offers an assistant that speaks Dutch. Most of the progress made in speech technology over the last decade has been due to the emergence of deep learning systems, which can be intensively trained using the improved processing power of computers and the increasing amount of speech data collected. To make further progress, and to teach computers new languages, for example, more data is needed: more human voice recordings and call logs. Another challenge is to have speech applications perform tasks in more difficult environments and communicate as empathetically as possible. To do this, developers want to combine speech with other data, such as facial expressions or lip movements.

5.1.2 Speech technology is being used more and more

In Chapter 3, we showed that various applications based on speech technology are already widely used in society. We also saw how technology providers and companies are starting to experiment with speech. We discussed speech technology that controls devices and speech technology that supports or provides services. The first category includes speech technology in the home, in the car, in industry and in wearables. The second category includes speech technology in the areas of travel and tourism, shopping, healthcare, education, call centres, news, language applications, identification and verification, online searches and

advertisements. In all these domains, speech technology is sold as a helpful assistant that opens up a wide range of services, talks to us, answers our questions and can even identify us.

All these applications collect data, by means of both call logs and audio recordings. Voice can therefore be considered as a new data source. The data is used by developers to personalise speech systems, and forms the basis of analyses in the field of emotion recognition and the diagnosis of diseases. These analyses are often not scientifically proven, but various companies are expecting a lot from the future possibilities of audio recordings.

Finally, we pointed out that speech technology is increasingly becoming our guide in the digital world. The objective of several technology giants, including Google and Amazon, is to get their hands on this portal. They do this by creating a broad platform of speech applications and link them to a voice assistant, such as Alexa and Google Assistant, which can perform a multitude of tasks. They buy up small start-ups and collaborate with players who have a lot of knowledge of a specific domain, such as healthcare. Although there are also small players operating in the speech technology market, the question is how these players will be able to hold their own in the "voice wars" and in light of the increasingly dominant position of the tech giants.

5.1.3 Societal and ethical aspects of speech technology

Chapter 4 set out the societal and ethical aspects associated with the rise of speech technology, making it clear that speech is one of the most important ways in which people express themselves, develop and relate to each other. It is therefore important that people's voices, conversations and social processes should be respected. We discussed five key aspects:

1. **Social relationships and norms** Speech technology intrudes into people's social lives. This involves several elements: the encouragement or, on the contrary, the elimination of bias, the blurring of boundaries that occurs when computers are treated as human beings, the social disciplining of users and the inclusion and exclusion of users. This raises questions about the relationship we consider desirable between humans and computers: do we, and should we, always know that we are talking to computers instead of a human being? At what point do we find it problematic when users consider their voice assistant to be their best friend? What tasks, for example parenting or providing care, do we want to entrust to computers? How do we ensure that speech technology respects existing social norms, for example with regard to equal treatment and disciplining?

2. **Autonomy** Speech technology helps to perform tasks, make decisions and give users an attractive experience. But there are concerns about losing skills, influencing and directing users as they search for information on the web, shop or consume news. Compared to screens, speech technology usually contains less scope for nuance and less information about the source. Who controls and decides on the answer the voice assistant gives? How do we ensure that the user is not unduly influenced? Finally, we mentioned the potentially addictive effect of speech technology.
3. **Privacy** Speech data is highly sensitive data: more than anywhere else, people reveal themselves in conversations at home, in the car and at work. Moreover, technology is being developed to extract additional information from our voices – not only what we say, but also *how* we say it. It can also be used to detect diseases. The idea of speech as a new data source therefore requires extra attention from developers and regulators to ensure that our private and family lives continue to be respected.
4. **Secure and healthy use** Speech technology can compromise people's security in different ways. Speech data can be stolen and abused, for example by cloning a voice. And, despite the improvements made, speech technology is not perfect – accidents can happen when controlling applications. Before speech technology is used in critical applications in healthcare, defence or manufacturing industry, the reliability of the technology will have to be beyond doubt. Finally, speech applications can lead to hearing loss if earphones are used too much and the sound is too loud.
5. **Power of technology companies** Speech technology has created a new and fast-growing market. The well-known tech giants dominate this market, using speech to further expand their existing dominant position. As they are the main developers, the major technology companies also play a key part in respecting social norms, our private lives, and creating secure applications that do not harm, confuse, improperly influence or addict individuals.

5.2 Speech technology demands societal and political action

This investigation builds on an extensive series of studies conducted by the Rathenau Instituut into trends in the digital society. For example, the essay "Intimate Technology" (2014) pointed out that technology knows more and more about us, is coming between us more and is more like us. The report "Human Rights in the Robot Age" (2017) highlighted the impact of digital technology on human rights such as privacy, security and non-discrimination. The overview study "Urgent Upgrade" (2017) charted the full breadth of the ethical and societal issues of the digital society and pointed out the responsibilities of governments, such as municipalities, provinces, central government, regulators and executive organisations, as well as industry and individuals themselves, to address these issues.

Many of the issues and actions raised in these studies are still topical. Speech technology adds a new dimension to the general task of managing digital technology effectively and shows that government and industry must take action. After all, our speech is at stake. Our voices and conversations are an essential part of who we are as human beings and the relationships we enter into with others. Speech technology gives us someone to talk to at all times – at home, in the car, at work and when shopping – and this will affect our speech and our relationships – both with each other and with computers. In addition, speech technology creates a new source of data, containing highly sensitive information.

These developments affect us as individuals but also have consequences for society as a whole. After all, speech technology not only influences the way individuals use computers, it also affects the behaviours we develop together. It changes not only the way in which individuals acquire knowledge, but also the knowledge on which public debate is based. And it has an impact not only on the relationship between customers and companies but also on the platform economy as a whole.

It is up to policy makers and politicians to provide frameworks within which speech technology can be responsibly developed. It is up to industry to create and manage socially responsible technology. And it is up to individuals to discuss the role of speech technology in their lives. The report therefore presents a series of actions for government and industry (see Table 1).

Table 1 Recommendations

Recommendations	Government	Industry
1. Ensure effective privacy protection	Introduce permit requirement for voice analysis. Monitor use of speech technology by law enforcement agencies.	Implement privacy principles rigorously.
2. Promote inclusive speech technology	Invest in a Dutch speech database. Call for industry to accept its responsibilities.	Beware of stereotyping in use of voice and promote recognition of diverse language use.
3. Create a fair market	Tighten up competition law. Provide opportunities for alternative suppliers.	Prioritise the rights of consumers.
	Government and industry	
4. Protect human dignity	Initiate an ethical dialogue on the use of speech technology and make agreements.	
5. Make sure speech technology is reliable	Address disinformation. Reduce error rates of speech systems.	
6. Invest in technological citizenship	Educate individuals to deal responsibly with speech technology, and boost research.	

1. Ensure effective privacy protection

Government: introduce a permit requirement for voice analysis, and monitor the use of speech technology by law enforcement agencies.
Speech technology makes it possible to collect sensitive data from people and use it to influence them. This also includes "special categories of personal data", such as biometric data that can be used to identify a person or data about a person's health. The collection of voice data therefore poses risks to people and their fundamental rights. Incorrect processing in healthcare can lead to misdiagnosis, the theft of a voice ID can lead to material damage and the increasing chance of identification can lead to "chilling" effects, where someone no longer dares to speak freely for fear that their voice will be analysed and/or identified.

It is no coincidence that the European Commission (2020) cited the use of biometric applications for remote identification as an example of "high risk" applications. Ethically responsible use of speech data is at the heart of "ethical AI", an important principle of current policy of both the Dutch government and the European

Commission on AI. Because of these sensitivities, the processing of special personal data is already subject to stricter safeguards under current legislation (the General Data Protection Regulation, GDPR).[22] The processing of special personal data is in principle prohibited, unless one of ten grounds for exception apply and the processing conforms to the principles of proportionality and subsidiarity. It is up to the supervisory authority to detect unlawful data processing – and this is a very demanding task given the numerous applications of special personal data.

In order to enforce more effectively the current ban on the collection and processing of special personal data, the Rathenau Instituut advocates the **introduction of a permit requirement for biometric analysis for verification and identification purposes, including voice analysis** (Rathenau Instituut, 2020). All players who would like to collect and process biometric data for verification and identification purposes would have to apply to the Dutch Data Protection Authority in advance for authorisation. The data supervisor already does so with regard to certain types of processing of criminal data. The GDPR Implementation Act provides scope for introducing such a permit requirement for biometric data. Governments also need to develop strategies to regulate sentiment analysis and health analysis based on voice data.

In addition to the introduction of a permit requirement, it is important for the government to **monitor the use of speech technology by law enforcement agencies**. For example, Interpol has already built up a considerable database of voice profiles, partly based on audio intercepted during investigations and partly based on social media messages. The use of speech technology by law enforcement agencies is on the rise worldwide. It is important to take a close look at the data collected by law enforcement agencies. For example, is it desirable for the police to "scrape" voice data from social media? Many individuals will not realise this when uploading their voice. It is also important to check whether the supervision system is adequate: is a law enforcement agency being monitored effectively to ensure that it complies with its own rules? With the rise of speech technology, these questions deserve the attention of politicians.

Industry: implement privacy principles rigorously.
At the moment, all kinds of requirements are imposed, principally by the GDPR, on the data use and algorithms of speech technology developers. It is important that industry does not follow these rules to the minimum extent. It must implement principles such as transparency, consent and data minimisation as rigorously as possible and focus its innovative capacity accordingly. For example, by investing in

22 For an explanation of the legal framework by the Dutch Data Protection Authority, see the General Data Protection Regulation Manual: autoriteitpersoonsgegevens.nl/sites/default/files/atoms/files/handleiding_avg.pdf.

techniques such as "federated learning", where AI systems are developed without the data leaving a device, by making the operation of voice assistants more transparent to consumers, and by improving the "wake word" function, which is essential for consent.

2. Promote inclusive speech technology

Government: invest in a Dutch speech database and call for industry to accept its responsibilities.
Speech technology can have excluding effects because it does not recognise voices from a variety of groups, such as people who speak a dialect, older people or small children and therefore does not work as well. This is worrying, as speech technology provides an opportunity to open up the digital world to less media-literate people, such as older people. Moreover, speech systems that recognise the language of some groups better than others are discriminatory – whereas everyone in the Netherlands is entitled to equal treatment. Finally, providing an inadequate service to specific groups can have a detrimental effect on the Dutch-speaking region because the technology forces people to start speaking a certain type of Dutch.

The government can promote inclusion and combat discrimination in various ways. First of all, the government can invest in a rich and diverse **Dutch language database**, which will be accessible to the public. Drawing on this database, all kinds of organisations, companies and the government itself can then develop inclusive speech technology. This kind of database can also support creating fair and accessible market conditions (see Action 3).

In addition, the government can call for industry to accept its responsibilities. Like other providers of services and products, speech technology developers must develop plans and implement policies to combat discrimination – it is the government's task to monitor whether this responsibility is being taken sufficiently seriously and to discuss it with developers.

Industry: beware of stereotyping.
Voice assistants are increasingly talking to users and taking on human profiles. It is therefore important for developers to take care that these **profiles do not encourage stereotypes**, such as the stereotype that women have to be submissive and cheerful. It would be good if there were as much diversity as possible in the range of voices and profiles on offer and if users could also choose their own. Given the increasing ease with which a voice can be synthesised, this is technically feasible.

3. Create a fair market

Government: tighten up competition law and provide opportunities for alternative suppliers.
There are currently major concerns in society about the dominance of a few large technology companies within the platform economy. Speech technology offers these large companies the opportunity to expand their dominant position even further, which can lead to monopolisation and the disadvantaging of other providers entering the market, such as "white label" platforms. It is therefore essential for the government to take steps to ensure that the speech technology market is fair.

This means first and foremost that **competition law must be tightened up**. At European level, the Dutch government is already seeking to tighten up competition law in order to create a fairer and more accessible market. This means that policy makers need to be extra alert to the possibility of speech technology further increasing the existing dominance of large platform companies. In addition, it is important that, as ***launching customers***, government and government-funded organisations look not only to the major technology companies but also to white label providers when purchasing speech technology. Open source technology may provide a solution in the future. For these two options to succeed, it is vital for the government to invest in a **public speech database** of the Dutch language, which all players can use – see Recommendation 2.

Industry: prioritise the rights of consumers.
A fair market is not only about relationships between companies but also about the relationship between companies and consumers. It is important that companies should be able to implement consumer rights, such as redress and information about the functioning of products and services, in a generous and decisive manner. When companies do not accept their responsibility, it is up to governments, and in particular regulators, to protect consumer rights. More opportunities have recently been opened up to address this issue in Europe.[23] But the most important thing is for businesses themselves to prioritise consumer rights in the first place.

4. Protect human dignity

Government and industry: initiate an ethical dialogue on the use of speech technology and make agreements.

23 https://ec.europa.eu/commission/presscorner/detail/nl/IP_18_3041

In view of the societal and ethical issues surrounding speech technology, and the role of government as guardian of fundamental and human rights, it is vital for government to establish a dialogue with industry on the responsible development and use of speech systems. Two topics in particular require attention.

First of all, speech technology can compromise people's **right to human contact**. The advent of speech technology makes it possible to have voice services provided not by people, but by computers. In future, this could mean that people will first have to deal with a computer when they contact the government and all kinds of companies by telephone. The Rathenau Instituut previously advocated a right to human contact, so that people can choose whether they want to be assisted by a human being (Van Est et al., 2017). The government can in any case adopt this principle for its own services, and it can encourage other players to do the same.

In addition, it is important that speech technology **does not confuse** people. This at least means that it will always be clear to users whether it is a computer or a human being talking. Government and industry can **enter into agreements about this**. It is also essential to have a dialogue about the extent to which computers should communicate with people empathetically, as this can be conducive to projecting human traits and even cause addiction.

5. Make sure speech technology is reliable

Government and industry: address disinformation and reduce the error rates of speech systems.
Speech technology has a lot to offer society, provided that it is reliable. A number of steps have to be taken to ensure this. First of all, it is important to **act decisively against disinformation** and voice cloning. Governments, and in particular the European Commission, have taken a number of measures in recent years to combat disinformation. One is to make platforms more aware of their responsibilities and the need to regulate themselves. For example, a Code of Practice on Disinformation has been agreed and signed by large platform companies. Nevertheless, there is persistent criticism of the platform companies and they are accused of not acting effectively enough. This is why the government should consider less non-binding measures and amend the Code to ensure that compliance can be enforced by coercive means. It could also invest in a system of digital signatures to verify the authenticity of audio recordings. Ultimately, it is up to the platforms themselves to detect disinformation. To this end, companies must continue to use innovative techniques to reveal and remove disinformation.

Secondly, it is essential for industry to **reduce the error rates of speech systems**. These rates are still significant: around 5% in good conditions and higher in poor conditions. This means that speech systems that perform all kinds of tasks can easily make mistakes and cause harm. Although developers are working to reduce these rates, government should monitor the extent to which the high error rates pose safety risks and support the development of safety standards – especially if speech technology gets used more frequently in critical applications such as a medical procedure. In addition, malicious parties try to fool speech recognition, for example using adversarial examples. It is therefore important for governments and companies to invest in new technology to prevent this type of abuse.

6. Invest in technological citizenship

Government and industry: educate individuals to deal responsibly with speech technology, and boost research about health and social effects.
Speech technology can be a solution in some situations – especially where people have difficulty controlling digital applications using reading and writing. But responsible and effective use of speech technology also requires knowledge and skills, for example in terms of searching for information and setting up routines. Government and industry can take a variety of steps to support technological citizenship.

Firstly, government should combine forces with companies and civil society organisations to **to** invest in education and training in media literacy, specifically regarding speech technology. Individuals need to learn to use speech technology and find out about risks such as voice cloning. Secondly, governments and companies can **boost ongoing research** into the influence of speech technology on our mental and physical health.

Individuals: make your voice heard and put speech technology on the agenda for public debate.
Individuals also have an important role to play. This report has identified many important societal and ethical issues that are worthy of public debate, such as the inclusiveness of speech technology, the relationship between people and talking computers, the impact of speech technology on our social norms and the use of voice analysis by companies and governments. But ultimately it is also up to individuals, based on their experiences with speech technology, to put these themes on the agenda for public debate. This report therefore ends with a call to action: share your experiences and speak out! Our speech is a vulnerable and meaningful commodity – and worthy of debate.

5.3 Final word

This research is part of a group of studies into digital technologies that are changing our relationship with the digital world. In addition to speech technology, the study "Responsible VR" also looked into Virtual Reality (VR), and the study *"Nep echt - Verrijk de wereld met Augmented Reality"* [Fake reality – Enrich the World with Augmented Reality] analysed the technology behind Augmented Reality. These technologies are not independent of each other but can be combined, for example if you are controlling AR glasses with your voice. It is therefore important to talk about the sum total of these technologies: the synthesis of speech technology, VR and AR. The Rathenau Instituut will organise and drive that conversation. It is precisely when we find out how individuals use these technologies and what effect these technologies have on us, that society can promote the use of digital technology that enriches our lives and counteract the use of digital technology that impoverishes our lives.

Literature

3PlayMedia. (2019). *State of Automatic Speech Recognition. An Annual Report*. https://go.3playmedia.com/rs-2019-asr

Abdulsatar, A. et al. (2019). 'Age and gender recognition from speech signals'. In: Journal of Physics: Conference Series Vol. 1410, nr. 1, p. 012073). IOP Publishing.

Alzantot, M., B. Balaji & M. Srivastava (2017). 'Did you hear that? Adversarial Examples Against Automatic Speech Recognition'. In: *31st Conference on Neural Information Processing Systems.*

Amazon (2020). 'Alexa for Hospitality'. https://www.amazon.com/alexahospitality.

Amazon (2020). 'Alexa Priza. Amazon'. https://developer.amazon.com/en-US/alexa/alexa-startups/alexa-fund

Amnesty International. (2019). *Surveillance Giants: How The Business Model Of Google And Facebook Threatens Human Rights*. Amnesty International. https://www.amnesty.nl/content/uploads/2019/11/20191119_FINAL_Surveillance-giants-online-FINAL.pdf?x68150

Anand, P. 'The Reality Behind Voice Shopping Hype'. In: *The Information*, 6 Augustus 2018. https://www.theinformation.com/articles/the-reality-behind-voice-shopping-hype

Arcon, N., P. Klein & J. Dombroski (2017). 'Effects of Dictation, Speech to Text, and Handwriting on the Written Composition of Elementary School English Language Learners'. In: *Reading & Writing Quarterly*, 33, nr. 6, pp. 533-548.

Arik, S., et al. (2018). 'Neural voice cloning with a few samples'. In: *Advances in Neural Information Processing Systems*. pp. 10019-10029. https://arxiv.org/pdf/1802.06006.pdf

Asimov, I. (2008). *I Robot*. New York City: Penguin Random House.

Baig, E. 'Say thank you and please: Should you be polite with Alexa and the Google Assistant?'. In: *USA Today*, 10 oktober 2019. https://eu.usatoday.com/story/tech/2019/10/10/do-ai-driven-voice-assistants-we-increasingly-rely-weather-news-homework-help-otherwise-keep-us-info/3928733002/

Baraniuk, C. 'How talking machines are taking call centre jobs'. In: *BBC News*, 24 augustus 2018. https://www.bbc.com/news/business-45272835

Barnes, E, N. Fandos & D. Hakim. 'White House Ukraine Expert Sought to Correct Transcript of Trump Call'. In: *The New York Times*, 19 november 2019. https://www.nytimes.com/2019/10/29/us/politics/alexander-vindman-trump-ukraine.html

Barocas, S. & Selbst, A. D. (2016). ‘Big data's disparate impact’. In: *104 California Law Review* 671.

Bharadwaj, R. ‘Voice and Speech Recognition in Banking – What’s Possible Today’. In: *Emerj*, 11 juli 2019. https://emerj.com/ai-sector-overviews/voice-speech-recognition-banking/

Bremmer, D. ‘Slimme spraakassistent houdt ouderen langer zelfstandig’. In: *Algemeen Dagblad*, 3 april 2019. https://www.ad.nl/tech/slimme-spraakassistent-houdt-ouderen-langer-zelfstandig~af0a3bea/?referrer=https://www.google.com/

Carmen, A. ‘Volkswagen now lets Apple users unlock their cars with Siri’. In: *The Verge,* 12 November 2018. https://www.theverge.com/2018/11/12/18087416/volkswagen-vw-car-net-app-siri-shortcuts

CBinsights (2019). ‘How Big Tech Is Battling To Own The $49B Voice Market’. In: *CBinsights Research*, 13 februari 2019. https://www.cbinsights.com/research/facebook-amazon-microsoft-google-apple-voice/

Chaudhari, S. & R. Kagalkar. (2014) ‘A review of automatic speaker age classification, recognition and identifying speaker emotion using voice signal.’ In: *International Journal of Science and Research* 3 nr. 11 pp. 1307-1311.

Chen, A. ‘Why companies want to mine the secrets in your voice’. In: *The Verge*, 14 maart 2019. https://www.theverge.com/2019/3/14/18264458/voice-technology-speech-analysis-mental-health-risk-privacy

Chiu, C. et al. (2018). ‘State-of-the-art speech recognition with sequence-to-sequence models’. In: *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 4774-4778). IEEE.

Chowdhury, G. (2003). ‘Natural language processing’. In: *Annual review of information science and technology 37* nr. 1, pp. 51-89.

Conner, K. ‘9 best Alexa games to play on your Amazon Echo’. In: CNET, 5 maart 2020. https://www.cnet.com/how-to/9-best-alexa-games-to-play-on-your-amazon-echo/

Cox, T. (2019). *Now You’re Talking. The story of human conversation from the Neanderthals to Artificial Intelligence.* Penguin.

Cox, T. (2020). Siri and Alexa Fails: Frustrations With Voice Search. The Manifest. https://themanifest.com/digital-marketing/resources/siri-alexa-fails-frustrations-with-voice-search

Danaher, J. (2018). ‘Toward an Ethics of AI Assistants: an Initial Framework’. In: *Philosophy & Technology*, 31, pp. 629–653.
Day, M., G. Turner, & N. Drozdiak. ‘Amazon Workers Are Listening to What You Tell Alexa’. In: *Bloomberg*, 11 april 2019. https://www.bloomberg.com/news/articles/2019-04-10/is-anyone-listening-to-you-on-alexa-a-global-team-reviews-audio

Dawar, N. ‘Marketing in the Age of Alexa’. In: *Harvard Business Publishing*, mei 2018. https://hbr.org/2018/05/marketing-in-the-age-of-alexa

Dujardin, A. 'Aan het eind van deze eeuw is het aantal talen op de wereld gehalveerd'. In: *Trouw*, 30 april 2017.

Dzieza, J. How Hard Will The Robots Make Us Work? In: *The Verge*, 27 februari, 2020. https://www.theverge.com/2020/2/27/21155254/automation-robots-unemployment-jobs-vs-human-google-amazon

Europese Commissie (EC) (2020). *Witboek. over kunstmatige intelligentie - een Europese benadering op basis van excellentie en vertrouwen*. Brussel. https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_nl.pdf

Faulkner, C. ‘Issa Rae is the next celebrity voice coming to Google Assistant’. In: The Verge, 10 oktober 2019. https://www.theverge.com/2019/10/10/20906823/issa-rae-google-assistant-celebrity-voice-cameo-android-ios-nest-hub-home-wavenet.

French, K. ‘Something in Your Voice’. In: *Proto*, 7 oktober 2019. http://protomag.com/articles/something-your-voice

Fussel, S. ‘Alexa Wants to Know How You’re Feeling Today’. In: *The Atlantic*, 12 oktober 2018. https://www.theatlantic.com/technology/archive/2018/10/alexa-emotion-detection-ai-surveillance/572884/

Gershgorn, D. ‘Here’s How Amazon Alexa Will Recognize When You’re Frustrated’. In: *Medium*, 27 september 2019.

Glynn, F. ‘What is voice picking?’ In: 6River systems 18 januari 2020. https://6river.com/what-is-voice-picking/

Gupta, M., S. Bharti & Agarwal, S. (2019). ‘Gender-based speaker recognition from speech signals using GMM model’. In: *Modern Physics Letters B*, 33 nr. 35, 1950438.

Habermas, J. (1997). *Between Facts and Norms*. Polity Press.

Hager, R. ‘Google confirms new voice-confirmation feature for purchases in Assistant’. In: *Android Police*, 25 mei 2020. https://www.androidpolice.com/2020/05/25/google-assistant-gets-new-confirm-with-voice-match-setting-for-payments/

Hamer, J. et al. (2019). *Cyberspace without conflict. The search for de-escalation of the international information conflict.* Den Haag: Rathenau Instituut.

Harris, D. ‘Baidu explains how it’s mastering Mandarin with deep learning’. In: *Medium*, 11 augustus 2015. https://medium.com/s-c-a-l-e/how-baidu-mastered-mandarin-with-deep-learning-and-lots-of-data-1d94032564a5

Harris, J. ‘Voice chat with Replika, a socialbot’. In: *Medium*, 7 januari 2019. https://medium.com/speaking-naturally/voice-chat-with-replika-a-socialbot-a9d4bcaacea8

Henzi, R. & O. Wright. ‘3 ways voice technology will change your life’. In: *World Economic Form*, 11 Juni 2019. https://www.weforum.org/agenda/2019/06/how-voice-technology-will-change-your-life/

Hern, A. ‘Apple contractors 'regularly hear confidential details' on Siri recordings’. In: *The Guardian*, 26 juli 2019. https://www.theguardian.com/technology/2019/jul/26/apple-contractors-regularly-hear-confidential-details-on-siri-recordings

Hirschberg, J., & C. D. Manning. (2015). ‘Advances in natural language processing’. In: *Science* 349 nr. 6245, pp.261-266.

Hoy, M. B. (2018). ‘Alexa, siri, cortana, and more: An introduction to voice assistants’. In: *Medical reference services quarterly*, 37 nr. 1, pp. 81-88

Howard, A & J. Borenstein (2017). ‘The Ugly Truth About Ourselves and Our Robot Creations: The Problem of Bias and Social Inequity’. In: *Science and Engineering Ethics*, 24, pp. 1521–1536.

Huang, X., J. Baker & R. Reddy (2014). ‘A historical perspective of speech recognition’. In: *Communications of the ACM*, 57 nr. 1, pp. 94-103.

iTalk2Learn (2015). Website: https://www.italk2learn.com/.

Jacobs, A. ‘Siri in het EPD: heeft spraakbesturing ook toekomst in de zorg?’ In: *Smarthealth*, 23 augustus 2018. https://www.smarthealth.nl/2018/08/23/siri-in-het-epd-heeft-spraakbesturing-ook-toekomst-van-de-zorg/

Jawad, U. ‘IBM beats Microsoft's word error rate in speech recognition, achieves 5.5%’. In: *Neowin*. 11 maart 2017. https://www.neowin.net/news/ibm-beats-microsofts-word-error-rate-in-speech-recognition-achieves-55

Jothilakshmi, S. & V. N. Gudivada. (2016). ‘Large Scale Data Enabled Evolution of Spoken Language Research and Applications’. In: *Handbook of Statistics* 35, pp. 301-340

Jyothi, P. *Automatic Speech Recognition - An Overview* [video]. Microsoft Research, 11 September 2017. https://www.youtube.com/watch?v=q67z7PTGRi8

Jyothi, P. *State-of-the-Art in Speech Technologies* [video]. Academic Research Summit 2018, 24 januari, 2018. https://www.youtube.com/watch?v=AHk51EsRlgg

Kamp, R. ‘Smartphone gebruiken in de auto’. In: Consumentenbond: 28 Maart 2019. https://www.consumentenbond.nl/smartphone/smartphone-in-de-auto

Kantar. ‘Gebruik smart speakers groeit explosief’. In: TNS-NIPO, 14 maart 2019. http://www.tns-nipo.com/nieuws/persberichten/gebruik-smart-speakers-groeit-explosief?lang=nl-NL

Kartalidis, N. (2018). *Speech recognition in construction equipment: Creating a voice assistant for an autonomous wheel loader*. Uppsala Universitet. https://pdfs.semanticscholar.org/c222/ba625f848912abf524c042e88ddc499a659e.pdf

Kepuska, V. G. & Bohouta (2018). ‘Next-generation of virtual personal assistants (microsoft cortana, apple siri, amazon alexa and google home). In: *2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC)* pp. 99-103

Kimmich, M. (2019). *Smart Homes: Strategic Opportunities, Business Models & Competitive Landscape* 2019-2024. Hampshire: Juniper Research.

Kinsella, B. ‘Amazon Files for Patent to Detect User Illness and Emotional State by Analyzing Voice Data’. In: Voicebot.ai, 10 oktober 2018.

Kinsella, B. & A. Mutchler (2018). *Smart speaker consumer adoption report march 2018. U.S: Voicebot.* https://voicebot.ai/wp-content/uploads/2018/10/voicebot-smart-speaker-consumer-adoption-report.pdf

Kinsella, B. (2019). *U.S. Smartspeaker consumer adoption report 2019.* U.S: Voicebot. https://voicebot.ai/smart-speaker-consumer-adoption-report-2019/

Kinsella, B. ‘Why Tech Giants Are So Desperate to Provide Your Voice Assistant’. In: *Harvard Business Review*, 7 mei 2019. https://hbr.org/2019/05/why-tech-giants-are-so-desperate-to-provide-your-voice-assistant

Kishore, S. P., & A. W. Black (2003). ‘Unit size in unit selection speech synthesis’. In: *Eighth European Conference on Speech Communication and Technology*. pp. 1317-1320.

Kodish-Wachs, J. et al. (2018). ‘A systematic comparison of contemporary automatic speech recognition engines for conversational clinical speech’. *In: AMIA Symposium*, pp. 683–689.

Kofman, A. Interpol rolls out international voice identification database using samples from 192 law enforcement agencies. In: *The Intercept, 25 juni 2018.* https://theintercept.com/2018/06/25/interpol-voice-identification-database/

Kohn, M. ‘Is the era of artificial speech translation upon us?’ In: *The Guardian*. 17 februari 2019. https://www.theguardian.com/technology/2019/feb/17/is-the-era-of-artificial-speech-translation-upon-us

Koksal, I. (2018). ‘Voice-First Devices Are The Next Big Thing -- Here's Why’. In: Forbes 1 februari 2018. https://www.forbes.com/sites/ilkerkoksal/2018/02/01/voice-first-devices-are-the-next-big-thing-heres-why/#361e22f36873

Kool, L., et al. (2017). *Urgent upgrade. Protect public values in our digitized society.* Den Haag: Rathenau Instituut.

Kool, L., J. Timmer. & R. van Est., (2014). *Sincere support. The rise of the e-coach.* Den Haag: Rathenau Instituut.

Kool, L. & R. van Est. (2015). *Working on the robot society. Visions and insights from science concerning the relationship between technology and employment.* Den Haag: Rathenau Instituut.

Koops, B. J. (2019). *Privacyconcepten voor de 21e eeuw*. Nijmegen: Ars Aequi, 68 nr. 7/8, pp. 532-544.

Kretzschmar, K et al. (2019). ‘Can Your Phone Be Your Therapist? Young People’s Ethical Perspectives on the Use of Fully Automated Conversational Agents (Chatbots) in Mental Health Support’. In: *Biomedical Informatics Insights*, 11, pp. 1–9.

Kritzler, M., et al. (2019). ‘Digital Companion for Industry’. In: *Companion Proceedings of The 2019 World Wide Web Conference,* pp. 663-667

Kuligowska, K., P. Kisielewicz & A. Włodarz. (2018). ‘Speech synthesis systems: disadvantages and limitations’. In: *International Journal of Engineering & Technology 7*, pp. 234-239.

Langley, H. & J. P. Tuohy. ‘Smart home privacy: What Amazon, Google and Apple do with your data’. In: *The Ambient*, 8 november 2019. https://www.the-ambient.com/features/how-amazon-google-apple-use-smart-speaker-data-338

Leswing, K. ‘People found a really easy way to make Siri curse’. In: *Business Insider,* 30 april 2018. https://www.businessinsider.nl/apple-siri-swears-when-asked-for-second-definition-of-mother-2018-4/

Leventon, W. ‘Alexa and Siri, meet Athena for machining’. In: *Cutting Tool Engineering*, 8 januari 2019.

Lens-FitzGerald, e.a. (2020). *Voice Verbindt: Slimme spraaktechnologie verbetert leefkwaliteit van ouderen*. Projectzilver.
https://projectzilver.com/onderzoeksrapport/Li, B., et al. (2020). ‘Towards fast and accurate streaming end-to-end ASR’. In: *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),* pp. 6069-6073

Lieberman, H. et al. (2005). ‘How to wreck a nice beach you sing calm incense’. In *Proceedings of the 10th international conference on Intelligent user interfaces*, pp. 278-280.

Liu, K. et al. (2019). *A review of the literature on computerized speech-to-text accommodations (NCEO Report 414).* Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

MacCartney, B. ‘Computational Linguistics (aka Natural Language Processing)’ [Presentation]. Stanford University, 26 mei 2011.
https://nlp.stanford.edu/~wcmac/papers/20110526-symsys-100-nlp.pdf

Manning, C. ‘Lecture 1 | Natural Language Processing with Deep Learning’ [video]. Stanford University School of Engineering, 3 april 2017.
https://www.youtube.com/watch?v=OQQ-W_63UgQ

Mattin, D. ‘Voice technology could soon be your new best friend. Here's why’. In: *World Economic Forum*, 10 Juli 2019.
https://www.weforum.org/agenda/2019/07/voice-technology-personalization/

Marmar, C. R., et al. (2019). ‘Speech-based markers for posttraumatic stress disorder in US veterans’. In: *Depression and anxiety*, 36 nr. 7, pp. 607-616.

Meeker, M. (2018). *INTERNET TRENDS 2018*. California: Kleiner Perkins.
https://www.kleinerperkins.com/perspectives/internet-trends-report-2018/

Meiling, C. ‘Demand for AI virtual anchors set to increase’. In: *China Daily*, 14 juni 2019. http://www.chinadaily.com.cn/a/201906/14/WS5d03013aa3103dbf14328348.html

Meridith, S. ‘A new era in commerce’: Amazon Pay VP says voice payments potential is ‘phenomenal’. In: *CNBC*, 4 juni 2019.

Meyer, D. ‘Introducing Alexa Conversations (Preview), a New AI-Driven Approach to Natural Dialogs through the Alexa Skills Kit’. In: Amazon Alexa, 5 juni 2019. https://developer.amazon.com/blogs/alexa/post/44499221-01ff-460a-a9ee-d4e9198ef98d/introducing-alexa-conversations-preview

Mittal, U. (2019). ‘Design of data virtual assistant using natural language processing’. In: *Journal of the Gujarat Research Society*, *21* nr. 8, pp. 786-791.

Muller, J. Improved voice controls could increase car safety. In: *Axios*, 29 mei 2019. https://www.axios.com/autonomous-vehicles-safety-voice-controls-ec6845d5-3480-4ed4-a15b-cd5641130982.html

Müller, C. (2006). ‘Automatic recognition of speakers' age and gender on the basis of empirical studies’. In: *Ninth International Conference on Spoken Language Processing*.

Multiscope. ‘Smart home markt groeit tot 2,5 miljard euro’. In: *Mutliscope*, 12 maart 2020 http://www.multiscope.nl/persberichten/smart-home-markt-groeit.html

Mwiti, D. A. ‘Guide for Automatic Speech Recognition’. In: *Medium / Heartbeat*, 4 September 2019. https://heartbeat.fritz.ai/a-2019-guide-for-automatic-speech-recognition-f1e1129a141c

Nass, C. I., & S. Brave. (2005). *Wired for speech: How voice activates and advances the human-computer relationship*. Cambridge, MA: MIT press.

Nederlandse Publieke Omroep (NPO) (2020). *De 10 lessen uit 15 smart speaker projecten*. https://innovatie.npo.nl/projecten/de-10-lessen-uit-15-smart-speaker-projecten

Newman, J. ‘Can Mycroft’s Privacy-Centric Voice Assistant Take On Alexa and Google?’ In: *Fastcompany*. 29 januari 2018. https://www.fastcompany.com/40522226/can-mycrofts-privacy-centric-voice-assistant-take-on-alexa-and-google

Olson C. & K. Kemery. (2019). *Voice Report*. Microsoft. https://about.ads.microsoft.com/en-us/insights/2019-voice-report

Oord, A. V. D. et al. (2016). ‘Wavenet: A generative model for raw audio’. London: Google DeepMind. https://arxiv.org/pdf/1609.03499.pdf

O'Shaughnessy, D. (2013). ‘Acoustic analysis for automatic speech recognition’. In: *Proceedings of the IEEE 101, nr.* 5, pp. 1038-1053.

Pardes, A. ‘Hey Alexa, Why Is Voice Shopping So Lousy?’ In: *WIRED*, 17 Juni 2019. https://www.wired.com/story/why-is-voice-shopping-bad/

Pichai, S. (2018) Keynote (Google I/O '18). Google Developers. https://www.youtube.com/watch?v=ogfYd705cRs&t=1014s

Pichai, S. (2019) Google Keynote (Google I/O '19). Google Developers. https://www.youtube.com/watch?v=lyRPyRKHO8M

Pisanski, K., et al. (2014). 'Vocal indicators of body size in men and women: a meta-analysis'. In: *Animal Behaviour* 95, pp. 89-99.

Place, S., Blanch-Hartigan, D., Rubin, C., Gorrostieta, C., Mead, C., Kane, J., ... & Azarbayejani, A. (2017). Behavioral indicators on a mobile sensing platform predict clinically validated psychiatric symptoms of mood and anxiety disorders. Journal of medical Internet research, 19(3), e75.Protalinski, E. (2019). ProBeat: Has Google's word error rate progress stalled?. In: Venturebeat. https://venturebeat.com/2019/05/10/probeat-has-googles-word-error-rate-progress-stalled/

Pundak, G., et al. (2018). 'Deep context: end-to-end contextual speech recognition'. In: *2018 IEEE Spoken Language Technology Workshop (SLT)* pp. 418-425.

Rathenau Instituut (2020). Reactie op verzamelwet gegevensbescherming. Brief aan Minister Dekker, 14 juli 2020

Reilly, M.. 'Should an Amazon Echo Help Solve a Murder?'. In: *MIT Technology Review,* 27 december 2016. https://www.technologyreview.com/2016/12/27/154864/should-an-amazon-echo-help-solve-a-murder/

Reuters. 'Germany planning to access voice assistant data to tackle crime'. In: *Reuters*, 5 juni 2019. https://www.reuters.com/article/uk-germany-crime-alexa/germany-planning-to-access-voice-assistant-data-to-tackle-crime-idUKKCN1T61IE

RTL Nieuws. 'Grote zorgen oorartsen: jongeren vaker met gehoorproblemen door koptelefoons en oortjes'. In: *RTL Nieuws*, 23 augustus 2019. https://www.rtlnieuws.nl/nieuws/nederland/artikel/4822536/gehoorschade-oren-kno-arts-dokter-tinnitus-piep-oorsuizen-muziek

Ruane, E., A. Birhane & A. Ventresque (2019). 'Conversational AI: Social and Ethical Considerations'. In: *AICS - 27th AIAI Irish Conference on Artificial Intelligence and Cognitive Science.*

Sankaranarayanan, P. (2017). *New Zealand Government Case Study*. Auraya Systems. https://aurayasystems.com/2017/04/28/new-zealand-government-case-study/

Sawers, P. 'How Duolingo is using AI to humanize virtual language lessons'. In: *Venture Beat,* 5 juli 2019. https://venturebeat.com/2019/07/05/how-duolingo-is-using-ai-to-humanize-virtual-language-lessons/

Schwartz, E. 'New Resemble AI Software Turns 3-Minute Records into Synthetic Speech Profiles'. In: *Voicebot.ai*, 30 december 2019.

https://voicebot.ai/2019/12/30/new-resemble-ai-software-turns-3-minute-records-into-synthetic-speech-profiles/

Sedaaghi, M. (2009). 'A Comparative Study of Gender and Age Classification in Speech Signals'. In: *Iranian Journal of Electrical & Electronic Engineering 5*.

Senseable Intelligence Group. (2019). Senseable Intelligence Group: Research. Cambridge: McGovern Institute for Brain Research. https://satra.cogitatum.org/group/research/

Shank, D., C. et al. (2019). 'Feeling our way to machine minds: People's emotions when perceiving mind in artificial intelligence'. In: *Computers in Human* Behavior 98, pp. 256-266.

Shulevitz, J. 'Alexa, Should We Trust You?' In: *The Atlantic*. November 2018. https://www.theatlantic.com/magazine/archive/2018/11/alexa-how-will-you-change-us/570844/

Simon, M. "The Genderless Digital Voice the World Needs Right Now". In: *Wired*. 3 november 2019.

Simonite, T. 'This Call May Be Monitored for Tone and Emotion'. In: *WIRED*. 19 maart 2018. https://www.wired.com/story/this-call-may-be-monitored-for-tone-and-emotion/

Smith, C. 'Alexa and Siri Can Hear This Hidden Command. You Can't'. In: *The New York Times*, 10 mei 2018.

Snijders, D. et al. (2020). *Nep Echt. Verrijk de wereld met augmented reality.* Den Haag: Rathenau Insituut.

Softengi (2020). 'Voice Recognition Technology Is a Must in Any Risky Industry'. Website: Softengi. https://softengi.com/blog/voice-recognition-technology-is-a-must-in-any-risky-industry/

Sokol, N., W. Chen & B. Donmez. (2017). 'Voice-Controlled In-Vehicle Systems: Effects of Voice-Recognition Accuracy in the Presence of Background Noise'. In: *Conference: Driving Assessment Conference*. pp. 158-164

Sondhi, S., et al. (2015). 'Vocal indicators of emotional stress'. In: *International Journal of Computer Applications*, 122, nr. 15. pp. 38-43

Staff, I. 'By voice or location, Israeli apps can determine your risk of coronavirus'. In: *The Times of Israel,* 31 maart 2020. https://www.timesofisrael.com/by-voice-or-location-israeli-apps-can-determine-your-risk-of-coronavirus/

Staff, I. 'Israeli startup aims to identify coronavirus carriers using their voice'. In: *The Times of Israel*, 25 maart 2020. https://www.timesofisrael.com/israeli-startup-aims-to-identify-coronavirus-carriers-using-their-voice/

Statt, N. "Amazon sent 1,700 Alexa voice recordings to the wrong user following data request". In: *The Verge*, 20 december 2018. https://www.theverge.com/2018/12/20/18150531/amazon-alexa-voice-recordings-wrong-user-gdpr-privacy-ai

Strayer, D & J. Cooper. (2015). *Up to 27 seconds of inattention after talking to your car or smartphone*. University of Utah. https://unews.utah.edu/up-to-27-seconds-of-inattention-after-talking-to-your-car-or-smart-phone/?doing_wp_cron=1586506141.6724519729614257812500

Stucke, M. & A. Ezrachi. 'The Subtle Ways Your Digital Assistant Might Manipulate You'. In: *Wired*, 29 november 2016.

Stinson, L. 'The Surprising Repercussions of Making AI Assistants Sound Human'. In: *WIRED*, 12 mei 2017. https://www.wired.com/2017/05/surprising-repercussions-making-ai-assistants-sound-human/

Streijl, R.C., S, Winkler, & D. Hands. (2014) 'Mean opinion score (MOS) revisited: methods and applications, limitations and alternatives'. In: *Multimedia Systems,* pp.213-227

Sundermeyer, M., R. Schlüter, & H. Ney (2012). 'LSTM neural networks for language modeling'. In: *INTERSPEECH-2012*, pp. 194-197

Synnaeve, G.et al., (2020). 'End-to-end ASR: from Supervised to Semi-Supervised Learning with Modern Architectures'. In: *The 37th International Conference on Machine Learning.*

Szymkowski, S. 'Apple's self-driving car system could use voice, gesture guidance'. In: *Roadshow,* 27 januari 2020. https://www.cnet.com/roadshow/news/apple-self-driving-car-system-voice-gesture-guidance/

Tatman, R. (2017). 'Gender and Dialect Bias in YouTube's Automatic Captions'. In: *Proceedings of the First Workshop on Ethics in Natural Language Processing*, pp. 53–59.

Tokuday, K., & H. Zen. (2015). 'Directly modeling speech waveforms by neural networks for statistical parametric speech synthesis'. In: *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* pp. 4215-4219

Turkle, S. (2017). *Alone Together: Why We Expect More from Technology and Less from Each Other (Third Edition).* U.S: Ingram Publisher Services

Turner, M. 'Smart Home 101: How to Develop for the Connected Home (Google I/O'19)' [video]. *Google Developers*. 8 mei 2019. https://www.youtube.com/watch?v=SJ2KYVKfURA

UNESCO (2019). *I'd blush if I could*. Equals skill coalition & Unesco. https://en.unesco.org/Id-blush-if-I-could

Van Bergen, W & Van Gelder, H. 'Shoppen via de slimme speaker'. In: *De Telegraa*f: 10 oktober 2018. https://www.telegraaf.nl/financieel/2659211/shoppen-via-slimme-speaker

Van Camp, J. 'My Jibo Is Dying and It's Breaking My Heart'. In: *WIRED,* 3 augustus 2019. https://www.wired.com/story/jibo-is-dying-eulogy/

Van Dijck, J., T. Poell & M. De Waal (2016). *De Platformsamenleving. Strijd om publieke waarden in een online wereld.* Amsterdam: Amsterdam University Press.

Van Est, R., J. Gerritsen, & L. Kool (2017). *Human rights in the robot age. Challenges arising from the use of robotics, artificial intelligence, and virtual and augmented reality.* Den Haag: Rathenau Instituut.

Van Keulen, I. et al. (2018). *Digitalisering van het nieuws. Online nieuwsgedrag, desinformatie en personalisatie in Nederland.* Den Haag: Rathenau Instituut.

Vernuccio, M., M. Patrizi & A. Pastore (2020). 'Developing voice-based branding: insights from the Mercedes case'. In: *Journal of Product & Brand Management*.

Vermeend, W. & J. W. Timmer (2016). *Internet of things. De nieuwste internet megatrend met een enorme impact op economie werken ondernemen.* Den Haag: Einstein Books en Ebooks

Vincent, J. 'Lyrebird claims it can recreate any voice using just one minute of sample audio.' In: *The Verge*, 24 april 2017. https://www.theverge.com/2017/4/24/15406882/ai-voice-synthesis-copy-human-speech-lyrebird

Vlahos, J. 'Amazon Alexa and the Search for the One Perfect Answer'. In: *WIRED*, 18 februari, 2019. https://www.wired.com/story/amazon-alexa-search-for-the-one-perfect-answer/

Vlahos, J. 'Barbie wants to get to know your child'. In: *The New York Times*, 16 september 2015. https://www.nytimes.com/2015/09/20/magazine/barbie-wants-to-get-to-know-your-child.html

Waarlo, N. "Het eerste woordje van de baby: 'mama' of 'papa'? Straks blijkt het 'Siri' of 'Alexa'". In: *De Volkskrant,* 9 augustus 2019. https://www.volkskrant.nl/wetenschap/het-eerste-woordje-van-de-baby-mama-of-papa-straks-blijkt-het-siri-of-alexa

Warren, T. 'Microsoft no longer sees Cortana as an Alexa or Google Assistant competitor'. In: *The Verge*, 18 januari, 2018. https://www.theverge.com/2019/1/18/18187992/microsoft-cortana-satya-nadella-alexa-google-assistant-competitor

Weizenbaum, J. (1966). 'ELIZA---a computer program for the study of natural language communication between man and machine'. In: *Communications of the ACM*, 9 nr. 1, pp. 36-45.

Welch, C. 'Waze now lets you record your own turn-by-turn audio'. In: *The Verge*. 8 mei 2017. https://www.theverge.com/2017/5/8/15585638/waze-record-custom-navigation-audio

Wenderow, T. 'Vocalis CEO Tal Wenderow Discusses Vocal Biomarkers, Healthcare, and Coronavirus – Voicebot Podcast Ep 145' [audio]. *Voicebot.ai*, 12 april 2020. https://voicebot.ai/2020/04/12/vocalis-ceo-tal-wenderow-discusses-vocal-biomarkers-healthcare-and-coronavirus-voicebot-podcast-ep-145/

Williams, R. 'Amazon expands Alexa with voice-powered wearables'. In: *Mobile Marketer*, 26 september 2019. https://www.mobilemarketer.com/news/amazon-expands-alexa-with-voice-powered-wearables/563740/

Wolters Kluwer. ‘Wolters Kluwer provides voice-enabled search powered by Nuance Dragon Medical One to millions of clinicians using leading clinical decision support’. In: *Wolters Kluwer*, 17 juni 2020. https://www.wolterskluwer.com/en/news/dragon-medical-voice-search-uptodate

Wu, H., et al. (2020). ‘Defense against adversarial attacks on spoofing countermeasures of ASV‘. In: *Electrical Engineering and Systems Science*, arXiv:2003.03065**.**

Yoffie, D. et al. (2018). ‘Voice War: Hey Google vs. Alexa vs. Siri." In: *Harvard Business School Case Collection*, pp. 718-519

Young, T., et al. (2018). ‘Recent trends in deep learning based natural language processing’. In: *IEEE Computational intelligence magazine*, 13 nr. 3, pp. 55-75.

Zen, H. Generative Model-Based Text-to-Speech Synthesis [video]. *Centre for Brain, Minds and Machines (CBMM),* 3 februari 2018. https://www.youtube.com/watch?v=nsrSrYtKkT8

Zhu, Z. et al. (2015). ‘Application of Speech Recognition Technology to Virtual Reality System.’ In: *2015 International Industrial Informatics and Computer Engineering Conference*. Atlantis Press. https://www.atlantis-press.com/proceedings/iiicec-15/16911

Contactgegevens
Anna van Saksenlaan 51
Postbus 95366
2509 CJ Den Haag
070-342 15 42
info@rathenau.nl
www.rathenau.nl

Het Rathenau Instituut stimuleert de publieke en politieke meningsvorming over de maatschappelijke aspecten van wetenschap en technologie. We doen onderzoek en organiseren het debat over wetenschap, innovatie en nieuwe technologieën.

Rathenau Instituut