ContentsList of Abbreviations  
1 Introduction 1  
2 Virtual digital assistants 1  
2.1 Speech recognition 2  
2.2 History of voice assistants 4  
2.3 Intelligence and responding to needs 5  
2.4 Problems 8  
2.5 The future of virtual assistants 9  
3 Personalized virtual assistant 10  
3.1 Mycroft – open source voice assistant 10  
3.2 Setting up a virtual machine 11  
3.3 Installing Mycroft 12  
3.4 Creating a skill from Template 14  
3.5 Creating a skill using Mycroft Skills Kit 17  
3.6 Fortune cookie skill using an API 18  
3.7 Conclusion 20  
4 Summary 20  
References 22

List of AbbreviationsAI Artificial Intelligence  
VDA Virtual digital assistant  
IVA Intelligent virtual assistant  
UNIX Hardware-independent operating system  
MSK Mycroft Skills Kit. Made for creating skills for Mycroft AI platform.  
API Application Programming Interface. Routines, protocols and tools for building software applications.  
JSON JavaScript Object Notation. A file-format with objects in arrays.

1 IntroductionWhen studying the history of virtual assistants, the past few years have shown a drastic  
change in the development and success of these nonhuman helpers. Nowadays most  
major technology companies have their own virtual assistant, or at least have started to  
develop one. The sci-fi fantasy of having a smart computer to talk to is no longer far in  
the future, but in the present. Modern AI assistants try to analyze and learn from the  
user’s behavior so that they can grow to be more efficient and user-friendly as time goes  
by.  
There are multiple popular smart assistants lined up to help us, but how can one tell  
which one is the most suitable for them? Finding the perfect AI assistant starts from  
deciding what the user wants to use the assistant for. After having a clear image about  
the required skills comes the research. Thus, what can one do if they cannot find the  
product they are looking for? The answer is to build one.  
When a technology becomes viral and the competition is severe, the chances are that  
open source equivalents exist. These open software designs can be used as a baseline  
for a personalized intelligent personal assistant. Creating a personal AI assistant takes  
time and effort, but it has its perks, such as customization and data privacy.  
The purpose of this thesis was to help the reader understand what virtual assistants do  
and how to create a personalized virtual assistant of their own. However, instead of studying only at the modern era, the thesis starts from the beginning of the voice assistants  
to really understand how long the journey has been to reach the current smart speaker  
revolution. This way the reader will grasp the concept of voice assistants, history and  
future included.

2 Virtual digital assistantsVirtual assistants are software applications that possess elements of modern artificial  
intelligence and have been created to interact with humans through voice. Large  
technology companies have their own intelligent virtual assistants for different purposes  
2  
and platforms. At first the main purpose of such an assistant was to answer the user’s  
questions, but a point has been reached, where it is time to enhance it further with more  
capabilities and personalities. [1.]  
2.1 Speech recognition  
Speech recognition is the ability to recognize and react to spoken commands, which  
enables hands-free control, and is one of the core skills of a virtual assistant. Speech  
recognition starts from a microphone, which translates the vibrations into a wavelike  
electrical signal, and is converted by a sound card into a digital signal. The speech recognition software will analyze the digital signal and separate the phonemes that are needed  
to form words. Words can sound alike; thus, some natural language processing is required to figure the correct word. One way to select the correct one is to determine the  
context through a trigram analysis, which uses a database of frequent three-word clusters. With statistical analysis of texts, it is possible to count the frequency of monograms,  
bigrams, trigrams, and word-level trigrams. [2.]  
Let’s use the previous paragraph as an example of natural language processing and  
cryptographic frequency analysis. Breaking the paragraph down to frequencies shows  
that the characters e, t and a are the most common ones, as seen on Figure 1. The  
occurrence of pairs of characters can be seen in Figure 2 and lastly the trigram counts  
on Figure 3. Trigram analysis is first figuring out the possible characters in groups of  
three to form words, and then finding out the sentence in groups of three to derive the  
phrases. For a virtual assistant to be considered intelligent, simple speech recognition is  
not enough. Reacting to the phrases and learning as time goes by are what makes an  
assistant good. Modern powerful personal devices allow the assistants to recognize  
speech patterns and increase accuracy. [2.]

Figure 1. Monogram frequency in speech recognition chapter’s first paragraph  
Figure 2. Bigram frequency in speech recognition chapter’s first paragraph  
Figure 3. Trigram frequency in speech recognition chapter’s first paragraph

2.2 History of voice assistants  
The first attempt at recognizing voice commands was Radio Rex in 1922, a bulldog toy  
that could respond to its name by jumping out of its house. However, it had a low utterance accuracy, and rarely reacted to the required command. [3] The first digital speech  
recognition tool was the Audrey system by Bell Laboratories in 1953. Audrey was fully  
analogic and could recognize strings of digits. [4] In 1961 IBM introduced Shoebox, a  
machine with a vocabulary of 16 words and ten digits. Shoebox could solve arithmetic  
problems on voice command and print the answers. [5] The first speech recognition system with a 1000-word vocabulary was the Harpy system in 1976. It could understand  
various speakers with over 90% speech accuracy and was developed at CarnegieMellon University. [6] Dragon Dictate, the first speech recognition product designed for  
consumers, was launched in 1990 and cost $9000. [7] In 1997 Microsoft introduced Windows Speech API on Windows 95 along with an office assistant that accepted speech  
input using Microsoft Speech Recognition Engine. [8] Apple introduced the well-known  
Siri assistant in 2011 for iOS 5. [9] In the same year IBM’s Watson system won a quiz  
competition Jeopardy, where contestants were presented with answers to which they  
had to guess the question for. [10] From this point on the competition begins between  
the large tech companies and moving onto smart speakers’ technology, as seen in Table  
1.  
Table 1. The smart speaker revolution

|  |  |
| --- | --- |
| **Voice Assistant/Smart Speaker Technology** | **Launched in** |
| Google launches Google Now | 2012 July |
| Microsoft introduces Cortana | 2013 April |
| Amazon introduces Alexa & Amazon Echo | 2014 November |
| Amazon launches Amazon Echo Dot & Amazon Tap | 2016 March |
| Google launches Google Home | 2016 November |
| LingLong launches DingDong | 2016 November |
| Samsung introduces Bixby | 2017 March |
| Cisco announces B2B Spark Assistant | 2017 November |
| Apple launches HomePod | 2018 February |

5  
Smart speakers are wireless speakerphones that can be controlled using voice commands. They possess artificial intelligence and can connect to the internet and various  
devices. These voice assistants are constantly waiting to hear the wake word, which  
triggers them to listen to a command. On a command the smart speaker can play music,  
set timers, control smart devices and even order pizza. Basically, if a speaker can do  
more than emit sound, it can be considered a smart speaker. [11.]

2.4 Problems  
The intelligent assistants do not come without complications. To be a useful assistant,  
storing information about the user is crucial, which leads to privacy issues: who has access to the user’s personal data and to the assistant in general? One issue that most  
manufacturers struggle with is controlling who has access to a smart speaker and its  
information. In the beginning of 2016 Google’s Google Home supported several users,  
but they were all linked to the same Google account, which means that the personalized  
assistant made for the user no longer remained personalized. It learned from other user’s  
as well and gave out private information. This was fixed and later Google Home could  
distinguish multiple users by voice and did not make all information available to everyone. [17] Amazon’s Alexa has had its own complications as well; when there was a story  
about a girl ordering a mansion dollhouse using Alexa, some Alexa speakers that heard  
it on the news and did not have a password to make purchases also ordered a dollhouse.  
[18] Apple’s HomePod that was released in 2018 did not learn from the competitors’  
earlier mistakes and does not recognize different voices. An assistant that cannot recognize its user is not very practical for personal requests. [19] However, even if the assistant can recognize voices, it is possible for someone to impersonate a user’s voice and  
use it to make voice commands to unlock doors or order items. The unwanted voice  
commands are the easiest way to hack the assistant; for example anyone could set an  
annoying alarm if they wanted to. There has been case where even a parrot managed  
to order gift boxes through Alexa. Televisions can also trigger smart speakers intentionally or unintentionally. [20, 13-14]  
Voice commands are not the only problem with IVA’s. An IVA can be connected to the  
internet and other devices, which leads to security vulnerabilities. Not to mention the fact  
that the assistant may record private conversations and send them to the cloud without  
the user knowing. Another problem with internet is hackers. If the wi-fi is not secured  
properly, someone may attack the assistant and change the settings. [20, 15.]  
There is also a bias problem with any kind of IVA that can learn things. The data fed to  
an assistant may not be filtered and neutral, and the level of intelligence these machines  
have does not recognize what is unfair, discriminating and unethical. The more advanced  
a system is the harder it is to predict what problems will occur in the future from the data  
9  
it has collected. Even if the creator has good intents, the AI system is only as good as  
the data it got from the user. [21.]

2.5 The future of virtual assistants  
For some it takes time to get used to a software with a human-like voice to know everything about them, while others have already fallen in love with their virtual assistant.  
Either way, the assistants are growing at a rapid speed, both in technology and userbase, as seen in Figure 4. However, experts say that current smart speakers are not  
intelligent, because they cannot fully understand their users yet. Some of them may ask  
follow-up questions after a command and understand context quite well, but it is not  
enough if the commands get complicated. The race of which one has the smartest assistant has begun, which means that as time goes by the assistants will also get more  
intelligent. [22.]  
Like smartphone apps, smart speakers have something similar called skills. For example, Amazon has a guide for developers to create skills for Alexa. The starting kit can  
convert speech to text, understand what the speaker means and even purchase products  
online. This way anyone can increase a smart speaker’s value and expand its  
knowledge. Not only is this good for the product, but it also is a way for developers to  
make money. [23.]

3.7 Conclusion  
Mycroft AI is fast to setup and simple to program with some basic understanding of UNIX  
environment and object-oriented programming. It is even easier to setup if the user does  
not want to customize it. The way skills work is easy to learn and there are not restrictions  
to what the skill can do. Starting from some very simple skills and working the way up to  
more complicated ones is a good method to determine what can be done, than proceeding straight into a complex algorithm. The speech recognition and response are fast with  
Mimic text to speech engine.  
How intelligent the assistant turns out is entirely depending on the developer’s skills, but  
the platform could be more flexible with its intents and keywords. Being forced to create  
an account to Home Mycroft and pairing it is an unpleasant step in the process. Regarding speech recognition, sometimes Mycroft has trouble understanding words and the  
user needs to speak slower and clearer than usual. Another problem is Mycroft reacting  
to its wake word way too easily and even a simple background noise may trigger it.  
All things considered; Mycroft AI is a fun platform to play around with. It allows the developer to create an assistant with the voice, personality and skills that the user prefers.  
It is a good project to work on as a hobby, especially if it is taught skills that make it learn  
something from the user and get smarter as time goes by.

4 SummarySpeech recognition and assistants have been around for a long time, but they have not  
shown much growth regarding intelligence yet. Natural language processing and reacting to phrases are not new technologies themselves, but the technology around them  
has grown rapidly, which has allowed assistants to expand their tasks. Fifty years ago,  
a voice command could only make the program solve arithmetic problems, while in the  
modern world one voice assistant can perform thousands of tasks. Only now that smart  
speakers are getting popular has voice and mood recognitions become relevant.  
What was learnt about artificial intelligence is that its definition can vary and be  
misleading in some cases. Some may refer to AI when talking about simple tasks like  
21  
decision-making or speech recognition that humans can do, while others refer to a  
machine simulating human intelligence. The definition is open because there is not one  
way to describe intelligence.  
It is a fact that the more data a virtual assistant gathers the smarter it is, but the data  
may end up in wrong hands. If the user is not careful enough anyone can access their  
personal information or even order products without them knowing. A smart speaker may  
also accidentally react to something thinking it is a wake word and record the conversation. Connection to the internet always adds risks and gives room to hackers. It is important to make sure that the wi-fi is properly secured.  
At first glance personal assistants may seem intelligent with their various skills. However,  
once the base is done, the actual skills are rather easy to program. Customizing a personal assistant may seem complicated, but with certain tools and basic programming  
knowledge it can be done in under a day

References

1 Techopedia dictionary [Internet]. Techopedia™; Intelligent Virtual Assistant;

[cited 2019 April 7]. Available from: https://www.techopedia.com/definition/31383/intelligent-virtual-assistant

2 Zwass V. Speech Recognition [Internet]. Encyclopædia Britannica Online: 2016;

[cited 2019 April 7]. Available from: https://www.britannica.com/technology/speech-recognition

3 Markowitz J. Toys That Have a Voice. Speech Technology Magazine [Internet].

2003 March [cited 2019 April 7]. Available from:

http://www.speechtechmag.com/Articles/PrintArticle.aspx?ArticleID=30031

4 ‘Audrey’ – New Electronic Device Developed by Bell Labs. Signal. 1953 Jan –

Feb; 52 [cited 2019 April 7]. Available from:

https://play.google.com/books/reader?id=IUEbAQAAMAAJ&hl=en&pg=GBS.PA52

5 IBM. IBM Shoebox. IBM Archives [Internet]. [cited 2019 April 7].

Available from: https://www.ibm.com/ibm/history/exhibits/specialprod1/specialprod1\_7.html