



jamk.fi

The design of voice user interfaces

Ella Väyrynen

Bachelor's thesis
November 2020
Business Information Technology

Jyväskylän ammattikorkeakoulu
JAMK University of Applied Sciences

Tekijä(t) Väyrynen, Ella	Julkaisun laji Opinnäytetyö, AMK	Päivämäärä Marraskuu 2020
	Sivumäärä 37	Julkaisun kieli Englanti
		Verkojulkaisulupa myönnetty: x
Työn nimi The design of voice user interfaces		
Tutkinto-ohjelma Tietojenkäsittelyn tutkinto-ohjelma		
Työn ohjaaja(t) Kiviaho, Niko		
Toimeksiantaja(t) Creoir Oy		
Tiivistelmä <p>Puhekäyttöliittymän suunnittelu on jossain mielessä samanlaista kuin perinteisten graafisten käyttöliittymien suunnittelu, mutta visuaalisuuden puuttuminen tuo uusia haasteita suunnitteluun. Tämä tutkimus antaa tapoja, joiden avulla puhekäyttöliittymiä tulisi kehittää. Verkosta löytyy runsaasti tietoa puhekäyttöliittymien suunnittelusta, joista monet ovat melko samanlaisia. Online-artikkelit ovat yleensä lyhyitä ja niissä on vain vähän tietoa, kun taas viralliset asiakirjat ovat usein pitkiä ja niistä tiedon löytäminen voi olla hankalaa. Lisäksi tutkimuksessa kosketetaan ääniohjattavien laitteiden historiaa pikaisesti.</p> <p>Tutkimuksen tarkoituksena on selvittää, mitä on otettava huomioon äänisovellusten luomisessa. Tutkimus tehdään kehitystutkimuksena kvalitatiivisilla tutkimusmenetelmillä. Tutkimuksen tarkoituksena oli luoda käsikirja, jota Creoir Oy voisi käyttää tulevien ääniohjattavien sovellusten suunnittelussa.</p> <p>Tutkimuksen ja annetun pohjamateriaalin perusteella rakennettiin käsikirja. Tämä käsikirja on tehty lyhyeksi, helposti luettavaksi ja ymmärrettäväksi. Tämän käyttöoppaan avulla on mahdollista luoda hyvin suunniteltu sovellus ääniohjattavalle laitteelle, seuraten sen ohjeita. Visuaalinen versio koko käsikirjasta lisättiin käsikirjan loppuun, mahdollistaen helpon ja nopean kokonaisuuden ymmärtämisen käsikirjan sisällöstä.</p>		
Avainsanat (asiasanat) Käyttöliittymä, puhekäyttöliittymä, suunnittelu		
Muut tiedot (Salassa pidettävät liitteet)		

Author(s) Väyrynen, Ella	Type of publication Bachelor's thesis	Date November 2020 Language of publication: English
	Number of pages 37	Permission for web publication: x
Title of publication The design of voice user interfaces		
Degree programme Business Information Technology -degree programme		
Supervisor(s) Kiviaho, Niko		
Assigned by Creoir Oy		
Abstract <p>The design for the voice user interface is in some sense similar to traditional graphical user interfaces we are used to, but the absence of visual cues brings new challenges to the mix. This thesis shows ways in which voice user interface design should be conducted. An abundance of information about voice user interface design can be found online, many of which are quite similar. The articles online tend to be short, with little information and the official documents can easily be lost into.</p> <p>The research starts with a brief look into the history of voice-controlled devices, so it can be easier to understand where we started from. It then starts with going through user interface design in a general sense, after which it goes into a deeper study of what needs to be concerned when the user interface is through voice.</p> <p>The purpose of the research was to find out what needs to be taken into accord upon creating voice applications. The research was conducted as development research using qualitative research methods. The purpose of the research was to create a manual which could be used by Creoir Oy, to design future voice applications.</p> <p>Based on the research and the original material, a manual was constructed. This manual was made to be short and simple to read and understand. With this manual, it is possible to create a well-designed voice user application following the steps in it. A visual version of the whole manual was added at the end of it for easier access for a quick look.</p>		
Keywords/tags (subjects) User interface, voice user interface, design		
Miscellaneous (Confidential information)		

Contents

1	INTRODUCTION	7
1.1	OBJECTIVES AND CONSTRAINTS OF THE THESIS	7
1.2	RESEARCH QUESTIONS	8
1.3	RESEARCH METHODS	8
2	LITTLE BIT ABOUT VOICE-ENABLED DEVICES AND APPLICATIONS	9
2.1	HISTORY OF THE DEVICES.....	9
2.2	TODAY’S MARKET	10
3	USER INTERFACE AND USER EXPERIENCE	11
4	VUI.....	13
4.1	MAKING THE EXPERIENCE EASIER FOR USERS	13
4.2	ALEXA	14
5	HANDLING OF USER EXPECTATIONS AND INTENTIONS	16
6	EXPECTING THE UNEXPECTED	18
7	THE PROCESS OF DESIGN	22
8	CREATION OF THE MANUAL.....	25
9	CONCLUSIONS	27
9.1	RESEARCH RESULTS	27
9.2	RELIABILITY OF THE WORK	28
9.3	FURTHER DEVELOPMENT	29
	SOURCES.....	30
	APPENDIX	32
APPENDIX 1.	MANUAL FOR UI AND UX DESIGN FOR AMAZON ALEXA SKILLS.....	32

Figures

Figure 1 Dr. E. A. Quade demonstrating Shoebox.....	10
Figure 2 Build page for Amazon Alexa Skills	15
Figure 3 Flowchart showing the fallback intent in the top left corner and example application start.	24

Glossary

UI	User Interface; anything that the user can interact with when using a digital product or service
UX	User Experience; means users emotions and view when using a product or service
VUI	Voice User Interface; specific naming of voice application UI
GUI	Graphical User Interface; interface where user can interact using graphical icons or audio indicators
HCI	Human-Computer Interaction; study of design and use of technology that happens between people and computers
NLP	Natural Language Processing; makes it possible for computers to understand humans in their language
Amazon	American multinational technology company
Alexa	Amazon's virtual assistant
Alexa Skills	Applications made for Amazons Alexa that can be used on the app or a web browser
ASK CLI	Alexa Skills Kit Command Line Interface; a tool to manage Alexa skills

1 Introduction

The thesis project has been made for Creoir Oy. The purpose of this thesis is to go through the concept of developing applications for smart speakers and create a manual to follow when creating an application for them.

In the theory part of the thesis, the basics of UI development are presented for voice applications and what is needed to be taken into consideration with VUI development in general, such as user expectations and how to improve user experience with error handling. In order to present the ways to design voice conversations on voice applications, examples will be given. For the purpose of the examples, my voice assistant is called “Anna”.

The output of the research is a manual, that can be used upon creating an application for Amazon Alexa Skills to ensure that the UI and UX have been designed as good as possible. Currently, there is an existing design guide made by Amazon, but I am to create a manual myself for Amazon Alexa Skills using my own findings and the existing design guide, as the one offered by Amazon is quite immense. The template of the manual mostly follows JAMK report guide. The manual will be added in as an attachment. In the research results part of the thesis, I will go through the creating process of the manual and why I decided to make it that way.

1.1 Objectives and constraints of the thesis

The purpose of the thesis is to research best practices in UI design for voice-controlled applications and what to take into consideration upon building such applications. In addition, on that research, a manual will be constructed that can be followed to build better voice-controlled software and applications for Amazon Alexa Skills. The objective is to build a simple and easily read manual.

As there exist different hardware and platforms for voice applications, I will limit myself to Amazon Alexa mainly, as it is the one, I am creating a manual for. I will mention other platforms as well to compare and because the basics of voice applications are similar cross-platform. I will not be going into hardware limitations of these devices, as it is out of the scope of this research.

1.2 Research questions

For my thesis I have selected the next research questions to help me focus my perspective:

What is VUI?

I will go through shortly what user interfaces are first and then delve into voice user interfaces and what differences it can with visual UI's. I will also touch shortly on HCI. The idea here is to acknowledge what is VUI first, before starting to build the manual.

What needs to be taken into consideration when building a user interface for voice-controlled applications?

This question will be acknowledged in the theory section in multiple chapters but will mostly be explored in the creation of the manual and the answer to the actual question in hand can be seen in the research results part of the thesis.

1.3 Research methods

As the objective of my thesis is to create a manual for voice applications the research method, I will be using is developmental research. According to Kananen (2015, 9 -

11.) it is research work that includes an integral part bringing about change or eliminating a problem. Development research itself is not a form of study, but part of a greater field of study, instructional development. Developmental research, as opposed to instructional development, has been defined as “the study of designing, developing and evaluating instructional programs, processes and products that must meet the criteria of internal consistency and effectiveness”. (Richey, Klein, & Nelson 2004, 2)

In a general sense, development means some kind of growth or progress, especially in technology development is used as a way to talk about progress. As an existing product exists, in my case its Amazons design guide, my work falls perfectly into developmental research, where I simultaneously do research and develop my manual. The manual will be an improvement from the original material. Additionally, to make this work developmental research I must not only depend on this source material but expand on it with my own research. The main idea is to confirm this previously used material and to compress its contents in an easier form of reading.

2 Little bit about voice-enabled devices and applications

2.1 History of the devices

Speech recognition for the longest time was only present in the science fiction movies and series, but in today’s world with Apple’s Siri and Amazon’s Alexa and the like, voice-controlled devices are pretty much in the customer’s hands ready to be used at a moment's notice. This previously futuristic technology started in the 1950s with Bell Laboratories “Audrey” system and then IBM’s Shoebox in 1961, that was able to recognise words. Shoebox was able to compute and print the answers to simple plus

and minus problems after getting input as numbers and command words (IBM History Exhibits, N.d.).



Figure 1 Dr. E. A. Quade demonstrating Shoebox (IBM History Exhibits, N.d.)

From the days of Shoebox, the technology has since greatly evolved, with multiple big and small companies pushing forward what is possible with voice.

2.2 Today's market

Nowadays, as much as smart speakers, such as Amazon Alexa, and integrated voice systems in cars are what is often shown and talked about in the media. According to Olson's and Kemery's (2019, 10) report the most popular way of using digital assistants with voice is through mobile phones. As a designer, it is important to take into

consideration the platform of the application. This can open different output possibilities, such as with mobile phones, there is the possibility for the addition of visual output along with voice output. This nevertheless could change in the future as people are getting more accustomed to using voice as an input method. Even though mobile is currently on the top, many reports now show smart speakers are 'eating' mobile as consumers find themselves putting down their phones to engage more with their smart speaker (Olson, C. & Kemery, K. 2019, 12). Because of this, a designer of voice applications should not solely rely on visual out-put upon designing a voice application.

Although with some systems, such as Siri that is integrated into Apple ecosystem of devices, one can control multiple devices and the input from the user's voice can act the tool with all this. This cross-platform way of control is something that upon users getting more familiar with using voice commands, manual keyboard input in simple home tasks might decrease, as using voice is much faster. Nevertheless, this connectiveness of devices will bring up questions of exclusivity, will these big companies keep their technology stuck inside their own market, such as it is with Apple, or will we be able to use these different devices from different companies accompanied together in the future? As always there are ways to make these devices work together, but with extra effort only. Meanwhile, as a designer of the applications we must concentrate on the UI and be ready to adapt when new changes come into play.

3 User interface and user experience

Usually, when using any application, you concentrate the most on the user interface (UI) of the application. It is the first thing you see. Anything else is tied to it. In some cases, the visuals of the UI itself can be enough for a customer to make an opinion of the application, either good or bad, no matter if the function of the app is working or

not. Hackos & Redish (1998, 1) define that good design happens only when the designers can understand the circumstances where users work with your design and how the users as groups collaborate to accomplish a goal.

A user interface is a bridge that brings the functions of the app to the customer in a way that the user can understand. As a designer of the UI, you must be able to understand the underlying functions of the applications to start working on the UI itself. According to Cabrera (2017, 6) designers should establish “the what and why” and leave developers to determine if it is possible and how it will be done. For that reason, the first thing to do in any UI development is to define the purpose of your application. This is the way you can start covering the question of why the application is needed. This helps the process when deciding all the features and functions of the application, as you will be able to pick out the unnecessary ones and have a clearer objective of what your UI should look like.

A big part of UI design is also knowing the user you are making the product for. Defining a user or a user group will help you to make decisions later on when thinking about personalization and how information should be given to the user itself. UI is part of a study called Human-Computer Interaction (HCI). HCI designers must consider a variety of factors: what people want and expect, what physical limitations and abilities people possess, how their perceptual and information processing systems work, and what people find enjoyable and attractive. (Galitz 2002, 4). The study of how your user group uses computers can be an integral part of your design process, as some groups of people like children, will have a different way of interacting with it than adults. Also, the language of the application should then reflect the audience. Of course, Alexa, Google Assistant and Siri on their own have a broader audience, but within those platforms, you can create applications, which this thesis is focusing on.

A good user experience (UX) with an application should feel effortless for the user, they should not be thinking of the UI that much, but instead of what the application gives to them. The UX is determined how well you are able to understand your audience upon designing the application. But not always everything can be done sitting and thinking about design, sometimes it takes some iterative actions where you test out your design on actual users you have defined to be your target audience, and then analysing the results to better your design. This can be done before any product is done by the developer and should be repeated in the design process multiple times. As new features are added in the application later on, the flow of the design should be looked over from the view of the user, as retaining your users is easier than gaining new ones.

4 VUI

4.1 Making the experience easier for users

With voice enabled devices, you must remember to think about the user experience. This time there is only a little or no visual ways to utilize, instead the experience must be made with audio capabilities alone. Things like chatbots have existed a while and people are used to seeing the chatbot box in the corner of the web page now or hearing them answer the phone e.g. when calling to a bank. The bots are created to appear like a human, to help the customer be more at ease, but sometimes this can cause problems. According to Dasgupta (Chapter 1 - Introduction to VUI) when something needs to be done, being too human can actually be restraint or an obstruction to the user. This shows especially with voice applications. Unless the application is made for conversational purposes, the extra words it takes to answer like a human might make the user impatient to wait, as unlike with visual text, it takes a while for information to be given by voice.

Using voice applications should be made easy for users. How many times you have been at the shop without a shopping list, trying to list all the things you needed aloud, and still forgotten one of the items on the list? Even with full attention, this can be difficult for some people. According to Dasgupta (2018, Chapter 2 - Principles of VUI) a human is not able to remember more than three options in a list when full attention is not given to the thing at hand. On visual user interfaces, lists can be an excellent way of presenting data, such as restaurant locations, but the same does not apply to voice user interfaces. That is why listing things should be limited. Voice applications should instead give a few options and then ask if the user wants more or new suggestion.

User: "What should I eat?"

Anna: "Nearby there is Pizzeria Romano, Pizzeria Mozzarella, Pizzeria Maria etc.?"

This example although good at showing all options to the users, is heavy to listen to and remember afterwards. Instead, the response should give an option and offer to find more options upon request. If you think of the responses as a human conversation it becomes easier to design the answers. A well-designed response would go as followed:

User: "What should I eat?"

Anna: "I heard Pizzeria Romano has high reviews! Or do you want something else?"

4.2 Alexa

As the manual will be created for Amazon Alexa Skills, it is good to mention something about the platform itself. Amazon Alexa is a voice-enabled software platform

that allows developers to create rich and engaging applications for the Amazon Echo and other voice-supported devices (Becker, 2017, Preface). These applications can be made on the developer console (Figure 2) provided by Amazon or by coding using the ASK CLI. The developer console is made in a way where you are able to make custom skills or use their templates, such as music player template, for easier building of applications. These templates can provide a great starting point for your design if your application fits in the same parameters as the template.

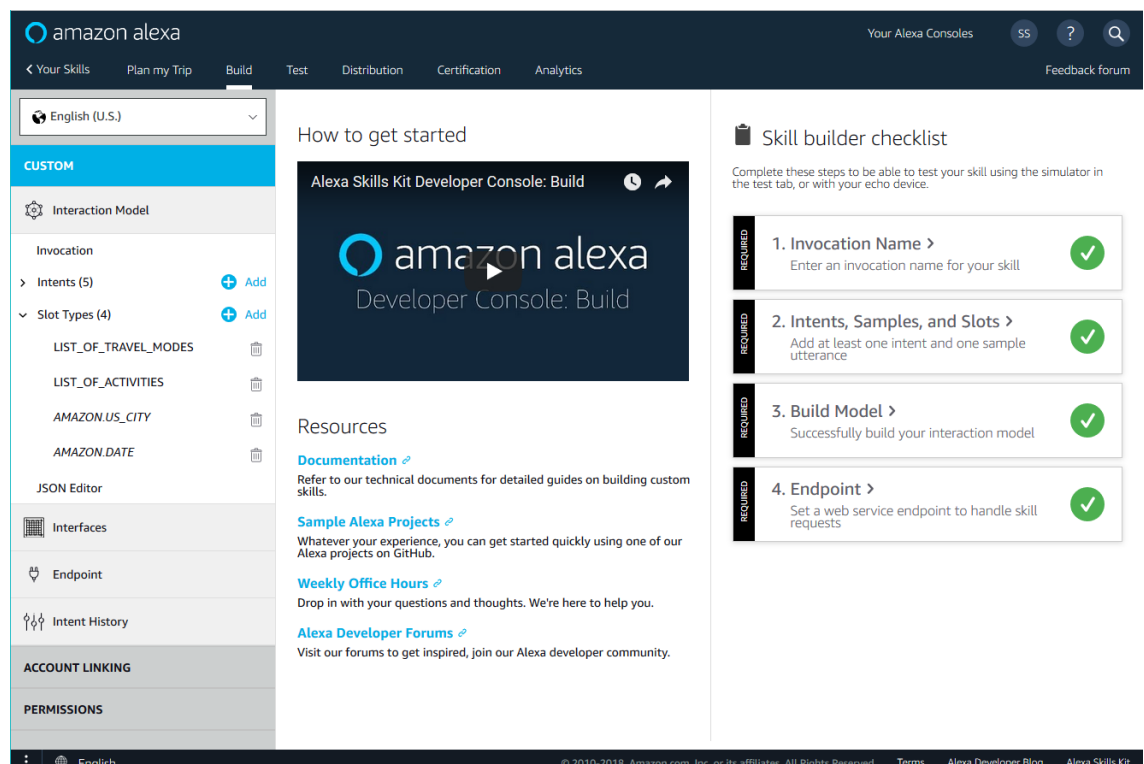


Figure 2 Build page for Amazon Alexa Skills (Manage Skills in the Developer Console. N.d.)

Amazon Alexa and their devices provide multiple ways to present output from your application. With the official products like Amazon Echo 5, you can present visual output along with your voice, this fact should be considered as it is possible the user expects something on the screen, and focusing only on the voice can hinder your application, but even then the focus should be in the design of the VUI. As you design

your responses with Alexa, it is important to remember the personality of the platform itself. As said on Amazon's tutorial on Alexa skills (Module 2: Design an Engaging Voice User Interface. N.d.), Alexa's personality is friendly, upbeat, and helpful. Your application should mirror this personality in most cases. Alexa talks to the user in first person, so your application should do the same. Exceptions can apply to certain type of applications, for example with games where responses can be done by playing audio than using Alexa's voice.

5 Handling of user expectations and intentions

The main reason for having a better understanding of people in the contexts in which they live, work, and learn is that it can help designers understand how to design interactive products that provide good user experiences or match a user's needs (Sharp, Preece & Rogers 2019, 15). A product made for a specific purpose, such as an application for a board game, has a different need from one targeted to be used in a professional setting such as a space station. Both rely on the same standards of voice user interfaces, but the approach of the output can be vastly different, depending on the context of the use, be it entertainment or formal. One might require more of a conversational output, like a human conversation, whereas the other one is used for direct instructions or answers. Additionally, recognizing things such as the user's location can help the designer with localizing the responses to the user's satisfaction. A resident in Helsinki asking about weather is most likely not looking for the weather in New York and going through the trouble of specifying their location each time is not enjoyable for the user. Ahead is an example of a voice application made to look for restaurants.

User: "I want some pizza, where should I go?"

Anna: "Closest pizzeria to you is Pizza Maria."

User: "Sounds good."

In this example, Anna uses location information to suggest the user a pizzeria. But not always are users as specific in what they want. To do so the application needs to be personal for the users. With Amazon Alexa, it is possible to recognise people by their voice, and so each user even if the device is for the whole household, can have their own user profile. On mobile, it is often assumed the user is always the same. With personalisation, you can make more assumptions and that is when the complexity of voice design increases.

User: "Anna I am starving."

Anna: "Pizza Maria has a sale on pepperoni pizzas today. Or do you want something else?"

User: "I don't want pizza today"

Anna: "What about sushi, Itomori is still open."

User: "Yes! How do I get there?"

Anna: "Walking there will take you about 6 minutes. I will open the map for you."

This is an example of how personalization works in practice. Here the voice assistant takes the intent of "starving" meaning the user is looking for food and upon finding what they want is able to guide them ahead. Using the user's previous activity or given input it can offer options that would more likely be helpful for the user. It still

asks a specifying question from the user “Or do you want something else?” to continue the conversation in case the option is given is not what they were looking for. In this example, the user rules out pizza as an option, and then receives another suggestion.

As much as the use of voice is increasing with new technology and people’s interaction with such technology becomes more commonplace in a household, one factor currently holding back some users is trust in the security of them. Misunderstood commands, unintended purchases and an Alexa listening controversy have all made it into the news (Olson, C. & Kemery, K. 2019, 24). Technology has had these hurdles before, such as with the rise of the early internet. Therefore, the makers of voice-controlled hardware and voice applications face a mission to ease the concerns of new users. Steps are being made in the right direction as Amazon has released more advanced voice purchase settings, including passcodes and voice recognition (Olson, C. & Kemery, K. 2019, 24). One other way to make the users more comfortable with the voice-controlled hardware and applications is to offer them something in value for the personal information they share. This can be done in a way of new features, various discounts, and giving clear indication where and how exactly the information is used.

6 Expecting the unexpected

When creating a design for UI be it visual or audio-based, your design might be the most polished and thought through of all time, but still, a user can misinterpret its use. One way that this can happen is through using a language or slang that was not taken into consideration. Voice control devices and programs use something called Natural Language Processing (NLP). According to Liddy E.D. (2001, 2) natural language processing can be defined as a computerized approach to analysing text that is

based on both a set of theories and a set of technologies. To put it more simply it is the means of the processing of a computer to understand human language. This combined with text-to-speech is the basis of all voice-enabled devices and programs. As a designer, it is important to look what your target audience is and then take into consideration how in your application you will handle things such as slang, accents, improper grammar, and the ambiguity of the language used. The broader the audience is you are trying to reach the more you must contemplate the language used in the application and what is to be expected or accepted as an input. Although constricting the accepted inputs to only certain ways can create an unpleasant user experience.

More complicated applications are able to remember the context and connect things a few sentences back, but that kind of level requires a lot of work on the designer's side. All this is great, but what happens when something goes wrong? When a sentence makes no sense, or some parts of the sentence weren't heard? An application cannot ever be a perfect one but should reach for a place where the user knows when they have misused the application. Worst case scenario is when a user does not even know what they have done wrong. Error handling of these situations is a huge part of the design process. These should be handled in a way where the user comes across the least amount of resistance and does not need to be the one clarifying everything. An answer can be correct, but not helpful. As an example:

User: "Can I play Spotify with this device"

Anna: "No."

This answer although correct is not inviting for the user and can leave a bad taste in users mouth as they interrogate the application what they know or don't. The same

situation could be handled better by showing alternatives or asking for more questions if it was too vague.

User: "Can I play Spotify with this device"

Anna: "You can't, but you can ask me to play a song from YouTube. What do you want to hear?"

Here the application gives an alternative and continues the conversation, prompting the user to keep using the application. Here it is important to make sure your answers sound like a human conversation where you use proper sentences and not just command names of the functions. When thinking about how long your sentences should be, consider a normal human conversation. In conversation people take turns, to reaffirm facts or ask for more information. As Yong (2016) explains in his article about human conversation, as humans we talk in turns, flipping back and forth in a regular pattern. This timing is almost universal, where each turn lasts approximately 2 seconds and the gap between the turns taken is about 200 milliseconds.

In voice application design this timing is good to remember as well. The response should be kept short, fitting around the 2-second timeframe in most cases. Additionally, if the pause between your question and the input from the user is long, it is advised to ask again, as the previous question might have been missed for various reasons. These responses and questions might arise multiple times, as the environment where these applications are used are not always perfectly quiet in the background, which is why you should in some cases ask for the users to repeat questions. This is also why it is good to add variation to your responses throughout the application. Well-thought variations of answers give life to the voice application and make it more human. These variations should also be added to what can be expected of the

user upon them asking questions, that way you are more prepared and less likely to encounter a situation where the application does not know what to answer.

In Graphical User Interfaces (GUI) somethings are easier than in voice user interfaces. One such thing is the back button. With GUI's it is easy, with a press of a button you can return to the previous page and change something you entered. With VUI's this can get a little complicated, but it is still manageable if the designer takes this into consideration. For the voice application not to break when the user decides to tell it to stop, we must handle these cases gracefully. There are many ways things can go wrong, be it connection issues, speech-to-text mistakes or simply the user saying something the computer was not expecting. Whatever the case is make sure the user knows what to do after encountering the error. If you've put in some time thinking about the possible interaction needs of the user, the best path is to tell them that the requested intent isn't supported and then turn to restate what is possible (Coates 2019, 151).

User: "Anna can you buy me tickets to the movie?"

Anna: "I don't think I can do that right now. Instead, I can get you tickets to the next show of The Wicked!"

Here it is acknowledged that the action requested is not possible, but a suitable option is given as an answer. But it is important to not do that if the same question is answered. To properly handle situations of repeat questions, the responses should change to be more specific. With this example, the application should recognise the intended action, watch movies, and next time the question is answered answer according to that.

User: "Anna can you buy me tickets to the movie?"

Anna: "I can't do that right now. What movie would you like to see?"

The third time the question is asked the application should return with a more guided way of telling what is possible with the application, as the function "buy me tickets to the movie" is not supported. These options of what the application can do should also be added to the action that triggers when the user asks for help.

7 The process of design

In this chapter, we'll go through a few things the designer can use to help themselves when designing VUI's. These should be done after the purpose, platform and the user group of the application has been defined.

Scriptwriting

Scriptwriting is a way to explore the ways a customer might interact with the application. Writing a script is a way to visualize any number of pathways that a customer might interact with in your experience (Write Out a Script with Conversational Turns. N.d.). A checklist to go through what needs to be scripted out before first testing should be done.

- Welcome message to the application for the first time and returning users
- In case the user needs help, know what to answer
- Exit and error criteria
- Check the shortest route to complete the flow of the application
- What alternative paths can be taken to get the same result

As a designer, you must think of the design flow of the conversations. Plan out your conversation flows visually, as these can be later used in the prototyping phase, to test your design. These scripts should be simple and just outline some of the turns, as

these can be then used to create storyboards. The following things should be outlined in the conversation flow:

- Keywords
- Conversation branching
- Example dialogues given to the user and the voice application

Storyboards

After some of these scripts have been constructed it is good to gather them into a storyboard. With storyboards, it can be easier to track the branching of possible conversations. Your visual representation of the conversation can be anything, below is an example made using a flowchart from Google's Dialogflow documentation page.

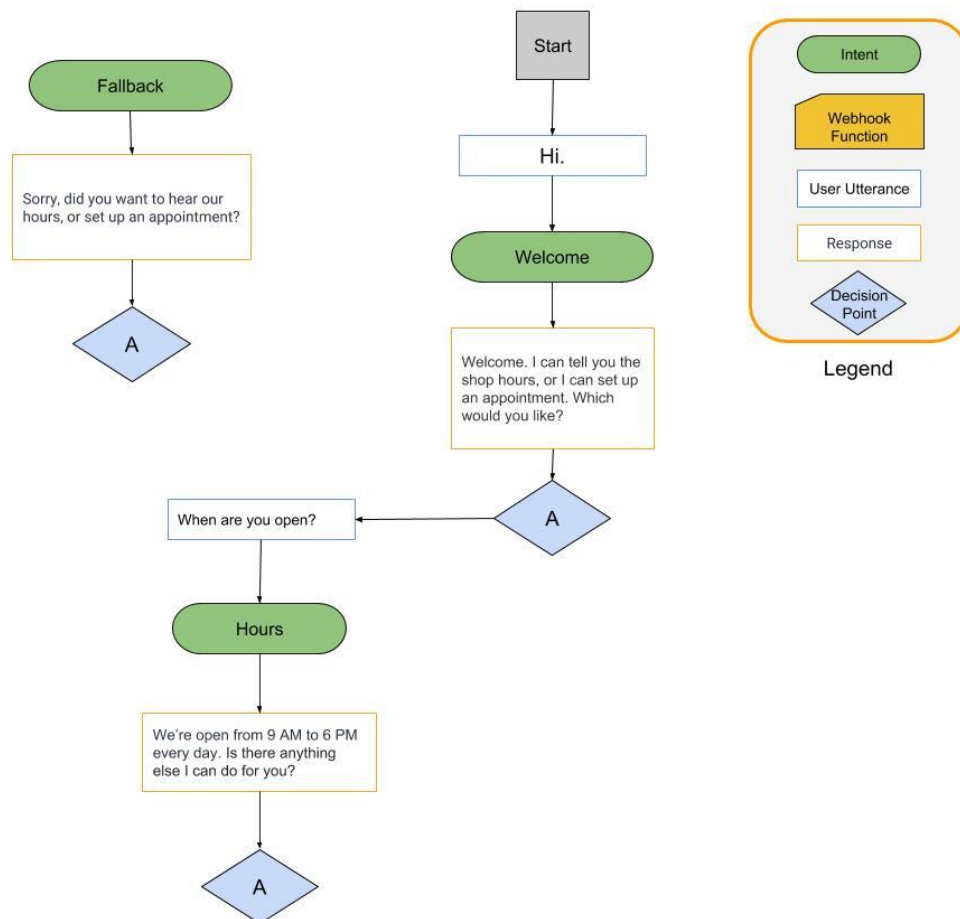


Figure 3 Flowchart showing the fallback intent in the top left corner and example application start. (Create and customize an agent N.d.)

With storyboards, you are to identify the different ways you can respond, and the user might interact with the application. The more robust you make your design, the better the experience will be for the customer using your skill (Build Your Design Artifact, N.d.). More you create these storyboards, more you will cover all the corner cases of the applications, making error situations less likely to happen.

Prototype

At some point, no amount of thinking will take you forward, and you have to test your design idea on actual people. One way is to create prototypes of your designs.

This can be simple or more defined, depending on how much time is dedicated to it. An example prototype of VUI would be a conversation with a user, where an actual person would work as the machine, working within the parameters given in the design guide, be it a flow chart or a list of responses. With this, the designers can see clear problems with UX that might arise during testing. This testing should be done at the design phase, as then there is still time to make sizable changes to the application. Additionally, many apps exist for prototyping VUI's. Amazon's Alexa Skill Builder offers a way to test out your application on your mobile phone or the browser if there is no Alexa device available otherwise.

8 Creation of the manual

Demand for the manual

There are a few manuals and guides available on the internet about Amazon Alexa Skills, for example, Amazon's official design guide. The length of these guides varies, and so does the contents. This guide was made for the assigner in mind, where I gathered the information from books, articles, and the official guide to give an updated guide on how to design a voice application. As the focus is on Alexa Skills, no other platforms like Google Assistant or Siri were brought into the manual.

Making of the manual

The design of the manual was started with compiling the table of contents. With this it was possible to organize what and where should everything be in the manual. The manual is being made as a Word-document and it was decided to keep the formatting of it close to JAMK-report style as no instructions on formatting was given by the assigner. The table of contents is ordered in the order of actions that should be

done when creating voice applications, making it easier to follow the manual when in the design process.

Table of contents

INTRODUCTION
FIRST STEPS.....
SCRIPTING
ADD VARIATIONS.....
CREATING A STORYBOARD.....
TESTING YOUR DESIGN
INTENTS AND SLOTS
VISUAL GUIDE.....

After compiling the table of contents, each chapter was filled in with a sufficient amount of information about each topic, keeping them short and simple. All these chapter names are then made into a visual guide that can be followed much easier than text.

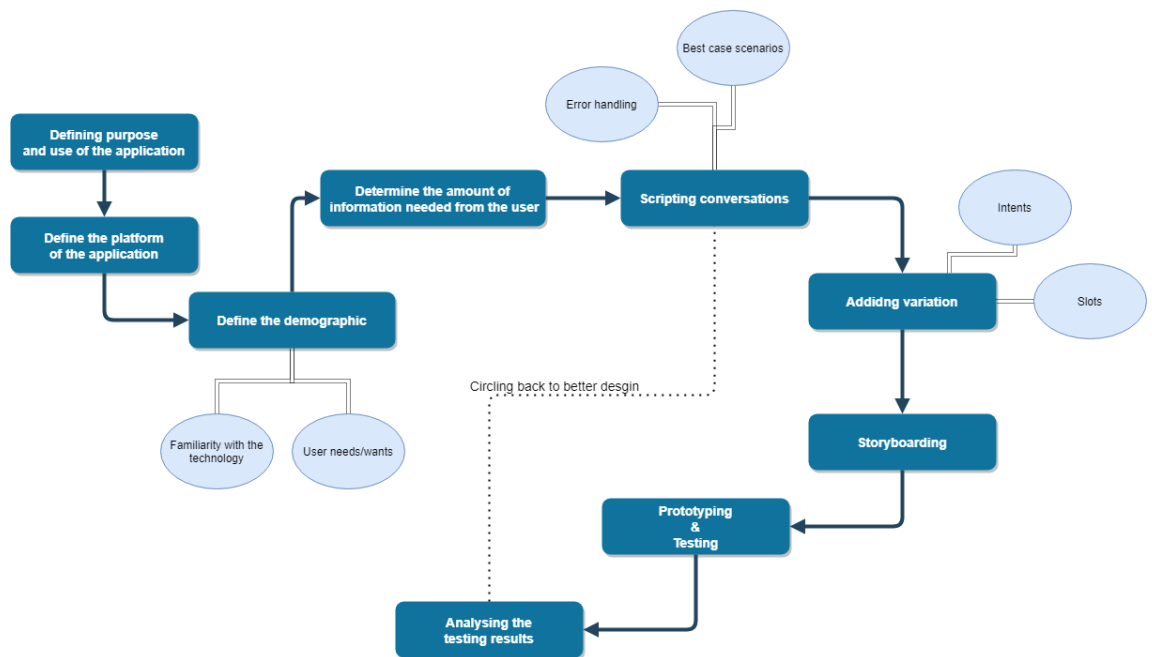


Figure 4 The visual guide from the manual

9 Conclusions

9.1 Research results

Some things should be done during the design process of voice applications in order to assure the user experience for the user is good. The objective of the research was to create a manual for designing voice of applications for Alexa. This manual was compiled with the research I did and by using Amazons design guide as the basis of the content in it.

During the research the problem that arose of what should be included in the manual. The manual was supposed to be short and easy to read as the previous material was quite difficult to read because of its larger size. I ended up making the manual quite general with the points made, keeping it simple as requested. The visual guide

made for the manual can be more closely followed for general VUI design, but even there are mentions of slots and intents, that are specific words used in Alexa Skills. The actual theory part of the thesis goes through voice application design in much more general sense than the manual, for the purpose of having the opportunity to see outside of Alexa's scope of VUI design.

As I was writing the manual, I had problems figuring out how to exactly format it. I had no previous experience in writing articles or manuals, so I decided to stick to a quite simplistic style. Alexa itself, and developing applications for it was already known to me, but the design part of it was completely new ground for me. Delving into the history of VUI and the specifics of the designing process with VUI taught me new skills, such as what I need to take into consideration when I start the design process. Before I would have jumped straight into developing, instead of looking at the design first.

I think the results of the research are sufficient and the manual created works well as a way to help design process of voice applications for Amazon Alexa. As I kept it simple and added the visual guide at the end, it should be clear what steps should be taken and in what order upon starting with a new idea. I will take this research and apply the use of the manual in future design work I do, even possibly using it outside of VUI design in graphical UI design.

9.2 Reliability of the work

Focusing on the manual, it is good to consider the reliability of the writer and therefore the reliability of the whole work. As a writer, I do not have any previous experience in writing manuals or articles. Additionally, the fact that the manual is specifically choosing Amazon Alexa Skills as the platform of voice application development to focus on should be seen as an issue if it were to be used in a more general way to

guide voice application design. Also, the reliability affects the use of Amazon's Design Guide as one of the main parts of the creation with the manual. The thesis part should have gone through the design guide in more detail and confirm its contents with other findings as a reliable source.

9.3 Further development

The manual created with this thesis can be further developed if needed by adding more updated information as it comes along the years, as the field is still fast-growing. As well as publishing the manual in article form in the future could reach more of an audience and that way also is in a more readable view. That though is up to Creoir Oy. Better visualisations could have been done for the work, by someone who has experience in a such a thing, so if anything, that is what should be next for this manual, should it be further developed.

Sources

Baker, J. 2018. Voice User Interfaces (VUI) — The Ultimate Designer’s Guide. Medium 25.11.2018. Referenced 9.10.2020. <https://medium.muz.li/voice-user-interfaces-vui-the-ultimate-designers-guide-8756cb2578a1>

Becker, K. 2017. Building voice-enabled apps with Alexa. Santa Rosa, CA: Bleeding Edge Press.

Cabrera, J. 2017. Modular Design Frameworks: A Projects-based Guide for UI/UX Designers. Berkeley, CA: Apress.

Coates, D. A. 2019. Voice Applications for Alexa and Google Assistant. Manning Publications. Referenced 18.11.2020. <https://janet.finna.fi>, Skillssoft Books.

Create and customize an agent. N.d. Documentation page in Google Cloud for Dialogflow. Referenced 18.11.2020. <https://cloud.google.com/dialogflow/es/docs/tutorials/build-an-agent/create-customize-agent>

Dasgupta, R. A. 2018. Voice User Interface Design: Moving from GUI to Mixed Modal Interaction. 1st ed. Referenced 19.11.2020. <https://janet.finna.fi>, Skillssoft Books.

Galitz, W, O. 2002. The Essential Guide to User Interface Design: an introduction to GUI design principles and techniques 2nd ed. New York: John Wiley & Sons.

Hackos, J. T., Redish, J. 1998. User and task analysis for interface design. New York: Wiley.

IBM History Exhibits. N.d. IBM Shoebox. Referenced 25.10.2020.

https://www.ibm.com/ibm/history/exhibits/specialprod1/specialprod1_7.html

Liddy, E.D. 2001. Natural Language Processing. In Encyclopedia of Library and Information Science, 2nd Ed. NY: Marcel Decker, Inc.

Manage Skills in the Developer Console. N.d. Documentation page from official Alexa Design Guide www-site. Referenced 18.11.2020. <https://developer.amazon.com/en-US/docs/alexa/devconsole/about-the-developer-console.html>

Module 2: Design an Engaging Voice User Interface. N.d. Tutorial on Amazon's site called Building an Engaging Alexa Skill. Referenced 19.11.2020. <https://developer.amazon.com/en-US/alexa/alexa-skills-kit/get-deeper/tutorials-code-samples/build-an-engaging-alexa-skill/module-2>

Olson, C., Kemery, K. 2019. Voice report. Referenced 11.10.2020. https://advertising-bing-blob.azureedge.net/blob/bingads/media/insight/whitepapers/2019/04%20apr/voice-report/bingads_2019voicereport.pdf

Sharp, H., Preece, J. & Rogers, Y. 2019. Interaction design: Beyond human-computer interaction. 5th edition. Indianapolis, Indiana: Wiley.

Write Out a Script with Conversational Turns. N.d. Documentation page from official Amazon Alexa Design Guide www-site. Referenced 18.11.2020. <https://developer.amazon.com/en-US/docs/alexa/alexa-design/script.html>

Yong, E. 2016. The Incredible Thing We Do During Conversations. The Atlantic 4.1.2016. Referenced 19.11.2020. <https://www.theatlantic.com/science/archive/2016/01/the-incredible-thing-we-do-during-conversations/422439/>

Appendix

Appendix 1. Manual for UI and UX design for Amazon Alexa Skills

Manual for UI and UX design for Amazon Alexa Skills

Table of contents

INTRODUCTION	32
FIRST STEPS.....	33
SCRIPTING.....	34
ADD VARIATIONS.....	34
CREATING A STORYBOARD	35
TESTING YOUR DESIGN	36
INTENTS AND SLOTS	36
VISUAL GUIDE	39

1. Introduction

This manual is created to help with designing of voice applications for Amazon Alexa Skills specifically. That is why hardware won't be taken into consideration. Manual

will be general so it can be followed upon creating any kind of Alexa Skill. I will be using the Amazon Alexa developer console to present some parts. At the end of the manual, there will be a visual guide of the steps introduced in text form below.

2. First steps

What purpose will the application be used for?

Defining a clear purpose for the application from the start will help you keep on track and from keeping features that might not be needed to complicate the application. This will also help customer to use the application if they know it is for a certain purpose.

What is your demographic?

Familiarity with voice applications beforehand will help you as a designer greatly, but always this question can't be answered if the customer base is wide and general. Therefore, at this point, it is recommended to figure out exactly what your demographic for the application is. If the audience you are trying to reach is specific, such as people more familiar with technology, you can make certain assumptions.

How much information is needed?

To be able to personalize answers to your users, access to their information is key. When the user uses the application for the first time, less we know about them. Some data is received from the user by Alexa itself, but some information must be collected upon the user's usage of the application. Define what is needed for the application to perform the best way possible.

3. Scripting

A way to start the design process is to start writing down a conversation. This is called scripting. This can be as simplistic as to writing the conversation on a piece of paper or visualising it on some way. A few things should be done at this point:

- Create a welcome message for new users
- Create a welcome message for returning users
- Figure out the shortest route to completion
- Create exit criteria
- Figure out what happens when the user asks for help
- Start looking into alternative ways the user can use application

4. Add variations




After you have added in your intents from your scripting is good to think about variation. The user won't always look for information the same way, which is why here you should add variations to the utterances that the user can use to arrive to the same place. Below is an example of a fact application where there are different ways

a user can ask for a fact.

Intents / GetNewFactIntent

Sample Utterances (14) ⓘ

 Bulk Edit  Export

What might a user say to invoke this intent?	
a fact	
a space fact	
tell me a fact	
tell me a space fact	
give me a fact	

< 1 – 5 of 14 > [Show All](#)

Adding variations makes your application seem more complex and human. Additionally, making it easier for users to use the application will better the user experience.

5. Creating a storyboard

With your scripts made before, you can start making storyboards out of them. Here these individual conversations are combined to a larger table, where you can see all the paths the user can take. These storyboards will help you to increase variation and it will help you on the next step.

6. Testing your design

Not everything can be thought through well, and this can be seen by testing. Testing should be done before any product has been developed, as at this point changing things is much easier. You can test your design by using different applications that exist with prototyping. Additionally, the created storyboards can be used with the testing. For example, a person can act as the voice assistant for the purpose of the testing, following the ways it can respond from the storyboard. Here the idea is to figure out the situations where the application would struggle with the answer or the response would be lacking.

7. Intents and slots

Alexa Skills developer console offers many tools to help on the design. It has a good number of built-in libraries, that will help you with design and eventually with the development of the application.

Intents

In Alexa Skills an intent means the action that happens upon the user's voice command. Alexa provides an existing built-in library, with the most commonly used actions, but adding your custom ones is also possible. These premade intents only include a few ways of arriving to that action, so it is good for the designer to edit these to fit their application. For example, the user asking for help might come in different forms. You can add sample utterances that activate the help intent. For example, you could add an utterance:

"I don't know what to do."

Upon which the help action would be activated, instead of it trying to complete another action. You can add these utterances in the field or import them in bulk using CSV file type. If any of these utterances conflict with each other, they will be marked under “Utterance conflicts”.

Intents / AMAZON.HelpIntent

Sample Utterances (0) ⓘ

[Bulk Edit](#) [Export](#)

What might a user say to invoke this intent? +

Slots

To catch what your user is looking for you need to understand the context of what they are looking for, as intents catch the action, slots catch the wanted target. With Alexa, you can define different slot types. Slot types are additional arguments that can be recognised. Alexa offers its own catalogue of pre-defined defined types. Such things include animals, cities, dates, and numbers. If possible, it is good to use these, as they are updated by Alexa.



NAME	DESCRIPTION	
<p>List Types 46 built-ins</p>	These slot types each represent a list of items. You can extend these slot types with additional values.	
AMAZON.Actor	Names of actors and actresses.	+ Add Slot Type
AMAZON.Airline	Names of a variety of airlines.	+ Add Slot Type
AMAZON.Airport	Names of a variety of airports.	+ Add Slot Type
AMAZON.Anaphor		+ Add Slot Type
AMAZON.Animal	Names of many different animals.	+ Add Slot Type
AMAZON.Artist	Full names of artists.	+ Add Slot Type
AMAZON.AT_CITY	Provides recognition for over 5,000 Austrian and world cities commonly used by speakers in Germany and Austria.	+ Add Slot Type
AMAZON.AT_REGION	Provides recognition for over 1000 geographic regions in Austria, Europe and the rest of the world commonly used by speakers in Germany or Austria.	+ Add Slot Type
AMAZON.Book	Titles of books.	+ Add Slot Type
AMAZON.City	Provides recognition for local and world cities commonly used by speakers in the locale in which the slot type is used. For example, Indian cities when used in an en-IN skill.	+ Add Slot Type

If the data you are targeting is something specific and does not fit the premade types, you are able to make your own custom slot type. These could be for example names of cards in a game application you are creating.

Slot Types / Add Slot Type / Spells

Custom slot types with values define a representative list of possible values, IDs and synonyms.

Slot Values (0) ⓘ

 Bulk Edit  Export

Search



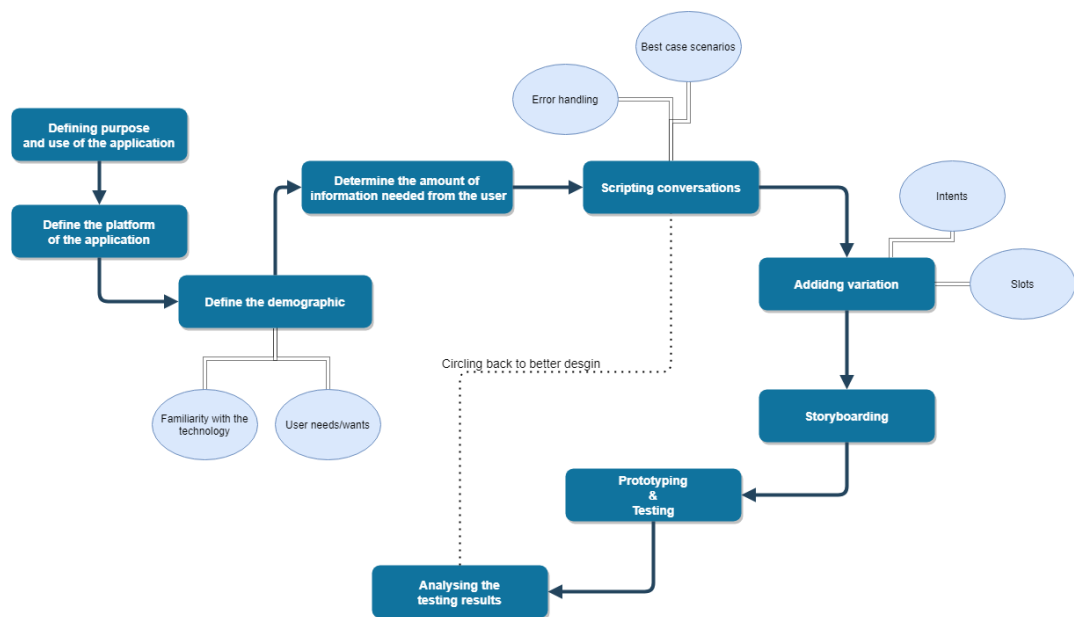
Enter a new value for this slot type



You can add these values in the field or import them in bulk using CSV file type.

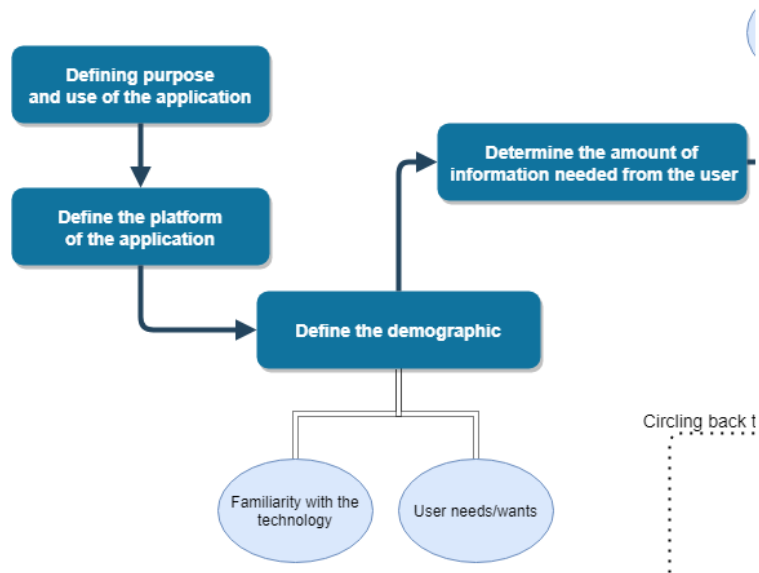
9. Visual guide

This visual guide is made so that it can be followed step-by-step when designing for voice applications. It has been made to be more general, including the part of defining the platform. If following this manual, that part will be Amazon Alexa.



Closer look at parts of the visual guide:

Part 1.



Part 2.

