FH Salzburg
MultiMediaTechnology

Voice interaction with fitness application keeping technology and health in balance

Bachelor Thesis 2

Author: Theresa Höck
Advisor: DI Michael Domhardt

Salzburg, Austria, 06.08.2018

Affidavit

I herewith declare on oath that I wrote the present thesis without the help of third persons and without using any other sources and means listed herein; I further declare that I observed the guidelines for scientific work in the quotation of all unprinted sources, printed literature and phrases and concepts taken either word for word or according to meaning from the Internet and that I referenced all sources accordingly.

This thesis has not been submitted as an exam paper of identical or similar form, either in Austria or abroad and corresponds to the paper graded by the assessors.

Date *Signature*

First Name *Last Name*

Kurzfassung

Im Rahmen dieser Bachelorarbeit wurde eine sprachbasierte Fitnessanwendung entwickelt. Diese Fitnessanwendung wurde als Amazon Alexa Skill umgesetzt. Sie soll motivieren mehr körperliche Aktivität in das tägliche Leben eines Menschen einzubeziehen, um die Überbeanspruchung der Mobiltelefone auszugleichen, oder sogar zu verringern. Dies soll dazu führen die Zeit die man mit der Familie oder Freunden, oder die Zeit in der man alltägliche Aufgaben löst effektiver zu nutzen. Mit Hilfe eines Experiemts und zweier Fragebögen stellte sich heraus, dass der Alexa Skill eine mögliche Lösung ist, das Gleichgewicht zwischen Technologie und Gesundheit zu gewährleisten.

Abstract

As part of this bachelor thesis, a voice-base fitness application was developed. The fitness application was created as an Amazon Alexa skill. The skill primarily aims at motivating users to include more physical activities in their daily life while helping to balance out the overuse of their mobile phones. By extension, it eases the reduction of mobile phone consumption, which can improve ones ability to concentrate, become more aware of social opportunities such as meeting with friends and family, and just being more aware of the surrounding environment. An experiment with a developed prototype and the results of two questionnaires show that the Alexa Fitness Skill is a promising way of balancing out technology and health.

Contents

1 Introduction

Regular physical activity is a key determinant of a health-conscious lifestyle and is therefore a key factor in promoting good health (Susanne Wolf, Bernhard Schwarz, and Bernhard Rupp 2013). The rapid development in the technology market and the use of smartphones, tablets and game consoles appears to be at odds with health-conscious behaviour. By the year of 2017, 5.62 millions monthly active smartphones where registered in Austria[1]. The average usage time of these smartphones is more than three hours a day in the country[2]. This makes smartphones usage an integral and important part of our conscious daily activities.

These mobiles devise which were once used for merely emitting and receiving phones call and text messages (SMS) have now gone far beyond those basic abilities. A major part of smartphone use as of today is concentrated on social interactions. A smartphone is a mean to connect with the globe and and it's culture. People can make use of these smartphones abilities to generate income, for example via social media apps like Instagram. They can promote themselves and their services for instance as "influencer" merely by publicly sharing media files (images and videos). Others get hold of romantic dates through dating platforms like Tinder. One may used online marketplace apps to sell goods and other services. Users are no longer obliged to download and save music on their device due to streaming apps such as Spotify. The same works with streaming videos content from services such as Amazon Prime on mobile phone. But more importantly for the sake of this work, smartphones are used as fitness trackers by means of tracking a user movement (steps walked, distance covered, speed per km, height climbed, and more) and fitness guides, for instance by providing nutrition and fitness plans.

Andrews et al. (2015) compared 23 american smartphone users over two weeks long with self-reported estimates of the Mobile Phone Problem User Scale. Their results in this study show that the participants used their phone around 85 times a day and spent more than 5 hours in front of their screen. Ward et al. (2017) conducted two experiments were they tested the hypothesis that the mere presence of one's own smart phone reduces available cognitive capacity. This phenomena is called "brain drain". They wanted people to put their devices nearby in sight (desk), nearby out of sight (bag), or in a separate room. Their results shows that the mere presence of the these devices has an impact on the working memory capacity and the fluid intelligence of participants. From their work, we can assert that smartphones participate in great importance to concentration problems and feeling unbalanced.

Booth, Roberts, and Laye (2012) recommends the inclusion of sports in one's lifestyle to compensate this mental and physical imbalance. Sport can produce important lifelong benefits to the physical and mental health (Booth, Roberts, and Laye 2012). The long standing and ongoing habits of sitting to watch television, work on computers or playing video games not involving much physical movement, have generally been observed as primary causes of mild to life threatening health issues. According to Garber et al. (2011), this lack of daily activity

1. https://www.statista.com/statistics/565947/predicted-number-of-smartphone-users-eu-austria/, 2018-07-04
2. https://www.mindtake.com/de/press-release/%C3%B6sterreicher-nutzen-mobiltelefon-mehr-als-drei-stunden-pro-tag, 2018-07-30

leads to chronic diseases. Booth, Roberts, and Laye (2012) describe these chronic diseases as "major killers in the modern era" while insisting that "physical inactivity is a primary cause of most chronic diseases". They further argue that 35 diseases, including heart failures, strokes, diabetes and mental conditions such as depression and anxiety could be prevented by doing more physical activities. The World Health Organisation (WHO) recommends that adults between 18-64 years old should do between 75-150 minutes of physical activity throughout the week depending on the intensity level and type of the activity [3].

The work described in this paper presents a voice-base fitness application developed as extension (a skill) of Amazon's voice based interaction device Alexa. This Alexa skill at term should:

- motivate users to include more physical activities in their daily life,
- help reducing the direct use of their mobile phone,
- improve concentration ability, social interactions and awareness of their surrounding.

An experiment with a developed prototype and two questionnaires should provide evidence as to whether the Alexa skill may be supportive to reach this goal.

2 Theoretical Background

The following section starts with an insight in the current development and relevance of fitness applications in the Austrian market. Then it explains the importance of conversation for humans. Finally, selected papers are exhibited to show in which areas voice interaction is a relevant topic.

2.1 Fitness application

As of today, multiple fitness mobile applications have been developed with the goal of encouraging people to start daily physical exercises while supporting them along the way. In neighbouring Germany, three main themes for these fitness apps use for former, current and potential users are: motivation (47%), physical optimization through muscle building or fat burning (36%) and healthier diet (33%) [4]. An investigation in the form of market research conducted by Statista (a statistic portal) for the potential of fitness applications in Austria shows that the two main technologies for fitness support are wearables and mobile applications. In the fitness applications segment of this research, sales in Austria amounted to 18 million euros. According to predictions, sales will rise to 19.5 millions euros by 2022. Currently there are about 800.000 fitness apps users. The number is estimated to rise to 1.2 million users by 2022 which is equal to a penetration rate of 13.4%. In comparison, the use of wearables (which already generates 33.2 million Euros

3. http://www.who.int/dietphysicalactivity/factsheet_adults/en/, 2018-07-04
4. http://www.marktmeinungmensch.de/studien/nutzung-von-gesundheitsapps-auf-dem-smartphone-in-/, 2018-07-04

annual revenues in Austria) should see a much slower increase in penetration by a value on only 0.5% between 2018 to 2022[5].

The establishment and relevance of fitness apps in the future is therefore becoming increasingly clear. One of those fitness apps is Freeletics. It has over 14 million registered users in more than 160 countries and available in 8 different languages. Initially they only provided introduction, guidance and tips on the correct techniques and approaches for single and isolated exercises for training the body and the mind. The founders came to realise that people not only possess willpower, but can build it up like a muscle. On this basis and with the aim to create a real value for all people in the world, they have developed a concept with Freeletics, which demands and promotes physical as well as psychological strength in humans. Training with Freeletics Bodyweight is based on short and high-intensity workouts that address the key components of physical fitness: strength, stamina, and muscular endurance. To make training even more efficient, the program includes additional elements such as functional training and HIIT (high intensity interval training)[6].

In the example of Freeletics Bodyweight a user goes through the weekly training plan by executing approximately 6 steps in a daily training, which takes between 30 to 40 minutes depending on the user's fitness level. A user can evaluate progress between each step: was it exhausting? Could more be done if they used better techniques? To accomplish each step, the user is required to interact with the mobile device to enter data via touch input. With this data the Training Coach (it can be virtual assistant or a human off-location) is updated.

These tactile interactions, like watching the exercises or giving feedback, with the device creates interruption which can lead to slow progress or abandon of the training. The user might be lead to check notifications from other apps and to procrastinate while entering the data via the mobile phone. This easily breaks the progress and makes the training harder to follow and finish. The users concentration will probably lower, hence exposing them to the aforementioned "brain drain" (Ward et al. 2017).

2.2 Voice Interaction

Why is conversation so easy for us? Garrod and Pickering (2004) are convinced that humans are designed for dialogue rather than monologue. They argue that it is easier for us to talk to someone next to us than holding a 15 minutes speech, a monologue, in front of an audience. The reason for that is called "interactive alignment". Within interactive alignment the interlocutors distribute the processing load between them. Each reuses information computed by the other. Kuhl (2007) explains that, as children, we learn our mother tongue primary through social interaction, so primary through dialogue. However our ability to learn to speak through conversations from an early age does not necessarily go without problems and creates other non intended problems. Kuhl (2007) continues by explaining that one of the problems posed by dialogue is for instance

5. https://de.statista.com/outlook/313/128/fitness/oesterreich#market-users, 2018-07-04

6. https://www.freeletics.com/en/press/wp-content/uploads/sites/24/2017/04/Freeletics_PressKit_DE_06.04.2017_web.pdf, 2018-07-04

that we are not able to plan far ahead what we will say. If someone asks us an unexpected question, we might be surprised, thus we may not respond on the spot. Furthermore, Kuhl (2007) adds that there is the problem of making what you say appropriate to the addressee. Appropriate in the sense of knowing if the addressee knows whom you are referencing to, for instance when talking about your neighbour named Bill. Probably the addressee knows several Bills or they can not remember all the neighbours names and thinks about someone else with the name Bill. In colloquial speech, this phenomenon is known as talking past each other. Sometimes the interlocutors realises the mistake within the discussion and it can be immediately corrected on the spot. In addition, Kuhl (2007) supposes,, conversation presents a whole range of interface problems. So as knowing when it is socially appropriate to come up with the answer, in average half a second before the addressee finished speaking. This means that we are planning what we are going to say while still listening to the addressee, also known as task-switching between listening and speaking. Our society developed rules on how to communicate in a natural way. Still, our language is the most natural flow, convenient and efficient way of information-sharing. According to Liu (2010), 75% of human communication is accomplished with voice. Also in Human-Computer-Interaction (HCI) the voice plays an important role in the natural user interface (NUI) and several researches have taken and continue to take place in order to integrate and turn over humanity's communication rules into technology. The upcoming paragraphs contains three cases where voice interaction with technology is or could become supportive for humanity.

Cronin and Doherty (2018) created a review about the relevance of voice interaction in hospitals. Their motivation was to keep non-sterile interaction with keyboards and mice low. According to them, it is difficult to keep computers and its peripherals sterile. Touch based interactions make them more likely to be contaminated. According to the WHO, contaminated operational utensils mostly lead to health care associated infections. In Europe for example, there are 16,000,000 extra days spent in hospitals, 37,000 attributable deaths and 7 Billion costs of euros every year [7]. Touch-less interaction methods have been extracted from different papers together with the used technology. Furthermore efficiency, effectiveness and user satisfaction of the systems have been taken into account for its real and successful applicability in the future. In the results they affirm that (with the example of gesture control implemented with Microsoft's Kinect) voice control with voice recognition systems and eye tracking could play a major role in surgeries. Suelze et al. (2013) installed a Kinect from Microsoft to develop a protocol allowing real-time interaction between the interventional cardiologist and the imaging software of a fluoroscopic image viewer. These main interactions consisted out of single arm gestures to limit disturbance with invasive procedures. If the cardiologist is unable to use his hands, voice commands were activated to execute the required actions. The resulting natural user interface influenced the efficiency positively in the catheterisation lab.

Another area for touch-less interaction is the automobile industry (Simmons, Caird, and Steel 2017). Distractions, like sending and reading text messages, dialling numbers, changing music on the mobile phone or just interacting with the automobile's display and other activities are the reason for 32,3% of Austrians deadly car accidents in 2017[8]. Regan and Strayer (2014) defines

7. http://apps.who.int/iris/bitstream/10665/80135/1/9789241501507_eng.pdf, 2018-03-26

8. http://www.bmi.gv.at/202/Verkehrsangelegenheiten/unfallstatistik_

distraction as “the diversion of attention away from activities critical for safe driving toward a competing activity”. Hands-free technologies allow drivers to do most of these tasks with letting the eyes on the streets to be able to react fast in difficult situations. Simmons, Caird, and Steel (2017) performed a meta-analysis to summarise the effects of using a voice-recognition (V-R) system while driving and performing various tasks. It showed barely effects on speed, speed variability, headway and headway variability compared to interacting without that system. All in all, V-R did not produce result as good as expected. There was too less verbal feedback from the system or an extra feedback on a display occurred, hence enabling distraction in the drivers. According to Simmons, Caird, and Steel (2017) it is necessary to be more aware of the importance of verbal feedback and the further design of the interaction methods.

Apart from the aforementioned adapted V-R systems to various technologies, there exist off-the-shelf voice-controlled intelligent personal assistants (IPAs). These are becoming increasingly common. For example, in recent years, Amazon and Google introduced smart speakers and home assistants to the market[9]. The interfaces installed in those smart speakers can be called speech-based natural user interfaces (NUI). The NUI focuses on human abilities like touch, vision, voice, motion and higher cognitive functions such as expression, perception and recall (Liu 2010). Apple and Microsoft respectively produced their own NUIs, the earlier naming theirs Siri and the latter Cortana. However, Siri and Cortana were not especially designed for users with disabilities. A home assistant such as an Amazon’s Echo (also known as Alexa) could become a more convenient helper in the daily live of physically and visually handicapped people. Tasks like querying the weather, reading out or adding things to a to-do list, shopping list or even online shopping itself, listening to the latest news or just asking random questions like: “What’s the capital of Australia?”, might become less challenging for this user group. Pradhan, Mehta, and Findlater (2018) conducted two studies to get to know more about how people with disabilities use IPAs. 37.9% of the participants had visual impairment, 30.6% motor or mobility impairment and 13.6% speech impairment. Blind participants for instance faced great difficulties in finding or discovering functionalities e.g. Skill in the Alexa Skills KID (ASK). Even though the NUI was not personalised for them, it was surprising that for instance 7 participants with speech impairments used the smart speaker for example as speech therapy. It helped them to talk slowly, clearly and loudly. In addition it was also seen as a support for caregivers. In other words it had reduced some caregiver demands, such as reading books, playing music, controlling the home environment or answering simple questions. Those surprising findings could have a huge impact on how NUI are designed in the future.

2.3 Commercially available voice-controlled IPAs

Those off-the-shelf voice-controlled intelligent personal assistants (IPA) have been released by different companies. An IPA is a software agent used to support users to carry out several daily actions (Santos et al. 2016). They come with smart phones or other devices that can play and

vorjahr.aspx, 2018-03-26

9. https://de.statista.com/infografik/12369/weltweiten-absatz-von-smart-speakern/, 2018-06-22

record audio. Since they underlay different systems they have different skills, different voices and they have different user interface designs. In the following section the most popular IPAs from Google, Microsoft, Apple and Amazon are presented.

The Google Assistant is a voice-controlled IPA. It helps users to find, organise and get things done in their daily lives. The smart speaker linked to it is called Google Home[10].Google Assistant's predecessor is Google Now. By developing apps, new actions can be build to enable users to get things done with different products and services. Users interact with apps for the Assistant through a conversation. Not in the same way as with traditional mobile or desktop apps. There are natural-sounding back and forth exchanges, and not traditional, computer-centric paradigms. Some of the promotional commands for Google Assistant include: "Ok Google, Remind me to pick up a birthday card", "Ok Google, Book me a table for 6 at Quartino for 8:30", "Ok Google, Who invented sushi?" [11].

Microsoft Cortana is the personal digital assistant of Microsoft for Windows 10[12]. Cortana searches the web, can find specific things on a windows devices, it can set reminders or dial phone numbers. It recognises natural language and can also answer. With a key combination it can be invoked, then certain commands could be fired by the user: "What's the weather like?", "Call Sarah", "What is five miles in kilometres?". Microsoft has not released an own speaker device so far.

Apple Siri is the personal assistant for Apple devices. First released in 2010, it is the oldest of the most used intelligent personal assistants[13]. The most common uses for Siri are web browsing and dictation. Currently, Siri is under a redesign process to add new functionalities and adapt it to new devices, so as the new HomePod that was recently released by Apple[14]. According to Apple, the HomePod has the highest-fidelity sound and brings Apple Music and Siri together. Furthermore it is a home assistant to control a smart home. Some of the promotional commands for Siri include: "Show my photos from Utah last August", "What movies are playing today?", "Find videos I took at Iva's birthday party", "Text Pete 'See you soon smiley exclamation point'", "Read my latest email", "Find a table for four tonight in Chicago"[15].

Alexa is a voice-controlled IPA developed by Amazon[16]. It is linked to the smart speaker called Amazon Echo. According to Amazon, Alexa offers tens of thousands of skills by now and capabilities to create a more personalised experience[17]. It can be used hands-free and via voice-control. Alexa is able to play music, read and write to do lists, streaming podcasts, playing audiobooks, and providing weather, traffic, and other real time information. Alexa can also control several smart devices using itself as a home automation hub. A common command for Alexa would be: "Alexa, set a timer for 10 min". This command could additionally be invoked

10. https://assistant.google.com/#?modal_active=none, 2018-06-22
11. https://developers.google.com/actions/extending-the-assistant, 2018-03-26
12. https://www.microsoft.com/en-us/download/details.aspx?id=47715, 2018-03-26
13. https://voicebot.ai/2017/07/14/timeline-voice-assistants-short-history-voice-revolution/, 2018-07-31
14. https://www.apple.com/homepod/, 2018-03-26
15. https://www.apple.com/ios/siri/, 2018-07-31
16. https://www.amazon.com/gp/product/B0749WVS7J/ref=ods_ac_dp_dr_ps, 2018-03-26
17. https://developer.amazon.com/alexa, 2018-07-31

by a mobile phone with the Alexa app, or other Amazon compatible devices[18].

2.4 Comparison of IPAs

Gustavo, Quesada, and Guerrero (2018) evaluated the four most popular assistants, Google Assistant, Cortana, Siri and Alexa, to become aware of the individual strengths and weaknesses in their natural feeling and correctness. A group of 8 people rated each response on every device. One of the researchers interacted with the devices by using the same voice tone and pace. The participants listened to the answers of the devices. They rated the responses with a 5-point Likert scale. Their focus was set on two categories: natural feeling and correctness. The statistical analysis did not find significant differences between the systems. That means, according to Gustavo, Quesada, and Guerrero (2018) there can not be set any preferences for any of the systems so far. Still the results show that Google Assistant is the most natural personal assistant, but suffered because it did not support some of the features evaluated in the research. On the contrary, Siri is the most correct personal assistant and the least natural.

Independently of the aforementioned evaluation, Consumer Intelligence Research Partners (CIRP) released an analysis on November 6th, 2017, of the results from its research on home automation devices, including Amazon Echo Dot from Amazon. CIRP analysis demonstrates that within the US installed base of devices, Amazon Echo has 73%, or 20 million units, while Google Home has 27%, or 7 million units[19]. For that reason it can be assumed that Amazon Echo is better known, used and integrated to the daily routine within households. The CIRP bases its findings on its survey of 500 US owners of Amazon Echo or Google Home. This is the reason why an Amazon Echo was chosen for the development of this prototype.

3 Development of the Amazon Echo Prototype

The prototype for this thesis is an Amazon Alexa skill. It should enable anybody, who is able to speak fluently German and has no physical disabilities, to rise the amount of activities throughout the daily life. Furthermore this skill is a possibility to decrease the general mobile phone usage, in order to gain more time for family and friends, but also to increase the ability for concentration.

The study, which will be carried out as part of this work, will use a separate developed project for the IPA Amazon Alexa. Therefore a fitness skill called “Theresas Fitness App” (TFA) was created. At the time of the constitution, some german-language fitness applications were published. So as 7 “Minuten Workout”[20], or ”5-Minuten Plank Training”[21]. In those applications

18. https://developer.amazon.com/alexa-skills-kit, 2018-03-26
19. https://www.voicebot.ai/wp-content/uploads/2017/11/cirp-news-release-2017-11-06-echo-home.pdf, 2018-03-29
20. https://www.amazon.de/gp/product/B0754C735Z?ref=skillrw_dsk_pnps_dp_3, 2018-06-29
21. https://www.amazon.de/cubic-ai-5-Minuten-Plank-Training/dp/B06XHTCB3Z, 2018-06-29

the user is lead through the workout challenges consisting of a range of exercises. The goal of the TFA is to offer the user a possibility to select a specific body part he wants to train and to give feedback on their training performance and intensity. Possible body parts are legs, belly, arms, back and buttocks. For every body part, four different exercises were chosen. They were picked by personal knowledge gained from sports club coaches and from the aforementioned fitness app called "Freeletics Bodyweight". This study did not take place under the supervision of fitness experts, so there is no guarantee for absolute accuracy of the exercises and workouts. The focus for this study lays on the voice interaction. After carrying out the four exercises users evaluate themselves. The feedback is structured into two questions. Firstly, users has to evaluate whether they could perform the exercises with the correct technique and secondly they have to rate how intense the workout was for them. The entered data, the body part and the feedback, is stored in an external database (DB) in order to provide persistence within the skill. Due to time constraints, only the main functions have been implemented, but in the future the skill could be extended based on the data collected. The source code to the skill can be found on a Gitlab repository[22]. Within this study the prototype was not published in the Alexa Skills Store. For the user testing it was executed by a proxy server and a local development console.

3.1 Skill Configuration and Creation

In order to develop a new skill for Alexa, it is required to create a free amazon developer account[23]. From there, the first step is to sign in to the Amazon Web Service (AWS). As the name already implies AWS offers different services, beginning with machine learning systems to train models, space for data storage, engines to analyse data and other services that enable a developer to manage and work with big sets of data. Within this work it will be taken advantage, as it can be seen in figure 1, of AWS Lambda, Amazon DynamoDB and Amazon CloudWatch Logs. AWS Lambda is a service that makes it possible to run code without the need of an extra server. Amazon only charges for the time when the code is running. Amazon DynamoDB is a NoSQL database. It is highly scaleable. Furthermore, it features integrated security, data backup and recovery. Amazon CloudWatch logs is a service to monitor, store and access logs from other services.

Hendrickson et al. (2016) asserts that working with Lambda functions has some advantages compared to conventional server-based approaches. For instance, there is no need for sever management. The handlers within the functions from different developers share the same pools of servers that are managed by the cloud providers. Those handlers are mostly written in JavaScript and Python, open source programming languages. The runtime environment can be shared. Finally, if the application is scaling up rapidly AWS Lambda can still manage the workload.

AWS offers the author to create a Lambda function from scratch, or to choose from one of many preconfigured examples, witch are templates to give developers an overview about the most important components within functions. AWS Lambda executes the code only when needed and

22. `https://gitlab.mediacube.at/fhs38521/bac2-alexa-skill`, 2018-06-29
23. `https://developer.amazon.com/alexa-skills-kit/alexa-skill-quick-start-tutorial`, 2018-04-12

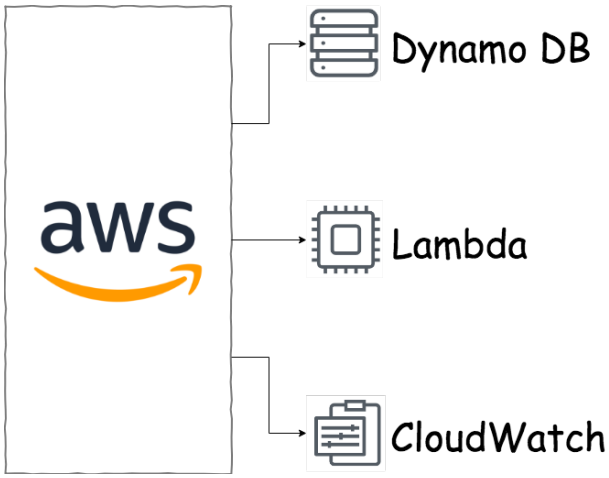

Figure 1: Amazon AWS Services

scales automatically, from a few requests per day to thousands per second. There are no charges by Amazon when the code is not running, only the compute time that is consumed is paid[24].

It is required to add a trigger in the Lambda Management Console (LMC) to the considered lambda function to ensure it is referenced with the new skill. In this case, the trigger is the Alexa Skills Kit (ASK). Secondly the skill ID, which is always available in the Alexa Skills Kit Developer Console (ASKDC), needs to be set in the LMC under the section of the added ASK. A skill ID, for instance 'amzn1.ask.skill.019fe2b6-9106-4534-b9a9-88de88959209' is generated and shown after the skill was created in the ASKDC [25]. Thus if the skill is invoked by a user, AWS Lambda knows what lambda function needs to be executed.

3.1.1 Alexa Skills Kid

The Alexa Skills Kid (ASK) is a Spoken Language Understanding (SLU) Software Development Kit (SDK) that makes it possible for Amazon developers to increase the range of skills of Amazon's Alexa. Through the ASK, over 25.000 skills have been developed by the end of the year 2017[26]. It has been created, inter alia, as a support for third party developers. In other words a skill developer does not necessarily need to work for Amazon to be able to develop a new skill. In order to understand the specific interaction flow of Alexa, Kumar et al. (2017) described and summarised the most important aspects of ASK SLU Service. As demonstrated in figure 2, first, a user can start the conversation with Alexa with a wake word. The standard wake words for Alexa are Echo, Computer, Alexa and Amazon. Second, the Echo starts to record the users intent. With the automatic speech recognition (ASR) the spoken words are converted into text[27]. Third, the users intent is classified and identified. In accord with the schema, which is designed by the

24. `https://docs.aws.amazon.com/lambda/latest/dg/welcome.html`, 2018-06-29
25. `https://developer.amazon.com/alexa`, 2018-04-12
26. `https://developer.amazon.com/de/alexa-skills-kit`, 2018-06-22
27. `https://developer.amazon.com/alexa-skills-kit/asr`, 2018-04-03

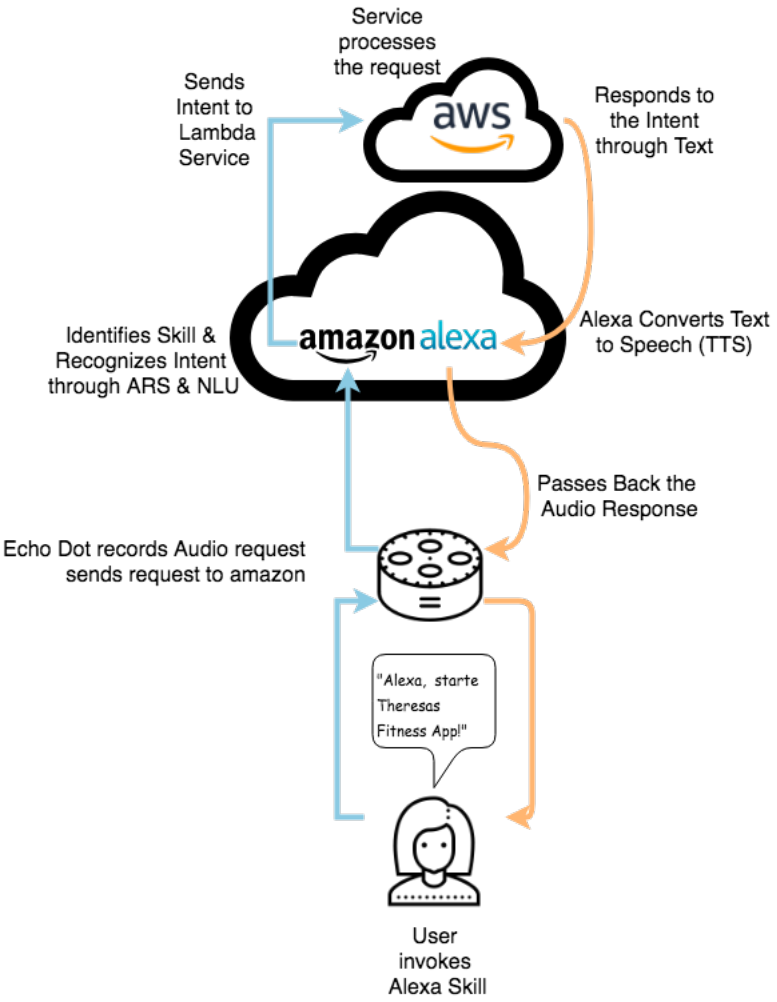

Figure 2: Alexa architecture after Kumar et al. (2017)

developer, the slots within the intent get filled. Besides the skill definition, a developer is also responsible for developing a web service. The web service receives a structured JSON request from Alexa including the intent. Then the Web Service is responsible for responding by text, sound stream or graphical response to Alexa. Alexa is able to convert the input and passes the results to the user.

3.1.2 Alexa Skills Kit Developer Console (ASKDC)

The ASKDC is beneficial for a developer in the first development flow. It leads the developer literally through the creation process of the skill step by step. First as previously mentioned, a skill needs to be invoked by the user. In this case, the prototype is called and invoked like this: “Alexa, starte Theresas Fitness App”. So, ’Theresas Fitness App’ is the Skill Invocation Name (SIN) and “Alexa” is the wake word. According to the documentation the SIN does not need to be unique across Alexa. Second, the intents can be added. By default, Amazon already

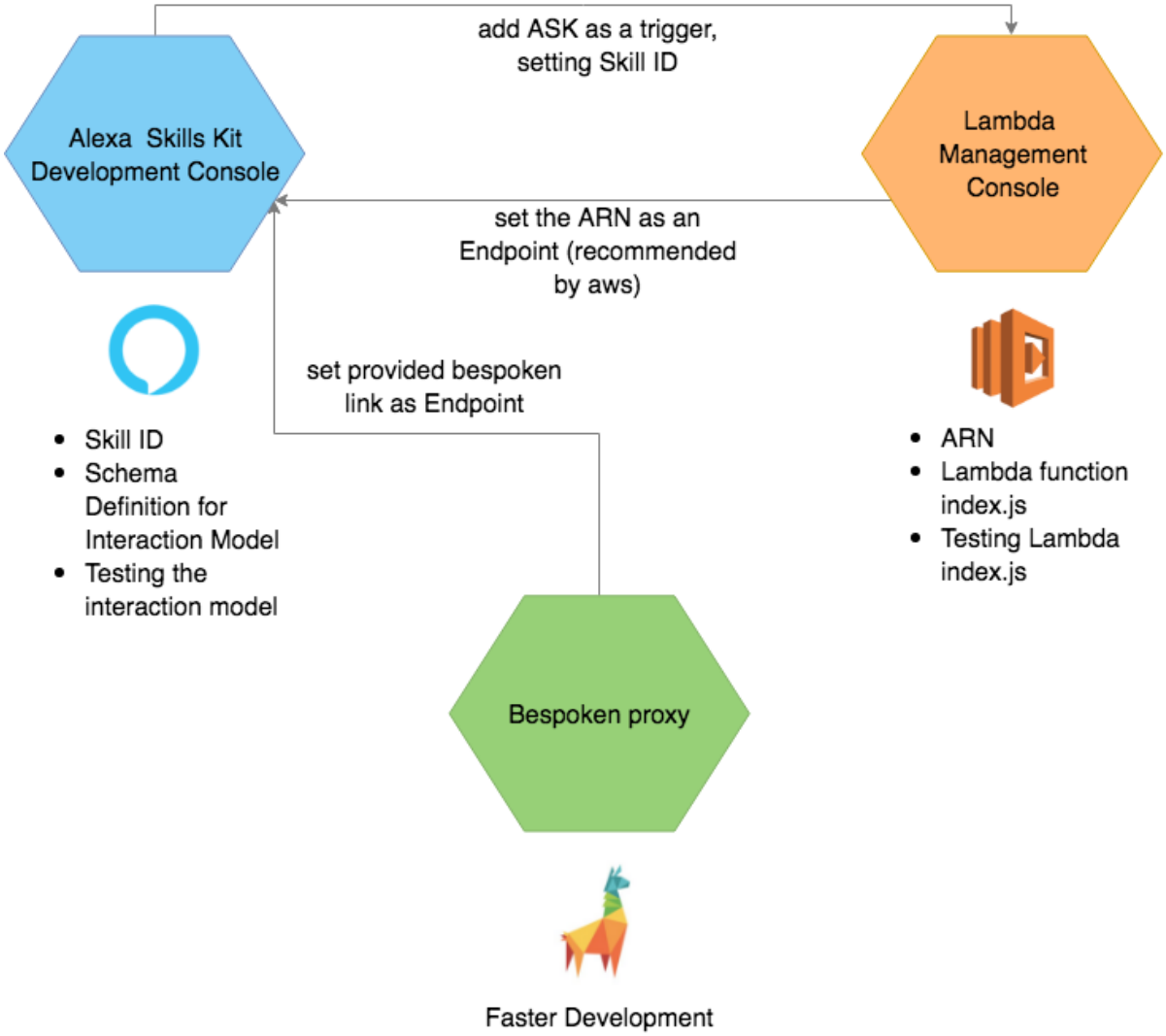

Figure 3: configuration of different aws services

added three intents: StopIntent, HelpIntent, CancleIntent. The names basically describe the purpose of these intents. The prototypes' key intents are the LunchRequest, the ExersiceIntent, the WorkoutIntent and the EvaluationIntent, visualised in figure 4. In every intent individual utterances can be defined and set, visible in figure 5.

The UI of the Alexa Console was built to introduce and help developers with creating the Intent Schema. The Intent Schema is a JavaScript Object Notation (JSON) schema that basically consists out of intents, slots and the respective, well considered and many different utterances. Only if a user uses the defined utterances while interacting with the skill, an intent can be identified and processed. Further, depending on the identified intent a same called event handler in the lambda function will be invoked and the application's logic can be processed. After creating the intents and required slot types it is necessary to add an endpoint to the skill. The endpoint is the AWS Lambda ARN. Later on it is recommended to use secure or "HTTPS"-Endpoint to

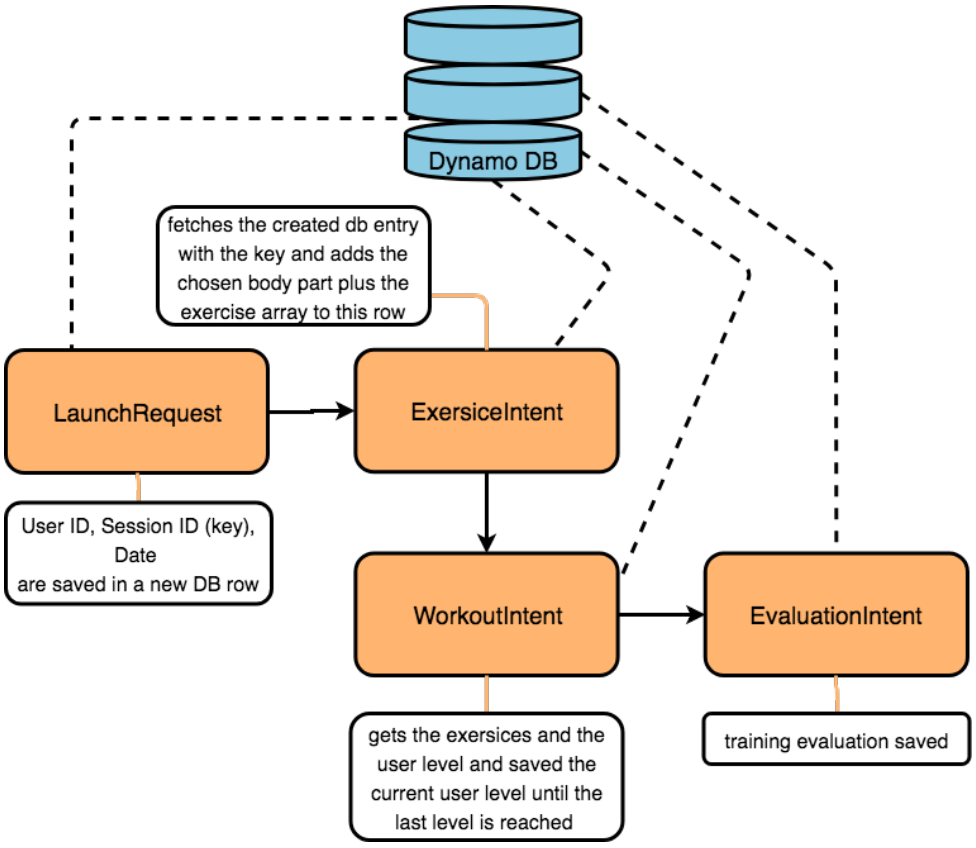

Figure 4: Intent flow and DB storage

support a smooth development flow.

Slot types and slots are the missing data that need to be entered by a user. The prototype firstly needs the name of the body part in order to get the list of four exercises. Secondly, after the workout the prototype wants to know the quality of the performed technique (evaluationTechniqueSlot) and the intensity (evaluationExhaustingSlot) of the workout. Those slots have to be defined within the dialog model in the intent schema, or they can be added via the console. Furthermore, it has to be defined, shown in 6, how Alexa elicits the missing data from the user. Generally a unequivocal question is asked to get the fitting answer.

The Dialog interface reference enables a developer to manage a multi-turn conversation between a skill and a user[28]. It is used to query the required information to fulfil missing slot values.

In order to test if the different environments work together and the skill can trigger handlers from the lambda function, the AWS Lambda ARN (Amazon Resource Names) must be set. The ARN is generated after adding a new function in LMC. Generally it uniquely identifies AWS resources. Depending on what resource it identifies it subject to a different pattern. The ARN for a lambda functions looks like this: “arn:aws:lambda:eu-west-1:516535208406:function:bleibFit” and is

28. `https://developer.amazon.com/docs/custom-skills/dialog-interface-reference.html`, 2018-04-15

Figure 5: Utterances

written and underlies that format: “arn:partition:service:region:account-id:resourcetype:resource”[29].

3.1.3 Local Development with Alexa Skills Kit Command Line Interface (ASK-CLI)

The Alexa Skills Kit Command Line Interface (ASK CLI) is a tool to manage Alexa Skills and related AWS Lambda functions. With the ASK CLI, a skill can be created and updated from the command line and local development environment[30]. The ASK-CLI provides high-level commands and low-level commands, that are either used to create, clone or deploy a skill, or are used to control individual parts of the Alexa Skill. The skill for this research was created, as described before, directly on the ASKDC. The ASK-CLI was used with the command “clone”. This command downloaded the existing script of the prototype to a local directory. The script was automatically structured by it’s interaction models and the lambda function. Additionally, to avoid over-complicated and manual testing routines, the endpoint was changed to a HTTPS Web Service within the ASKDC. The default region “https://cheap-lennon.bespoken.link” had to be

29. `https://docs.aws.amazon.com/general/latest/gr/aws-arns-and-namespaces.html`, 2018-06-07
30. `https://developer.amazon.com/docs/smapi/quick-start-alexa-skills-kit-command-line-interface.html`, 2018-06-10

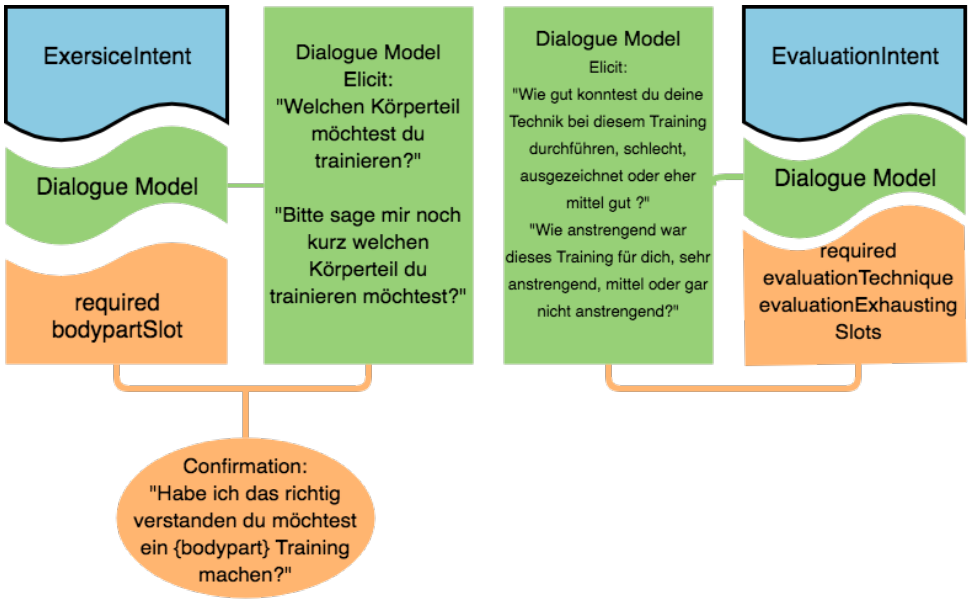

Figure 6: Slots and Dialogue Model

set to the category: “My development endpoint is a sub-domain of a domain that has a wildcard certificate from a certificate authority”. Every Skill needs a certification in order to be published. If a skill is going to be published in the amazon skills store, it must pass a certification process[31]. The Bespoken.io[32] offers their certificate by providing a proxy. So the skill can be tested with a echo device and not only by simulation programs. The bespoken tool can be integrated to the project by a NPM package. A NPM (Node Package Manager) package is a script of javascript code written by other developers for developers. NPM packages are often necessary to increase the speed of the development progress of an application. After the installation process bespoken provides a public URL for accessing my local service through a proxy server. This link has to be set in the skills manifest as the new endpoint, instead of the previously set ARN.

The advantage of working from the local console is to avoid errors like: “An Error occured: The trigger setting for the Lambda arn:aws:lambda:eu-west-1:516535208406:function:bleibFit is invalid. Error code: SkillManifestError”. This means that the correct endpoint or lambda function version was not found. The Lambda function version needs to be set every single time the lambda function is updated and tested from the ASKDC testing section. In figure 3 it can be seen how the aforementioned consoles and services relate to each other.

3.1.4 Dialogue Management

Jeffress (2018) published a tutorial about Dialogue Management to capture missing slot values. After his example a model was created to elicit the bodypartSlotType, the evaluationTechniqueS-

31. `https://developer.amazon.com/docs/custom-skills/certification-requirements-for-custom-skills.html`, 2018-07-31
32. `http://docs.bespoken.io/en/latest/`, 2018-06-07

lotType and the evaluationExhaustingSlotType. In the interaction model schema a dialogue model is defined besides the languageModel. In the dialogue model the intent with a conversation is defined. If there are defined slots they need to be referenced in the dialogue model as well. As shown from line 321 to 330, the bodypart slot has to be elicited and confirmed by the user, because the values are set to true. As another option, the evaluationTechniqueSlot has to be elicited but not confirmed, by setting “confirmationRequired” to false.

```
321                     "slots": [
322                         {
323                             "name": "bodypart",
324                             "type": "bodypartSlot",
325                             "confirmationRequired": true,
326                             "elicitationRequired": true,
327                             "prompts": {
328                                 "confirmation": "Confirm.Slot
                                       .470869105580.24874554309",
329                                 "elicitation": "Elicit.Slot
                                       .1421064436695.1232511755537"
330                             }
331                         },
332                         ...
333                     ]
334                 },
```

Listing 1: Slot for bodypart in de-DE.json

Further the confirmation prompts need to be set. The confirmation prompt is what Alexa says to get a confirm, generally a ‘yes’ or ‘no’, to ensure that she understood the right slot value. It can be seen on the left side in figure 7 in the section ExersiceIntent, that the dialogue state is set to “STARTED”.

When the dialogue is started, the bodypart slot is not yet filled. Therefore the requests intent is picked and passed to the next call in line 8 in listing 3, with the “:delegate” parameter. Then the intent is updated and the missing slot gets filled. The delegateSlotCollection() function can be used in both dialogue situations, because it only depends on “this”.

3.1.5 State Management

In this project states are essential. There are three different types of states that have to be differentiated. First, when registering the handler in the index.js the modules where wrapped in a Alexa.CreateStateHandler() function, allowing a state to be applied. Second, the users level in the workoutIntent had to be remembered. To provide persistence here, a DynamoDB table was created. Third, Alexa SDK provides to save states in “this.attributes”, which was used to save the previous answers of Alexa for the repeatIntent.

Every registered event handlers subjects another intent. Still there need to be an AMAZON.stopIntent in every handler. If a user says “stop” in the exercise handler, the right stopIntent has to be handled by the application. Statehandlers are helpful to keep the code structured.

```
368                     "prompts": [
369                         {
370                             "id": "Elicit.Slot.1421064436695.1232511755537"
                                    ,
371                             "variations": [
372                                 {
373                                     "type": "PlainText",
374                                     "value": "Welchen Kr perteil m chtest
                                        du trainieren?"
375                                 }
376                             ]
377                         },
378                         {
379                             "id": "Confirm.Slot.470869105580.24874554309",
380                             "variations": [
381                                 {
382                                     "type": "PlainText",
383                                     "value": "Alles klar, du m chtest ein {
                                        bodypart} Workout machen, stimmts?"
384                                 }
385                             ]
386                         },
387                         {
388                             "id": "Elicit.Slot.485474712600.243705635324",
389                             "variations": [
390                                 {
391                                     "type": "PlainText",
392                                     "value": "Wie anstrengend war dieses
                                        Training f r dich, sehr anstrengend,
                                         mittel oder gar nicht anstrengend?"
393                                 }
394                             ]
395                         }
396                     ]
```

Listing 2: Prompts for Dialogue in de-DE.json (code shortened)

The second possibility to save the sate is to keep persistence in a DynamoDB. DynamoDB is a NoSQL database service. For the prototype the SessionID (key), the UserID, the date, the user level, the bodypart and the exercises where saved in the table. For every new Session a new row was created in the database. As shown in figure 4, with every LaunchRequest the row was created. In the next intent the data was firstly fetched from the existing row and new data was added. The read and update process of the database was managed in every intent for different values. The table fitnessappDB was directly created in the dynamoDB console. It required a table name and a primary partition key (userID). The primary sort key (sessionID) was also set. In order to have permission to read and update the database, those three values always have to be set. Updating the table without deleting the already passed data requires to set the passed data once again.

```
3  module.exports = {
4    delegateSlotCollection() {
5      if (this.event.request.dialogState === "STARTED") {
6
7        var updatedIntent=this.event.request.intent;
8        this.emit(":delegate", updatedIntent);
9
10     } else if (this.event.request.dialogState !== "COMPLETED") {
11
12       this.emit(":delegate");
13
14     } else {
15
16       return this.event.request;
17     }
18   }
19 }
```

Listing 3: Function to delegate slot values in dialogue.js

The third possibility in keeping the state is to add a new attribute to this.attributes.lastResponse “Say hello!”. This was used to always save the previous respond by Alexa. If a user wants to repeat something Alexa already said, the system can easily fallback on this set attribute[33].

3.2 Testing the Skill

Integration tests and pretesting has improved the development speed. In the beginning of the development AWS CloudWatch[34] was used to debug the very first version of the index.js file. Amazon CloudWatch is a monitoring service for AWS Cloud Resources. It can be used to gain system-wide visibility into resource utilisation, application performance, and operational health. Additionally, these insights help to react quickly and keep the application running smoothly. With CloudWatch all the skill activity can be logged, but also console logs for debugging. After integrating bespoken tools, the debugging process could be done via the local console.

For testing the dialogue flow, visualized in figure 7, the Alexa simulator in the ASKDC was used. Besides the Alexa simulator, there is also the possibility to test manual entered JSON and voice and tone from Alexa. Amazon provides the Sound Library[35] for words that are pronounced differently in other english or german speaking countries. They even provide animal, ambience or office sounds and many more. Since bespoken tools were used for the development the skill can even be tested via a real echo dot, to check the time for the response window for example.

One of the most intense testing parts was quite in the beginning of the development, how to invoke the skill correctly. A skill can be invoked either by using only the invocation name or by

33. https://developer.amazon.com/blogs/post/Tx213D2XQIYH864/announcing-the-alexa-skills-kit-for-node-js, 2018-07-02
34. https://aws.amazon.com/cloudwatch, 2018-07-02
35. https://developer.amazon.com/docs/custom-skills/home-sounds.html, 2018-07-02

```
5     const Alexa = require('alexa-sdk');
6     const APP_ID = 'amzn1.ask.skill.019fe2b6-9106-4534-b9a9-88de88959209';
7     const newSessionHandler = require('./newSessionHandler');
8     const exersiceHandler = Alexa.CreateStateHandler('_EXERSICE',require('
          ./exersiceHandler'));
9     const workoutHandler = Alexa.CreateStateHandler('_WORKOUT', require('./
          workoutHandler'));
10    const evaluationHandler = Alexa.CreateStateHandler('_EVALUATION',
          require('./evaluationHandler'));
11
12    exports.handler = function (event, context, callback) {
13        const alexa = Alexa.handler(event, context, callback);
14        alexa.appId = APP_ID;
15        alexa.registerHandlers(newSessionHandler, exersiceHandler,
              workoutHandler, evaluationHandler);
16        alexa.execute();
17    };
```

Listing 4: State Management in index.js

using the invocation name plus the desired intent. As shown in table 1 something like “Alexa starte Theresas Fitness App und ich möchte meinen Bauch trainieren” can be said in order to start the skill and immediately accessing the exercise intent.

4 Empirical Evaluation of ”Theresas Fitness App”

The following section describes how the experimental study of the voice-base fitness application prototype was conducted. The prototype was designed and developed only for the purpose of this thesis.

4.1 Study design

On the basis of the above mentioned data, it will be investigated if voice interaction with fitness application could help keep technology and health in balance. The aim of this thesis is to conduct an experimental study to show, that a gain in physical activity can be achieved by the use of a voice based Alexa skill. Further it should raise awareness about mobile phone usage and if possible, it should be motivating enough to minimise the usage with voice-based technology. It will give some general information about the age of the participants and also examines an experience perspective on how intuitive the interaction with the voice-based Alexa skill is. As an investigation method within this bachelor thesis, an empirical examination was selected by means of two questionnaires after conducting a user test with the prototype.

The prototype provides a range of exercises depending on five different body parts. A user is able to choose one of those body parts in the beginning of the workout. Then the user is required to perform the suggested exercises in order to fulfil the training part. In addition, it is possible

for the users to reflect on their own performance. Therefore, Alexa asks them two question regarding the technique and the intensity. During the workout, the user has no graphic support or distraction over a display. The whole application is designed for only voice based interaction in the German language.

4.2 Theoretical background of the study

The idea for this prototype came up during a Freeletics workout, while trying to focus on physical and psychological strength improvement. The interaction with the mobile phone felt unsupportive for the user. Notifications and simply starring at the display interrupted the training process. The idea was then to make it possible to interact with a fitness application only with voice interaction to keep the mind and the body in focus, with the intention to increase the user's training morality.

As aforementioned in section 2.1 fitness applications are a promising, well accepted and increasingly strong option to keep the human body healthy. The master thesis of Susanne Wolf, Bernhard Schwarz, and Bernhard Rupp (2013) with titled "Motivationsmöglichkeiten zu gesundheitsförderndem Verhalten unter Einbeziehung neuer Medien" was inspiring and a good basis for this thesis. She examined how mobile phones and other entertaining technologies could contribute in motivating people to live a healthier lifestyle. In her results she says that most of the participants want to change something in their lifestyle and that they would consider smartphones and apps as an option to get more motivation to reach their goal. Based on this thesis a lot of questions for the questionnaire were developed and should be insightful for the hypothesis that voice based fitness application could rise motivation for keeping technology and health in balance.

Moreover Cordasco et al. (2014) developed a prototype for voice controlled assistive care and communication services for elderly people's home to investigate how those people interact with the system. Measuring intuitivity was one part of the study. Therefore the INTUI questionnaire, developed by Diefenbach and Ullrich (2015), was used. The INTUI model aims to answer these questions and explore the phenomenon of intuitive user interaction from a User Experience (UX) perspective. It combines findings from psychological research on intuitive decision-making and user research in HCI, as well as findings from interviews on subjective feelings related to intuitive interaction. This phenomenological approach recognises the multi-dimensionality of the concept and thus opens up important influencing factors and starting points for the interface design. The INTUI model suggested five components of intuitive interaction, Gut Feeling (G), Verbalizability (V), Effortlessness (E) and Magical Experience (X) and Intuitivness (I). Those five components have been extracted by different intuitive interaction theories. While Gut Feeling is a result of feeling-based decision-making driven by a "cognitive process that somehow produces an answer, solution, or idea without the use of a conscious, logically defensible step-by-step process" (Klein, Orasanu, and Zsambok 1994), Magical experience represents a novel aspect - intuition is hard to grasp for someone who makes automatically the right decision. People are often unaware of the actual source of their intuition, that is probably stored as memories or cognitive processing, and might refer to it with supernatural power or magic. Verbalizability plays a role in measuring

intuition, because often people can not explain how they made something, they literally just did it without conscious thinking. In the questionnaire, they measure Verbalizability with questions like “knowing how to operate the product but not knowing why” or “I don’t remember the single steps but it must have been obvious. There was no need to read the manual”. Finally, they also took Effortlessness into account, because intuition appears to be quick and effortless. So in the segment of Effortlessness according to the INTUI Model it should show up that the user could interact without any strains. Those four components seem to be promising for measuring the intuitivity of this prototype.

4.3 Participants

On this experimental, laboratory study 13 people took part. They were between 20 an 38 years old, whereby the average age was 22 years. Nine of them were male and four of them were female. The participants were not chosen by a specific pattern. They were more driven by curiosity to participate in this study. The author asked in the surroundings for people who would like to contribute to this study. The only requirements were speaking fluently German and having no physical restriction. One participant was pregnant and successfully passed the process. Children or underage persons were not allowed to contribute to the study without their parents agreement, due to privacy reasons.

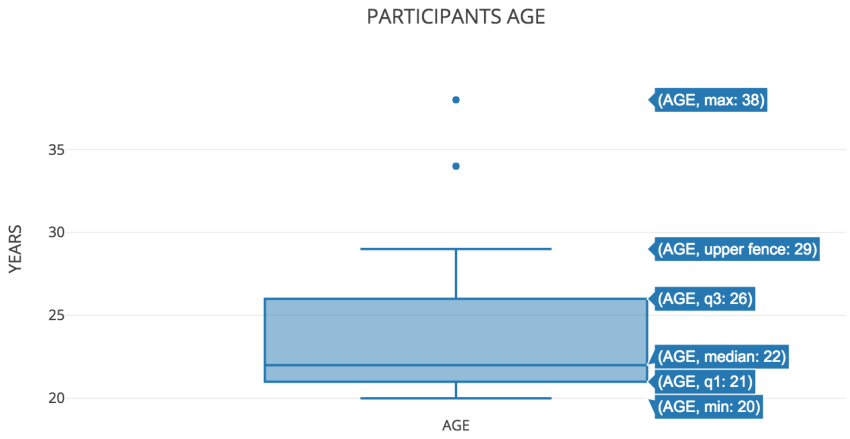

Figure 8: Participants Age

4.4 Study flow

The User Testing consisted out of two parts. First, the participants had to interact with the voice-based fitness application in form of a workout. Second, after finishing the workout they still had to interact with Alexa by giving feedback on their own training performance. Third after the interaction with Alexa was finished they where asked to complete two questionnaires. One to measure whether they can be motivated to include more physical activity into their live styles

within 23 questions and another one to measure the intuitivity of the voice-based prototype with 16 question in a seven score scale.

All participants were invited to a labor in order to ensure the same conditions for everyone to get comparable results. None knew what the study was about in advance. They did not know that they had to interact with an Alexa Skill. They had to pass the study alone, only with a supervisor, in a room. Before starting the experiment the supervisor told them how to start the prototype. Furthermore, they were told not to talk and that they only had 8 seconds time to respond to Alexa, if she asked something.

Every participant invoked the skill with the sentence "Alexa, starte Theresas Fitness App", then they were lead through the skill by Alexa. The fastest walkthrough requires to respond to Alexa ten times, where one of the responses is only a confirmation response. The walkthrough can be seen in Figure 7, what takes approximately 15 minutes. Completing the questionnaires took the participants another 10 minutes. So for every participant it was an effort for around 25 minutes in total.

4.5 Outcome measures

The key concepts of the thesis is to get to know information about the user experience on a specially developed voice-base prototype. The focus was set on the intuitiveness of the system. In order to get details about the intuitiveness different aspects like Gut Feeling, Effortlessness, Verbalisability, Magical Experience and Intuitiveness were analyzed. Diefenbach and Ullrich (2015) developed a questionnaire were those aspects were taken into account. For the reason to get to know whether the participants are potential users of the developed prototype a second questionnaire was developed. It can be separated into two different sections so as fitness and cell phone usage or fitness application usage. Important variables in there were, whether they gained some experience with Alexa already, or if they are satisfied with their current fitness behaviour. Further it gives information about how many hours they use their cell phones and if they feel unbalanced with it.

4.6 Methods for data acquisition

This section informs about the data acquisition of the thesis. First, it shortly explains how the user testing was completed and what data was extracted during the testing. Second, it gives a small overview about the intuitiveness questionnaire, how it is structured and defined. Third, this section gives some reasons why the second questionnaire was included in this thesis.

The first part of the data acquisition was the user testing of the voice-base prototype. Each participant had to start the fitness application by their voice and were lead through the app by Alexa. It took about 20 minutes for a participant to complete the user testing. The participants were supervised by one person of the development team of this thesis. Notes, concerning the usage of the system and some difficulties during the user testing have been taken down. Additionally, with the connection of a database to the system it was possible to save time or single answers the participants gave during the user testing.

The second part was to complete two questionnaires, one for measuring intuitiveness and the other one to get to know if the participants would be potential users. In order to get results regarding Gut Feeling, Effortlessness, Verbalisability, Magical Experience and Intuitiveness, Diefenbach and Ullrich (2015) created a seven-likert scale. The questionaire is separated into different sections like, 'While using the product ...', 'Using the product...' and 'In retrospect...'. This questionnaire consists out of 17 questions relating to the different aforementioned aspects. The first section includes for example a scale between "... I acted deliberately" and "...I acted on impulse". The second section offers scales like "...was inspiring" to "... was insignificant". The participants where allowed to mark always only one of the scored of the 17 questions.

In order to get an idea about the potential market for the prototype, another questionnaire with 23 questions was developed. Susanne Wolf, Bernhard Schwarz, and Bernhard Rupp (2013) conducted a research to a similar topic in the context of her Masterthesis. The developed questionnaire in her thesis was inspiring for this questionnaire. This questionnaire is about fitness and mobile phone usage. It gives information whether people feel unbalanced due to a misuse of their cell phone. It investigates in whether they want to change something about it, how they would like to change something and if they think that the developed system could help them to reach a new fitness goal. Further they could say in what way mobile phone misuse unbalance their life. For better understanding seven diagrams have been developed.

4.7 Methods for data analysis

In this section it is explained how the data was extracted, visualized and finally analyzed. The data was manually extracted form the questionnaires. In the beginning it was entered to Google's tables. First the questionnaire for intuitiveness was analyzed in more detail. Every answer of the 17 answers was given an unique code. For example "G_01 - G_04, M_01 - M_05, X_01 - X_04, V_01 - V_03 and INT_01" also visible in figure 17 on the right side in every row of the questionnaire. Those single codes define the groups of the aforementioned aspects. At eight different questions there is a "P" written before a code. The "P" stands for "polarity reversal" and is really important for a correct analysis. Each code is subject to a 7-point Likert scale. In order to get a value the scores were added up and divided by the amount of questions within one group. If one question's polarity had to be reversed the number eight had to be subtracted from the actual score. The resulting negative number had to be set to a positive number and the new score had to be added up as well. For every group and every participant the average values were calculated. For better understanding the results, a chart called box plot was created. It was created with the web application "plotly"[36] and gives primary information about the variation between the single groups and the participants. The weighting of the different groups regarding intuitiveness can be seen clearly.

Then the second questionnaire was analysed. It consists out of two major categories, fitness and mobile phone usage, and also some questions about the sex, age or previously acquired knowledge about Alexa. It is important to know if the participants have ever tried to talk to Alexa and if they have general experience w0ith voice-based systems, for acquiring serious

36. https://plot.ly/~TheresaElisa/3/, 2018-08-06

results. Since the prototype is a fitness skill it is also important to know how the participants think about their own fitness behaviour and whether they use fitness apps for achieving fitness goals. Furthermore, questions about mobile phone usage were chosen, because it is important to know how the participants feel about their own mobile phone usage. Do they feel unbalanced? How and why would they like to change their behaviour? In the end, questions were written, whether the participants think that a voice-base fitness application can help them to become more active in their daily life and to balance out their mobile phone usage. To attract interest on those questions, several diagrams were created. Four diagrams support the comprehension about the fitness behaviour. Figure 9 therefore is a pie chart to get an idea about how active the participants are.

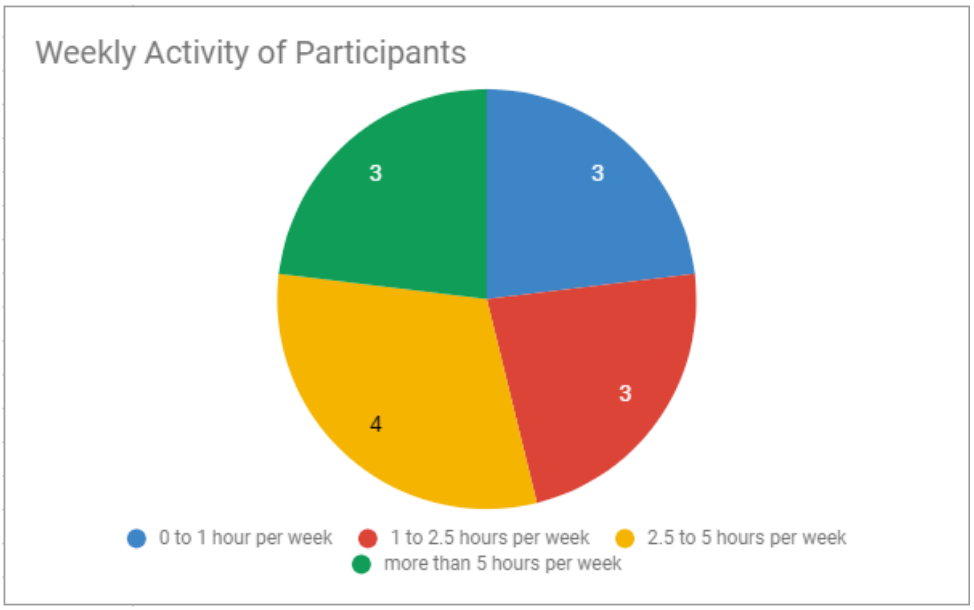

Figure 9: Physical Activity per Week in hours

Further figure 10 is a bar chart showing the type of sports participants would like to perform or try out. In figure 11, the primary reasons for changing the fitness behaviour are demonstrated with a bar chart again. A pie chart then shows how the participants think about support by common fitness apps like a pedometer. Two diagrams support the understanding about the mobile phone usage. Figure 13 therefore shows as a pie chart how many hours the participants are active per day. Finally in figure 14 reasons are listed and weighted why the participants would like to change their phone usage. All the charts, excepted the box plots, were created with Google's tables

5 Results

The results in this section can be separated in fitness, phone usage and intuitiveness. For better understanding several diagrams were created. In sum four diagrams visualise information about

the participants fitness level and their attitude towards connecting fitness and technology. In addition there are two diagrams about cell phone usage and the well-being with it. Finally there is a box plot that gives information about the intuitiveness about the voice-base system. The empirical study indicates, that 8 out of 13 participants have ever heard something about Alexa, whereby only two of them have ever tried to talk to Alexa in person. Five participants have experience with other voice-based intelligent assistants, four mentioned in this regard Apple's Siri and one person refereed to a smart phone, but did not give further information about a special application.

5.1 Results in Fitness

In the average the participants are rather satisfied with their physical activity. Six out of 13 would like to change something in their movement behaviour even though more than 50% of them perform sports more than 2.5 hours per week (visualized in figure 9). This is exactly within the recommendation of WHO mention in section 2.1. The three activities, that wanted to be increased and which were named the most often were more endurance training, more strength training and more exercise in everyday life.

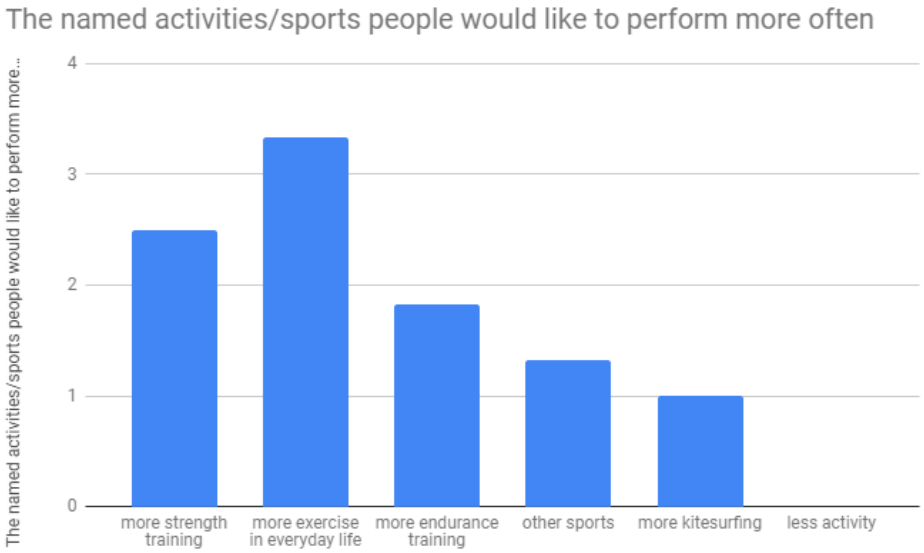

Figure 10: Activities people would like to perform

According to the resulted ratings more exercise in everyday life was chosen the most often, visible in figure 10. Personal reasons for a change are mostly health, the look and that the participants wanted to try out something new, what is the leading answer and can be seen in figure 11.

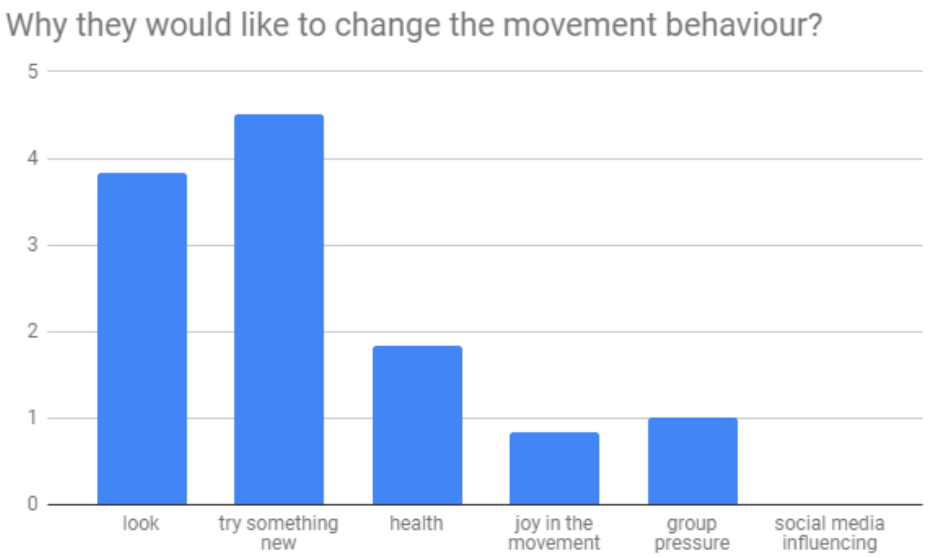

Figure 11: Reasons for change

For reaching the goal three of the participants do not think that supportive or motivating applications could be helpful. Three of the participants think a fitness app could be helpful and seven of them can not tell if it would be helpful or not, visualized in figure 12. One of the participants mentions an app called STRAVA that is tracking runs and bike tours. The participant uses it for motivation and self tracking. It is not a voice interaction supported application.

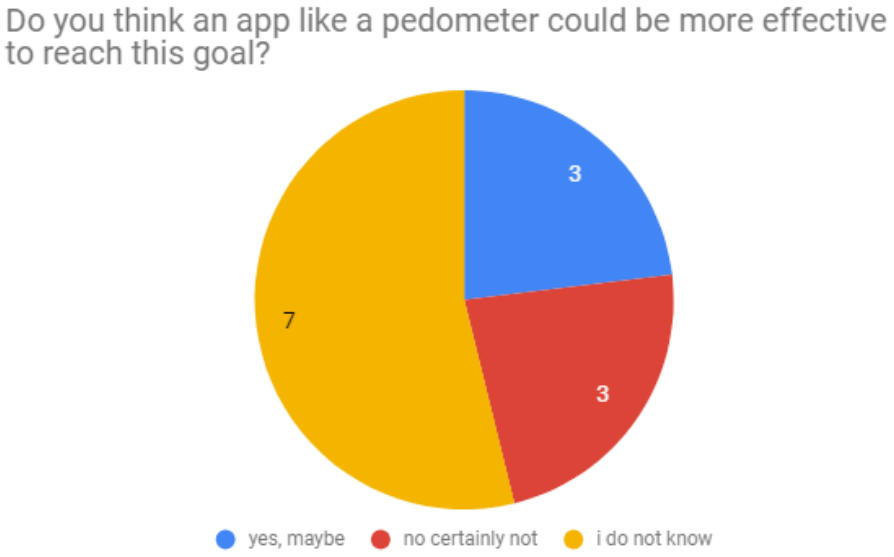

Figure 12: Support by apps

5.2 Results in Mobile Phone Usage

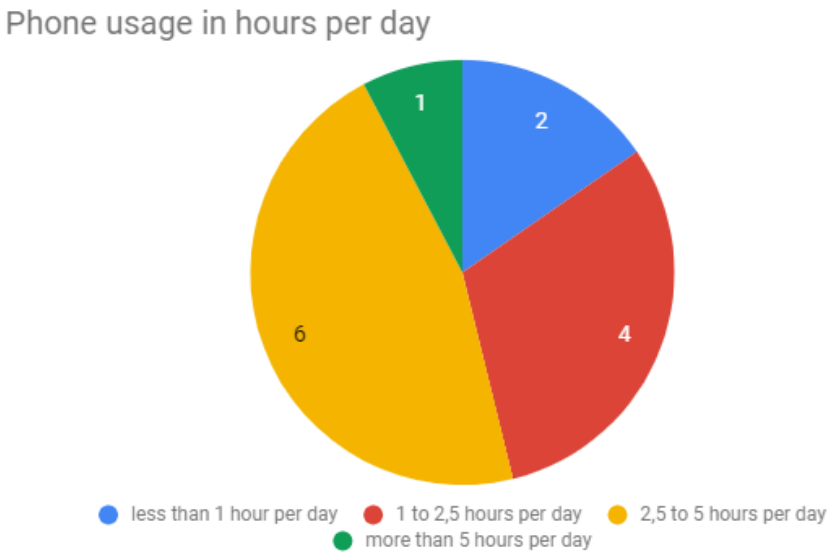

Figure 13: Phone usage per day in hours

In figure 13 it can be seen that six participants use their phone between 2.5 and 5 hours per day. Hence six people felt unbalanced. Also six feel therefore unbalanced. Eight people said that they want to reduce or change their usage behaviour. Most of the participants want to use their cell phone with the same functionalists but less often. Others said that they want to use the cell phone only for phoning and messaging but without internet. According to the questionnaire most of the participants want to lower the phone usage to perceive their environment better again, such as family and friends. Others think that they lose to much time to fulfill more meaningful and important tasks in their life. Furthermore visualised in figure 14 they also refereed to the gain in their ability to concentrate, or that the constant phone use might not be healthy.

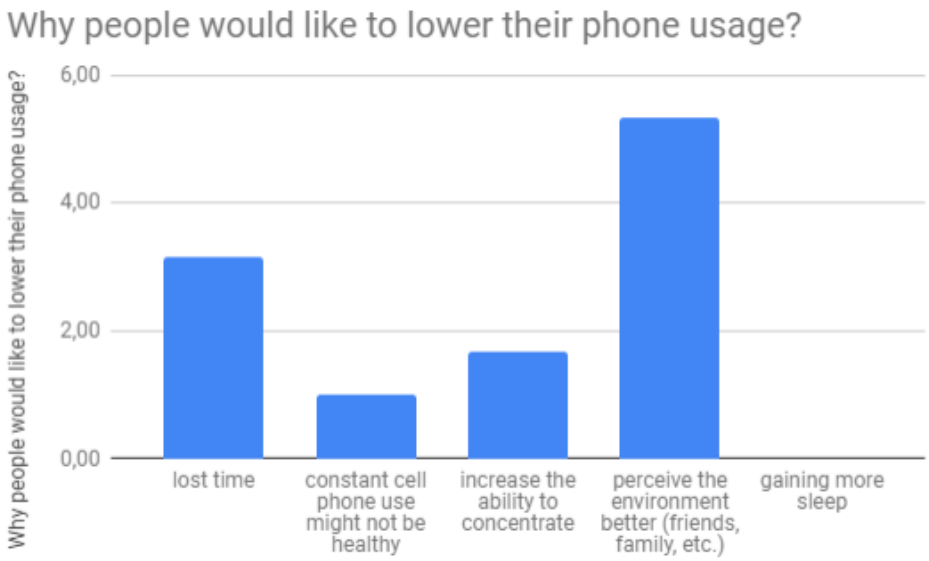

Figure 14: Reasons for phone usage change

When asked if they think that a language system like Alexa could help them to lower the phone usage, four answered with ”no”, three answered with ”yes maybe” and six answered with ”i don’t know”. Finally, when asked whether they would use the tested system also at home, seven of them answered with no, because of reasons like fear of private data misuse or dependence on the application. Five said yes because they think they would be less distracted by their cell phone, because the voice could be motivating, or because the exercise technique was well explained by the voice.

5.3 Results in Intuitiveness

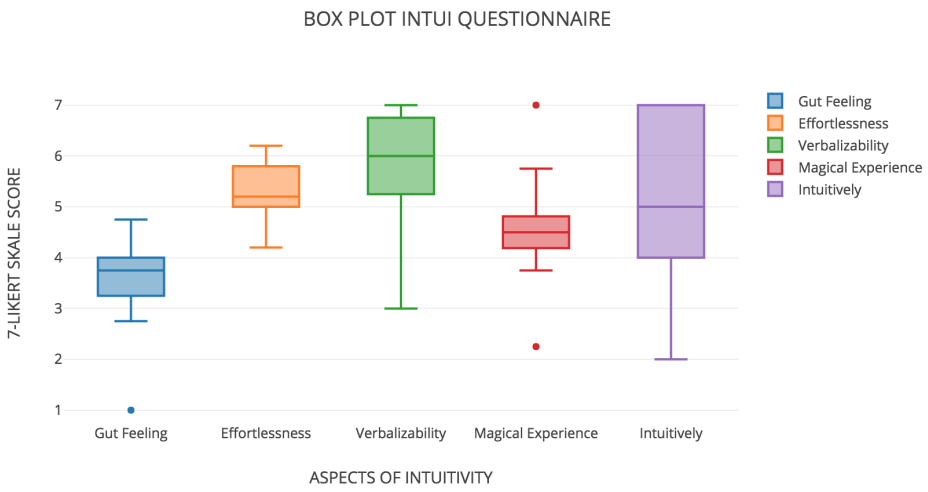

Figure 15: INTUI box plot (n=13, N=221)

In the box plot, figure 15, the weighting of the five aspect describing intuitiveness can be seen. On the y-axis is the seven likert scale. The x-axis are the different aspects summarized as single box plots. The first box plot describes the aspect 'Gut Feeling(GF)', what takes feeling based decision making into account. It can be seen that this box plot is the lowest in the graphic compared to the others. For 75% of the participant GF got a score between 1 and 4, where 50% chose close values between 3.35 and 4. The maximal value for Gut Feeling lays at 4.75. Effortlessness(E) is shown as a short box plot. It has the highest minimal value of all the others with 4.20 points. 50% of the participants set the score for effortlessness from 5.00 to 5.80 points. The maximal value of it is 6.20. The next box plot is Verbalisability(V) reach as a maximal value 7.00 points, but also a minimal value of 3.00 points. It can be seen that the variance in this box plot is rather high. Still 50% of the participants chose scores from 5.33 to 6.67, what is in general really close to the maximal value of seven. The fourth box plot rating the Magical Experience(X) of the prototype has also a high variance reaching from 2.25 points to 7.00 points. Though 50% of the participants are quite in complete agreement with 4.35 to 4.75 points. The last box plot describes intuitivness and was created out of only one question per each participant. It clearly has the highest variance of answers starting from 2.00 points to 7 points. Also the 1 quarter (4.00 points) and the 3 quarter (7.00 points) are rather far from each other. Nevertheless this means that 50% of the participants chose a value that describes the system as rather intuitive.

6 Discussion

During the intensive study of this topic, others also invested in the research of a fitness application for Alexa. Smolyak, Lee, and Choe (2018) created an app called TandemTrack for the smart

phone and Alexa. It's capabilities include exercise guidance, data capture, feedback, and reminder. In order to measure the effectiveness of the app and the skill, they compared both applications. It came out that Alexa has an important role in social interaction for the participants. It rises motivation of the participants, because it was seen as a training partner.

Furthermore, Anup, Mohit, and Kumar (2018) created a system to recognize exercises and count the repetitions only by the resulting sound, that appears while performing these exercises. They used a circular microphone array, that most smart speakers have to provide hands free, voice-only interaction from a distance. The microphone is similar to the Amazon Echo microphone. They measured the Doppler shift on a pilot tone caused by a gesturing human body, and used beamforming of the mic array to increase the range of the detection. They trained a deep neural network to recognize 10 different exercises and counting repetitions by peak-finding heuristics. In their study 17 participants were tested. 96% of the performed excersises could have been identified. The count accuracy is 91%. This is a new approach for tracking or supervising people while doing sports with a voice-based fitness application.

Therefore it is obvious that there is an ongoing trend in expanding the existing fitness skills of Alexa. The relevance for this study and for future studies is still given. In the following the theoretical background, the development, the questionnaires and the findings are going to be examined critically. In the end there will be a confrontation with the future development of voice-based apps and the ethical issues that may need to be considered.

6.1 Critical examination of the theoretical background

In the section 2.1, the theoretical background of fitness applications is discussed. In the example of the fitness applications the app Freeletics was chosen as an inspiring application to develop an Alexa fitness skill. While choosing this app the scientific correctness in terms of training without a real coach was not taken into account. It is not sure that the app users perform the recommended exercises of Freeletics in a physically friendly way. Although they point in their general terms and conditions[37] out that they go with the latest science, but they do not assume liability for any other development. The use of the app is at the users own risk. So if someone performs exercises with the wrong technique, because no coach was there to give instant feedback for correction, it is the users own fault. That could be in general a critical issue for using fitness apps.

In the section 2.4, Gustavo, Quesada, and Guerrero (2018) conducted a study on the four most popular voice interaction systems, Alexa, Cortana, Siri and Googls Assistant. Therefore a team member was giving every device the same commands in same tone and pace. Then the participants rated it with a 5-point Likert scale in natural feeling and correctness. The statistical analysis did not find significant differences between the systems. The relevance of this study might be given, but implementation should be discussed here. First of all, it might not be possible to repeat commands in exactly the same tone and pace as a human. Second the perception of the participants, who rated the results, is relative towards each other. It depends on their origin and

37. `https://www.freeletics.com/de/pages/terms`, 2018-08-01

their living environment. Therefore the natural feeling and correctness in languages might be different.

In the next paragraph the study of CIRP was taken into account in order to find a suitable device for developing the prototype. CIRP demonstrates it's findings on the basis of 500 US owners of an Amazon Echo or Google Home. Transferring the results on all existing devices in the US, what are according to CIRP 27 million installed units, might be daring. Furthermore it is not sure if those results are applicable for Europe. The market for IPA's in Europe might be different. At the beginning of the study, however, there was no better comparison available.

6.2 Critical examination of the development

In the development more time than expected was invested. First of all, the connection of the Dynamo DB was not really necessary in order to implement the prototypes functionalities. The main reason for adding a DB was actually to provide persistence within the application. During the development it resulted that keeping the app persistence was not required, because every user had only one walk through and the different walk throughs did not depend on the data of the previous one. The database usage in the app is reduced to state management, but this state management could also have been solved with another sate management tool.

Furthermore it was confusing, that it is possible to enter expected slot values. When testing the app and the slot values where elicited, the user could say anything and Alexa would record it. For a developer this is an unexpected behaviour of the system. With further research it was found out, that the slot values are just possible values, but do not restrict the users input. The correctness of the slot value was then checked within the code.

As it can be imagined the development of the Alexa Skill was time consuming. It has to be remembered that there are still three other platforms providing the development of voice-based applications. For example as a company like Freeletics, it would be prerequisite to implement the same skill for every platform to assure complete user satisfaction. Since every platform is different and relies on different designs and code, this would require four different development teams with different skills and knowledge. Therefore, if Amazon or others would like to reach a higher number of skill publications it might be useful to provide a cross platform service, so as React Native with mobile apps, to rise motivation for adapting existing apps to a voice-based system.

6.3 Critical examination of the questionnaires

The second questionnaire was created too premature. Although interesting questions were chosen, they had barely references to any paper. The solution for that problem was to find a paper, that fit to the existing and already filled out questions. This procedure was not optimal to receive meaningful results in the end. Fortunately a qualitative paper was found.

The INTUI questionnaire is small and manageable. The corresponding paper was also easy to find and ensured the scientific reference to the topic of intuitiveness. The first impression

was deceptive, because afterwards it was hard to find out how the questionnaire needed to be evaluated. There was no reference about the evaluation on the website neither in the paper. Only after asking the publisher personally, the assumption became true. Diefenbach and Ullrich (2015) confirmed that there were some values that had to be inverted in order to get out correct results.

6.4 Critical examination of the results

The idea of the thesis was to show that it might be possible to compensate the feeling of being unbalanced and too little active due to mobile phone overuse, with a voice-based fitness application. The results of this study tend to be promising for supporting this hypothesis. Most of the people would like to use their mobile phone less often but with same functionality, due to reasons like being not able to perceive the environment, friends and family good enough anymore. Furthermore the feeling of loosing time or the inability to concentrate make people think more and more about lowering the mobile phone usage. The concentration problem, due to mobile phones, was already mentioned in section 1. Building on that, Alexa would be the perfect possibility to leave the mobile phone shut down in a different room. Alexa still allows to use the advantages of the internet. There would be no risk of being distracted by the phone, while searching the weather or doing the daily fitness routine. Even thought the mobile phone usage per day is rather high, most of the participants are physically active enough, in order to keep their body healthy. The median age is 22 years, probably that is why they are rather sportive and open for trying out new kinds of sports. The young age might be also the reason why look is more important than health.

However, because only two of the participants mentioned that they had experience with Alexa, and only four of them have ever talked too Apple's Siri, it can be said that they have little to no experience with voice-based systems. Moreover, only one participant stated that he is in use of a fitness application to support his fitness behaviour. This is the reason why it can be assumed as well that they have no idea about using fitness apps. This inexperience can be traced back to the reason why most of the participants would rather not use or never thought about using Alexa in their daily life. On the contrary, after testing the fitness application prototype five participants could imagine to use the system at home. Reasons for that are again primary to avoid being not distracted by the phone. This contradiction could have probably been avoided, if the participants had the chance to not only use Alexa for the time of the experiment, but to use it for several weeks.

The outcomes of the INTUI questionnaire are surprising as well. Even though the participants had little to no experience with Alexa the system seems to be rather intuitive. The aspects like Effortlessness, Verbalizability and Intuitively are highly rated. This might result out of the dialogue design of the system. Alexa always provides the user what he is able to say to reach the next step. Therefore it was not high effort to walk through the app and people could remember better what they said afterwards. On the contrary one participant mentioned that this would be exactly the reason why the app would not be used. He felt automated by just repeating Alexas words and being limited in the freedom of choosing own words. It can be assumed that the participants also rated Magical Experience rather high, since it was for most of them the first

encounter with Alexa.

6.5 Future development

For future development it can be assumed that there could come up several ethical aspects. First of all as mentioned in section 2.2, the social interaction is really important for humans to learn speaking. It might be daring, if voice-base systems are tend to be developed to support nursing stuff by reading out stories or just telling handicapped or sick people the latest news. Those people might become emotionally stunted without interpersonal contact.

Furthermore, people might become suspicious concerning data misuse. Tracking the own fitness could give a lot of information about someones lifestyle or health. There might be possibilities to enter the exact weight, high, bust, waist, and hip measurement. Furthermore whether they smoke, how much they smoke and what they eat in detail for example. Another thing would be recording and analysing voices. Google implemented for their voice-based system a feature called Voice Match, which makes it possible that only certain people can talk to Google Home.

Besides the ethical issues the development of voice-based applications is going to increase, as already mentioned in section 6. Supervising fitness application users only by its performing sounds would make the usage of voice-based fitness applications (or sound-based fitness application) even more attractive. It is a wide filed, where further studies could be conducted.

7 Conclusion

In conclusion, as part of this bachelor thesis it has been shown that voice interaction is a relevant topic in Human-Computer-Interaction. To contribute to a positive development of this technology, towards the human health a voice-base fitness application was developed. The fitness application was created as an Amazon Alexa skill. The skill primarily aims at motivating users to include more physical activities in their daily life while helping to balance out the overuse of their mobile phones. By extension, it eases the reduction of mobile phone consumption, which can improve ones ability to concentrate, become more aware of social opportunities such as meeting with friends and family, and just being more aware of the surrounding environment. An experiment with a developed prototype and the results of two questionnaires show that the Alexa Fitness Skill is a promising way of balancing out technology and health. Due to the amount and the missing categorisation of the participants the results are not completely generalisable. They would have been more meaningful, if other participants had been chosen according to their experience in using fitness apps and Alexa. However, most of the participants would like to use their mobile phone less often, but with same functionality, due to reasons like being not able to perceive the environment, friends and family good enough anymore. Furthermore the feeling of loosing time or the inability to concentrate make people think more and more about lowering the mobile phone usage. Alexa would be the perfect possibility to leave the mobile phone shut down in a different room. Although, it still allows to use the advantages to grasp the latest information

from the internet. There would be little to no risk of being distracted by the phone anymore, while searching the weather or doing the daily fitness routine with a fitness application.

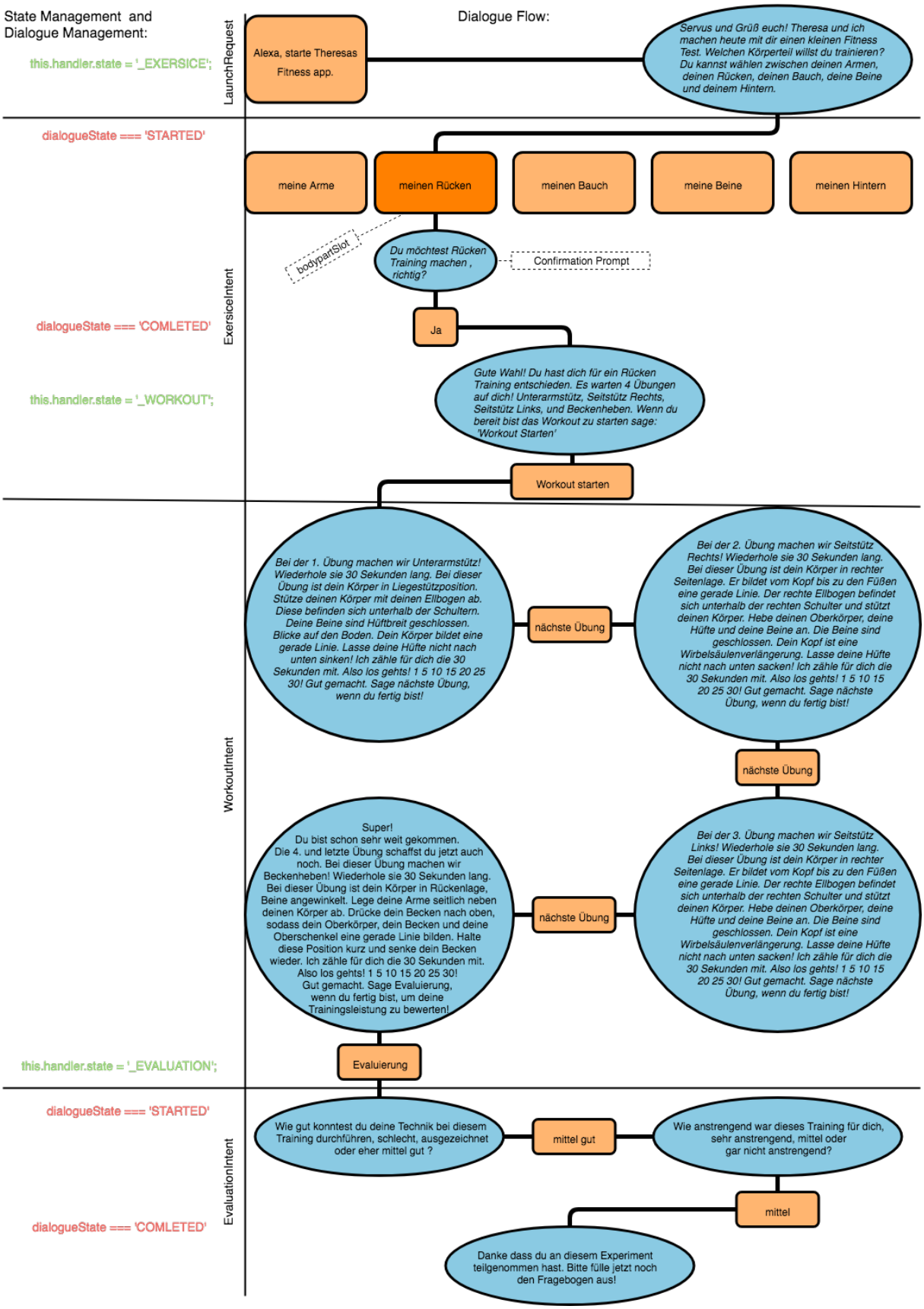

Figure 7: Dialogue

	Description	*Case*	*Answer*	*Result*
1	Invoce TFA without Intent	”Alexa starte Theresas Fitness App”	”Servus und Grüß euch zu fit bleiben. Theresa und ich machen heute mit dir einen Fitness Test. Sage mir einfach was du machen möchtest.”	Success
2	Invoce TFA with Intent	”Alexa starte Theresas Fitness App und Bauch”	”Du möchtest bauch Training machen , richtig?”	Success
3	Invoce TFA with Intent	”Alexa starte Theresas Fitness App und ich möchte meinen Bauch trainieren”	”Was möchtest du trainieren?”	Dialogue started successfully, but did not get body-part within the invocation
4	Invoce TFA with Intent	”Alexa starte Theresas Fitness App und schlage mir ein Training vor”	Audio only response	Failed
5	Invoce TFA with Intent	”Alexa starte Theresas Fitness App mit Leibesübung”	”Bitte sage mir noch kurz welchen Körperteil du trainieren möchtest?”	Success
6	Invoce TFA with Intent	”Alexa starte Theresas Fitness App und empfehle mir ein Training”	”Ich konnte keine aktivierten Video Skills finden. Gehe in den Musik, Video und Buch-Bereich der Alexa App, um den gewünschten Videoanbieter zu aktivieren. Verknüpfe dann dein Konto und schließe die Einrichtung ab.”	Failed
7	Invoce TFA with Intent	”Alexa starte Theresas Fitness App mit Leibesübung”	”Bitte sage mir noch kurz welchen Körperteil du trainieren möchtest?”	Success
8	Invoce TFA with Intent	”Alexa starte Theresas Fitness App und sage mir welches Training ich machen könnte.”	”Welchen Körperteil möchtest du trainieren?”	Success
9	Invoce TFA with Intent	”Alexa starte Theresas Fitness App und meinen Bauch.”	”Du möchtest bauch Training machen , richtig?”	Success, but case does not make sense
10	Invoce TFA with Intent ’improved test 3’	”Alexa starte Theresas Fitness App und ich will meinen Bauch trainieren.”	”Du möchtest bauch Übungen machen, richtig?”	Success, changed ’möchte’ to ’will’ and she fills the bodypart slot with ’bauch’

Table 1: Testing Skill invocation

Abbreviations

ARN	Amazon Resource Names
ASK	Alexa Skills Kid
ASKDC	Alexa SkillsKit Developer Console
ASK CLI	Alexa Skills Kit Command Line Interface
ASR	Automatic Speech Recognition
AWS	Amazon Web Service
CIRP	Consumer Intelligence Research Partners
DB	Database
E	Effortlessness
G	Gut Feeling
HCI	Human-Computer-Interaction
HIIT	High Intensity Interval Training
HTTPS	Hypertext Transfer Protocol Secure
I	Intuitivness
INTUI	Name of the Questionnaire
IPA	Intelligent Personal Assistent
JSON	JavaScript Object Notation
LMC	Lambda Management Console
NPM	Node Package Manager
NUI	Natural User Interface
SDK	Software DevelopmentKit
SIN	Skill InvocationName
SLU	poken Language Understanding
TCP	Transmission Control Protocol
TFA	Theresas Fitness Application
V	Verbalizability
VR	Virtual Reality
V-R	Voice Recognition
WHO	World Health Organization
X	Magical Experience

List of Figures

Listings

List of Tables

References

Andrews, Sally, David A. Ellis, Heather Shaw, and Lukasz Piwek. 2015. "Beyond Self-Report: Tools to Compare Estimated and Real-World Smartphone Use." *PLOS ONE* 10, no. 10 (October): 1–9. doi:10.1371/journal.pone.0139004. https://doi.org/10.1371/journal.pone.0139004.

Anup, Agarwal, Jain Mohit, and Patel Kumar. 2018. "OPPORTUNISTIC SENSING WITH MIC ARRAYS ON SMART SPEAKERS FOR DISTAL INTERACTION AND EXERCISE TRACKING," no. Signal Processing for Internet of Things: IOT Systems: 3–7. https://homes.cs.washington.edu/~mohitj/mic-icassp2018.pdf.

Booth, Frank W., Christian K. Roberts, and Matthew J. Laye. 2012. "Lack of exercise is a major cause of chronic diseases." *Comprehensive Physiology* 2 (2): 1143–1211. ISSN: 20404603. doi:10.1002/cphy.c110025.

Cordasco, Gennaro, Marilena Esposito, Francesco Masucci, Maria Teresa Riviello, Anna Esposito, Gérard Chollet, Stephan Schlögl, Pierrick Milhorat, and Gianni Pelosi. 2014. "Assessing Voice User Interfaces: The vassist system prototype." *5th IEEE International Conference on Cognitive Infocommunications, CogInfoCom 2014 - Proceedings*, no. November: 91–96. doi:10.1109/CogInfoCom.2014.7020425.

Cronin, Seán, and Gavin Doherty. 2018. "Touchless computer interfaces in hospitals: A review." *Health Informatics Journal*. ISSN: 17412811. doi:10.1177/1460458217748342.

Diefenbach, Sarah, and Daniel Ullrich. 2015. "An experience perspective on intuitive interaction: Central components and the special effect of domain transfer distance." *Interacting with Computers* 27 (3): 210–234. ISSN: 09535438. doi:10.1093/iwc/iwv001.

Garber, Carol Ewing, Bryan Blissmer, Michael R. Deschenes, Barry A. Franklin, Michael J. Lamonte, I. Min Lee, David C. Nieman, and David P. Swain. 2011. "Quantity and quality of exercise for developing and maintaining cardiorespiratory, musculoskeletal, and neuromotor fitness in apparently healthy adults: Guidance for prescribing exercise." *Medicine and Science in Sports and Exercise* 43 (7): 1334–1359. ISSN: 01959131. doi:10.1249/MSS.0b013e318213fefb. arXiv: arXiv:1011.1669v3.

Garrod, Simon, and Martin J. Pickering. 2004. "Why is conversation so easy?" *Trends in Cognitive Sciences* 8 (1): 8–11. ISSN: 13646613. doi:10.1016/j.tics.2003.10.016.

Gustavo, L, Luis Quesada, and Luis A Guerrero. 2018. "Advances in Human Factors and Systems Interaction." 592. doi:10.1007/978-3-319-60366-7.

Hendrickson, Scott, Stephen Sturdevant, Tyler Harter, Venkateshwaran Venkataramani, Andrea C Arpaci-dusseau, and Remzi H Arpaci-dusseau. 2016. "Serverless computation with openlambda." *Elastic* 60:80.

Jeffress, Justin. 2018. *Using Dialog Management to Capture A and B or C Slots.* Accessed July 2. https://developer.amazon.com/blogs/alexa/post/44dd83f4-4842-40c5-91f8-3868b9f4608c/using-dialog-management-to-capture-a-and-b-or-c-slots.

Klein, Gary a, Judith Orasanu, and Caroline E Zsambok. 1994. *acta psychologica Book reviews.* Vol. 85. Elsevier Science B.V. ISBN: 000l-6918(93)E0049-8.

Kuhl, Patricia K. 2007. "Is speech learning 'gated' by the social brain?" *Developmental Science* 10 (1): 110–120. ISSN: 1363755X. doi:10.1111/j.1467-7687.2007.00572.x. arXiv: NIHMS150003.

Kumar, Anjishnu, Arpit Gupta, Julian Chan, Sam Tucker, Bjorn Hoffmeister, and Markus Dreyer. 2017. "Just ASK: Building an Architecture for Extensible Self-Service Spoken Language Understanding": 1–13. arXiv: 1711.00549. http://arxiv.org/abs/1711.00549.

Liu, Weiyuan. 2010. "Natural user interface - Next mainstream product user interface." *2010 IEEE 11th International Conference on Computer-Aided Industrial Design and Conceptual Design, CAID and CD'2010* 1:203–205. doi:10.1109/CAIDCD.2010.5681374.

Pradhan, Alisha, Kanika Mehta, and Leah Findlater. 2018. "" Accessibility Came by Accident ": Use of Voice-Controlled Intelligent Personal Assistants by People with Disabilities." *Chi 2018.* doi:10.1145/3173574.3174033.

Regan, Michael A, and David L Strayer. 2014. "Towards an Understanding of Driver Inattention: Taxonomy and Theory." *Annals of Advances in Automotive Medicine* 58:5–14. ISSN: 1943-2461. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4001671/.

Santos, João, Joel J.P.C. Rodrigues, Bruno M.C. Silva, João Casal, Kashif Saleem, and Victor Denisov. 2016. "An IoT-based mobile gateway for intelligent personal assistants on mobile health environments." *Journal of Network and Computer Applications* 71:194–204. ISSN: 10958592. doi:10.1016/j.jnca.2016.03.014. http://dx.doi.org/10.1016/j.jnca.2016.03.014.

Simmons, Sarah M., Jeff K. Caird, and Piers Steel. 2017. "A meta-analysis of in-vehicle and nomadic voice-recognition system interaction and driving performance." *Accident Analysis and Prevention* 106 (April): 31–43. ISSN: 00014575. doi:10.1016/j.aap.2017.05.013.

Smolyak, Daniel, Bongshin Lee, and Eun Kyoung Choe. 2018. "TandemTrack." *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18:* 1–6. doi:10.1145/3170427.3188536. http://dl.acm.org/citation.cfm?doid=3170427.3188536.

Suelze, B, R Agten, P B Bertrand, T Vandenryt, R Thoelen, P Vandervoort, and L Grieten. 2013. "Waving at the Heart: Implementation of a Kinect-based real-time interactive control system for viewing cineangiogram loops during cardiac catheterization procedures." *Computing in Cardiology 2013:* 229–232. ISSN: 0276-6574.

Susanne Wolf, Verfasserin, Univ Bernhard Schwarz, and Hon Bernhard Rupp. 2013. "Verhalten unter Einbeziehung neuer Medien." PhD diss. http://www.fgoe.org/projektfoerderung/gefoerderte-projekte/FgoeProject%7B%5C_%7D3496/75894.pdf.

Ward, Adrian F., Kristen Duke, Ayelet Gneezy, and Maarten W. Bos. 2017. "Brain Drain: The Mere Presence of One's Own Smartphone Reduces Available Cognitive Capacity." *Journal of the Association for Consumer Research* 2 (2): 140–154. ISSN: 2378-1815. doi:10.1086/691462. arXiv: arXiv:1011.1669v3. http://www.journals.uchicago.edu/doi/10.1086/691462.

Appendix

A Git-Repository

Link to the repository:

https://gitlab.mediacube.at/fhs38521/bac2-alexa-skill

B Archived Websites

https://www.statista.com/statistics/565947/predicted-number-of-smartphone-users-eu-austria/, 2018-07-04

https://www.mindtake.com/de/press-release/%C3%B6sterreicher-nutzen-mobiltelefon-mehr-als-drei-stunden-pro-tag, 2018-07-30

http://www.who.int/dietphysicalactivity/factsheet_adults/en/, 2018-07-04 http://www.marktmeinungmensch.de/studien/nutzung-von-gesundheitsapps-auf-dem-smartphone-in-/, 2018-07-04

https://de.statista.com/outlook/313/128/fitness/oesterreich#market-users, 2018-07-04

https://www.freeletics.com/en/press/wp-content/uploads/sites/24/2017/04/Freeletics_PressKit_DE_06.04.2017_web.pdf, 2018-07-04

http://apps.who.int/iris/bitstream/10665/80135/1/9789241501507_eng.pdf, 2018-03-26

http://www.bmi.gv.at/202/Verkehrsangelegenheiten/unfallstatistik_vorjahr.aspx, 2018-03-26

https://de.statista.com/infografik/12369/weltweiten-absatz-von-smart-speakern/, 2018-06-22

https://assistant.google.com/#?modal_active=none, 2018-06-22

https://developers.google.com/actions/extending-the-assistant, 2018-03-26

https://www.microsoft.com/en-us/download/details.aspx?id=47715, 2018-03-26

https://voicebot.ai/2017/07/14/timeline-voice-assistants-short-history-voice-revolution/, 2018-07-31

https://www.apple.com/homepod/, 2018-03-26

https://www.apple.com/ios/siri/, 2018-07-31

https://www.amazon.com/gp/product/B0749WVS7J/ref=ods_ac_dp_dr_ps, 2018-03-26

https://developer.amazon.com/alexa, 2018-07-31

https://developer.amazon.com/alexa-skills-kit, 2018-03-26

https://www.voicebot.ai/wp-content/uploads/2017/11/cirp-news-release-2017-11-06-echo-home.pdf, 2018-03-29

https://www.amazon.de/gp/product/B0754C735Z?ref=skillrw_dsk_pnps_dp_3, 2018-06-29

https://www.amazon.de/cubic-ai-5-Minuten-Plank-Training/dp/B06XHTCB3Z, 2018-06-29

https://gitlab.mediacube.at/fhs38521/bac2-alexa-skill, 2018-06-29

https://developer.amazon.com/alexa-skills-kit/alexa-skill-quick-start-tutorial, 2018-04-12

https://docs.aws.amazon.com/lambda/latest/dg/welcome.html, 2018-06-29

https://developer.amazon.com/alexa, 2018-04-12

https://developer.amazon.com/de/alexa-skills-kit, 2018-06-22

https://developer.amazon.com/alexa-skills-kit/asr, 2018-04-03

https://developer.amazon.com/docs/custom-skills/dialog-interface-reference.html, 2018-04-15

https://docs.aws.amazon.com/general/latest/gr/aws-arns-and-namespaces.html, 2018-06-07

https://developer.amazon.com/docs/smapi/quick-start-alexa-skills-kit-command-line-interface.html, 2018-06-10

https://developer.amazon.com/docs/custom-skills/certification-requirements for-custom-skills.html, 2018-07-31

http://docs.bespoken.io/en/latest/, 2018-06-07

https://developer.amazon.com/blogs/alexa/post/44dd83f4-4842-40c5-91f8-3868b9f4608c/using-dialog-management-to-capture-a-and-b-or-c-slots, 2018-07-02

https://developer.amazon.com/blogs/post/Tx213D2XQIYH864/announcing-the-alexa-skills-kit-for-node-js, 2018-07-02

https://aws.amazon.com/cloudwatch, 2018-07-02

https://developer.amazon.com/docs/custom-skills/home-sounds.html, 2018-07-02

https://www.freeletics.com/de/pages/terms, 2018-08-01

C Templates for the study material

Einverständniserklärung für die Teilnahme an einer Benutzerstudie

I. Zweck der Teilnahme

Ich nehme freiwillig an der Benutzerstudie durchgeführt von Theresa Höck der Fachhochschule Salzburg GmbH teil. Die Studie wird ca. 20 Minuten dauern.

Ziel des Forschungsprojektes ist es, die Funktionalität sowie die Benutzerfreundlichkeit des Alexa Skills BleibFit zu messen.

II. Regelungen und Durchführung

1. Meine Teilnahme an dieser Studie ist freiwillig. Mir ist bekannt, dass ich keine Bezahlung für meine Teilnahme erhalte. Ich kann meine Teilnahme an der Studie jederzeit widerrufen oder die Studie beenden, ohne dass mir hierdurch ein Nachteil entsteht. Wenn ich mich dazu entscheide, der Teilnahme zu widersprechen oder die Studie zu beenden, wird niemand außerhalb der Forschungsgruppe / Studiengruppe von dieser Entscheidung erfahren.

2. Ich kann jederzeit auch nach Beendigung der Studie meine Zustimmung zur Verwendung meiner erhobenen Daten ohne Angabe von Gründen zurückziehen und eine Löschung derselben verfügen.

III. Datenschutz

4. Die während der Studie gesammelten Daten werden nach Abschluss statistisch ausgewertet, wobei keine persönlichen Daten der TeilnehmerInnen, die eine Identifizierung mit legalen Mitteln ermöglichen würden, gespeichert werden. Die ausgewerteten Daten werden für wissenschaftliche Zwecke (z.B. Publikationen) genutzt, wobei weiterhin eine vollständige Anonymität aller TeilnehmerInnen gewährleistet ist.

Ich habe die Einverständniserklärung gelesen und verstanden. Alle meine Fragen wurden zu meiner Zufriedenheit beantwortet und ich stimme hiermit der freiwilligen Teilnahme an der Studie sowie der Verwendung der anonymisierten Daten zu Forschungszwecken zu. Weiters habe ich eine Kopie dieser Einverständniserklärung erhalten.

Name der Testperson in Blockbuchstaben	Datum, Unterschrift
Datum, Unterschrift der Interview- / TestleiterIn	**Weitere Informationen:** Office Studiengang MultiMediaTechnology Fachhochschule Salzburg GmbH office@fh-salzburg.ac.at

Figure 16: Consent Form

Bitte vergegenwärtigen Sie sich jetzt noch ein Mal die Nutzung des Produkts und **beschreiben Sie Ihr Erleben der Nutzung mit Hilfe der folgenden Aussagenpaare.** Die Paare stellen jeweils extreme Gegensätze dar, zwischen denen eine Abstufung möglich ist.

Vielleicht passen einige Aussagen nicht so gut, kreuzen Sie aber trotzdem bitte immer an, welcher Begriff Ihrer Meinung nach eher zutrifft. Denken Sie daran, dass es keine "richtigen" oder "falschen" Antworten gibt - nur Ihre persönliche Meinung zählt!

	1	2	3	4	5	6	7			
Bei der Nutzung (des Produkts)...										
...handelte ich überlegt	☐	☐	☐	☐	☐	☐	☐	...handelte ich spontan		G_01
...erreichte ich mein Ziel nur mit Anstrengung	☐	☐	☐	☐	☐	☐	☐	...erreichte ich mein Ziel mit Leichtigkeit		M_01
...handelte ich unbewusst, ohne lange über die einzelnen Schritte nachzudenken	☐	☐	☐	☐	☐	☐	☐	...führte ich bewusst einen Schritt nach dem anderen aus	P	G_02
...ließ ich mich von meinem Verstand leiten	☐	☐	☐	☐	☐	☐	☐	...ließ ich mich von meinem Gefühl leiten		G_03
...war ich orientierungslos	☐	☐	☐	☐	☐	☐	☐	...konnte ich mich gut zurechtfinden		M_02
...handelte ich ohne dabei nachzudenken	☐	☐	☐	☐	☐	☐	☐	...konnte ich jeden Schritt genau begründen	P	G_04
Die Nutzung (des Produkts)...										
...erforderte viel Aufmerksamkeit	☐	☐	☐	☐	☐	☐	☐	...ging wie von selbst		M_03
...war begeisternd	☐	☐	☐	☐	☐	☐	☐	...war unbedeutend	P	X_01
...war einfach	☐	☐	☐	☐	☐	☐	☐	...war schwierig	P	M_04
...war nichts Besonderes	☐	☐	☐	☐	☐	☐	☐	...war ein magisches Erlebnis		X_02
...war sehr intuitiv	☐	☐	☐	☐	☐	☐	☐	...war gar nicht intuitiv	P	INT_01
...war belanglos	☐	☐	☐	☐	☐	☐	☐	...war mitreißend		X_03
...fiel mir leicht	☐	☐	☐	☐	☐	☐	☐	...fiel mir schwer	P	M_05
...war faszinierend	☐	☐	☐	☐	☐	☐	☐	...war trist	P	X_04
Im Nachhinein...										
...fällt es mir schwer, die einzelnen Bedienschritte zu beschreiben	☐	☐	☐	☐	☐	☐	☐	...ist es für mich kein Problem, die einzelnen Bedienschritte zu beschreiben		V_01
...kann ich mich gut an die Bedienung erinnern	☐	☐	☐	☐	☐	☐	☐	...fällt es mir schwer, mich zu erinnern, wie das Produkt bedient wird	P	V_02
...kann ich nicht sagen, auf welche Art und Weise ich das Produkt bedient habe	☐	☐	☐	☐	☐	☐	☐	...kann ich genau sagen, auf welche Art und Weise ich das Produkt bedient habe		V_03

Figure 17: INTUI Questionnaire

Juni 2018 - Interactive Sprachinteraktion mit intelligentem, persönlichem Assistenten (IPA), um Technologie und Gesundheit in Balance zu halten - Fragebogen zur Bachelorarbeit von Theresa Höck

Dieser Fragebogen wird im Zuge der Bachelorarbeit, verfasst von Theresa Höck mit dem englischen Titel: “Voice interaction with fitness application keeping technology and health in balance”, durchgeführt. Alle Antworten bleiben anonym und dienen allein dazu, die oben genannte Annahme zu beweisen. Hierbei handelt es sich um eine wissenschaftliche Arbeit, daher bestätigen Sie mit dieser Teilnahme, dass alle Angaben vollständig und richtig sind.

1) Bitte geben Sie Ihr Geschlecht an:

▢ weiblich ▢männlich ▢keine Angabe

2) Wie alt sind Sie? _____

3) Haben Sie schon einmal von einem Amazon Echo Dot gehört?

▢ ja ▢ nein ▢ weiß nicht

4) Haben Sie schon einmal einen Amazon Echo Dot verwendet?

▢ ja schon öfter ▢ ja einmal ausprobiert ▢ nein ▢ weiß nicht

5) Besitzen Sie einen Echo Dot ?

▢ ja ▢ nein ▢ weiß nicht

6) Haben Sie andere sprachgesteuerten Geräte oder Systeme verwendet?

▢ ja ▢ nein ▢ weiß nicht

5.1) Wenn ja, welche sind das (max. 3)? _____

7) Sind Sie der Meinung, dass Sie sich momentan ausreichend bewegen? (bitte nur eins ankreuzen)

trifft sehr zu	trifft eher zu	weder noch	trifft eher nicht zu	trifft nicht zu
▢	▢	▢	▢	▢

8) Wieviele Stunden pro Woche sind Sie durchschnittlich regelmäßig körperlich aktiv, also z.B. - schnell Spazieren gehen, Laufen, Schwimmen, Radfahren?

0 - 1 h/Woche	1 - 2,5 h/Woche	mehr als 2,5 h/Woche	mehr als 5 h/Woche
▢	▢	▢	▢

Figure 18: 2nd Questionnaire p.1

Juni 2018 - Interactive Sprachinteraktion mit intelligentem, persönlichem Assistenten (IPA), um Technologie und Gesundheit in Balance zu halten - Fragebogen zur Bachelorarbeit von Theresa Höck

9) Möchten Sie derzeit prinzipiell etwas an Ihrem Bewegungsverhalten ändern?

- ▢ ja
- ▢ nein, weil ich mit meinem Bewegungsverhalten zufrieden bin und es so beibehalten möchte
- ▢ nein, obwohl ich mit meinem Bewegungsverhalten eigentlich nicht zufrieden bin

10) Was möchten Sie ändern? (Mehrfachnennung möglich)

- ▢ mehr Bewegung im Alltag
- ▢ mehr Ausdauertraining
- ▢ mehr Krafttraining
- ▢ andere Sportart
- ▢ weniger Bewegung
- ▢ sonstiges: _____

11) Warum möchten Sie das ändern?

- ▢ gesundheitliche Gründe
- ▢ Aussehen
- ▢ Freude an Bewegung
- ▢ um Neues auszuprobieren
- ▢ Gruppenzwang
- ▢ 'Influencing' auf Social Media
- ▢ sonstiges: _____

12) Denken Sie, dass Ihnen eine App wie z.B. ein Schrittzähler oder ein persönliches Fitness-Tagebuch am Handy helfen bzw. Sie motivieren könnte, Ihr Bewegungsverhalten zu ändern?

- ▢ ja, vielleicht
- ▢ ich weiß nicht
- ▢ nein, sicher nicht

13) Verwenden Sie zu diesem Zweck derzeit eine App?

▢ ja ▢ nein

14) Welches App ist das? (bei mehreren bitte das am häufigsten verwendete angeben)

_____

15) Können Sie in diese Apps Daten allein mit Ihrer Stimme speichern?

▢ ja ▢ nein

Figure 19: 2nd Questionnaire p.2

Juni 2018 - Interactive Sprachinteraktion mit intelligentem, persönlichem Assistenten (IPA), um Technologie und Gesundheit in Balance zu halten - Fragebogen zur Bachelorarbeit von Theresa Höck

16) Können Sie in diese Apps allein mit Ihrer Stimme navigieren?

▢ ja ▢ nein

17) Wieviele Stunden pro Tag nutzen Sie durchschnittlich Ihr Mobiltelefon (z.B. - Telefonieren, Whats App, Instagram, Facebook, Fotografieren etc.)

▢ weniger als 1 Stunde pro Tag
▢ 1 - 2,5 Stunden pro Tag
▢ 2,5 - 5 Stunden pro Tag
▢ mehr als 5 Stunden pro Tag

18) Fühlen Sie sich durch die Nutzung, oder bei der Nutzung ihres Mobiltelefons unausgeglichen z.B.: gestresst, gelangweilt?

trifft sehr zu	trifft eher zu	weder noch	trifft eher nicht zu	trifft nicht zu
▢	▢	▢	▢	▢

19) Möchten Sie derzeit prinzipiell etwas an dem Nutzungsverhalten Ihres Mobiltelefons ändern?

▢ ja
▢ nein, weil ich mit meinem Nutzungsverhalten zufrieden bin und es so beibehalten
▢ nein, obwohl ich mit meinem Nutzungsverhalten eigentlich nicht zufrieden bin

20) Was möchten Sie ändern? (Mehrfachnennung möglich)

▢ mein Mobiltelefon gar nicht mehr nutzen
▢ mein Mobiltelefon nur noch zum Telefonieren und SMS schreiben nutzen
▢ mein Mobiltelefon öfter ohne mobile Daten (Internet) verwenden
▢ mein Mobiltelefon mit gleichen Funktionen nutzen jedoch weniger oft.
▢ sonstiges _____

21) Warum möchten Sie das ändern?

▢ verlorene Zeit die ich anders nutzen könnte z.B. Hobbies, Arbeit
▢ Konzentrationsfähigkeit steigern
▢ um mehr Schlaf zu gewinnen
▢ um meine Umwelt wieder besser wahrzunehmen
▢ sonstiges _____

Figure 20: 2nd Questionnaire p.3

22) Denken Sie, dass Ihnen ein Sprachsystem wie z.B. Alexa helfen würde, Ihr Nutzungsverhalten zu ändern?

- ▢ ja, vielleicht
- ▢ ich weiß nicht
- ▢ nein, sicher nicht

23) Denken Sie, dass Sie das getestete System privat verwenden würden?

- ▢ ja, weil ich nicht von meinem Smartphone abgelenkt werden würde (z.B. Whats App Nachrichten, Instagram Benachrichtigungen, E-mails etc.)
- ▢ ja, weil mich die Stimme motiviert weiter zu machen
- ▢ ja, aus einem anderen Grund und zwar weil:

  _____
- ▢ nein, weil ich die Stimme als Trainer nicht mögen würde.
- ▢ nein, weil ich die Anweisungen nicht visualisiert vor mir hätte
- ▢ nein, aus einem anderen Grund und zwar weil:

  _____

Figure 21: 2nd Questionnaire p.4