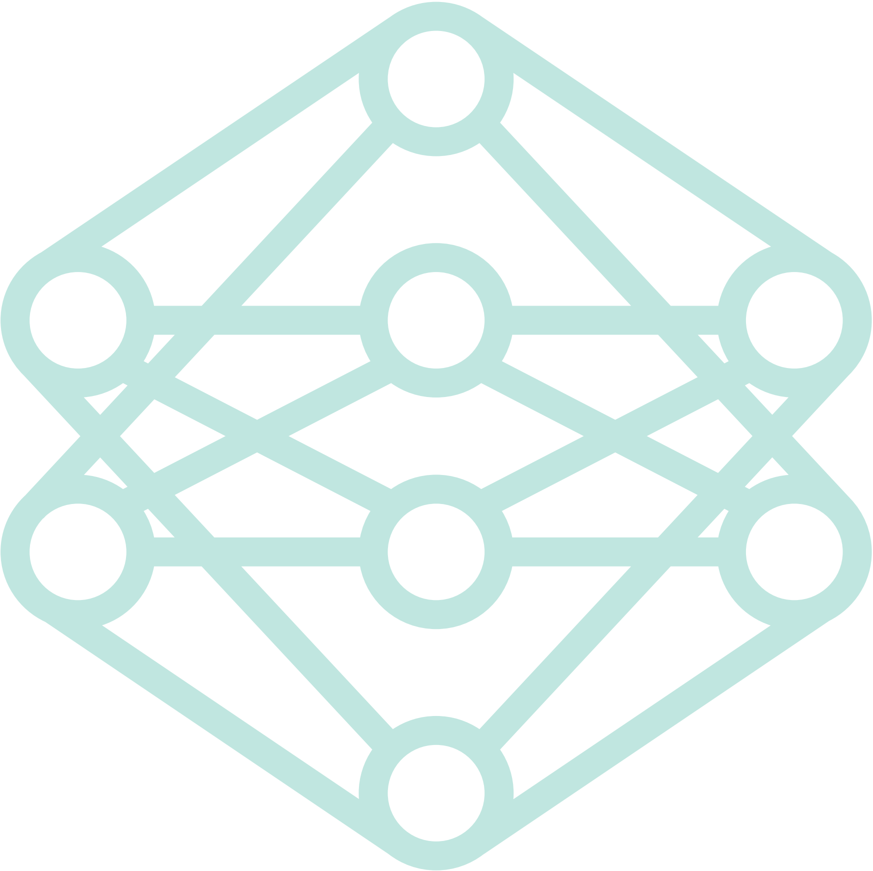
Defining Ally's Interactions

A Deep Learning framework to create a personalised interaction between users and a medical pod



Sathya Ranjani Rangarajan

Delft University of Technology

Defining Ally's Interactions

A Deep Learning framework to create a personalised interaction between users and a medical pod

Sathya Ranjani Rangarajan

Master Thesis
MSc Design for Interaction

Faculty of Industrial Design Engineering
Delft University of Technology

in association with
Cardiolab

Mentor: Quiel Beekman
Chair: Maaike Kleinsmann

March 2018

"Don't Panic"
-Douglas Adams,
(The Hitchhiker's Guide to the Galaxy)

“When something is important enough,
you do it even if the odds are not in your
favour.”
- Elon Musk

Acknowledgements

I would like to start by thanking my chair, Dr.ir.Maaike Kleinsmann for your support, constructive criticism and guidance throughout this thesis. I am thankful for the hours you spent helping me with my research. Next, I would like to thank my mentor i.r.Quiel Beekman. I am grateful to have had you as my mentor, because you have been a constant support and positive influence for me throughout this thesis. You have been not only a good coach, but also a great friend over the course of this thesis.

Now coming to my JMP team. Dino, the seed we sowed in the name of Ally has manifested itself in the form of my thesis. I can only thank you guys for it. Ana, Marta, Tijs, Lukas and Gaurav, I could not have asked for a better JMP team. Thanks for all the lame jokes and incessant gifs on the group.

Next I would like to thank Bhavna, Sricharan, Vaibhav, Pari, the people who brought back a bit of home in Delft. Harshini, Gayathri and Siva, thanks for letting me bounce ideas off you guys. And thanks for all the emotional support. I would also like to thank Elon Musk for his existence.

Lastly, Amma, Appa and Vijay, thanks for being the backbone of my life. This Master was only possible because of your love, support and encouragement...and the fees you paid for me.

Abstract

Ally, is an intelligent, voice activated medical device concept that users talk to on a routine basis to log details regarding their health and well-being. The goal of this thesis is to create a personalised interaction between users and Ally. The user group is segregated into three generations; Baby Boomers, Generation X and Millennials. A digital prototype of Ally is used to understand how different generations interact with Ally. A questionnaire to identify the voice characteristics each generation prefers in a medical device, is created. The results from the user test and questionnaire are used to design a Deep Learning framework to generate a WaveNet TTS voice. This framework is a foundation for a personalised interaction between Ally and users, based on the generation they belong to. By creating a framework to cater to specific generations, this model sets the ground rules for personalisation.

Glossary

Boundary Object

A boundary object is a 'thing' that is both defined enough that several communities can recognise it as the same thing, yet flexible enough that each community can use it according to their own needs. In the conceptual sense they can be abstract or concrete, but either way they exist outside of peoples' heads (Cooper-Wright, 2012).

Emotional Toolkit

A generative toolkit aimed at expressing emotions (Sanders & Stappers, 2012).

Explicit Knowledge

It can be stated in words, and is relatively easy to share with others (Sanders & Stappers, 2012).

Exploring Interactions

EI is a Master specific design projects for MSc Design for Interaction. The project involves analysis and design of interactions, the way people use, understand and experience products and situations. It involves designing experiential interactions scenarios for specific people and situations, exemplified through a product, a service, an environment, a thing, or any combination of these.

Latent Knowledge

This refers to thoughts and ideas that we haven't experienced yet, but on which we can form an opinion based on past experiences (Sanders & Stappers, 2012).

Levels of Knowledge

Knowledge refers to thoughts and ideas that have already been experienced and have been stored in memory. There are four levels: Explicit, Observable, Tacit and Latent. (Sanders & Stappers, 2012)

Observable Knowledge

It refers to thoughts and ideas that can be obtained by watching how things happen or how people behave (Sanders & Stappers, 2012).

Social Identity

Social identity is a person's sense of who they are based on their group membership(s) (McLeod, 2018).

Tacit Knowledge

This refers to things we know but are not able to verbally communicate to others (Sanders & Stappers, 2012).

Text to Speech

TTS refers to the ability of computers to read text aloud. A TTS Engine converts written text to a phonemic representation, then converts the phonemic representation to waveforms that can be output as sound. (Allen, Hunnicutt, Klatt, Armstrong & Pisoni, 1987)

Vocoders

A synthesizer that produces sounds from an analysis of speech input (Flanagan & Golden, 1966)

Waveforms

A waveform is the shape and form of a signal such as a wave moving in a physical medium or an abstract representation (Wei & Zhang, 2012).

Wizard of Oz

"In the field of human–computer interaction, a Wizard of Oz experiment is a research experiment in which subjects interact with a computer system that subjects believe to be autonomous, but which is actually being operated or partially operated by an unseen human being" (Harrington & Martin, 2012, p.204).

Table of Contents

Orientate

Ideate and Iterate

Create

About Cardiolab

Cardiolab is one of the Delft Design Labs, with its focus on the health continuum, from their pre-event (before the occurrence of a disease) through an acute event, diagnosis, treatment and post-discharge experience. It is a consortium between the IDE faculty of TU Delft, De Hartstichting and Philips Design. Within CardioLab, research of the various phases in the health continuum is conducted by IDE master students in collaboration with the members of the consortium. Their main area of focus is Cardiovascular Diseases (CVDs). This thesis, is a continuation of a Joint Master Project (JMP) done by Dino Design for Cariolab. Although focus of the JMP was on the pre-event part of the health continuum., the concept developed by Dino can be applied through all the stages of the health continuum. The focus of this thesis will be on detailing parts of the concept developed by Dino Design.

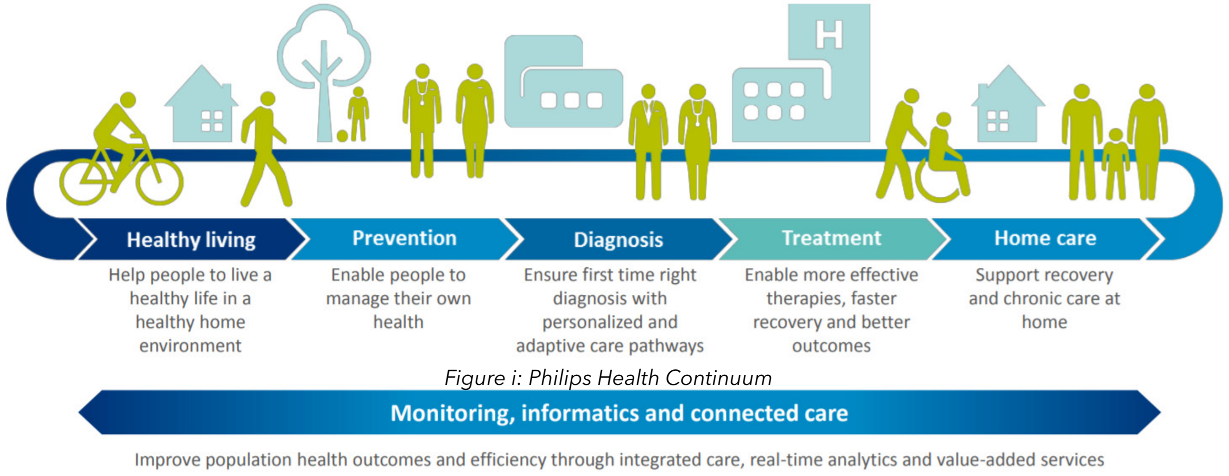

H
Healthy living
Help people to live a healthy life in a healthy home environment
Prevention
Enable people to manage their own health
Diagnosis
Ensure first time right diagnosis with personalized and adaptive care pathways
Treatment
Enable more effective therapies, faster recovery and better outcomes
Home care
Support recovery and chronic care at home
Monitoring, informatics and connected care
Improve population health outcomes and efficiency through integrated care, real-time analytics and value-added services

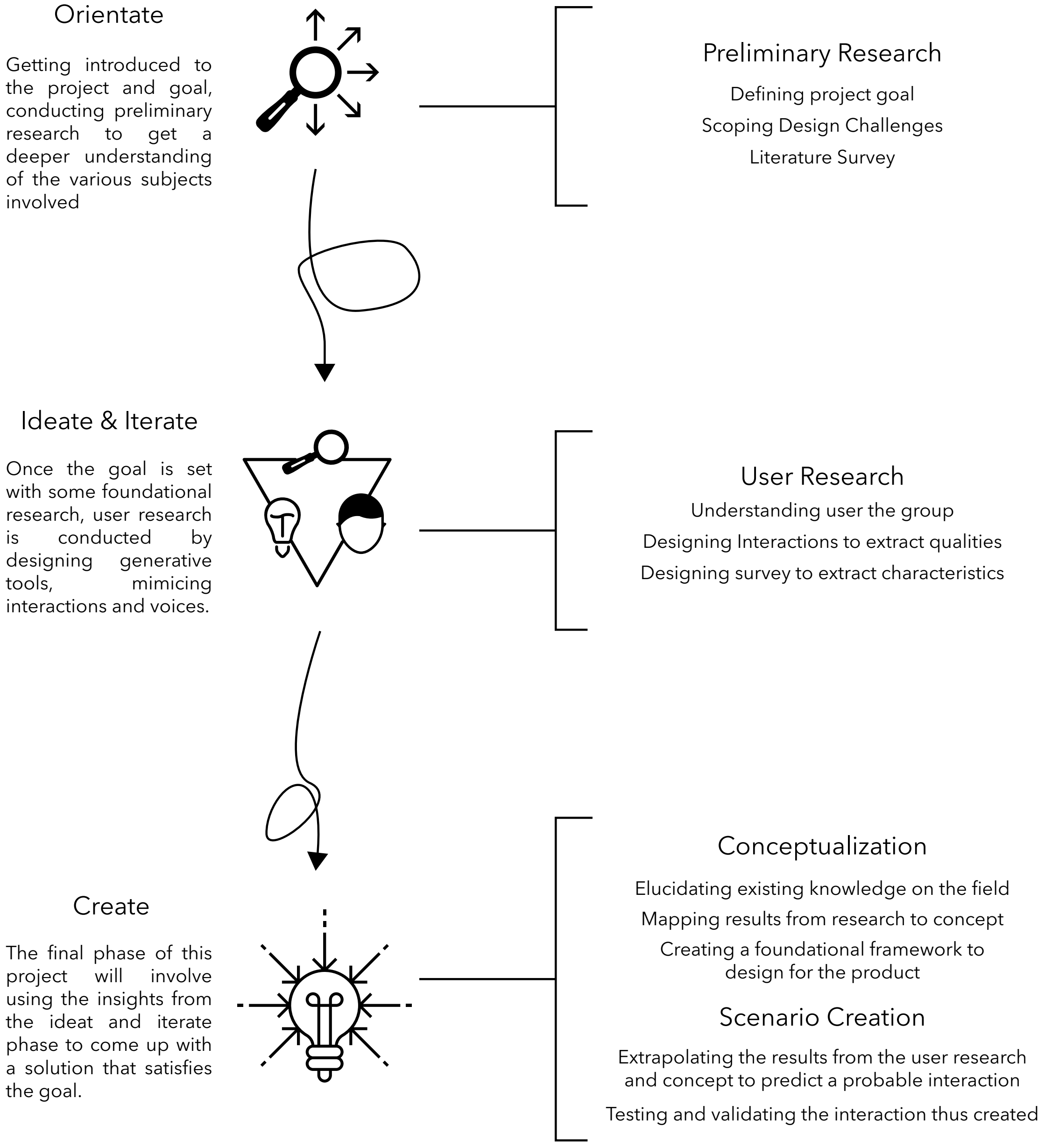
Figure i: Philips Health Continuum

About Dino Design

Dino Design was a six-member group comprising two students from each track (DfI, IPD and SPD) of the Master course at the faculty of IDE, TU Delft. The Joint Master Project (JMP) was carried out by Dino Design in association with de Hartstichting and Philips Design, as part of Cardiolab. Over the course of 20 weeks the team made an effort to try and find a way to detect strokes at its onset. It started with getting to know as much as possible about strokes through literature research, interviews with doctors, general practitioners and stroke survivors. Soon it became apparent that the product needed to be more than just a device that is able to detect strokes. This lead to the creation of Ally, an Artificial Intelligence device which uses Natural Language Processing to talk to the users about their health on a routine basis. The idea behind this is the assumption that by getting the users to interact with Ally on a routinely basis, the device can learn about their health by experience and predict risks and emergencies. Ally, as designed by Dino is at a very conceptual level, and the exact interaction is unknown. This thesis will discover the requirements to create an interaction between Ally and different users.

Project Process

Project Process: Inspiration based design research



Orientate

Getting introduced to the project and goal, conducting preliminary research to get a deeper understanding of the various subjects involved

Preliminary Research

Defining project goal

Scoping Design Challenges

Literature Survey

Ideate & Iterate

Once the goal is set with some foundational research, user research is conducted by designing generative tools, mimicing interactions and voices.

User Research

Understanding user the group

Designing Interactions to extract qualities

Designing survey to extract characteristics

Create

The final phase of this project will involve using the insights from the ideat and iterate phase to come up with a solution that satisfies the goal.

Conceptualization

Elucidating existing knowledge on the field

Mapping results from research to concept

Creating a foundational framework to design for the product

Scenario Creation

Extrapolating the results from the user research and concept to predict a probable interaction

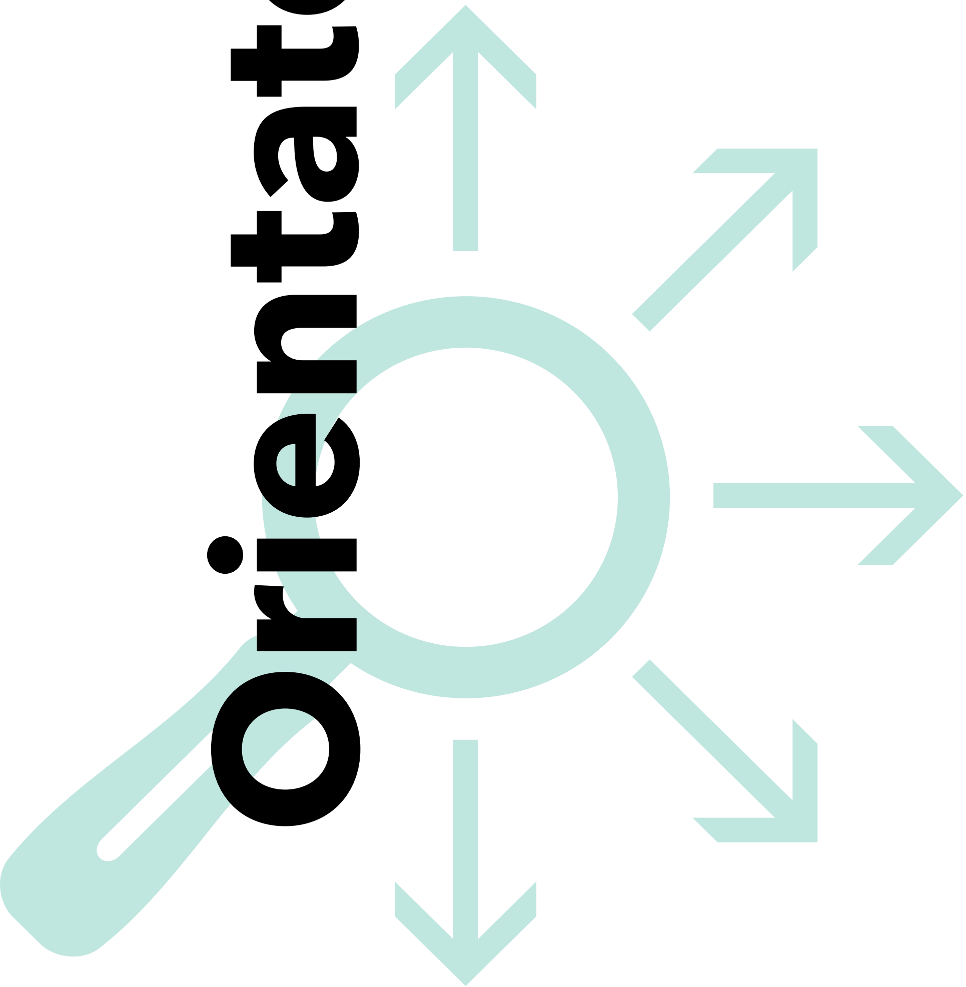
Testing and validating the interaction thus created

Research through design

In this project, the research through design approach is adopted. According to Zimmerman, Forlizzi and Evenson (2017) "Research through design is an approach to research that leverages the design process of repeated problem reframing as a method of scholarly inquiry. The work can result in the conceptual frameworks for design and evidence of the value of guiding philosophies for design." 493-502. Although adequate literature research has been conducted to get a deeper understanding of the topic we are dealing with, the user research conducted in this project uses a more designerly approach, where in generative toolkits, interactions and surveys are created to get a deeper understanding of the primary goal: defining an interaction.

Corroborating with literature, this approach in this project has indeed lead to the development of a framework which creates the ground rules when the product's interactions are created. Hence the approach throughout this project has been intuitive and explorative, rather than sticking to standard methodologies in evaluating user tests. This is reflective in both the project process mentioned here.

Orientate



01
Scoping and Preliminary Research

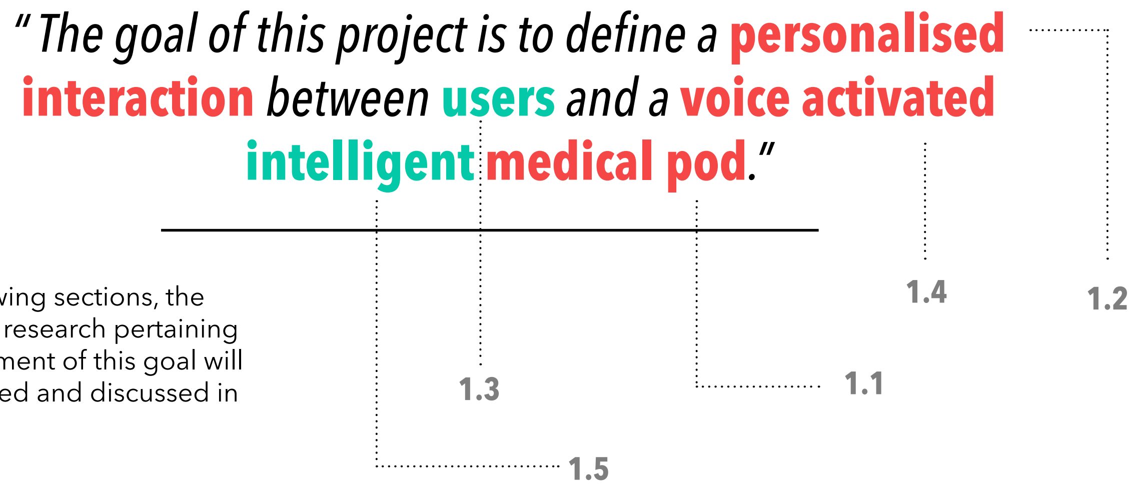
The goal of this project is established in this chapter. Then, a preliminary research with respect to the goal is conducted thereby leading to a hypothesis based on which further research is conducted. The various elements of the goal, like voice user interfaces, human robot interaction, users and Artificial Intelligence is discussed here.

Introduction

Dino Design was given the assignment to help users identify a stroke at its onset. After a lot of research, they conceptualised an intelligent voice activated medical device called Ally. Ally is a device users would interact verbally with on a routine basis. While voice activated intelligent devices like Amazon Alexa, Google Echo and Siri already exist, they serve more a more general purpose, like playing music, controlling smart homes, running searches etc. When it comes to confronting or speaking about health related issues, the barrier to communicate with a device might be higher. How comfortable people are to talk to a device about their health might vary based on factors such as age, tech savviness and other values. Since a verbally communicative intelligent device for medical care is a relatively new to the field of artificial intelligence, it is important to establish a foundation on which these interactions can be created by the device.

This brings us to the goal of this project.

Goal definition



"The goal of this project is to define a **personalised interaction** *between* **users** *and a* **voice activated intelligent medical pod**.*"*

In the following sections, the preliminary research pertaining to each element of this goal will be conducted and discussed in detail.

1.1
Medical pod: Ally the concept

In the context of this project, a medical pod refers Ally, the concept designed by Dino Design (JMP Final Report, 2017) as part of their JMP project.

Ally is a three piece package comprising a pod, a tracker (Figure 1) and a mobile application. The user interacts (verbally) with the pod at home, wears the tracker constantly and uses the application to gain an overview of the data collected. More details about each piece is given below.

The Ally pod creates a physical grounding in the house of the user. It can be placed anywhere they feel comfortable using the product. ***It is conceptualized as a personal device that the users talk to, to log their health and wellness related details. The pod helps with general and urgent medical questions and most importantly, can detect and act fast in case of an emergency. The system uses deep learning to predict patterns in users' health, which enables it to understand user needs and respond appropriately.*** Sharing through speech feels natural and we tend to share more information when we speak. Part of a conversation involves being heard, which is increasingly important for people that live alone. Thus, the pod captures subjective data (about their health) which can be linked to the objective data the bracelet measures (like heart rate, stress). These data are then analysed and predictions and conclusions about the users health are made by the device. Once the user gets used to sharing health related information with the device, the threshold of sharing such information decreases, and eventually in the case of emergency, Ally will be the first to be contacted.

The Tracker constantly collects data which can be used for personalized feedback through the pod or the application. ***It tracks the user's activity, by means of a HR sensor, movement using a tri-axis accelerometer, as well as their emotions using electrodermal sensors.*** These enable the collection of reliable data about the user's vitals. This, combined with the logs made by the user help in predicting and identifying patterns in the user's lifestyle.

The Application was conceptualised ***to help give the user an overview of their interactions with the Ally pod and the data collected by the tracker.*** The application provides a means of easy access to data and trends in a visualised manner. The user can also share the data with whoever they feel comfortable with, at the snap of a finger.

Why not just an App?

According to Dino Design, the Ally pod acts as a physical grounding, which has a specific task: To listen to the user and converse with them about their health. If this was made into an application, then the exclusivity would be lost, and it becomes a more generic product, which in turn might make users take it less seriously. Dino Design's assumption is that pod's physicality makes more trustworthy. Another aspect is that an App would be too heavy on the technicality front, as this uses Artificial Intelligence and NLP and might need more processing power.

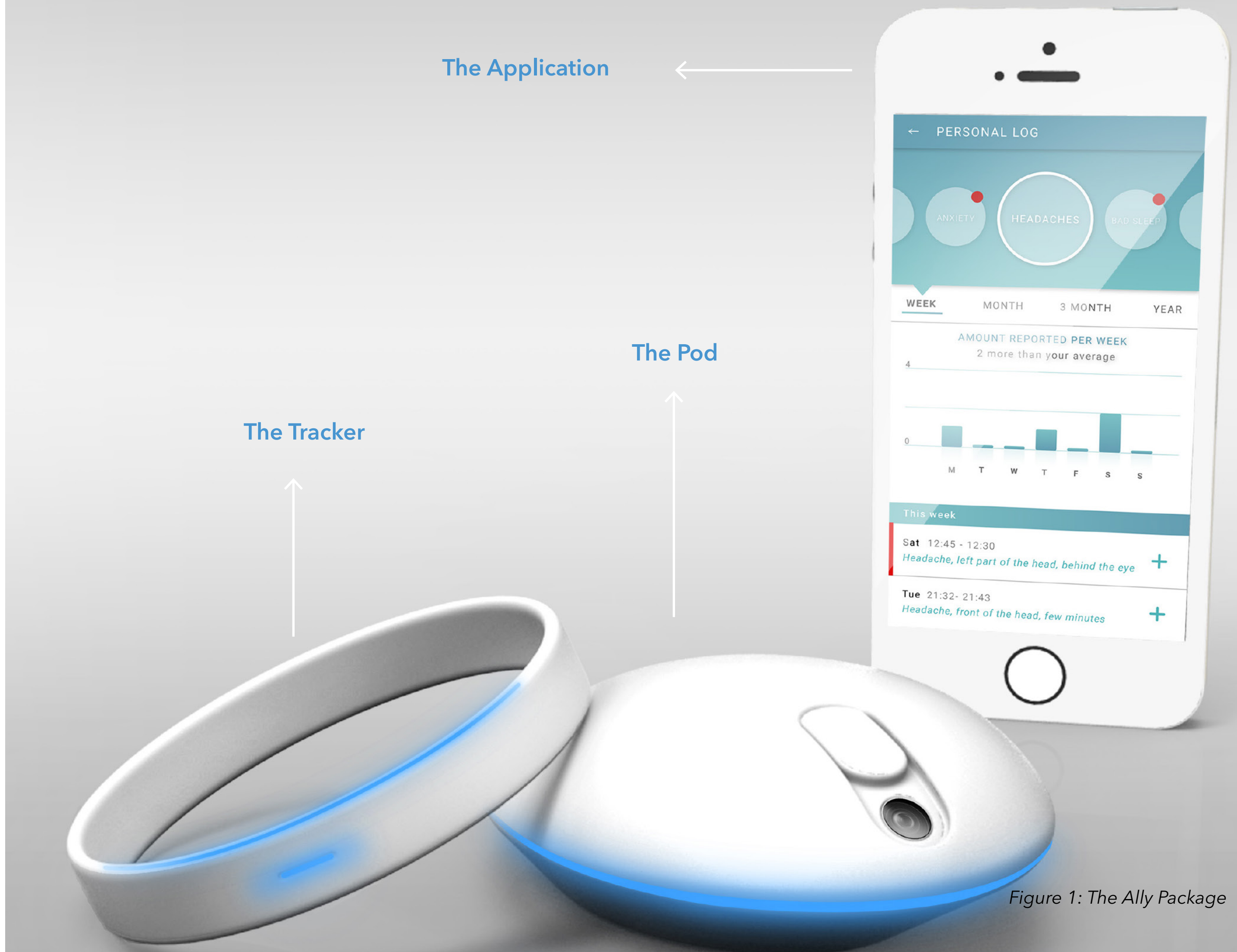


Figure 1: The Ally Package

An ecosystem consisting of all the elements surrounding this concept is given in Figure 2. As can be seen in the ecosystem, Ally pod is the center of the concept. The bracelet is constantly tracking heartbeat, activity and emotions to get a good overview of the day and accurate information as input for the Ally pod. Based on this information Ally pod can initiate certain questions to get more subjective data from the user, to learn their patterns. Ally pod can also compare the objective data from the bracelet with the information that is being shared with the pod. This deepens how the user feels and might make them more aware of their feelings. Looking critically at what the user feels increases the chance of discovering new things faster and action can be taken quicker. This action can be towards a family member, a medical professional or in case of emergency the ambulance. The vocal interaction is made possible by voice recognition and Natural Language Processing (NLP) which translates the speech to usable data which is useful information for Cardiolab. The current concept and drawback of elements in this ecosystem will be discussed in the following section.

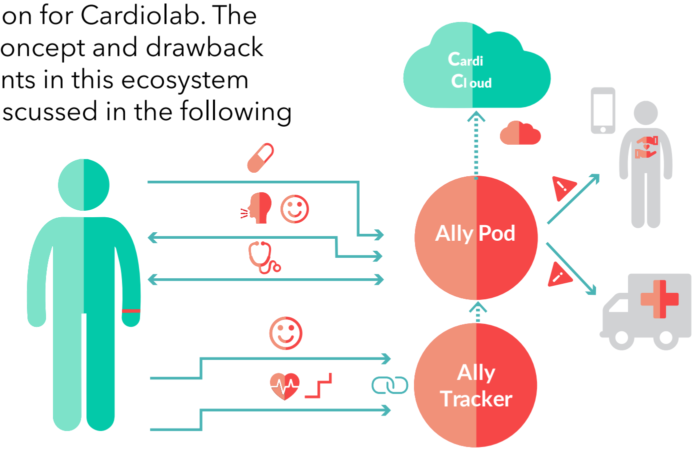
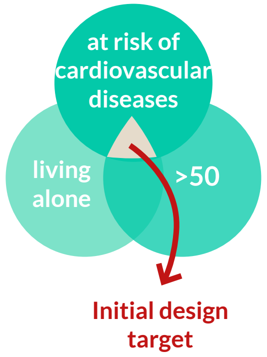
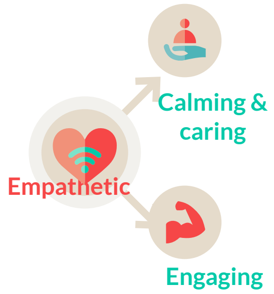


Figure 2: Ally's ecosystem

Drawbacks in Dino Design's concept

	Design Challenge	Dino's Concept
Users	Ally is a concept that can be used by a wide range of people and not just people with Cardiovascular Diseases (CVDs) who are over 50 and live alone. ***The target group needs to be broader and more accommodative, especially because De Hartstichting focuses on prevention of CVDs at an earlier age.*** This way, a wider range of diseases can be prevented across a larger target. Such a device has the capability to be used by multiple people, alone and together.	 at risk of cardiovascular diseases living alone >50 Initial design target Dino's concept is designed with the assumption that the target group is : people who are over 50 years old, at the risk of CVDs and live alone.
Interaction	The exact interaction between the user(s) and the device is unclear. How they approach it, talk to it, what they talk to it is not defined. There are no definitive approaches to how the conversation should proceed and no conversational design, leaving a huge gap in the entire interaction. ***Ally's design was at a very conceptual level, and hence, the focus of this project would be in detailing it ate a more grassroot level.***	3 use case scenarios of how the user might use the device was given by Dino Design 1. A scenario where the user walks up to Ally to talk 2. The user "vents" their frustration to Ally 3. Ally initiates a conversation when the user walks by.
Voice	Voice is the backbone of this device. There is not enough research or testing on how such qualities can be brought in a voice or ***what personalities people associate with voices.***	 In Dino's design, it is mentioned that the voice of the device should be empathetic, calming caring and engaging.

The design challenges in Dino's concept of Ally, with respect to the users, interaction and voice has been established. In the forthcoming sections, we will discuss what exactly interaction means in this context, how user group plays a role and what voice activation means in this project.

1.2
Defining the term “interaction”

Personalised Interaction refers to catering to a conversation or communication based on who the communicator is. To understand what it means in this project, let us split the “interactions” into three parts.

Human - Human Interaction
Human - Robot Interaction
Human - Inanimate Interaction

Human - Human Interaction:

To give context to this interaction in this project, we consider the interpersonal communication between physicians and patients. There are two contrasting styles of communication that is reflected by physicians during medical visits. These are: Affiliation and control (Buller & Buller, 1987). The different characteristics of both styles is given on the right.

"Affiliation is composed of communication behaviours designed to establish and maintain a positive relationship between the physician and patient."

"Control is to establish upper hand"

Research has proven that a physician who adopts a more affiliative communication style receives more favourable evaluations. Friendly treatment, positive affect and warm concern for patient’s worries produced greater overall satisfaction (Korsch & Negrete, 1972). Physicians who exerted high controlling communication style were seen as non-satisfactory. Patients who were more satisfied with physician’s communication are generally more satisfied with the healthcare (Buller & Buller, 1987).
Keeping this in mind, the user test in section 2.2 of Chapter 2, designed to test the interaction across different generations of users and Ally has been created with qualities that evoke affiliation. ***We will find out if what makes an interaction affiliative varies based on who is interacting.*** The following questions are addressed in that section.

- ***What are the qualities different user groups (generations) find affiliative and affective, when conversing with Ally?***
- ***How do users interact with Ally generally?***
- ***How do different users react when an emergency is predicted by the product?***

More about why the user group is segregated into generations will be discussed in section 1.3. Since Ally is not human, it is important to identify what they seek in a medical device, that makes their experience positive and satisfactory. This leads us to the next segment of this section, Human - Robot Interaction. Section 2.2.1 of Chapter 2 discusses the affiliative qualities of different generations that make their interaction satisfactory.

Interaction refers to the verbal communication between two entities (Dix, 2009)

Communication refers to “the way one verbally or paraverbally interacts to signal how literal meaning should be taken, interpreted, filtered or understood”. (pg 99-122, 1978).

Interpersonal Communication is an exchange of information between two or more people. Interpersonal communication occurs in every context of our life, at home, in school, at workplace etc. (Berger & Charles, 2008).

Affiliative Communication

Friendly
Encouraging
Verbally acknowlegement
Open and Honest
Empathetic
Attentive
Relaxed

Controlling Communication

Dominating
Verbal exaggration
Argumentative

Human - Robot Interaction:

The interaction between users and Ally can be classified as human robot interaction. To be more specific, a verbal interaction between a human and a machine.

While it is important to consider the elements that contribute to a satisfactory experience between a doctor and a patient, we must remember that Ally is not a doctor. Ally is only a machine that prompts the user to log their daily health related concerns and gives suggestions.

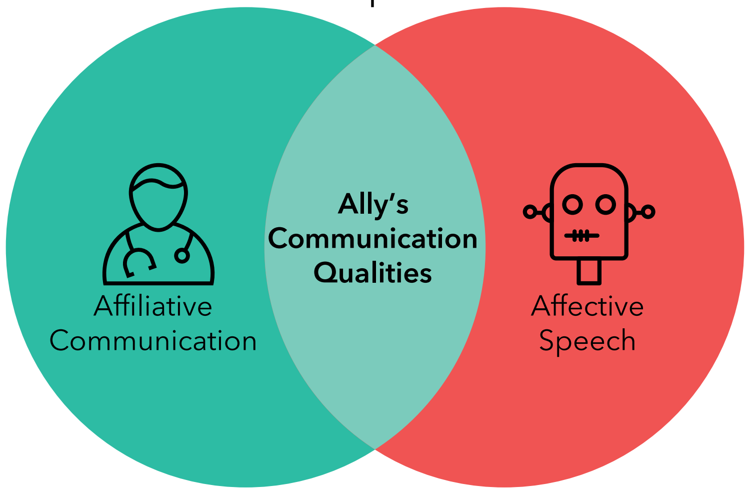
Humans are emotional beings. Affect/emotion is an important dimension of cognition. Speech generated by Ally needs to be affective.
Affective speech becomes relevant when talking about affective computing (sometimes called artificial emotional intelligence, or emotion AI) which is the study and development of systems and devices that can recognize, interpret, process, and simulate human affects.
(Kaliouby, 2017)

Why is affective speech important?

According to Mavridis (2015), “the affective dimension is very important in human interaction, because it is strongly intertwined with learning, persuasion, and empathy, among many other functions. Thus, it carries over its high significance for the case of human – robot interaction. For the case of speech, affect is marked both in the semantic/pragmatic content as well as in the prosody of speech: and thus both of these ideally need to be covered for effective human–robot interaction, and also from both the generation as well as recognition perspectives.” (p. 27)
The interaction between users and Ally or any intelligent

Affect is as defined in this context refers to the human characteristics that govern the various subtleties of intonation and phrasing, which reveal extra-linguistic and paralinguistic information about the speaker, about the speaker’s relationships with the hearer, and about the progress of the discourse and the degrees of mutual understanding attained throughout its progress (Campbell, 2008).

Prosody refers to “the patterns of stress and intonation in a language (Prosody, n.d.)



Computers showing emotions and expression.
+
Doctors’ communicating style that makes it friendly

Figure 3: Ally’s communication qalities

machine that is used by a layman needs to have a natural flow of language and conversation. From flexible manufacturing robots, household robotic assistants, assistive robots to companion robots, one common requirement in all of them is the desirability of natural, fluid interaction with humans. There is a need for supporting natural language and nonverbal communication. According to According to Mavridis (2015), non-expert humans, i.e users are used to interacting with other humans through a mixture of natural language and nonverbal signs. ***Thus building robots that let users interact naturally and fluidly collaborate with other humans would be easier for humans and also help capitalize people's ability to teach and interact with robots that are constantly learning and adapting.***

Thus ***to create a seamless interaction between users and Ally, the communication needs to be affective and affiliative.*** To create such a communication, it is important to identify the qualities that comprise such a speech. Figure 3 illustrates how Ally's communication forms when affective and affiliative communication are brought together. The qualities will be identified by the user tests conducted in section 2.2.

Before identifying the qualities in the interaction between the users and a robot (Ally), it is interesting to zoom out from the context and look at the interaction qualities between humans and inanimate objects. This would help identify affective elements in products, that make people talk to them. This bring us to the next segment, Human-Inanimate product interaction.

Human - Inanimate object Interaction:

In the previous segments, it was explained what sort of a communication users should have with Ally. Ally, is a product that talks. Without the talking aspect, Ally, is an inanimate object. The following section acts as a starting point to understand how users think of products in general, in terms of how they invest emotions. The results from this section are used to incorporate affective communication in user tests of section 2.2 The research approach in this section is illustrated in Figure 4. ***By identifying the kind of inanimate products people feel comfortable expressing to, and the qualities they attribute to such products, we can gain more insight on how Ally can be made to reflect such characteristics.***

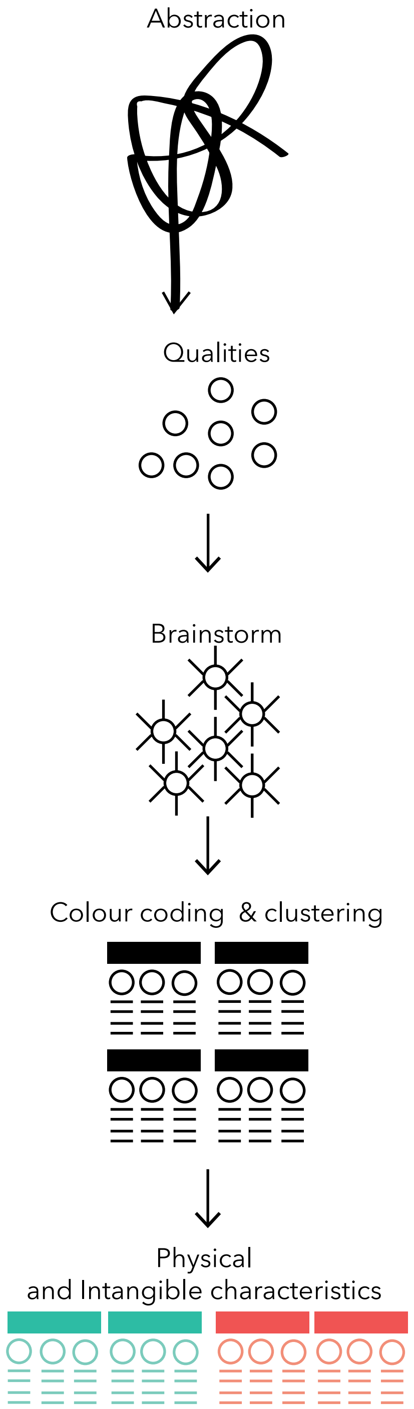
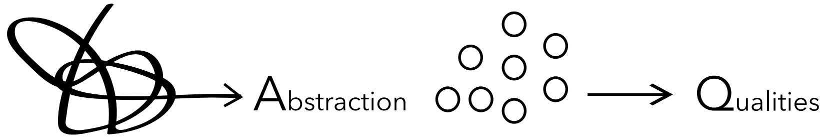


Figure 4: Process of discovering interaction qualities between humans and inanimate objects

In this approach, the problem at hand is abstracted and the context is broadened. The results from the abstraction as later converged through a brainstorm and clustered to get useful insights. This method is often used in Exploring Interactions.



As discussed, we zoom out of the current goal and try to discover elements that make people comfortable with interacting with inanimate products in general. 10 people were asked the question given on the left hand side. The answers to the question ranged from soft toys, mirrors, cell phones, specific corners in their homes, where people felt comfortable to express themselves in solitude. Some quotes by the people are to the right.

It was found that people started attributing personalities to these objects overtime, breaking the barrier of talking and feeling at ease in front of the inanimate object. From the answers received for the question, qualities mentioned by the participants were extracted. The qualities are highlighted in the quotes given. The product qualities as described by the people is given in Figure 5.

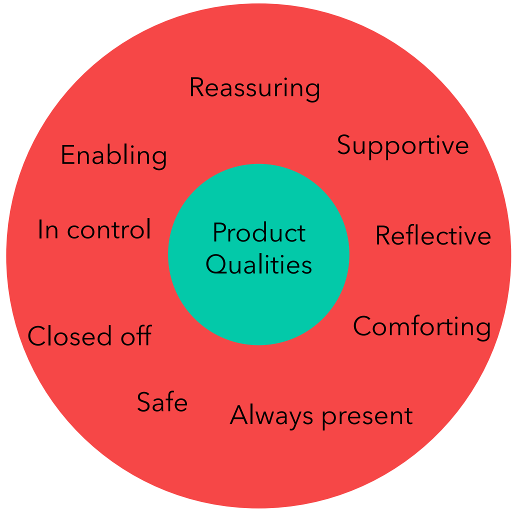
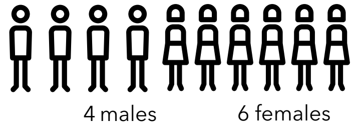


Figure 5: Product qualities as given by users

Question: *"As kids, we've had soft toys or toys that we could just talk to, feel comfortable in expressing things to around them, even if they are inanimate. Are there any such products you can think of? It could be relevant even now. And it need not necessarily be a product, it could even be a place."*

Age Range: 20-30



"I started giving him (tiger soft toy) personalities. Richard parker makes me feel safe, like everything's alright"

"I could hug it without being poked, was comforting and gave me a a sense of reassurance"

"I liked resting my back on that small spot and thinking because it gave me support for the back and was closed off"

"I cry in front of the mirror, and I don't like to see the sad face and that stops me from crying"

"The mirror is reflective in a literal sense but also in a figurative sense. It gives me power to control my emotions"

"The phone is always present, and is enabling, selflessly at my disposal"

Brainstorm

Using the qualities given in the previous section as a starting point, a brainstorm session was conducted which resulted in creating possible product features that could reflect these qualities. This brainstorm can be found in Appendix A.1

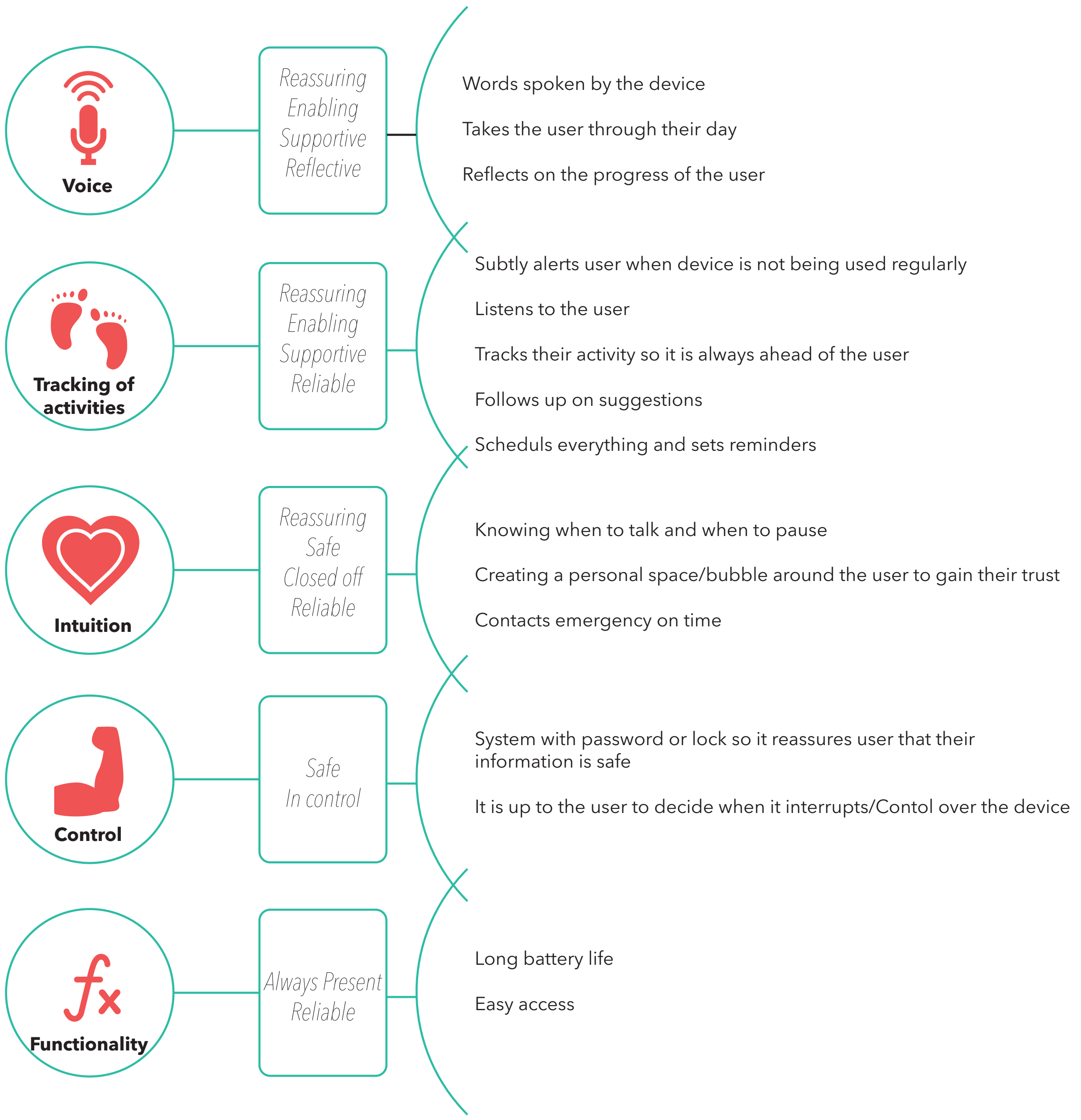
Colour coding and clustering

The results of the brain storm were colour coded in order to cluster them into refined product characteristics. After this, a new chart was made with all the characteristics and qualities placed in their respective clusters. This can be found in Appendix A.1.

Possible Intangible & Physical characteristics

As a last step and end result of this baseline, the clustered qualities and characteristics were classified into possible Physical and Intangible characteristics. Physical characteristics refer to the possible product features that had to do with the physical grounding, surrounding and tangible aspects of the Ally pod. This can be found in Appendix A.1. Intangible characteristics refer to the product functionalities that aren't physical, but relating to functional aspects (Figure 6). Since this project deals only with creating an interaction, only the parts of "Possible intangible characteristics" is considered.

Possible Intangible Characteristics



Voice

Reassuring
Enabling
Supportive
Reflective

Words spoken by the device

Takes the user through their day

Reflects on the progress of the user

Tracking of activities

Reassuring
Enabling
Supportive
Reliable

Subtly alerts user when device is not being used regularly

Listens to the user

Tracks their activity so it is always ahead of the user

Follows up on suggestions

Scheduls everything and sets reminders

Intuition

Reassuring
Safe
Closed off
Reliable

Knowing when to talk and when to pause

Creating a personal space/bubble around the user to gain their trust

Contacts emergency on time

Control

Safe
In control

System with password or lock so it reassures user that their information is safe

It is up to the user to decide when it interrupts/Contol over the device

Functionality

Always Present
Reliable

Long battery life

Easy access

Figure 6: Possible Intangible characteristics

1.3 Who are the users?

In the previous section, we discussed what kind of interaction we need to establish between the users and Ally. When we say users, who exactly are we talking about? It is clear from section 1.1, that Dino Design's user group is limited and that it needs to be extended. Since Ally is a medical device which can be used for various purposes, from tracking health to fitness and wellness, the user group can be expanded to almost all adults.

But, in order to create personalised interactions, based on who the user is, it is important to segregate the user groups in some way. According to Stewart & Blanchard (2008), ***, “language use varies by age, sex, and other socially-defined groups, as well as contexts within which communication takes place.”***. Technology has been constantly evolving and how different people adapt to it varies. ***Technology acceptance and adoption is influenced by age, and this difference can also be classified generationally, due to the difference in values*** (Morris & Venkatesh, 2000).

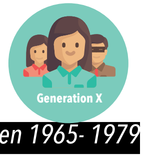
When speaking of generations, the most relevant at present are 3 generations: Baby Boomers, Generation X and Millennials. Before we go into to understanding different generations, we must keep in mind that there might be differences in attitudes, values, behaviors, and lifestyles within a generation as there are between generations. But that does not diminish the value of generational analysis; it merely adds to its richness and complexity (Taylor and Keeter, 2010). Figure 7 illustrates the values of each generation.

Since the scope of this project is related to healthcare, let us look at the acceptance of healthcare technologies across different ages. Arning and Ziefle (2009) conducted an experiment to test the acceptance of an E-Health system across young (Millennials and Gen X) and old people (Baby boomers); it was found that it gained a lot of acceptance in both categories, with the older age group being more open to using such technology for themselves, and the younger group realizing the importance of the E-Health system for others, rather than themselves. ***According to Gaul and Ziefle, reliability of technology is a barrier that is common across all generations of users in using E-Health devices (2009). Can this be tackled by Ally through effective communication?*** In healthcare, communicating effectively is perceived as a core competency for patient-centred collaborative practice (Suter et al; 2009). Effective communication helps

“Generation” is also often used synonymously with cohort in social science; under this formulation it means “people within a delineated population who experience the same significant events within a given period of time”. (Pilcher, 1994).

Born between 1946 - 1964

Born after World War II, boomers are widely associated with privilege, as many grew up in a time of widespread government subsidies in post-war housing and education, and increasing affluence (Wikipedia, Baby boomers, 2018)



Born between 1965- 1979

Generation X-ers were children during a time of shifting societal values and as children were sometimes called the “latchkey generation”, due to reduced adult supervision as children compared to previous generations (Wikipedia, Generation X, 2018)

Born between 1980 - 1995

This generation is generally marked by an increased use and familiarity with communications, media, and digital technologies. In most parts of the world, their upbringing was marked by an increase in a liberal approach to politics and economics (Wikipedia, Millennials, 2018)

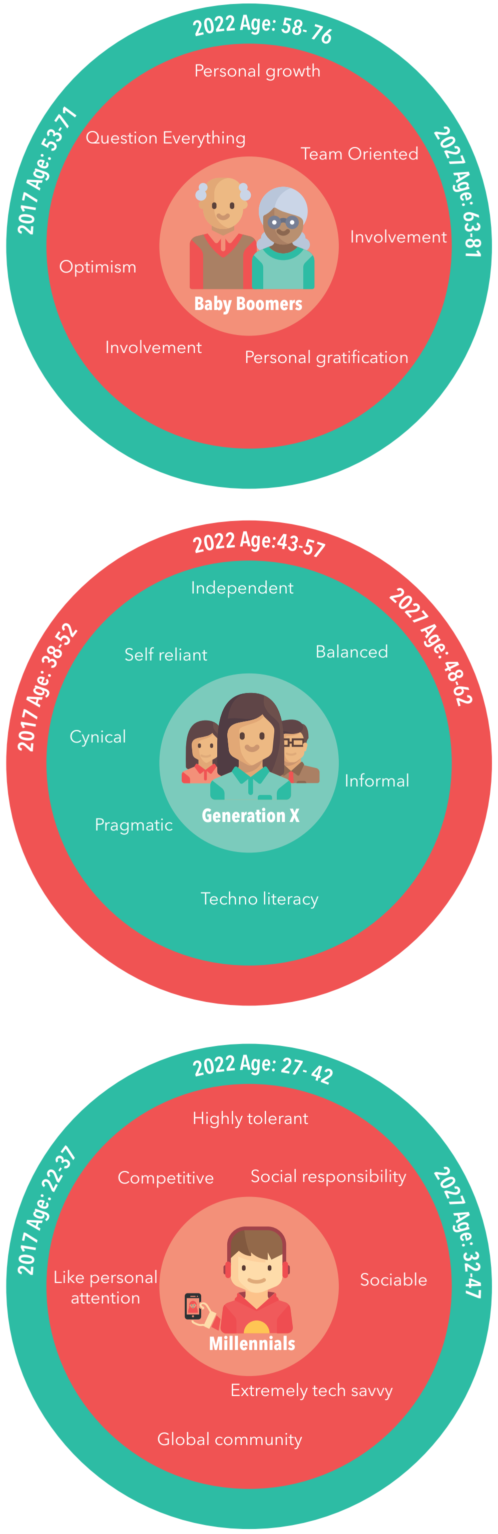


Figure 7: Generational Values (Wmfc.org, 2018)

“The technologies available as a generation matures influence their behaviors, attitudes, and expectations. People internalize the technologies that shape information access and use, as well as the ways they communicate. Matures (born 1946–1964) were exposed to large vacuum-tube radios, mechanical calculators, 78 rpm records, dial telephones, and party lines. Baby Boomers grew up with transistor radios, mainframe computers, 33? and 45 rpm records, and the touchtone telephone. Gen-Xers matured in the era of CDs, personal computers, and electronic mail. For the Net Generation (Millennials), the prevailing technologies are MP3s, cell phones, and PDAs; they communicate via instant messaging, text messaging, and blogs. For each successive generation “technology is only technology if it was invented after they were born.” (Oblinger et al., 2005)

build trust and reliability among patients and doctors (DeLemos et al, 2010 and Ommen et al, 2008). This also corroborates a reason to use affiliative speech as discussed in section 1.2

How each generation perceives technology is an important marker to understanding what they would seek from a product. As given by Oblinger et al; (2005) it is technology only if it was invented after they were born. For example, for Baby Boomers, the internet and mobile phones are technology that was invented while they were growing up. While millennials are born to such technology, thus having no problem learning to adapt it. This gives rise to the need for understanding how different users (generations) perceive technology and related products. ***To understand what each generation values in a product or technology and identifying their relationship with it, a generative session interviewing all three generations is conducted. This is discussed in section 2.1 of Chapter 2.***

This brings us to the hypothesis made to identify different characteristics required for a smooth, personalised interaction between users and Ally pod. The hypothesis made for the user research in Chapter 2, is that:

Research Hypothesis

“Different generations (baby boomers, generation X and millennials) have different styles of interactions based on their values and what they seek in products. This will have an impact on their preference of conversation.”

1.4 Voice User Interfaces

The previous sections discussed the type of communication required and how it needs to be formulated, based on who they are talking to. Communication is one thing, but the carrier of this communication or dialogue is through a voice.

According to Markowitz (2016), a human’s cooperation with a robot that provides advice or instruction is influenced by its speaking style. Considering that one of Ally’s main function is to provide suggestions, it is very important to consider this aspect. Developing dialogues that cater to such things is a challenge but the situations involved are also difficult for relationships between humans. (Markowitz, 2016).
In section 2.2, various sets of dialogues are developed to test the overall interaction qualities of Ally. The dialogues in this project act as a medium to understand the interaction.

Speech or spoken dialogue can be interpreted in various ways, depending upon the tone of voice, the pitch, the phrases used (Abelin and Allwood, 2000). When talCking about voices, there are a lot of aspects that can influence the perception of a voice. This varies from the gender of the voice, the content, the stereotyping of the voice for the product etc. A few of these factors are discussed in this section.

Gender of voices:

When discussing the voice of a product, one of the first things to consider is the gender of the voice. People have a strong preference of classifying and categorizing people (Cantor and Mischel,1979). Hence when a voice is presented, even if it is ambiguous (neither masculine nor feminine), people are inclined to classify it as male or female (Coleman, 1976). Anyone or anyone with an ambiguous voice is classified as dislikable, dishonest and unintelligent (Cantor and Mischel,1979). Hence, it is very important to keep in mind that the voice of Ally, irrespective of gender, needs to be unambiguous. According to Nass and Brave (2005), males liked the male voice better while females liked the female voice better. This is because of social identification relating to their social identity.

Ambiguous voices: If a female voice is too masculine or a male voice is too feminine. A voice- whether male or female, synthetic or recorded can be made to sound more feminine and less masculine by cutting off the low frequencies and increasing the volume of the high frequencies. Conversely, a voice can be made to sound more masculine and less feminine - regardless of perceived gender - by cutting

“They found that the congruence between the task and the robot’s demeanor strongly influenced the subjects’ willingness to respond to the robot’s instructions. The playful demeanor produced more willingness to perform the light-hearted jelly bean-sorting task but the serious demeanor engendered greater willingness to perform the more serious exercising task. When it used the playful demeanor the robot was seen as being enjoyable and witty, albeit sometimes obnoxious. When the robot used the serious demeanor it was seen as being more intelligent and much more conscientious. ” (Markowitz, 2016)

off the high frequencies and increasing the volume of the low frequencies. (Voelker, 1995)

Gender stereotyping of voices and products:

In the previous section we discussed about the gender association and preferences of voices. Gender preference and interpretation in an isolated setting is different from the influence of gendered voices in context (Nass and Brave, 2005).

Gender stereotypes are prevalent not only with respect to the role of men and women in this society, but also in products. A research studying the role of gender in E-commerce by letting participants listen to two stereotypically male and female products, each described by both male and female voices. Product descriptions were seen as more credible when the gender of the voice matched the gender of the product described (Nass and Brave, 2005)

Even with respect to teaching and learning, gender stereotyping runs so deep, that people are often unaware that they harbour them. For example, in an experiment conducted by (Nass, Moon, and Green, 1997) female-voiced computers were seen as a better teacher of love and relationships and a worse teacher of technical subjects than a male voiced computer.

But this does not mean we need to design products that adhere to such stereotyping. Rather, we need to create products which destroy them.

Personality of voices

As with gender, determining another's personality may be so important that when people hear any voice, ***no matter how clearly not human, they automatically and unconsciously use their voice analysis skills to assign a personality to voice*** (Nass and Brave, 2005).

In a study where participants had to identify if a synthetic voice or introverted or extroverted, participants had no trouble in identifying them and the extroverted voice was rated clearly more extroverted than the introvert voice (Harmon et al, 1985). Going by the similarity attraction theory, it was found that extroverted participants liked extroverted voices more and introverted participants liked the introverted voices more even though the content of what they heard was exactly the same. But several practitioners argue that, "… this research sounds compelling on its face … [but] when you examine the details, it [becomes] much less solid" (Lewis, 2017).

Similar to extroversion, the personality of voices and the preferences of users can be tested with various traits.

Gender Stereotypes are preconceived ideas whereby females and males are arbitrarily assigned characteristics and roles determined and limited by their gender (EIGE, 2018)

Personality is *the combination of characteristics or qualities that form an individual's distinctive character [Dictionary, 2018]. Personalities can be described as introverted, extroverted, judging or intuiting, kind or unkind and a host of other traits that provide a powerful framework for understanding how people think, feel and behave (Nass and Brave, 2005).*

> *"Regardless of their appropriateness or accuracy, stereotypes serve a powerful and consequential role in all aspects of life, whether one interacts with people or with technologies. On one hand, conforming to stereotypes seems to create more natural and effective interfaces - doing so simply acknowledges and leverages' expectations. But at the same time, mindlessly designing interfaces to conform to every stereotype is often unjustified and even detrimental to the society at large. There is no easy answer to this dilemma, but designers must make conscious and considered decisions when choosing to follow, counter, or ignore gender stereotypes as they build computers that talk and listen" (Nass and Brave, 200)*

While section 2.2 deals with the communication part of the user research, from the above instances,we can say that ***understanding the qualities or personalities of voices for a medical device needs to have to elicit affective speech is vital as well. In section 2.3 of chapter 2 an online survey is designed to identify the qualities/ characteristics of voices for the context of a medical device and the difference in perception of voices across (if any) different ages is studied.*** The gender of voices is not considered for this research since the goal isn't to identify the "perfect voice" but rather to identify the characteristics of it. These characteristics can apply for both male and female voices.

The current assumption is that these qualities vary based on age group (Hypothesis). While standardized VUI test questionnaires exist to analyse voices based a given set of characteristics (Lewis, 2016), this is not applied in this project because as mentioned in the previous paragraph, ***the goal is not to identify the perfect voice, but to identify the qualities of different voices that make it desirable for medical instruction use.***

Dialogue delivery

With respect to language, women tend to be more "involved" in their speaking and highlight interpersonal aspects and personal feelings more than specific, detailed information (D.Biber, 1988). Men, tend to be more "informational" focusing on the details of the things being mentioned. Females also express more concern over male listeners. For a device like Ally, a the language style needs to be in-between: Adequately informational yet nuanced with empathy. Thus, such a voice is picked for the user test in section 2.2, as there, both dialogue and tone is important.

1.5
Defining "intelligence"

In the context of this project, the term "intelligence" refers to Artificial Intelligence (AI). While there are many definitions for what Artificial Intelligence is, in this project, the definitions given by Kurzweil, and Rich and Knight perfectly fit the bill:

> *"The art of creating machines that perform functions that require intelligence when performed by people." (Kurzweil, 1990)*

> *"The study of how to make computers do things at which, at the moment, people are better." (Rich and Knight, 1991)*

Thus, AI is the simulation of human intelligence processes by machines, especially computer systems. These include learning, reasoning and self correction.

These definitions are classified as a computer "Acting Humanly" by (Russell and Norvig, 2016). For a computer "act humanly", it needs to possess the following capabilities. This is also called the Turing Test as it was designed by Alan Turing (1956).

Natural Language Processing: natural language processing to enable it to communicate successfully in English

Knowledge Representation: To store what it knows or hears

Automated Reasoning: to use the stored information to answer questions and to draw new conclusions

Machine Learning: to adapt to new circumstances and to detect and extrapolate patterns.

Figure 8 illustrates with an example, what the above capabilities mean.

Ally as per Dino's concept (JMP Final Report, 2017) uses NLP to speak to the user humanly. Knowledge representation and automated reasoning are beyond the scope of this project, since we are not focusing on programming the AI itself. ***Machine learning is relevant because in Chapter 3, we will discuss more about a subject that is a facet of machine learning, which acts as an integral part of the final model.*** When speaking of machine learning, and having a machine with a "brain", it is important to acquaint ourselves to the concept of neural networks. This would help us gain a deeper understanding of how an AI system can function.

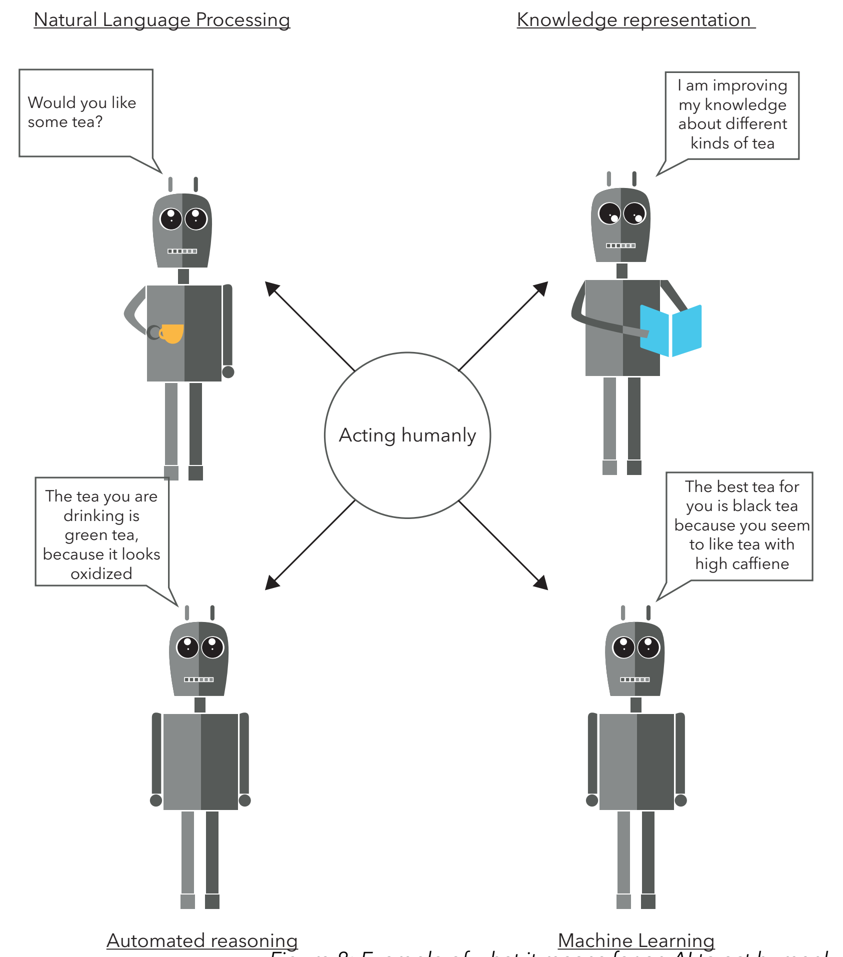

Natural Language Processing
Knowledge representation
Would you like some tea?
I am improving my knowledge about different kinds of tea
Acting humanly
The tea you are drinking is green tea, because it looks oxidized
The best tea for you is black tea because you seem to like tea with high caffeine
Automated reasoning
Machine Learning

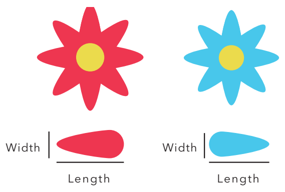
Figure 8: Example of what it means for an AI to act humanly

What is a neural network?

A neural network, is similar to the network of neurons in the human brain. Our brain has the capability to process tonnes of information within milliseconds. It can make connections between past experiences and the present and arrive at conclusions. ***The neural network of a machine is capable of performing data processing and analysis similar to the human brain. It can predict patterns, compare with previous experiences and draw decisions.*** The prime difference between a neural network of a machine and a neural network of a brain is that the machine lacks consciousness. With the advent of neural networks, there has been tremendous increase in accuracy handwriting recognition, image recognition etc. Before going into further details, let us take a look at the following example.

Let's assume there's a farmer likes to measure her flowers. She has two types of flowers, the red and the blue. She measures the length and width of each petal of the flower and notes the color down. She does the same for the blue flower. She sells all her flowers after drawing this table (Figure 9). She realises that she has forgotten to note the colour of one of the flowers.
So in order to identify the colour of the missing entry, she draws a graph with the length and width of the petals on the X and Y axis respectively. As seen in Figure 10, if the length of a red petal is 3 and width is 1.5, a red dot is placed at (3,1.5). This is done for all the values till the graph 10 appears. Now, the length and width of the mystery flower is mapped with a black dot.

As you can see, the mystery flower is surrounded by red flowers in the graph. So she can take a guess and predict that the mystery flower is red.*



Length	Width	Colour
3	1.5	🔴
2	1	🔵
4	1.5	🔴
3	1	🔵
3.5	0.5	🔴
2	0.5	🔵
5.5	1	🔴
1	1	🔵
4.5	1	?

Figure 9: Classification of colour based on petal length and width

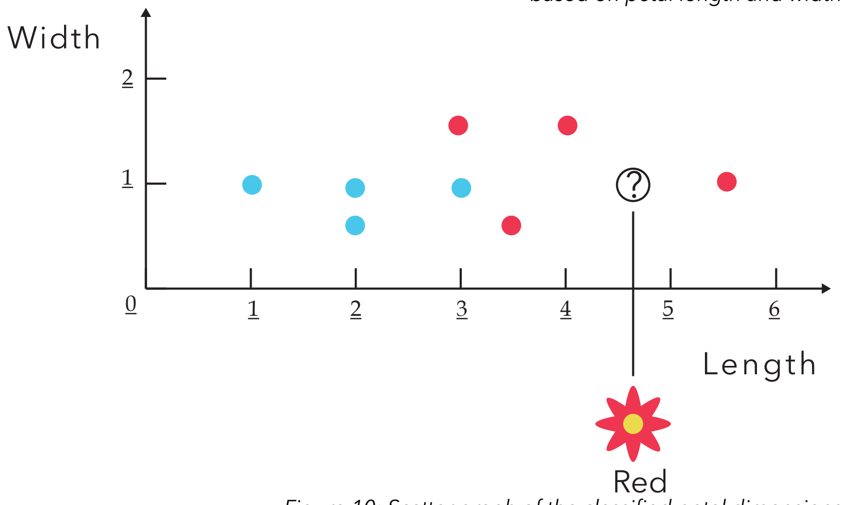


Figure 10: Scatter graph of the classified petal dimensions

Example Source: https://goo.gl/EnuzAE

The same problem can be solved by using a neural network. ***Neural networks are organized in layers and have 3 main layers: The input layer, the hidden layer and output layer.***

Neural network is a complex subject and we do not need to go into the mathematical details of it. It can be explained in simple terms are follows: ***Neural networks are made of nodes. The input node is fed with data (also known as features), the hidden nodes process this data, detecting patterns or extracting attributes and send it to the output node.*** A neural network for the above example is given in Figure 11.

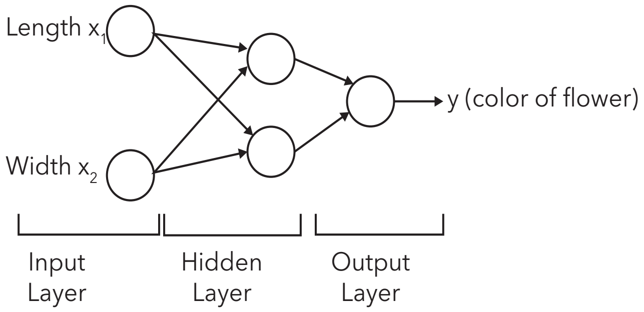


Figure 11: Neural networks of the flower problem

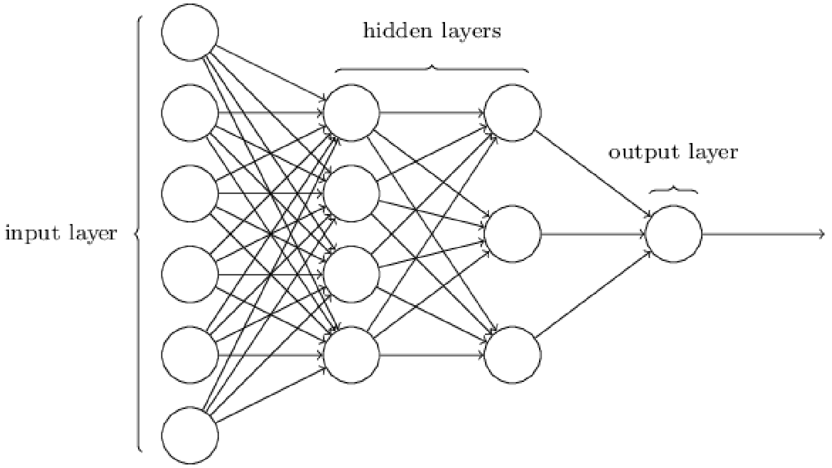


Figure 12: Layers of a neural Network

The input for the network in the example given above would be the table with the values and the output would be the colour of the flower. While in the above example, we plotted a graph to estimate the answer, in neural network, this step is done in the hidden layer, so how the prediction is made is not known.

This particular example is a simple one. But neural networks can solve very complicated problems with multiple features (inputs) and predict the correct answer. The more complicated the problem and the larger the data available, the denser the network. A representation of a general neural network is given in Figure 11.

This section discussed neural network in detail because, the concept in Chapter 3, involves deep learning (section 3.1), which uses neural networks. More about this will be discussed in Chapter 3.

Summary and conclusions

Goal

The goal of this project is to define a personalised interaction between users and a voice activated intelligent medical pod."

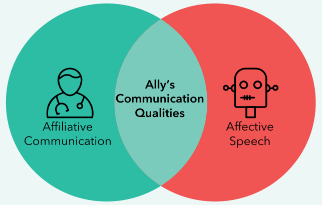
Design Challenges

Broaden user group

Define interactions

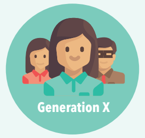
Explore voice user interfaces

Interaction with Ally



The interaction should also be reassuring, enabling supportive, In control, reflective, closed off, comforting, safe

Users



The user group is split into three generations as technology acceptance and communications can vary based on generational values

Research Hypothesis

"Different generations (baby boomers, generation x and millennials) have different styles of interactions based on their values and what they seek in products. This will have an impact on their preference of conversation."

Voice User Interfaces

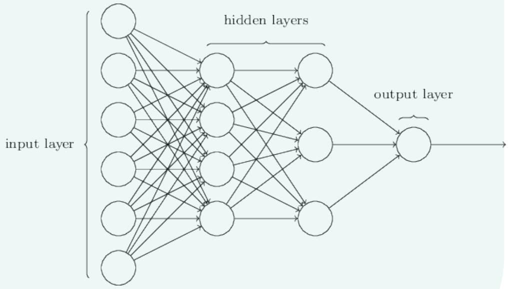
People automatically assign personality to voices
It is important to understand the characteristics that a voice elicits, in order to have an affiliative and affective communication.

Artificial Intelligence

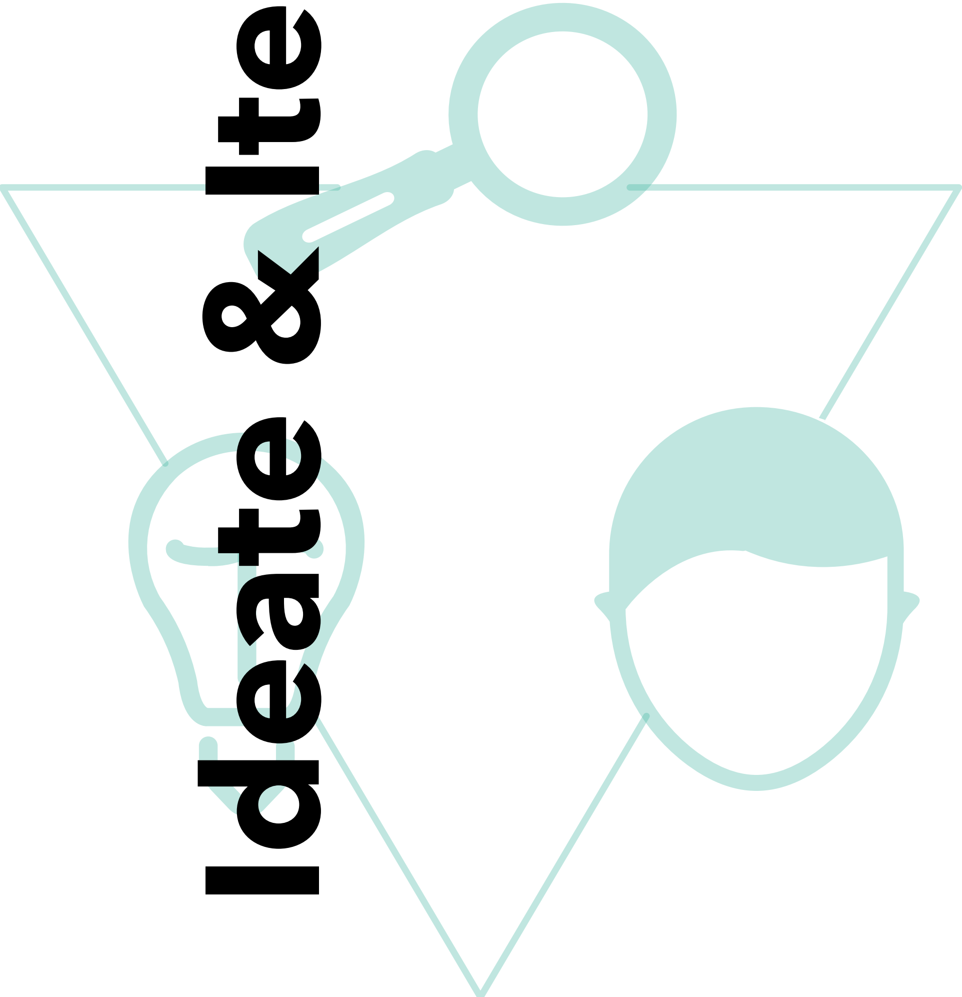
For a machine to "act humanly", it should have the following skills:

Natural Language Processing
Knowledge representation
Automated reasoning
Machine Learning

Neural networks are efficient networks that process data in hidden layers to predict accurate outputs



Ideate & Iterate



02

User Research

The design challenges in Ally's concept, preliminary research with respect to the goal of this project and the hypothesis based on which further research is conducted have been established in the previous chapter. The chapter begins with establishing the research methodology, and goes to the user researches focusing on understanding the user group, qualities of interaction with Ally and personalities different users seek in a voice for a medical device.

Research Goals

The goal of this project is to design a personalised interaction between users and a voice activated intelligent medical pod. The collection of data from users is conducted by means of generative sessions, online questionnaires and interactive user tests. The objective of this user research is:

• To gain a better understanding of each user group (generation) with respect their perception of consumer electronics
• To explore the interactions of different generations with an intelligent voice agent
• To extract voice characteristics based on generation preference.

To organize and articulate each research objective, a set of research questions were defined for each section.

Research Methodology

As mentioned in page 19, the approach to this project has been research through design. Hence, ***the user tests conducted in this chapter involves designing interactions based on intuition and emotion (of the user) and to extract qualitative insights.*** This follows the approach learnt during the course Exploring Interactions, where the design is created and iterated upon to explore the qualities of interaction and the characteristics various users seek in voice for a medical device.

2.1
Perception of home appliances by different generations:

In section 1.4 of Chapter 1, the reasoning for picking the three generations: Baby Boomers, Generation X and Millennials is discussed. From the discussion it was concluded that the perception of technology by different generations needs to be studied further. In this section, a generative session is conducted to understand what each generation values in a product or technology and what their relationship with it is.

According to (Sanders and Stappers, 2012) there are 4 levels of knowledge: Explicit, Observable, Tacit and Latent. ***In this section, we will try to extract tacit and maybe latent knowledge about people's feelings towards consumer products (Figure 13).*** For this, a generative session is conducted where what people say, use and what they feel are extracted. Hence, in this section, we try to extract some knowledge from all the levels, using a mixture of observation and interviewing in the process of conducting a generative session. ***It is to be noted that there is no sensitizing involved and the participants are approached directly on the day of the interivew.***

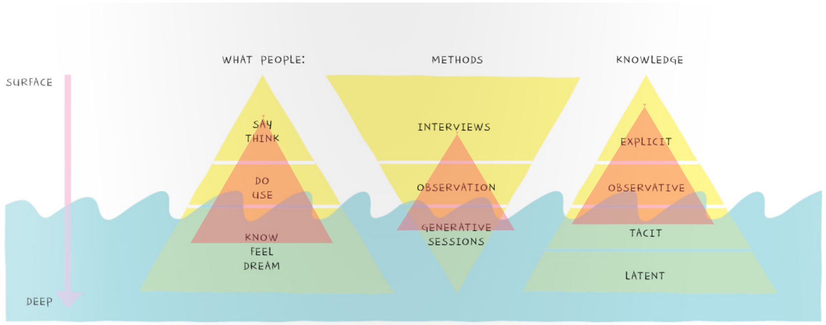


Figure 13: Red regions are what this research intends to focus on. (Sanders and Stappers, 2012)

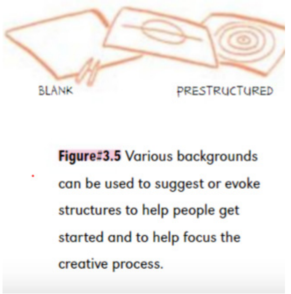
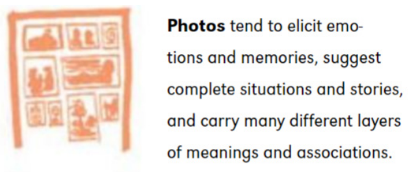

BLANK
PRESTRUCTURED

Figure:3.5 Various backgrounds can be used to suggest or evoke structures to help people get started and to help focus the creative process.

Figure 14: Excerpt from Convivial Toolbox (2012)



Photos tend to elicit emotions and memories, suggest complete situations and stories, and carry many different layers of meanings and associations.

Figure 15: Excerpt from Convivial Toolbox , (2012), about using photos as a tool

Research Questions:
The following research questions to organize to articulate the objective of this research segment.

- *What are the values or characteristics that different users look for in house appliances? (looking back at the hypothesis to get a better understanding of it)*

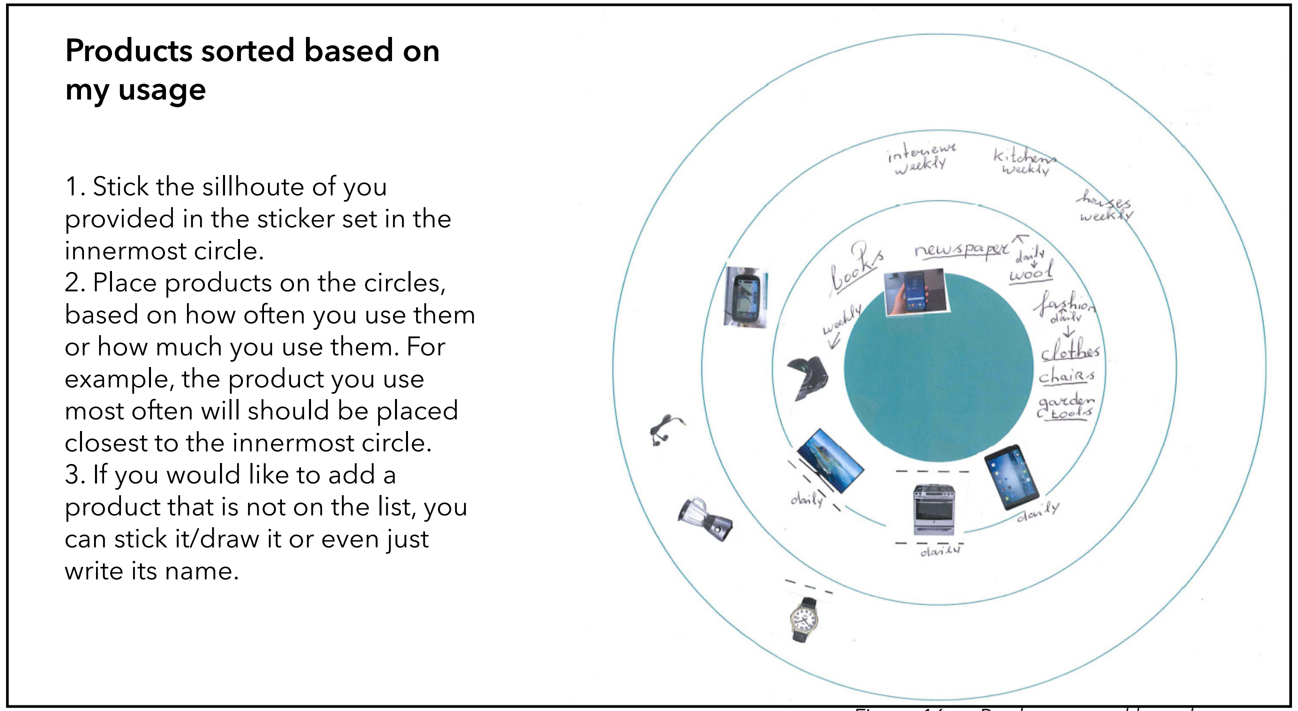
- *How do they perceive or what is their relationship with technology? (adaptation, perception etc in terms of how they humanize it or see its use, etc)*

Research Method:
Generative toolkit + Interview

Sample Set:
Four people of each generation, i.e Baby Boomers, Generation X and Millennials were interviewed for this generative session

A toolkit comprising three worksheets and a set of product stickers was designed. The product stickers images of products that are used most often at home. Photos are a common ingredient in creating such toolkits (Figure 14) (Sanders and Stappers, 2012). The background in the toolkit consisted of concentric circles, suggestive of the symbolic distance between the product and the user (Figure 15). The entire generative session toolkit can be found in Appendix B.1.

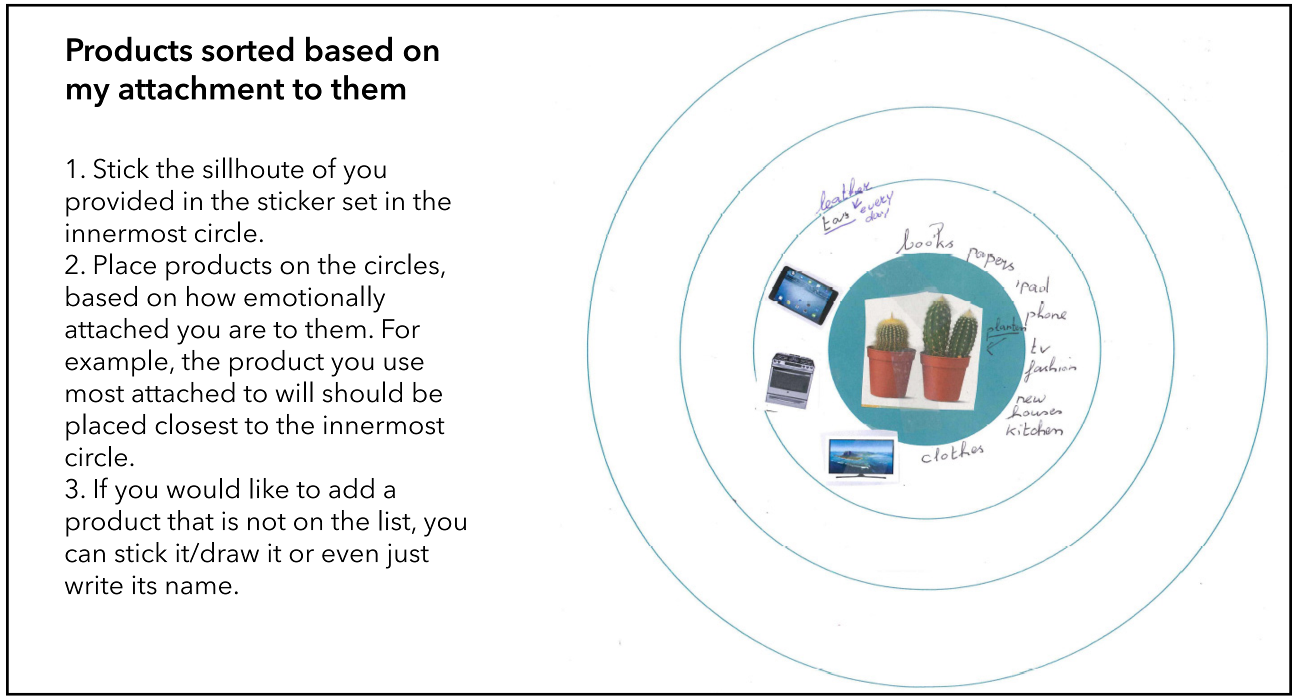
Setup:
• A task sheet as shown in Figure 10 is given, along with a set of stickers comprising daily use consumer electronics.
• On the first page, the participants have to place the products they find relevant, based on how often they use them. Participants could also add products that are not given in the stickers.
• On page two, they have to place the stickers based on how attached they are to each product
• On the last sheet, they have to place the products on an x-y graph, where x and y are the time it took to get used to the product vs. the liking towards the product respectively.
Figure 16 a, b and c shows the task sheet filled out by one of the participants.



Products sorted based on my usage

1. Stick the sillhoute of you provided in the sticker set in the innermost circle.
2. Place products on the circles, based on how often you use them or how much you use them. For example, the product you use most often will should be placed closest to the innermost circle.
3. If you would like to add a product that is not on the list, you can stick it/draw it or even just write its name.

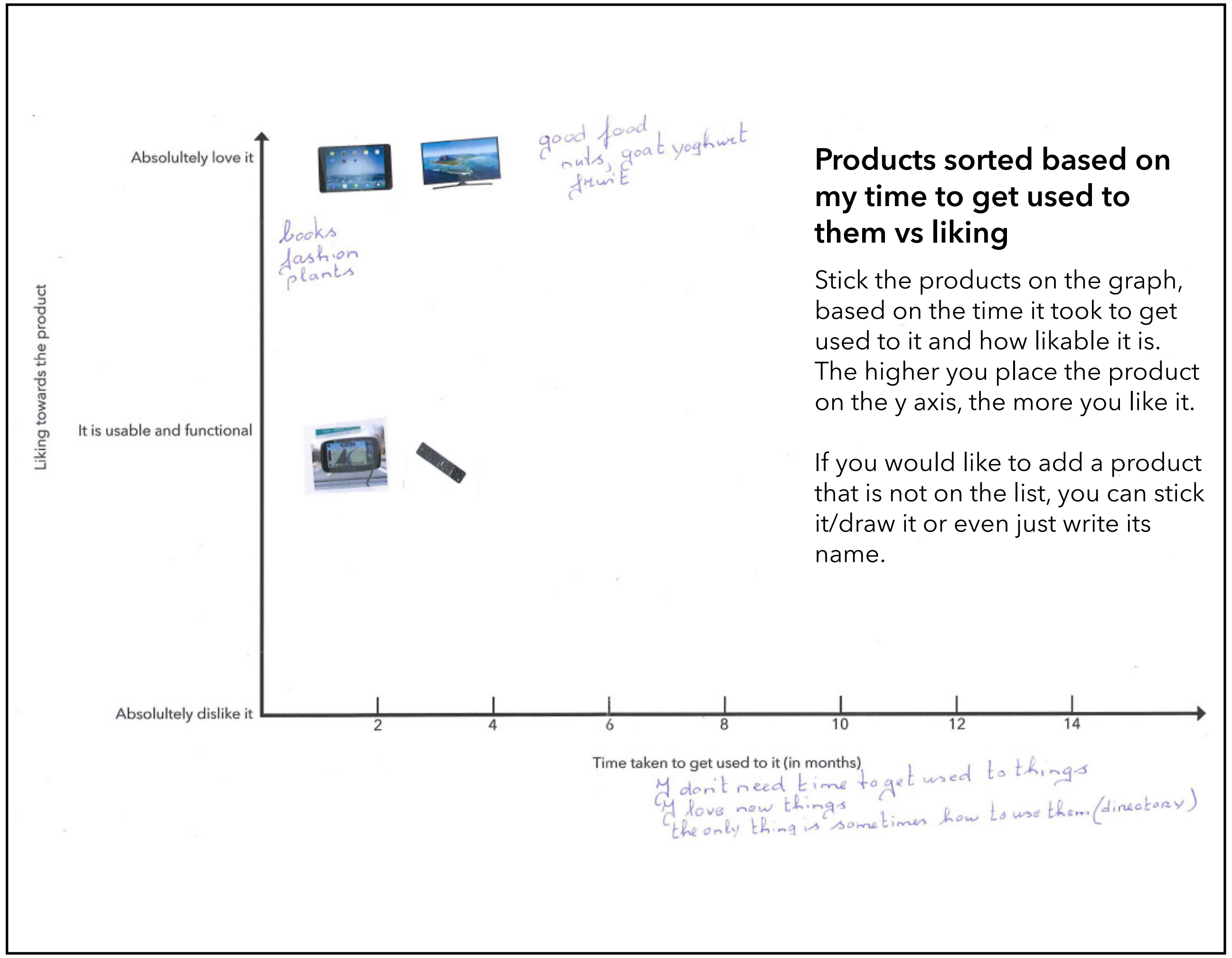
Figure 16.a : Products sorted based on usage



Products sorted based on my attachment to them

1. Stick the sillhoute of you provided in the sticker set in the innermost circle.
2. Place products on the circles, based on how emotionally attached you are to them. For example, the product you use most attached to will should be placed closest to the innermost circle.
3. If you would like to add a product that is not on the list, you can stick it/draw it or even just write its name.

Figure 16.b : Products sorted based on attachment



Products sorted based on my time to get used to them vs liking

Stick the products on the graph, based on the time it took to get used to it and how likable it is. The higher you place the product on the y axis, the more you like it.

If you would like to add a product that is not on the list, you can stick it/draw it or even just write its name.

I don't need time to get used to things
I love new things
the only thing is sometimes how to use them (directory)

Figure 16.c : Products sorted based on adaptation vs liking

2.1.1 Result and Analysis

As each user used the toolkit, their casual conversation about each product was also noted down. The documentation of all the raw insight gained can be found in Appendix B.2. Every insight that was gained during the generative session was digitized into an excel sheet and then cut out to place different insights on the wall. The spread sheet analysis is given in Appendix B.2.

From this, the most relevant insights were taken out and the insights were converted into infographic 17.

What are the values or characteristics that different users look for in house appliances?

For three of four baby boomers, functionality and ease of use was most important in a product. ***From the interview and analysis it was also concluded that they value kindness and respect.***

The generation X seek usability, functionality and relevance in a product. They valued ***honesty and would like the product to be "human".***

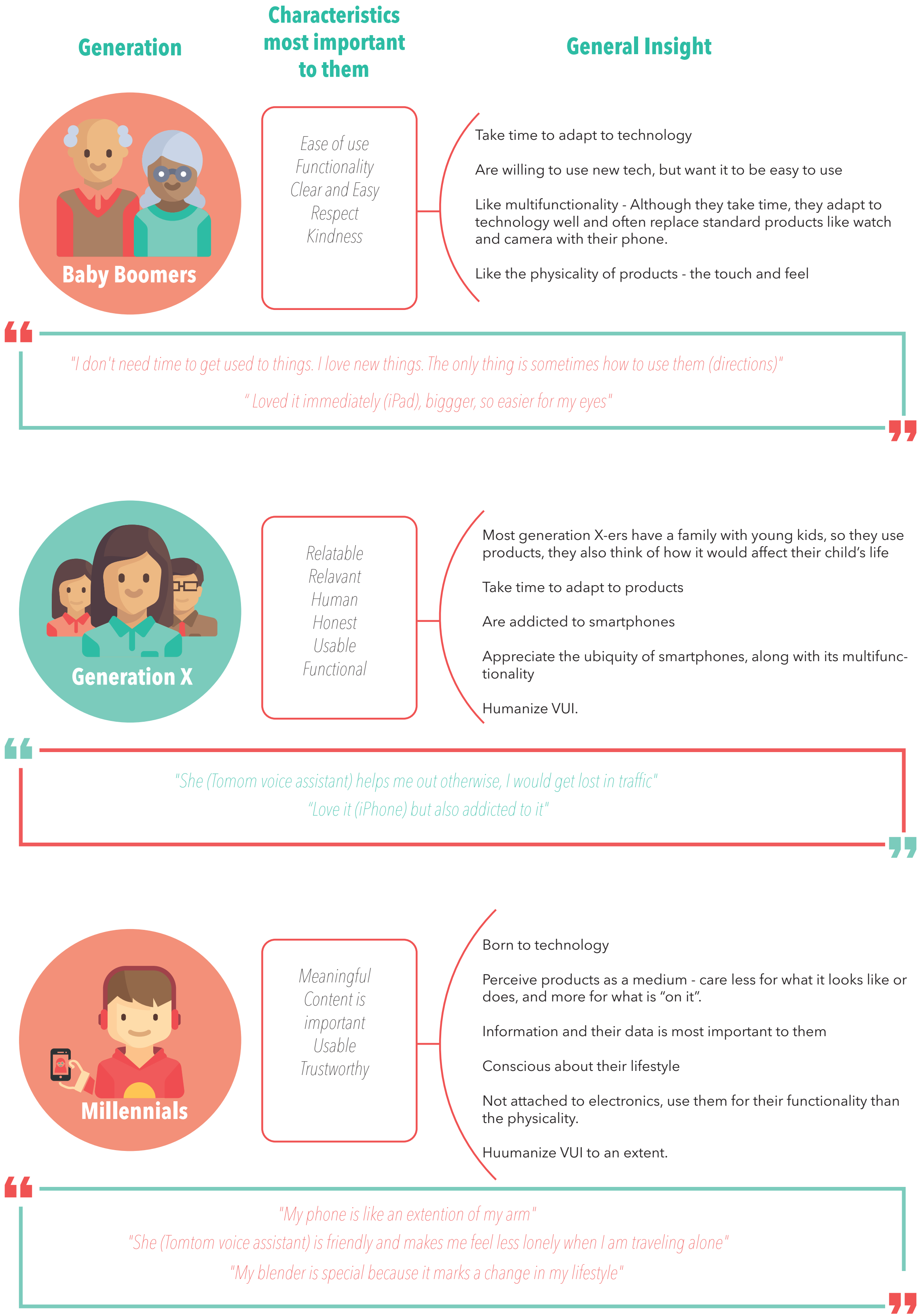
Millennials look for content or the information a product has to give, and usability. ***They also want the product to be trustworthy and have a meaning.***

How do they perceive or what is their relationship with technology? *(adaptation, perception etc in terms of how they humanize it or see its use, etc)*

With respect to perception, most ***baby boomers*** understand the need and usefulness of products even though they personally might take time to adapt to it. Only one baby boomer out of five said they would not be interested new technology, but appreciates the value of it. Although he said he wasn't interested in tech products, he was excited to show off the functionalities and usage of his iPhone 2G. Most of the baby boomers interviewed were willing to use products as long as they are easy to learn or adapt to.

Generation X are the first generation who use their smartphones on a regular basis. Since a lot of them have children at home, their perception of a product is shaped by how useful, safe and functional it is for them and their kids. They require some time to adapt to technology, but lesser than baby boomers.

Millennials, literally born to most technological advancements of today, think of products as a medium for a message. They place their value for a product on what it has to convey, than on the product itself. They are conscious about their lifestyle and try to improve it by means of technology.



Generation | Characteristics most important to them | General Insight

Baby Boomers

Ease of use
Functionality
Clear and Easy
Respect
Kindness

Take time to adapt to technology

Are willing to use new tech, but want it to be easy to use

Like multifunctionality - Although they take time, they adapt to technology well and often replace standard products like watch and camera with their phone.

Like the physicality of products - the touch and feel

> "I don't need time to get used to things. I love new things. The only thing is sometimes how to use them (directions)"
>
> " Loved it immediately (iPad), bigger, so easier for my eyes"

Generation X

Relatable
Relavant
Human
Honest
Usable
Functional

Most generation X-ers have a family with young kids, so they use products, they also think of how it would affect their child's life

Take time to adapt to products

Are addicted to smartphones

Appreciate the ubiquity of smartphones, along with its multifunctionality

Humanize VUI.

> "She (Tomom voice assistant) helps me out otherwise, I would get lost in traffic"
>
> "Love it (iPhone) but also addicted to it"

Millennials

Meaningful
Content is
important
Usable
Trustworthy

Born to technology

Perceive products as a medium - care less for what it looks like or does, and more for what is "on it".

Information and their data is most important to them

Conscious about their lifestyle

Not attached to electronics, use them for their functionality than the physicality.

Huumanize VUI to an extent.

> "My phone is like an extention of my arm"
>
> "She (Tomtom voice assistant) is friendly and makes me feel less lonely when I am traveling alone"
>
> "My blender is special because it marks a change in my lifestyle"

Figure 17: Generative session analysis

Conclusion:

The generative session gives a more realistic insight on how people of different age group feel about technology. We can see that some values of each generation given in section 1.4 have a direct match with the respective generation's outlook. The data gathered from this will be used to create the user test for testing the interaction between users and Ally. The qualities and values generated in this section will be cross checked in section 2.2.

2.2
Interactions of different generations with an intelligent voice agent specialized for medical advice:

In section 1.3 , it was concluded that Ally's communication qualities should be affiliative and affective in nature. From section 1.2, we know that an affiliative approach to communication leads to a satisfactory experience.

After researching the different aspects of interaction and user groups, the hypothesis is that different generations would interact differently based on their values. In the previous section, what different generations value in a product and their perception about technology was studied. With the results from the previous section and the insights gained from section 1.2 about how the communication should be, in this section we will study how different user generations interact with Ally, and what qualities make their interaction satisfactory. The research through design methodology is applied here and thus, an "AI" with affective qualities is prototyped and tested to extract certain characteristics.

Set up:

• Participants are asked to give consent to be recorded.
• A pre-test questionnaire is given to test their knowledge and experience with voice activated devices and preferences to phraseologies (Figure 18).
• Next, the participants are given a setting where they are either "healthy and feeling good" or "have a headache" and are asked to interact with Ally. This step is repeated thrice with 3 different phraseologies and style of speech. After completing each scenario, they have to rate them and are then asked questions about the scenarios.
• Finally, participants are given a set of symptoms and asked to interact with Ally, and an emergency is detected. Participants are asked questions about this final scenario.

The first part of the pre-test questionnaire was created to gain an understanding of the users' past experiences (or lack of) with voice activated devices, while the reasoning for the second part is explained as follows.

When a participant interacts with Ally, both the dialogue delivery and the phraseology might influence their experience. In order to have a better overview of each generation's preference of words/sentencing, this questionnaire was added to the user test.

Participants had to rate 3 sentences that conveyed the

Research Questions:
The following research questions to organize to articulate the objective of this research segment.

- What are the qualities different user groups (generations) find affiliative and affective, when conversing with Ally?

- How do users interact with Ally generally?

- How do different users react when an emergency is predicted by the product?

Research Method:
Qualitative, interactive user test + interview

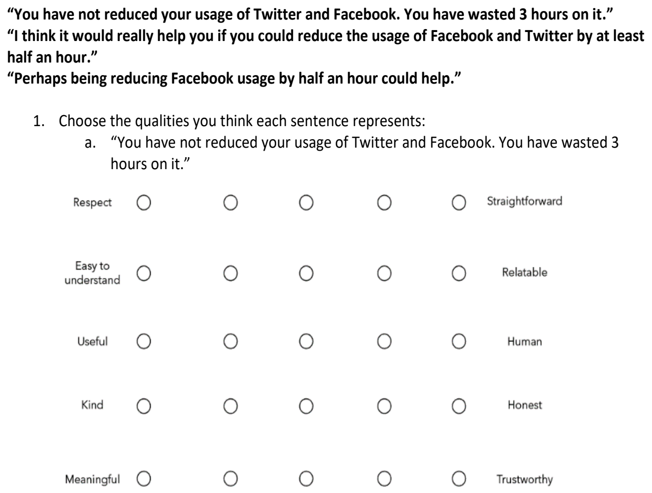
Sample Set:

Five participants (all Indians), from each generation, i.e, Baby Boomers, Generation X and Millennials were interviewed. The criteria was selection was that they were well versed in English.

same message but had different phraseologies. ***Their choices were a tradeoff between two qualities. This technique was adopted over having a likert-like rating scale or an antonyms scale because, the goal was to align users towards the quality that was most striking to them in the sentence. An antonym scale or a likert scale would not yield such a result.*** Each value described in the scale is taken from the output of the generative session given in section 2.1.1 (Figure 18).

For the interaction part, various scenarios were created for the purpose of this user test, along with the possible conversation flow. Out of these, the two most important ones, i.e the general interaction and the stroke scenario were picked, iterated upon and detailed. The scenarios and the conversational flows can be found in Appendix C.1.

Three different scenarios with varying phraseologies were created. These were conceived to reflect how the system should possibly interact when speaking to different generations. Values of each generation were incorporated into the dialogues. For example, Baby boomers value kindness and respect (from Figure 17). Hence the first conversation was kind and respectful, with a lot of diplomatic words and sentences. For Generation X, it was a straightforward, but polite while the third scenario, was to the point but enthusiastic.



"You have not reduced your usage of Twitter and Facebook. You have wasted 3 hours on it."
"I think it would really help you if you could reduce the usage of Facebook and Twitter by at least half an hour."
"Perhaps being reducing Facebook usage by half an hour could help."

1. Choose the qualities you think each sentence represents:
   a. "You have not reduced your usage of Twitter and Facebook. You have wasted 3 hours on it."

Respect ○ ○ ○ ○ ○ Straightforward

Easy to understand ○ ○ ○ ○ ○ Relatable

Useful ○ ○ ○ ○ ○ Human

Kind ○ ○ ○ ○ ○ Honest

Meaningful ○ ○ ○ ○ ○ Trustworthy

Figure 18:The sentences to be rated against various qualities

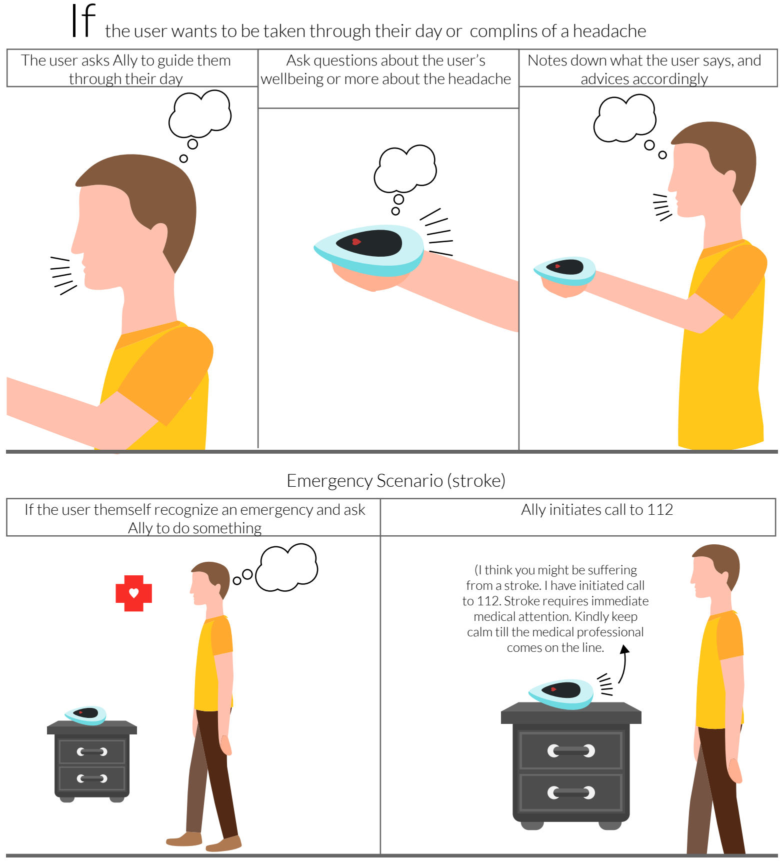

If the user wants to be taken through their day or complins of a headache
The user asks Ally to guide them through their day
Ask questions about the user's wellbeing or more about the headache
Notes down what the user says, and advices accordingly
Emergency Scenario (stroke)
If the user themself recognize an emergency and ask Ally to do something
Ally initiates call to 112
(I think you might be suffering from a stroke. I have initiated call to 112. Stroke requires immediate medical attention. Kindly keep calm till the medical professional comes on the line.

Figure 19:Storyboard of the scenarios created for user test

Note: It is important to note that the conversations are used only as a tool to identify the different characteristics each generation prefers. Participants might prefer different scenarios for the same reasons. The reasons are important, not the scenarios themselves.

In an ideal situation where the AI is fully functional, instead of creating dialogue cues manually, the AI would be trained to understand the context and who it is talking to, enabling the algorithm to learn the pattern using natural language processing and adapt to the phraseology automatically. How this can be achieved will be sen in Chapter 3. But to get an understanding how how the interaction would be, and what they seek in these phrases, each dialogue and phrase were carefully chosen to reflect the communication style of the respective generations.

Initially, a basic AI system that would respond to the user's questions with the phrases that were fed into it was created. But the AI's response time was very slow. Since the goal of this user test is to test the experience and understanding of phrases and not the AI's usability itself, the AI was shelved and the Wizard of Oz technique was used.
A suitable female voice was chosen and the dialogues were recorded. The users would interact with the Ally pod by talking and the pod would respond to them accordingly. The gender factor of the voice is not taken into consideration, and the mere interaction is given more focus. In Section 1.4, it was mentioned that women tend to speak with "involvement" and highlight interpersonal feelings while men tend to be more informational and specific. The interaction of this user test needs an equal mix of both. Since using an ambiguous voice is not a good solution (Section1.4), a female voice that elicited adequate emotion coupled with informative phraseology was used.

In the acutal device an option to choose between a male and female voice should always be available, as it is a frequently voiced opinion. But for the scope of this user test, such an option was deemed unnecessary.

Scenario 1:
In the first scenario, the user is told that they are either "Healthy and feeling good" or are having a "headache" and asked to interact with Ally.The participants were asked to act as themselves. They merely guided to interact the way they would, but with certain parameters. The scenarios are made specific in order to have more specific results.
After each scenario, the participants were asked to rate the scenarios on a scale of 1 to 6

Figure 20: Participant given a bluetooth speaker as Ally

Scenario 2:
In this scenario, the participants were given a profile of a person with certain symptoms of a stroke and were asked to interact with the Ally pod. We did not explain what the symptoms mean The goal was to test the reactions of different age groups to a device informing them that they are facing a medical emergency, and how seriously they perceive the gravity of the situation. This is given because Ally is a device that was conceptualised for this very purpose. After the scenario, the users are interviewed to document their experiences. (Figure 20 illustrates a participant interacting with Ally)

2.2.1
Result and Analysis

In this section, the results and insights from the user test will be discussed. The analysis process is depicted in Figure 21. The transcripts from the user tests can be found in Appendix C.2

The qualities described by the participants for the scenarios they preferred the most were clustered based on the generation they belong to. The rating of scenarios was inconsistent with the findings from the verbal feedback because participants tend to rate the feedback based on the content of the conversation than the dialogue delivery. But their reasonings for liking a scenario aligned with their generational values. There seems to be an **overlap between the values of baby boomers and millennials.**

Since the objective feedback does not correlate with the subjective feedback. This is because even though different users rate different scenarios as their favourite, their reasoning is consistent i.e the scenarios merely act as a tool for them to assign interaction qualities. Hence, the ratings (objective feedback) are not displayed here.

The qualities of scenarios most preferred by participants is given in Figure 22. The qualities described for all 3 scenarios by each generation can be found in Appendix C.3.

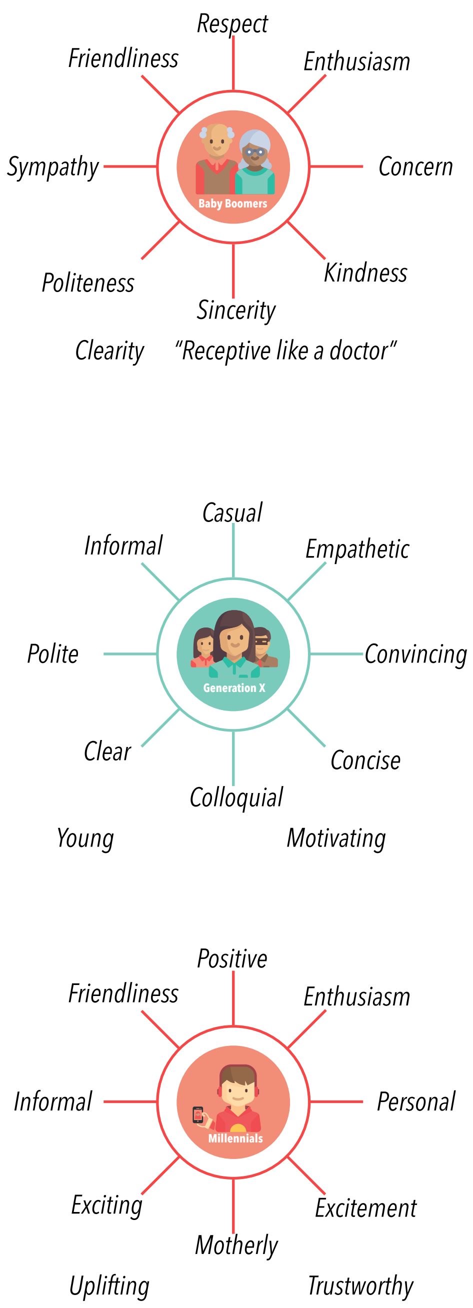


Figure 22: Qualities of most preferred scenarios

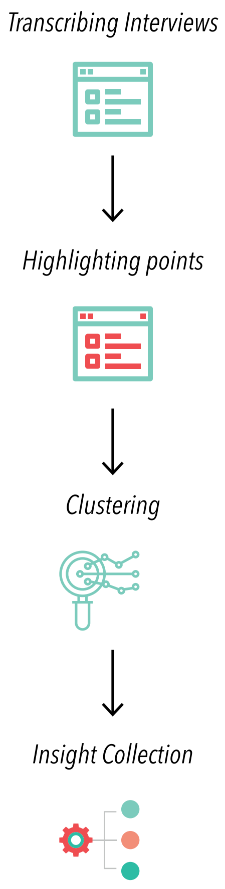


Figure 21: Analysis Process of the interaction user test

	Baby Boomers	Generation X	Millennial
Users' basic interaction	All baby boomers seem to give respect to Ally while talking to it. They use phrases like with "Thank you very much", "Wish you the same", "No dear". They seem to treat the device as they would treat humans, and feel like the device is doing a favour by advising them, hence, they converse in a certain way.	This is not the same with Generation X. While they do use "thanks" there is a clear difference in how they interact with the product in terms of salutation and often give one word replies. They are often curt. They are very focussed on the purpose they approach the device for, and want to optimise its functionality.	Millennials fall in between baby boomers and Gen X where in they're neither too friendly nor too straightforward.
Their tone while interacting	They conversed in a friendly manner with Ally and seemed to consider Ally as a person than a device.	They were very monotonous and they came to Ally for a purpose, and when the purpose was fulfilled, they often ended the conversation with a "bye" or "no that's enough".	Neither too friendly nor too monotonous. They treated Ally as a device that was friendly.
Reference to Ally	Referred to Ally as "she" or "her"	Referred to Ally as "it"	Referred to Ally as "she" and "her"
The concept of self	When talking or giving feedback, the feedback is from a general perspective, about the device	Generation X talk about their perspective and how they like or dislike it.	Their responses are centered around their preferences and often they used words like "I felt that she could have been happier", "I am a very shy person". It was more of their opinion and how Ally's interaction was relevant to them as individuals. They have lesser inhibitions while talking to Ally.

After clustering the results, each generation's interaction with Ally was segregated into 4 catogeris. This was created when by looking at the differences in pattern of interaction. This is given above.

Phrases and their characteristics:

As part of the pre-test questionnaire, participants were given a set of three phrases and asked to rate them.

For each phrase, the average value of each charactesrtics was noted.

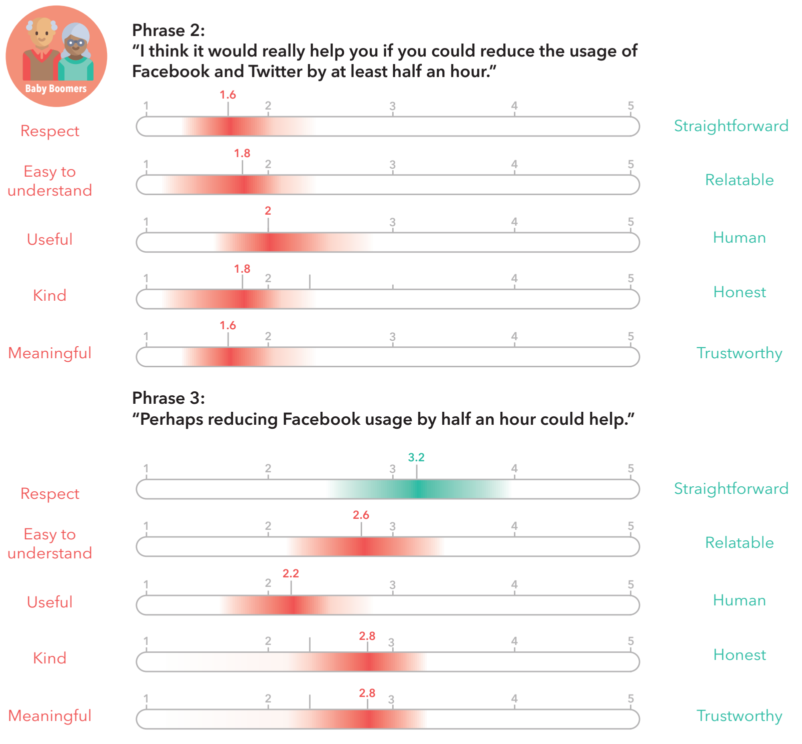
If the score for a characterstic is less than three, then it means that people are leaning towards the characterstics on the left for that phrase. If the value is above 3, then people are leaning toward the characteristic on the right. If the score is 3, then it means they find it to be neutral between the given characterstics.
The analysis and illustration of all the ratings by all 3 groups can be found in Appendix C.4.

Baby boomers:

60% of the baby boomers prefer the second phrase, and it can also be seen that they have rated this phrase high on respect, kindness, usefullness, meaning and understanding (Figure 23).

The rest prefer the third sentence, which they believe inches towards straightforwardness, relitability, kindness, usefulness and meaning

It is interesting to see that there is an overlap between the behaviours / preferences of the baby boomers and mllennials, in some places. For example, both seem to prefern similar qualities in their interactions, like friendliness and enthusiasm, when conversing with Ally.


Phrase 2:
"I think it would really help you if you could reduce the usage of Facebook and Twitter by at least half an hour."

Phrase 3:
"Perhaps reducing Facebook usage by half an hour could help."

Figure 23: Baby boomers' preferred phrases

Generation X:

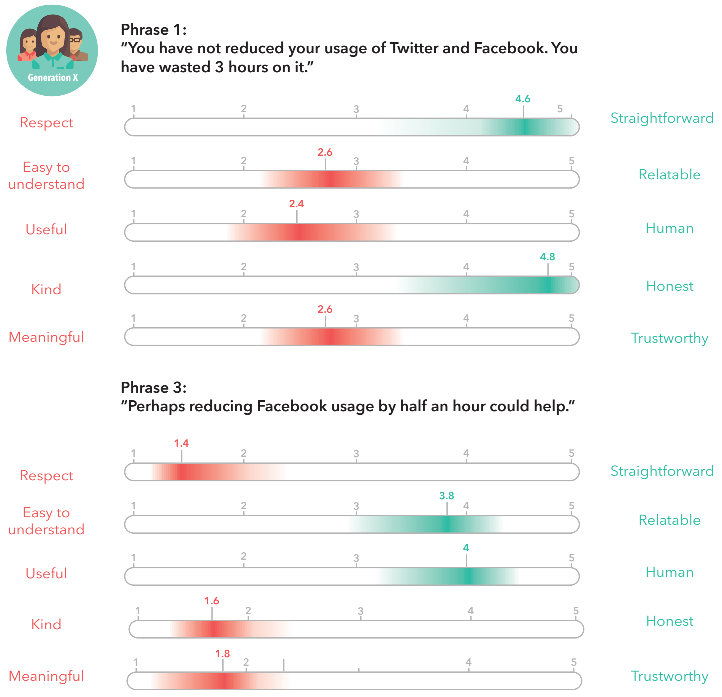
This generation has an equal mix of prerences where 40% prefer the first sentence, 40% the third and 20% prefer the second sentence. (Figure 24)

The first sentence is prefered for the fact that it is direct and honest while the third is preferred for its suggestive nature which is found to be more convincing. The third sentence is found to be respectful, relatable, human, kind and meaningful

Millennials:

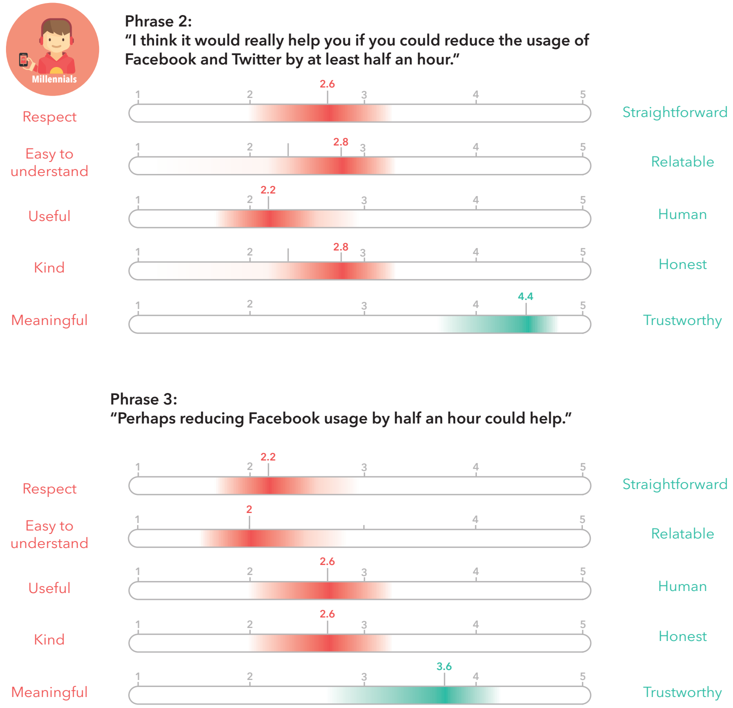
The millennials seem to prefer the second and third sentence more and the first less. This is an interesting result, as it was expected that the millennials would prefer the first sentence. It was quoted that the first sentence is preferred only in certain contexts by millennials. (Figure 25)

The second sentence is seen as more respectful, easy to understand, useful, kind and trustworthy. It is interesting that they find this sentence trustworthy while the other two generations don't. This could also be owing to the context of the sentence and how relevant it is to millennials compared to the other


Phrase 1:
"You have not reduced your usage of Twitter and Facebook. You have wasted 3 hours on it."

Phrase 3:
"Perhaps reducing Facebook usage by half an hour could help."

Figure 24: Generation X' preferred phrases


Phrase 2:
"I think it would really help you if you could reduce the usage of Facebook and Twitter by at least half an hour."

Phrase 3:
"Perhaps reducing Facebook usage by half an hour could help."

Figure 25: Millennials' preferred phrases

generations. The third sentence is seen as the same as the second sentence, with the same characterstics. This ascertains the preferences in characterstics of the millennials as well.

It is clear from this section that baby boomers' and millennials' preferences match their characterstic traits. It seems like the millennials and baby boomers are decisive about what they like and have preferences that correlate.

Reacting to an emergency:

In this scenario, Ally predicts a stroke after participants are given certain symptoms to complain about. The general reaction, their openness to such a prediction and how they percieved the gravity of the situation was analysed and plit into 4 categories.

- Being informed of a "stroke"
- Their feelings
- Tone of Ally's voice
- Recogntion of need for calling emergency

The analysis is given on page 63. Please note this is a consensus of overall opinion.

	Baby Boomers	Generation X	Millennial
Being informed of a "stroke"	Baby boomers defintely liked being informed and think it is a great test that they would take very seriously	Most Gen-Xers feel the word "stroke" is too heavy and it should just call the emergency and say the situation requires attention.	Millennials have no problem with being informed of the stroke as it is. But they don't want to be asked to "keep calm"
Their feelings	Being informed of a stroke made them nervous, tensed and shocked. But they would reconcile because the call is being made.	They said they would be shocked and not so happy with the fact that it went for a diagnosis	They said it felt intense, stressful and a bit confusing because of the tone
Tone of Ally's voice	No comments or problems with the tone of voice	Most Gen-Xers feel the tone needs a sense of urgency and seriousness so they could recognize the impact of the situation	The millenials also feel that the tone is a bit too calm and needs to sound more urgent
Recognition of the need for calling the emergency	Baby boomers are glad it takes prompt action on their behalf.	Gen-Xers are okay with the emergency being contacted, as long as they are convinced of the need for it	Millennials have no problem with the emergency being contacted.

Conclusions:

Thus in this section, we analysed the interactions of different generations with the users and the how they characterised these interactions. This is important and is used as a benchmark in the concept presented in Chapter 3. The phrase characteristics are a good indicator of the kind of language the users prefer when talking to a medical device. More about how this is put to use will be discussed in Chapter 3.

2.3
Voice characteristics of a medical device:

The user researches conducted in sections 2.1 and 2.2 comprised exploring the users and the general interaction between users and Ally. One of the main components of Ally's interaction is its medium, i.e Voice. In section 1.4 of Chapter 1, the various aspects that influence the perception of voices is discussed. People tend to like voices that elicit personalities similar to their own. By means of this user test, we will try to identify the kind of personalities users of different age groups associate with different voices, and which they find most suitable for a medical device.

Setup:

- ***A questionnaire was created in which people had to answer different questions after listening to different voices. Five different voices were chosen and certain phrases were recorded.***
- ***In the first set of questions users had to select different qualities they thought each voice had and purpose they would prefer each voice for. The qualities are shown in Figure 26***
- ***In the second section, they were asked to map the voices to different faces if their minds made such an association (Figure 27)***
- ***In the final section, they were asked to map a given set of qualities to the faces (Figure 27).***

The script of the questionnaire can be found in the Appendix D.1

Rationale for setting up the questionnaire is given as follows:

The aim of this user test is to understand the kind of characteristics different people attribute to different voices and then understanding which ones suit a medical device the best. The questionnaire conception can be split into three parts of reasoning: The sentences chosen, the voices chosen and the faces chosen.

The Phrases:

The phrases for this questionnaire were carefully constructed that reflected intangible qualities of voice given on the right. They needed to be reassuring, supportive, enabling and reflective. The following sentences shown on the right were recorded.

Sentence #1 This elicits authority and assertion in the first half of the statement and support in the second half.

Sentence #2 relates closer to Ally's functionalities. The first sentence is assertive and informative while the next sentence is suggestive of the outcome if the assertion made in the first half is reflected upon.

Research Questions:

The following research questions was formulated to organize and articulate the objective of this research segment.

- What are the voice characteristics different users preferred for the medical device?

Research Method:

Online research by means of a questionnaire

Sample Set:

A total of 50 participants consisting of 12 Baby Boomers
12 Generation X
26 Millennials
across different countries and continents.

#1 I wouldn't advise you go ahead with that script (authoritative)...I could proofread it for you after you make the changes I've mentioned (supportive)

#2 Your tracker shows that you have not been getting sufficient exercise. Exercise is vital to keep your body healthy. (Assertive, reflective)

#3 How are you today? (Friendly)

#4 Everything will be fine, don't worry! (Reassuring)

Sentence #3 s just a casual sentence with a friendly overtone, but its interpretation can vary based on the tone of speech.

Sentence #4 shows care, ressurance and positivity

The voices:

Five male voices with different pitches and tones were chosen for this questionnaire. The personalities of each voice are different. From Figure 6, it was concluded that the voice of a device like Ally needs to be reflective, supportive, reassuring and enabling. From videos of voice experts, it was found that people tend to listen to voices that are full and deep, but their tone also influences the listener. Therefore, it was concluded that not all voices need to be full and deep in order to elicit the characteristics chosen in the options section.

The voices chosen can be described as follows:
1. Middle aged, assertive yet friendly
2. Deep but friendly
3. Deep, full authoritative, but friendly
4. Slightly deep, yet friendly
5. Youthful and friendly

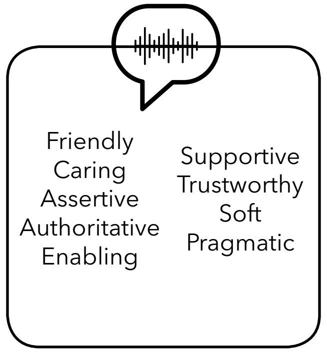
Male voices were chosen because gender of the speaker is irrelevant in this user test. We are not looking to find the most suitable voice for Ally, but trying to identify the most suitable personalities in a voice, for Ally. None of the sentences chosen are also gender neutral for the same reason.

The faces:

The section where the participants have to map faces with voices acts as an intermediary tool for participants to think of the qualities of each voice. The faces are a mere boundary object, and do not provide any insight by themselves. They only act as a medium to trigger the users' implicit needs, presumptions, preference or emotions. They are ambiguous tools that will help provide inspiration for the designer. They are provided as an emotional toolkit.

The faces were also provided to find out if people picture a person when they hear a voice. For example, when they have a car navigation system, do they picture a face to the voice provided by the GPS? Although this does not deal with or influence the goal of this research, it would still be an interesting insight.

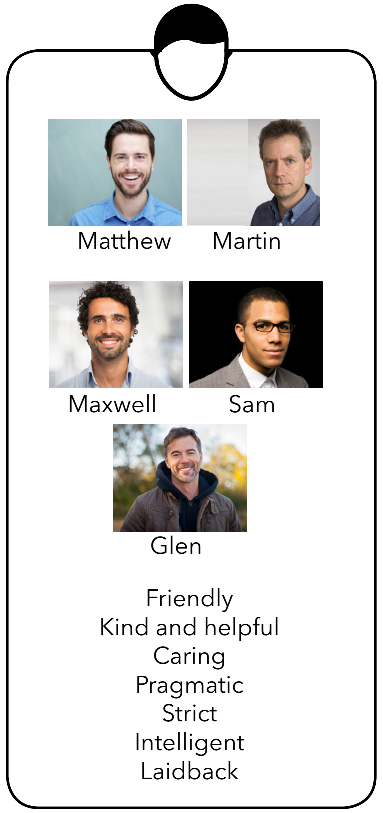
The options provided as qualities range from positive to negative and cover all the qualities given in the intangible characteristics of Figure 6. If a person feels that they need to add more qualities, they can also do so in the questionnaire.



Friendly
Caring
Assertive
Authoritative
Enabling

Supportive
Trustworthy
Soft
Pragmatic

Figure 26: Characteristics given as option for voice



Friendly
Kind and helpful
Caring
Pragmatic
Strict
Intelligent
Laidback

Figure 27: Faces and the characteristics given as options.

In order to find out which voices users prefer for medical advice the most, three other options are provided along with "medical advice conversations" in order to give the user a broader idea to compare the voices for different roles and make a suitable conclusion.

The personalities provided for the faces are also similar to the ones provided for the voice test, to keep the users close.

The result and analysis section will follow in the next page.

2.3.1
Result and Analysis

This section discusses the results and insights from the user test. The analysis process is depicted in Figure 28. The data of all the steps included in the analysis can be found in Appendix D.2.
The analysis was done for:
1. All the data sets together
2. Splitting the data set generation wise.

Since the faces acted only as a medium to get people to think deeper and choose more qualities, the voice characteristics and the characteristics of the faces mapped to the respective voices are combined.

Of the five voices, only three were chosen for the final analysis as the other did not receive high scores for "Medical device conversations" as a usage", thus becoming irrelevant for this project. The most optimal way to describe the voices that have been narrowed down on paper are sound as follows:

Voice 1: Middle aged, assertive yet friendly
Voice 2: Deep but friendly
Voice 3: Slightly deep, yet friendly

It is important to note that the participants of this user test are from various countries, including Netherlands, Belgium, India, Spain, Romania and others. The age group also ranges from 22 to 72. Hence, every similarity found in the following section is applicable universally.

Figure 29 represents the qualities present in each voice, of the total participant group. We can see that on a generic level, people prefer a voice that is always friendly and pragmatic. Two out of three voices are considered caring, supporting, kind and helpful and intelligent, thus making it a secondary preference.

We will now move on to individual i.e generational preferences of voices. Figure 30 illustrates the characteristics each generation prefer the most in a voice for a medical device.

Characteristics	Voice 1	Voice 2	Voice 3
Friendly	🟩	🟩	🟩
Caring		🟦	🟦
Supportive		🟦	🟦
Kind and helpful		🟦	🟦
Intelligent	🟦	🟦	
Pragmatic	🟩	🟩	🟩
Strict	🟩		

Figure 29: Voice characteristics of entire dataset

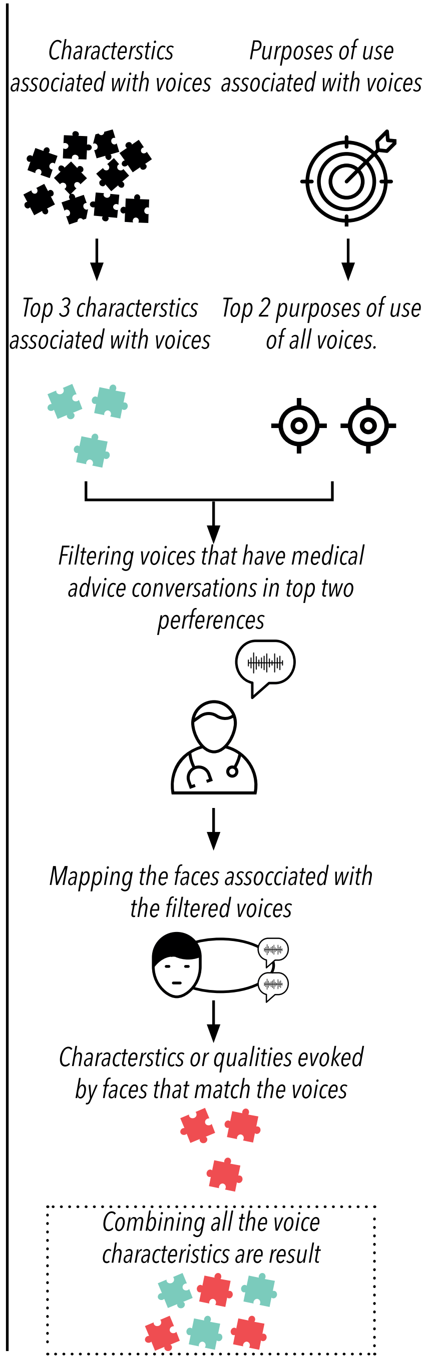
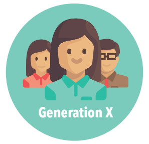

Characterstics associated with voices
Purposes of use associated with voices
Top 3 characterstics associated with voices
Top 2 purposes of use of all voices.
Filtering voices that have medical advice conversations in top two perferences
Mapping the faces associated with the filtered voices
Characterstics or qualities evoked by faces that match the voices
Combining all the voice characteristics are result

Figure 28: Analysis process of the user test questionnaire

🟩 Occurs in all 3 voices
🟦 Occurs in 2 of 3 voices
🟥 Occurs in 1 of 3 voices

Figure 30: Preferred voice characteristics of each generation

Characteristics	Voice 1	Voice 2	Voice 3
Friendly	🟩	🟩	🟩
Caring	🟦	🟦	
Supportive			🟥
Kind and helpful		🟦	🟦
Intelligent	🟦	🟦	
Pragmatic	🟥		
Strict	🟥		
Authoritative			
Assertive	🟥		
Trustworthy			



Characteristics	Voice 1	Voice 2	Voice 3
Friendly	🟦	🟦	
Caring		🟦	🟦
Supportive		🟦	🟦
Kind and helpful		🟦	🟦
Intelligent	🟦	🟦	
Pragmatic	🟩	🟩	🟩
Strict	🟥		
Authoritative	🟦		🟦
Assertive			
Trustworthy	🟥		

Characteristics	Voice 1	Voice 2	Voice 3
Friendly	🟩	🟩	🟩
Caring		🟥	
Supportive		🟦	🟦
Kind and helpful		🟦	🟦
Intelligent	🟩	🟩	🟩
Pragmatic	🟥		
Strict	🟥		
Authoritative	🟥		
Assertive	🟥		
Trustworthy	🟥		

As seen in the figure, two more features have been added to the list. This is because, while selecting the top characteristics and combining them, the characteristics with a 1/3 probability (occurring in 1 of 3 voices) per generation were compared with the other generations. If they occured in one of the other generations, resulting in a probability of 2/9, they were added to the list. If they had a probability of 1/3 in one generation and existed in only that generation's data, the resultant probability is 1/9. Such characteristics were eliminated. Thus, characteristics with a 2/9 probability in the generation-wise data are not that prominent in the overall generalised data, but are nevertheless important for the respective generation's opinion.

Baby boomers:

From the above figure we can say that baby boomers like the voice for a medical device to be always friendly. This correlates with that they prefer in an interaction, given in Figure 22. The voice should sound caring, kind and helpful, intelligent two out of three times. And one in three times it shound sound supportive, pragmatic, assertive and strict.

Generation X:

Gen X want a voice that sounds pragmatic. This kind of correlates to the insights from section 2.2.1 where they are curt and to the point with the device. Gen-Xers seem to prefer more qualities in the voice than baby boomers.

Millennials

Millennials want the voice to sound friendly and intelligent. The other relatively prominent features they want is supportiveness and kindness.The other characteristics should exist one out of three times.

It is important to note that although all generations prefer similar characteristics, it is the probability in which they prefer it that matters. For example, while both Baby Boomers and Millennials do want pragmatism in the voice, it is only a fraction compared to Gen X. So, this might change the way the voice sounds and delivers a dialogue.

More about how these fractions of qualities is applied will be discussed in the next chapter.

Conclusions:

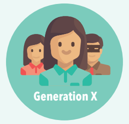
Thus in this section, we analysed the voice characteristics preferred by each generation and in the probability in which they like it to occur. The relevance of this analysis can be understood better in the next chapter.

Summary and Conclusions

Understanding perception of users

A generative session is conducted to get a deeper understaning how consumer electronices is perceived by different generations. The outcome as resulted in specific qualities preferred by each generation in a product

Ease of use
Functionality
Clear and Easy
Respect
Kindness



Relatable
Relavant
Human
Honest
Usable
Functional

Meaningful
Content is
important
Usable
Trustworthy

Interactions between different generations and Ally

Going by the hypothesis, 3 scenarios is created for users to interact with Ally. A phrase questionnaire to rate the phrases based on qualities is created. The outcome is a set of interaction qualities each generation likes in Ally and the qualities of phrases each generation prefers. The results from this will be used in section 3.1.1.

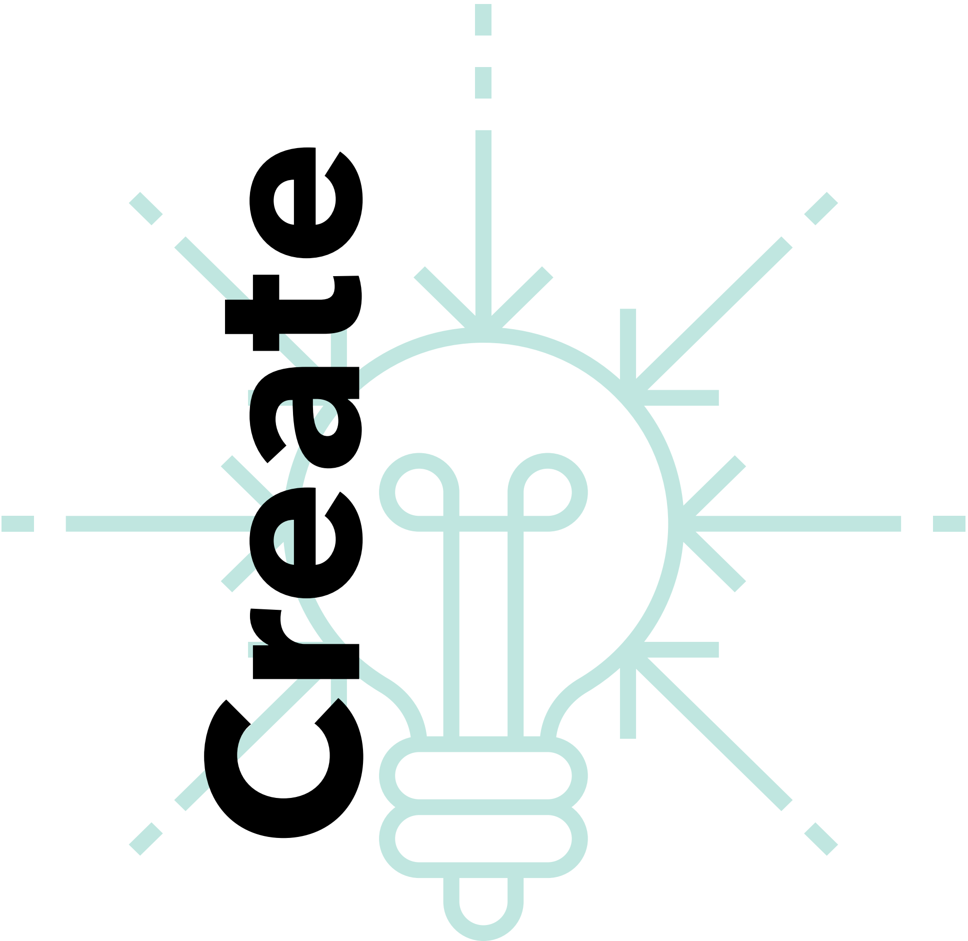
Voice characteristics of a medical device

To identify the kind of characteristics different users prefer in the voice of a medical device, an online questionnaire is designed.
The probabilities in which each generation prefers different characteristics is determined. This will be used in section 3.1.1.

Going back to the hypothesis

The assumption that every generation has different ways of interaction holds, based on the insights of this section 2.1.1, 2.2.1 and 2.3.1. It can also be said that the values of what they look for in Ally differs generation to generation, although there are some similarities between Baby Boomers in Millennials

Create



03 Conceptualization

The final concept, that is a deep learning framework to define an interaction between users and a medical pod is conceptualised in this chapter. Results from the previous chapter is used as input to some elements of the framework. The chapter begins with an introduction to Machine learning and goes to explain what features are, how to train a neural network in the context of this project, voice synthesis and finally, the framework for interaction.

3.1 Machine learning and deep learning

In section1.5, various aspects of "intelligence" and what artificial intelligence means was discussed. It has been established Ally is an intelligent device, and a lot of insights to make the interaction between Ally and the users more personalised and easy has been collected in the previous chapter. ***In this chapter, we will create a framework or the foundation that would help build and define what an interaction between Ally and different generations should be like.***

Before moving on to the concept, it is first important to get acquainted to the the field of machine learning and its applications at present. The topic was briefly introduced in Chapter 1, while discussing different capabilities a computer needs to have to "act humanly".

We use machine learning technology in our daily life without actually realising what it is. The examples of daily usage are given on the right. Machine learning is a field grown out of Artificial Intelligence. (Ng, Andrew (2017)
Machine learning is applied to the field of medicine at present already. One of Andrew Ng.'s lecture on Machine Learning (Ng, Andrew (2017) states that

"With the advent of automation, we now have electronic medical records, so if we can turn medical records into medical knowledge, then we can start to understand disease better."

This strengthens our reasoning to use Machine Learning or more specifically, ***Deep Learning for Ally, as the system can learn to understand diseases better, and more importantly in our concept, create a seamless interaction to discuss medical issues with the user.***

But what exactly is machine learning?

There is no one single definition for machine learning. But for simplicity, let us consider the definision by Tom Mitchell (1999). He defined a well posed learning problem as

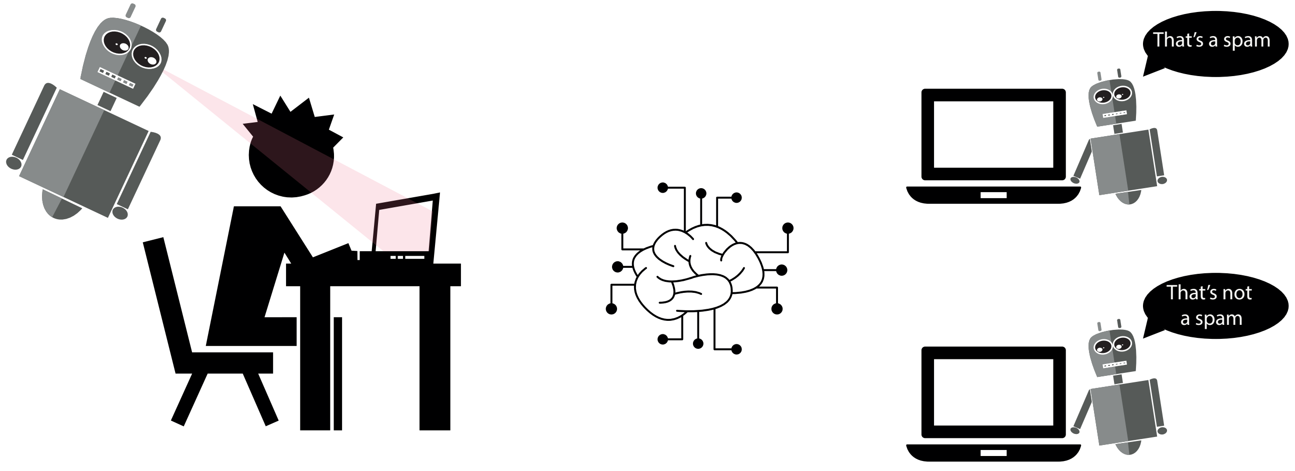
"A computer program is said to learn from experience E, with respect to some task T, and some performance measure P, if its performance on T as measured by P improves with experience E."

While the definition may sound a little tricky, it can be understood better by the example given below.

An example of the above definition given by Andrew Ng, can be seen as follows: Assume there is an email program that

"Every time you use a web search engine like Google or Bing to search the internet, one of the reasons that works so well is because a learning algorithm, one implemented by Google or Microsoft, has learned how to rank web page"

"Every time you read your email and your spam filter saves you from having to wade through tons of spam email, that's also a learning algorithm."



Sees user mark mails as spam or not spam

Learns what makes an email a spam or not

Predicts if a certain email is a spam or not from experience.

Figure 31: Spam classification using a machine learning algorithm

marks emails as spam or not. How does it emails to mark as spam and what emails to not mark? It learns by experience. As shown in Figure 31, it first learns by noting what emails users mark as spam and what as not-spam. This is the experience E. It then learns from the experience E and performs the task T, of marking emails as spam and not-spam by itself. The performance measure P, in this case is the fraction of emails that it classifies correctly. Thus by performing a machine learning algorithm, the program has learnt to classify emails.

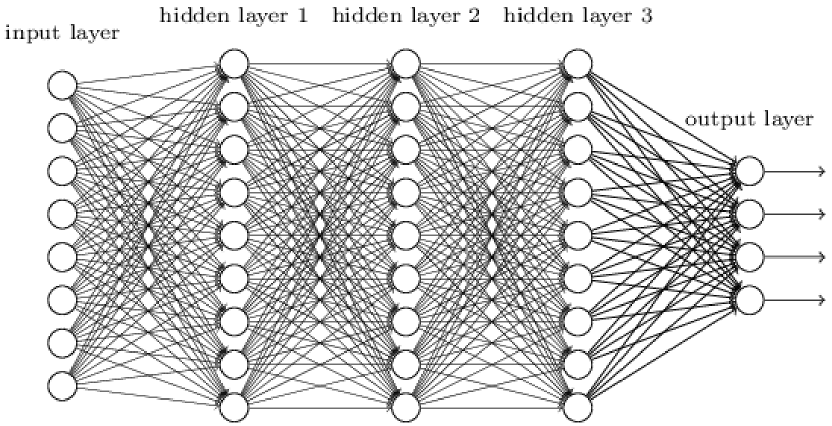


Figure 32: Deep learning algorithm with its dense hidden layers

The above example of spam classification can be made even more accurate by using a Deep Learning framework.

Deep learning is a subset of Machine Learning, that uses Neural Networks with multiple hidden layers. (Section 1.5).

This means that the neural network in a deep learning algorithm is dense (Figure 32).

There are three main types of deep learning methods. Supervised learning, unsupervised learning and reinforced learning. In supervised learning, the network is initially taught to perform a certain action while in unsupervised learning, the network learns by itself. In this project, we will focus on supervised learning.

Supervised Learning: In supervised learning, you have a set of input x for which you have output y. Recollecting the for understanding neural networks in Chapter 1, section 1.5 (Figure 9), we had two input features, the length and width of the petal, that gave the flower type as the output. That is an example of supervised learning.

We now know that by providing input to a well trained supervised deep learning algorithm, we can get a desirable output.

3.1.1
Features

The goal of this project is to create an interaction between users and a medical device based on who is interacting with it. In this project, this will be done by creating a theoretical for a deep learning framework to produce the ideal voice interaction.

Thus, our output is an interaction or a dialogue. From the results from various user tests in Chapter 2, our "inputs" can be split into voice inputs and phrase inputs.

In this context, inputs can also be referred to as "Features", as these are the attributes that we required for a deep learning algorithm to give a prediction. ***One of the toughest processes in creating a deep learning algorithm for a task is identifying the right features.*** From the results of the various user tests conducted in Chapter 2, sections 2.2.1 and 2.3.1, the most appropriate results have been formulated into features. These will be discussed in this section.

A feature is an attribute or property shared by all of the independent units on which analysis or prediction is to be done. Any attribute could be a feature, as long as it is useful to the model. (Wikipedia, 2018)

"Coming up with features is difficult, time-consuming, requires expert knowledge. "Applied machine learning" is basically feature engineering."
–Andrew Ng, Machine Learning and AI via Brain simulation"

Voice features:

From the user tests conducted in section 2.3 of Chapter 2, the voice characteristics preferred by each generation for a medical device was identified. These will act as the features to generate the right voice based on the user.

Figure 30 in section 2.3.1 represents the various characteristics different generations prefer in the voice for a medical device, with the probability they like it in. These, will act as features for the final framework.

Figure 33 illustrates the voice characteristics or features as preferred by each generation. These are grouped into the probability in which they should occur. For example, Baby boomers want a caring, intellegent and a kind and helpful voice two out of three times.

So when a voice is generated by Ally for baby boomers, it will be friendly 100% of the time, while it will be try to incorporate a caring, kind and helpful, and intelligent voice, 2 out of 3 times it generates a statement. These probabilities later change based on user feedback. This is discussed in section 3.2.1.

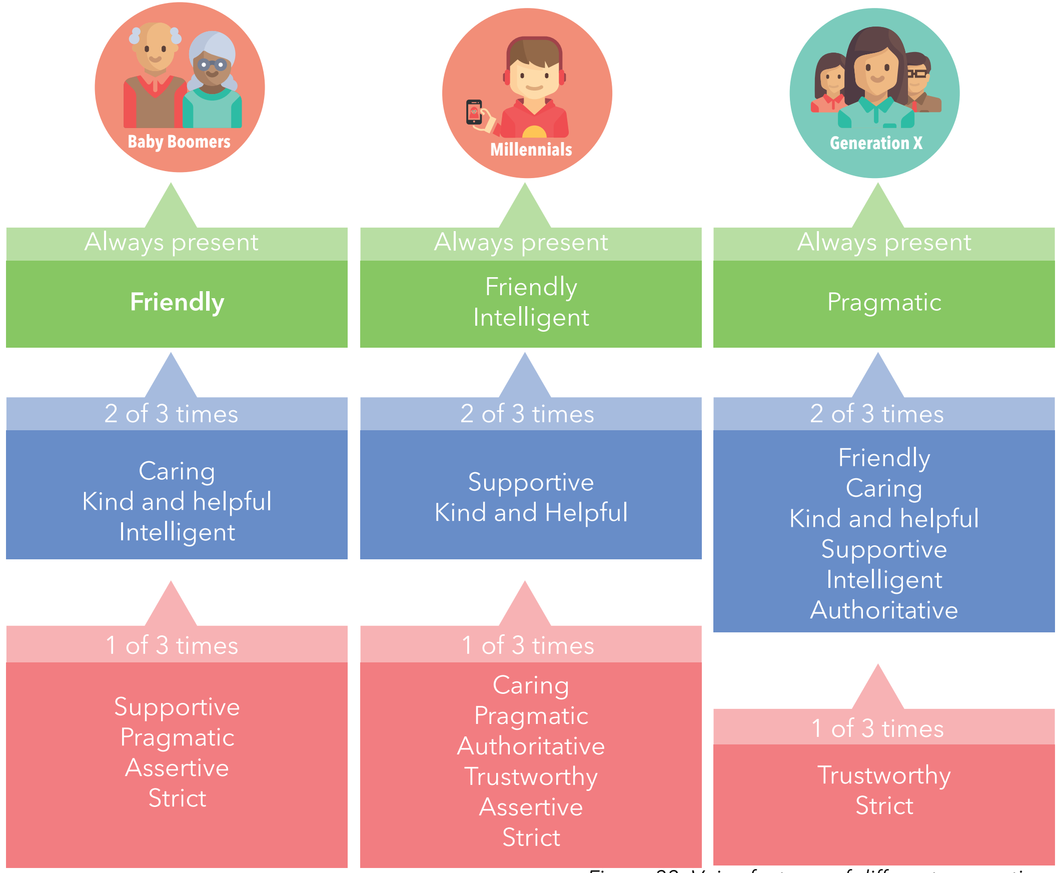


Figure 33: Voice features of different generations

Phrase features:

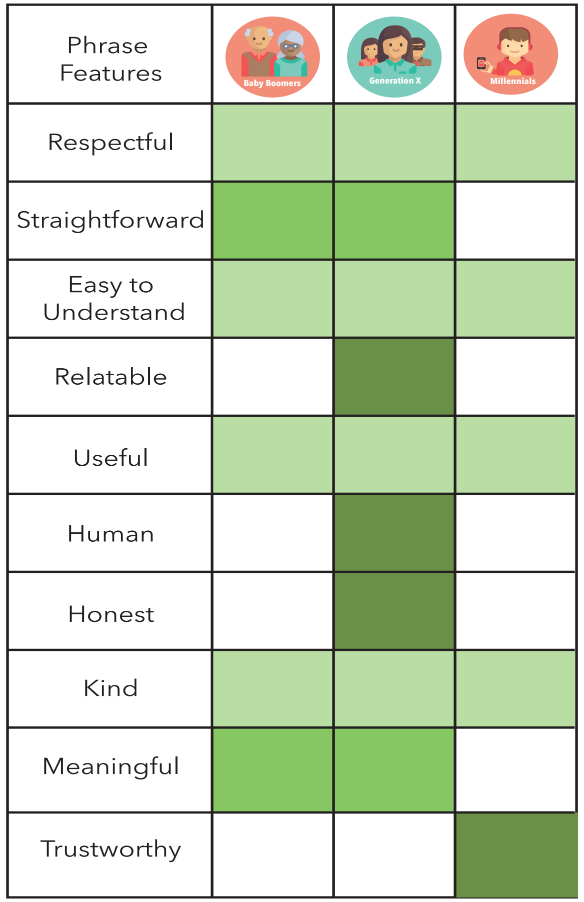
While voice is one part of the interaction, we have also collected insight about the kind of sentences different generations like and prefer. These are classified by the characteristic nature of the sentences. Thus a set of features each sentence is created. Each generation likes to have the characteristics given in Figure 34.

These features were extracted from Figures 23, 24 and 25 of section 2.2. The qualities each generations thought the phrase was inching towards was taking as the features. Thus, from the three figures mentioned above, since Generation X had a mixed opinion, the number of features are relatively more as opposed to millennials, who were consistent in the way they rated the phrases they preferred the most.

Unlike voice features, these are not based on probability as the given dataset produces a 1 or ½ chance of a characteristic.

Hence, instead of choosing to make it a probability the entire set of features (created for that respective generation) should exist for each phrase generated.

Thus we now know the kind of voice and phrase characteristic the interaction needs to have. But, how will a machine know what friendly or caring means? This brings us to the next section, the training of a network.



Phrase Features	Baby Boomers	Generation X	Millennials
Respectful	🟩	🟩	🟩
Straightforward	🟩	🟩	
Easy to Understand	🟩	🟩	🟩
Relatable		🟩	
Useful	🟩	🟩	🟩
Human		🟩	
Honest		🟩	
Kind	🟩	🟩	🟩
Meaningful	🟩	🟩	
Trustworthy			🟩

Figure 34: Phrase features of different generations

3.1.2
Training a neural network

For an deep learning algorithm to understand what it means for a voice or phrase to be friendly or caring or any of the other characteristic, we need to feed it a lot of friendly voices/phrases so it can "learn" what makes the voice or phrase have that characteristic.

Voice classification:

Let us see how an algorithm is trained to learn about voice characteristics. For simplicity, let us now assume that we need the machine to learn what a friendly voice is. The first step in this process involves collecting a dataset of friendly voices.

Step 1: Let's say 50 voices are chosen, of all genders. ***The sample size here is only an assumption***. Users or participants need to listen to this voice and identify which voice they find friendly and which they find as unfriendly. Let's say, for example 26 voices are classified as friendly and 24 are unfriendly by the users. Figure 35 illustrates this process.

Thus we now have a data set comprising of friendly and unfriendly voices. Like in the flower example, where the length and width were features, here friendly and unfriendly are the features and the voices are the outputs.

Step 2: The next step in the process is feed the deep learning algorithm with the friendly and unfriendly voices. The algorithm will compare and try to build patterns on what makes a voice friendly and what makes a voice unfriendly. It will identify the attribute that makes a voice friendly (Figure 36). This is called a trained network.

Note that the process where the machine "figures" out the attributes that make it friendly or unfriendly is the hidden layer (as mentioned in Figure 32). It is not possible for us to know what the attribute is, neither is it important. This is the biggest advantage of neural networks. The machine figures it out for us. This is a concept termed as a Blackbox approach. Where the inputs and outputs are known, and are processed through the box, but what happens inside this hypothetical box is unknown.

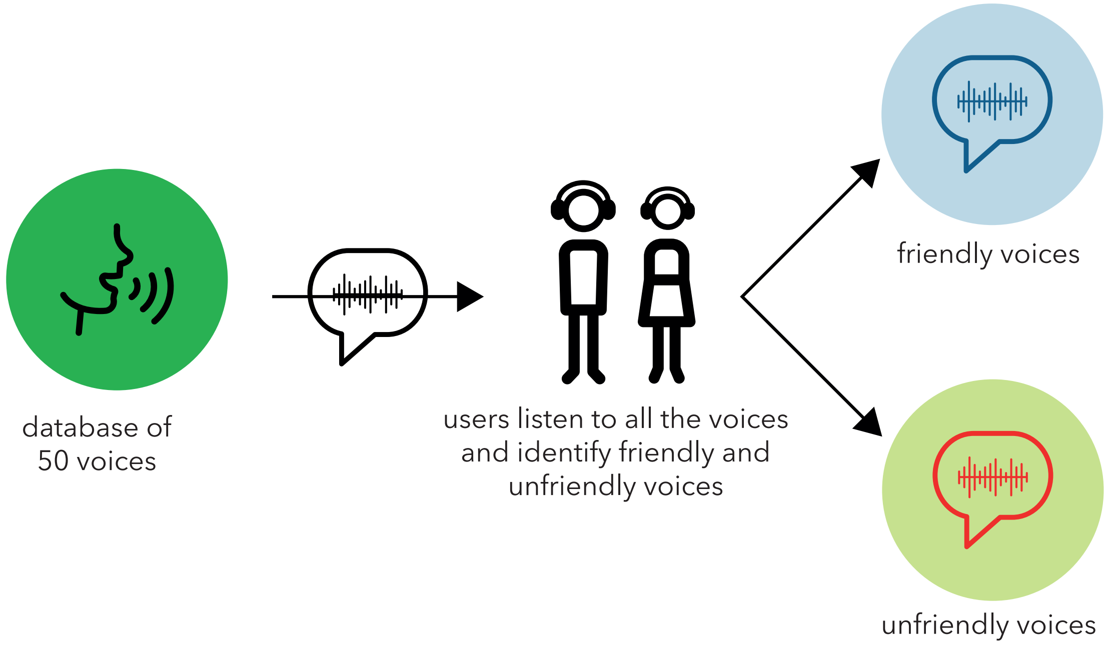

friendly voices
database of 50 voices
users listen to all the voices and identify friendly and unfriendly voices
unfriendly voices

Figure 35: Manual classification of friendly and unfriendly voices

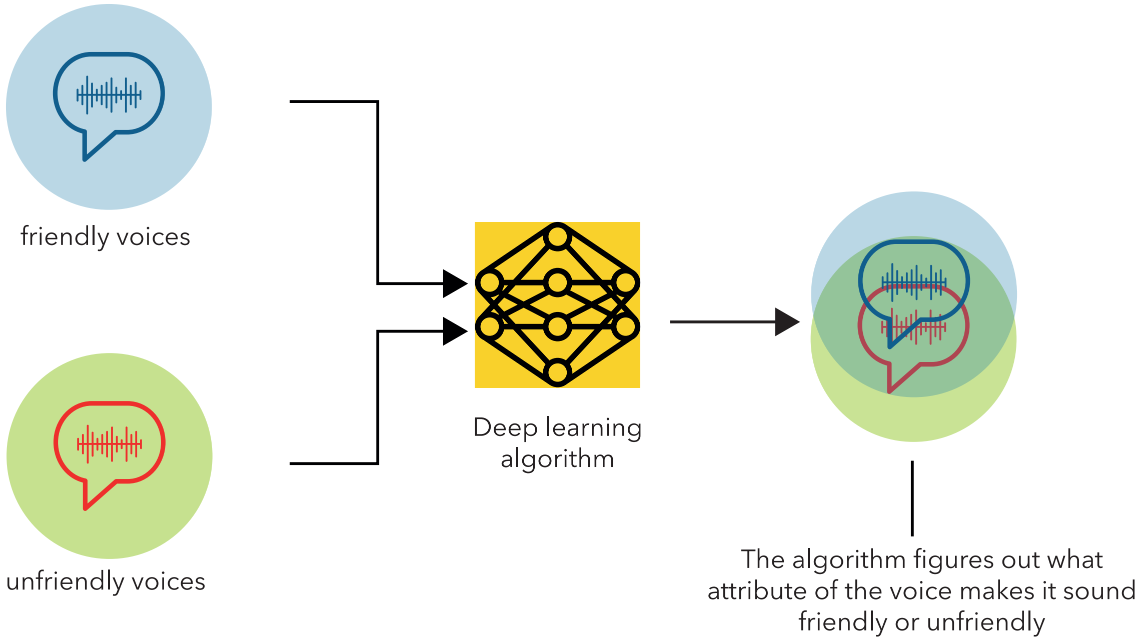

friendly voices
Deep learning algorithm
unfriendly voices
The algorithm figures out what attribute of the voice makes it sound friendly or unfriendly

Figure 36: Training the deep learning algorithm to learn friendly and unfriendly voices

Thus the trained network identifies the underlying attributes that make the voice have a certain characteristic. ***Similarly, all the features (that are created) of the voice will be run through the users and then through the network, for training.***

Once all the features are inputted and the algorithm is trained, we have what is called a developed network**.** ***The developed network contains the collection of all the attributes responsobile for the various features.*** The figure 37 shows the steps of the voice feature classification, at a generalised level (i.e for all feature training).

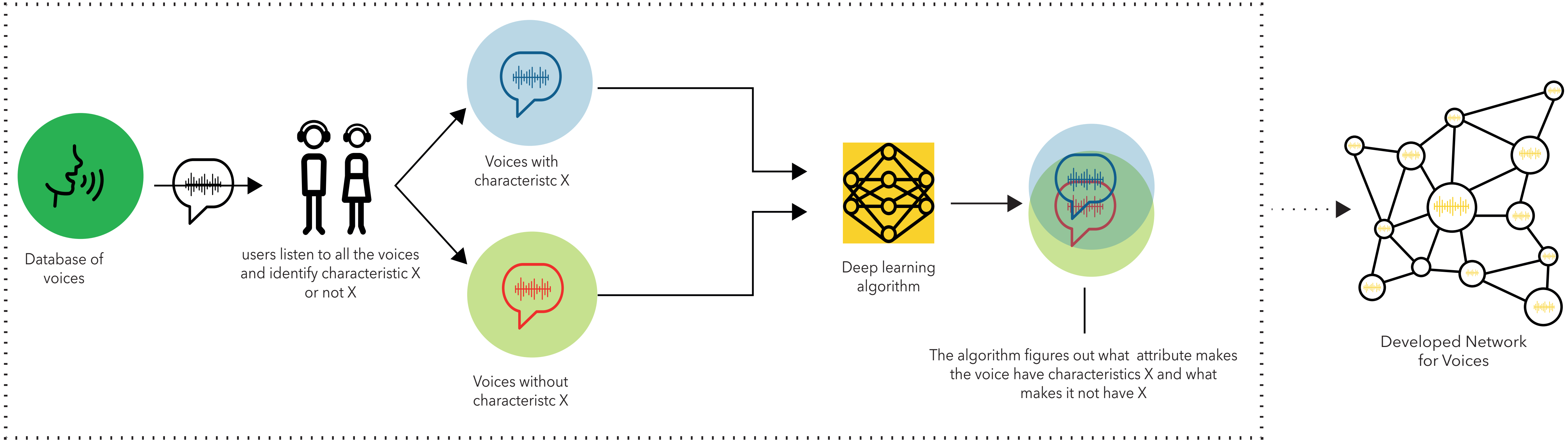

Database of voices
users listen to all the voices and identify characteristic X or not X
Voices with characteristc X
Voices without characteristc X
Deep learning algorithm
The algorithm figures out what attribute makes the voice have characteristics X and what makes it not have X
Developed Network for Voices

Figure 37: A developed network comprising all voice features

Phrase classification:

As in the previous section, the phrases that have the features of respect, straightforwardness etc need to be classified and the machine needs to be trained. The steps for this is also the same as in the previous section. Let us now assume that we need the machine to learn what a respectful sentence is.

Step 1: To create a dataset of respectful sentences, we need to get participants to rate many sentences to be respectful/ not respectful. Note that the phrase that is not-respectful doesn't necessarily make it disrespectful. These sentences are then classified so (Figure 38)

Step 2: The next step would be to feed these sentences to the deep learning algorithm. Like in the previous example, the algorithm will compare the dataset of respectful and disrespectful sentences and identify the attribute(s) that are responsible for their nature. It will now know how to produce a respectful sentence. Thus a trained network for classifying the phrase features is created (Figure 39)

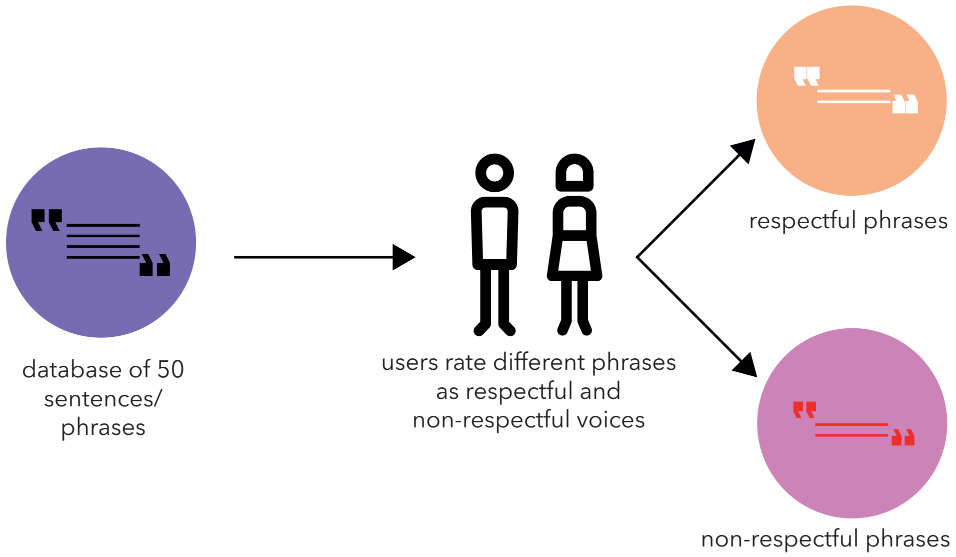

database of 50 sentences/ phrases
users rate different phrases as respectful and non-respectful voices
respectful phrases
non-respectful phrases

Figure 38: Manual classification of respectful and non-respectful phrases

All the features (that are created) of the phrase will be run through the users and through the network, for training. Thus another developed network for phrases is created. The figure 40 shows the steps of the phrase feature classification, at a generalised level (i.e for all feature training).

This completes the training network segment of the concept. This will later be implemented into the a larger network.

It is important to note the final model will generate is completely new voice(s) and sentence(s). It will not necessarily generate one of the voices or phrases that are inputted for training. For example, for the training, 2 different voices with one known to be "friendly" but not assertive while one known to be "assertive" but not friendly can be fed into the algorithm and it picks the attributes that make it friendly and assertive and generates a whole new voice with both the features. So how is this entirely new voice synthesized? This is what we will deal with in the next section.

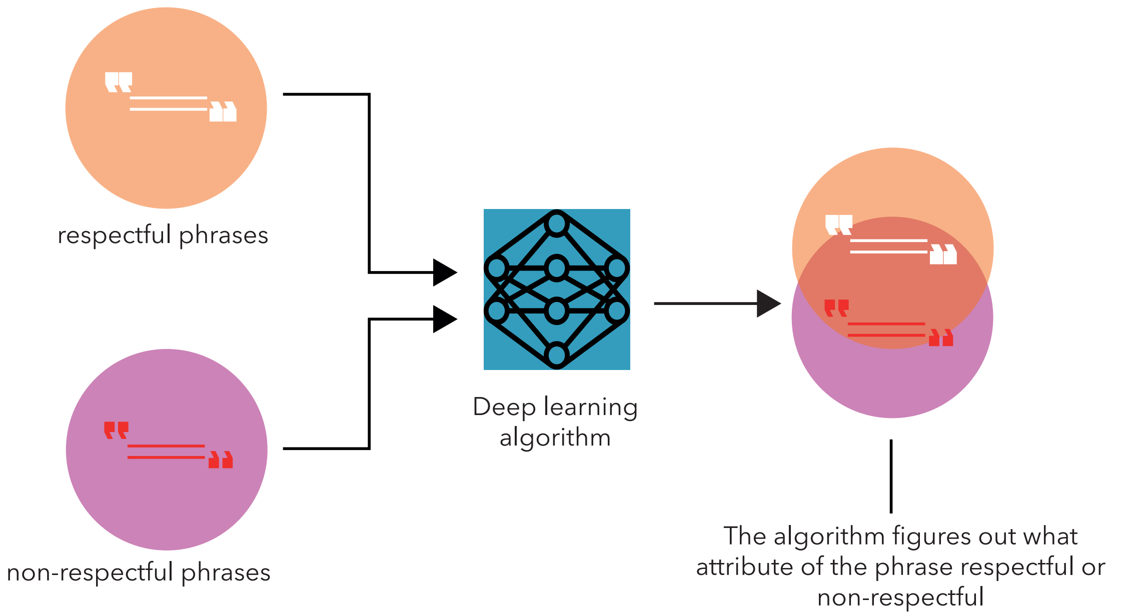

respectful phrases
non-respectful phrases
Deep learning algorithm
The algorithm figures out what attribute of the phrase respectful or non-respectful

Figure 39: Training the deep learning algorithm to learn respectful and non-respectful phrases

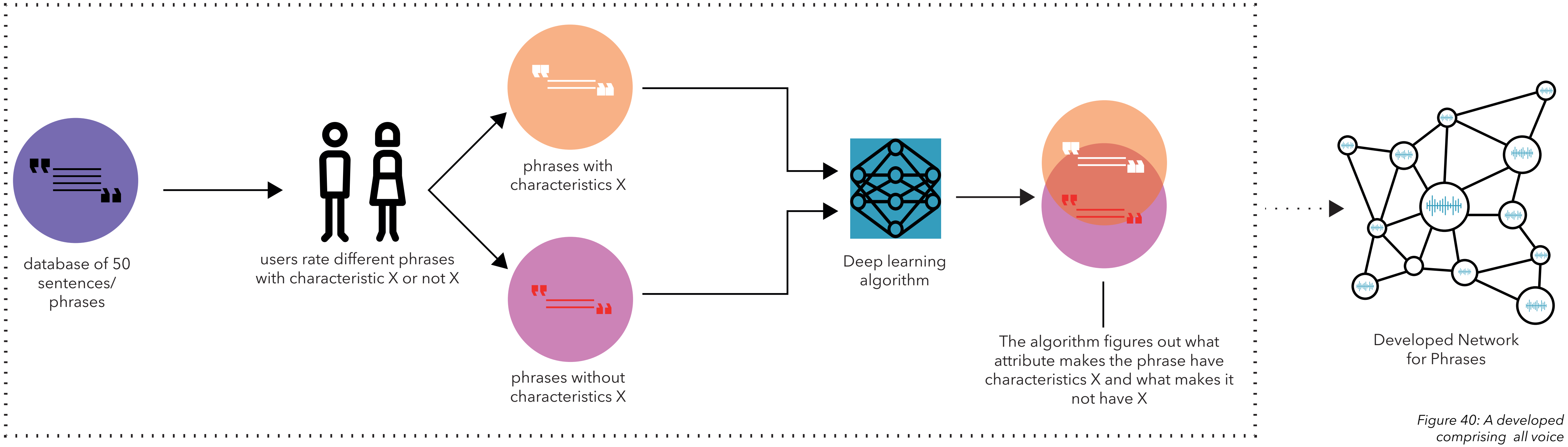

database of 50 sentences/ phrases
users rate different phrases with characteristic X or not X
phrases with characteristics X
phrases without characteristics X
Deep learning algorithm
The algorithm figures out what attribute makes the phrase have characteristics X and what makes it not have X
Developed Network for Phrases

Figure 40: A developed network comprising all voice features

3.1.3
WaveNet

According to Aaron van den Oord et al 2016, while the ability of machines to understand human speech or natural language has been revolutionised in recent years (eg. Google voice search), the synthesis of speech by computers a.k.a text to speech (TTS) still uses a method called “Concatenative TTS” or ”Parametric TTS”, which tends to be robotic and un-humanlike

An artificial intelligence company called DeepMind (DeepMind, 2018), has created a new tool called WaveNet to synthesis voice using neural networks. Wavenet is a deep generative model of raw audio waveforms. In the research conducted by Aaron van den Oord et al 2016, when people were asked to rate different speeches on a scale of 1 to 5, it was found that WaveNet speech synthesis was rated closest to human speech, compared to parametric and concatenative (Figure 41). Mean Opinion Scores were used to measure the ratings of speech.

For Ally’s proposed interaction framework, it is recommended to use WaveNet to synthesize speech. Now that we have the method for speech synthesis, the next step is to understand the process of speech synthesis as a whole.

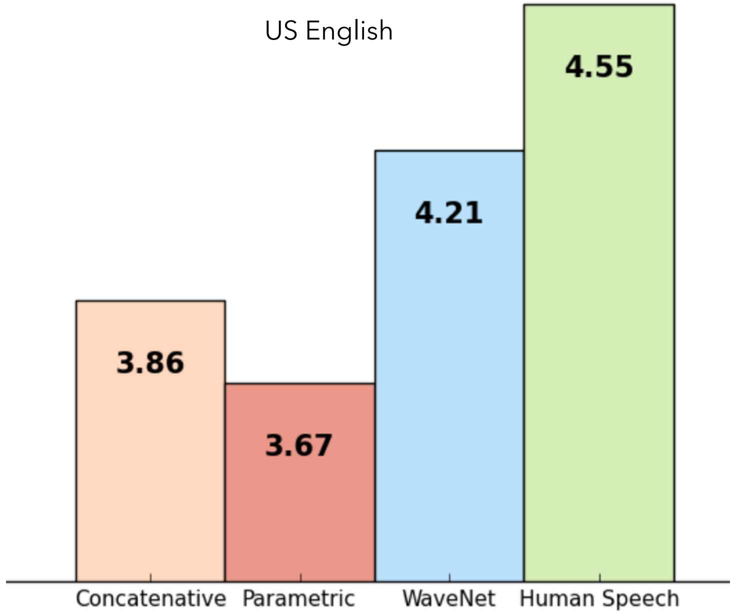


Figure 41: Mean opinion score of different TTS voice synthesis (DeepMind, 2018)

Concatenative TTS: In this kind of speech synthesis technique, a very large database of short fragments of speech are recorded by a single speaker. This is then recombined to form whole sentences and utterances. Such a method does not leave room for alterations in emphasis of speech or emotion and an entire new database of speech fragments have to be recorded again to obtain a different result. *(DeepMind, 2018)*

Parametric TTS: In this technique, all the information required to generate the data are stored in the parameters of a synthesis model. The existing models generate audio signals by passing their outputs through signal processing algorithms known as vocoders. This model is known to sound more synthesized and unhuman-like than the concatenative model *(DeepMind, 2018).*

3.1.4
Speech Synthesis Model

Let us see the components that are involved in a model that synthesizes speech. The explanation given below is a simplified one, without delving in the technical details of the process.

In a speech synthesis model, a ***database of raw text is first analysed. This is where all the “content” of the conversation or interaction is processed***. The text “content” is sent to ***linguistic analysis next, where the “voice” part of the conversation is processed.*** Here, ***the phasing, intonation and duration are generated.***

Thus this segment is responsible for synthesising the different elements of the speech synthesis. The next part of this synthesis is ***phoneme generation.***

A phoneme is one of the units of sound (or gesture in the case of sign languages, see chereme) that distinguish one word from another in a particular language (Wikipedia, 2018)

Thus, each word occurrence is framed and is sent to the waveform generation unit where they are put together and finally, the final waveform of speech is generated.

With each step of the speech synthesis process established, we can now move on to the final step of this design, creating the theoretical model to generate the interaction between the users and the device, based on who is talking to it.

Phasing of the signal tends to play a very important role in human speech synthesis and recognization. The phase spectrum provides useful information that contribute to speech intelligibility and specifying intervocalic and stowap consonants (Shi, Shanechi and Aarabi, 2006)

Intonation refers to the rise and fall of voice during speech. It is variation of spoken pitch that is not used to distinguish words; instead it is used for a range of functions such as indicating the attitudes and emotions of the speaker, signalling the difference between statements and questions, and between different types of questions, focusing attention on important elements of the spoken message and also helping to regulate conversational interaction (Wikipedia, 2018).

Duration refers to how long or short a note, phrase is. It is the “ ”Duration is the length of time a pitch, or tone, is sounded.”(Wikipedia, 2018)

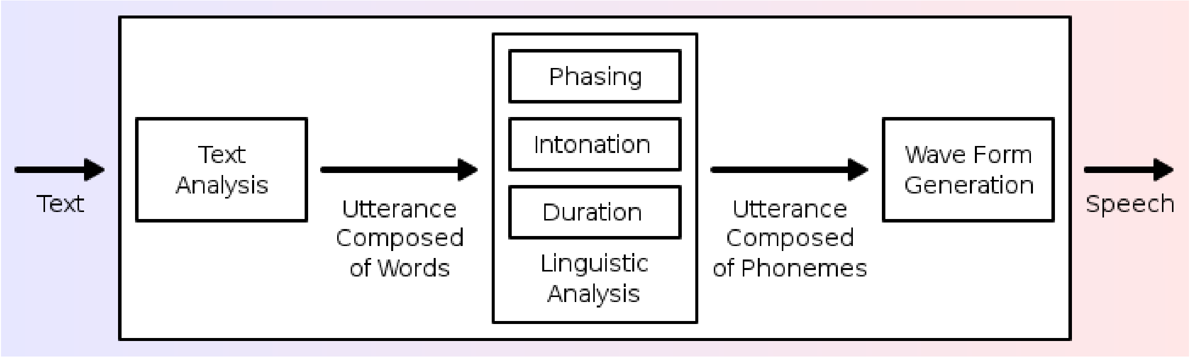


Figure 42: Voice synthesis model (Allen et al; 1987)

3.2
A Deep Learning Framework to create a personalised interaction between Ally and users

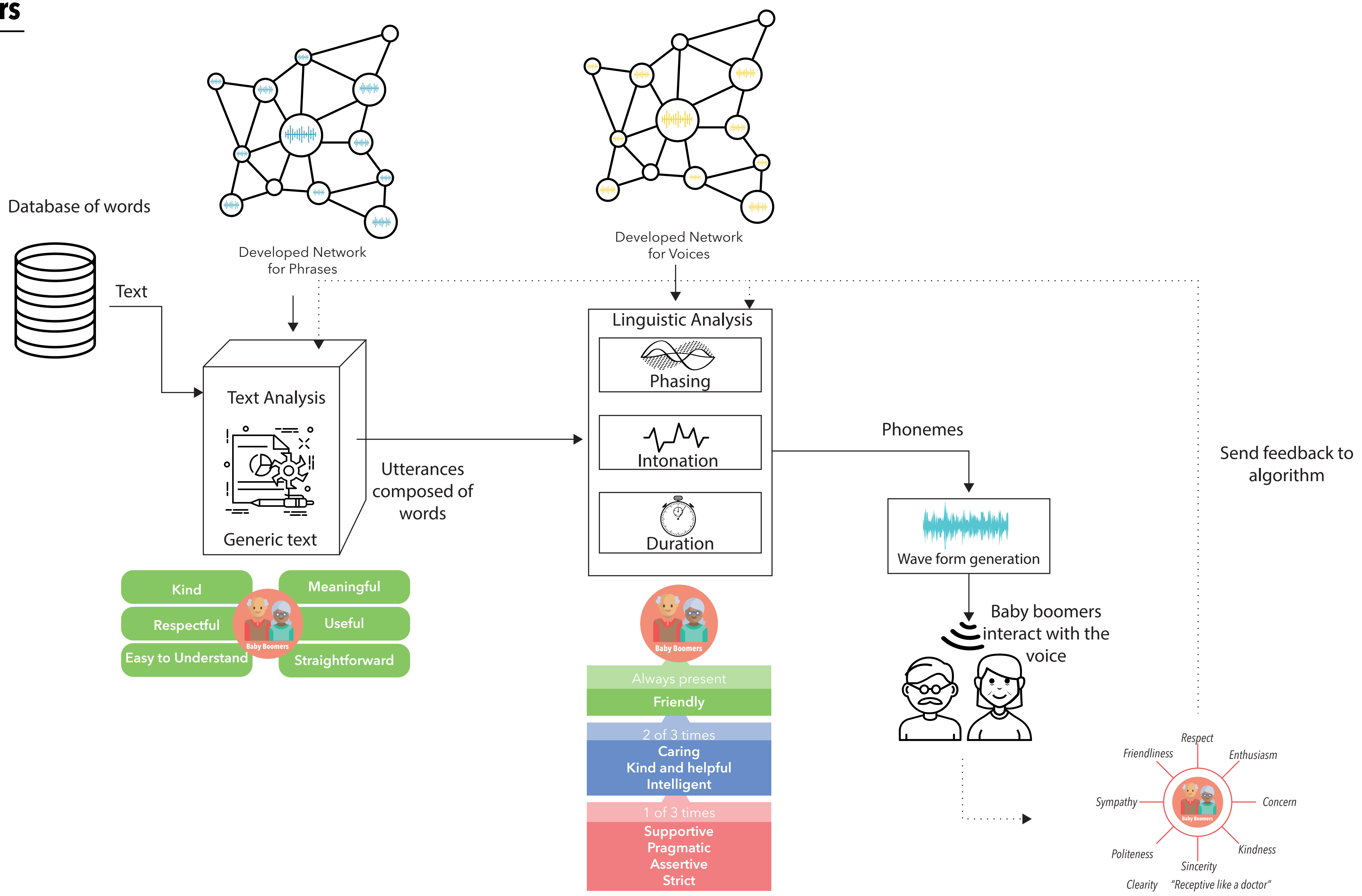
Since this framework personalizes the interaction betweel Ally and the user based on their generation, let us assume that the deep learning network is creating an interaction between the machine and a baby boomer.

The first part of the model is ***collecting a database of words and text and feeding it to the text analysis block.***

Like mentioned in the speech synthesis section, ***the text is analysed and the "content" of the interaction is created. Here, the developed network to classify attributes of different phrases from section 3.1.2 (Figure 38), is fed in to the text analysis.*** These attributes are for all the features that are created. To have a user or generation specific interaction, only some phrase features are required. Hence, ***the phrase features developed (Figure 34) for the respective generation is fed in, based on the respective users.*** For example, in the model given in Figure 41, the phrase features of baby boomers is given to the text analysis segment.

Once the content of the interaction is identified, ***the utterances composed of words are sent to the linguistic analysis where the phasing, tone/ intonation and speed/duration are processed.*** Since this is the

Figure 43: A deep learning framework for personalised interaction between Ally and different users



place where the voice linguistics are created, ***the developed network of the voice features (Figure 37) is fed in here.*** Like in the developed network for phrase, the developed network for voice contains all the attributes for features. ***So, in this example where the model has to create an interaction with a baby boomer, the voice features required for baby boomers from section 3.1.1 (Figure 33) are fed into this unit.***

From this block, phonemes are created and these are sent to the waveform generation block which produces the final output, i.e the voice with various voice features, speaking a sentence that is both meaningful and has the phrase features that baby boomers have known to prefer.

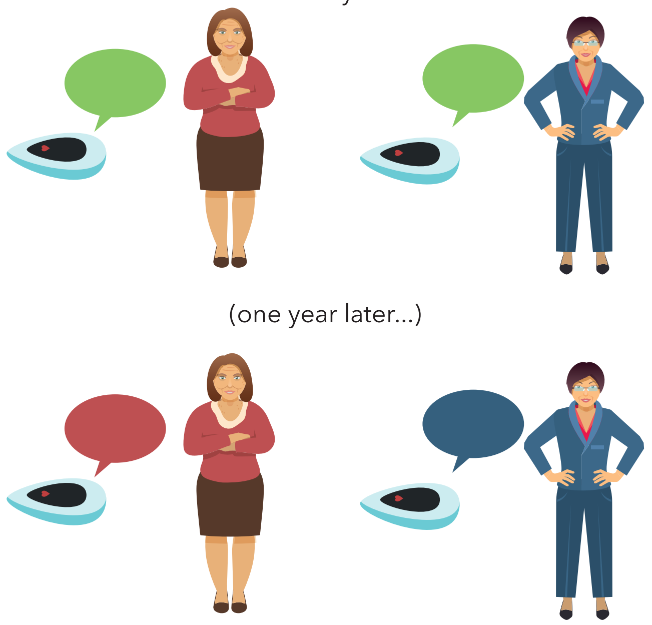
Now, the goal of creating the interaction based on the user is achieved, ***but it is also important to know if the entire interaction is in accordance to the preference of the generation.*** From section 2.2.1, we have classified the qualities each generation like most about an interaction with Ally (Figure 22). Hence, ***a feedback loop is created where the AI occasionally asks the user if the interaction checks all the boxes of the qualities preferred by each generation. Depending on the feedback of the user, the machine learns to change and adjust to make the interaction more seamless.***

3.2.1 Personalization of Interaction

The goal of this project is to create a personalised interaction between users and Ally. ***By identifying and creating a model to cater to specific generations, this model has set the ground rules for personalization.*** It is a more generalized model for personalization. It sets some ground rules to target 3 large groups of people. So, if say there is a baby boomer who prefers a more pragmatic interaction than a respectful or kind one, then the machine will either figure it out because the user says their opinion upfront, or by using the feedback loop mentioned in the previous paragraph, it can ask the user if they like the interaction or they need it to change.

Thus, if two baby boomers with different personalities start using it at the same time, a few months later, the personalities of the two devices might be different (Figure 44).

Both baby boomers start Ally at same time.
Ally interacts with the interaction features suited for baby boomers.



(one year later...)

Ally interacts in a more personalised manner adjusting its dialogue based on the person.

Figure 44: Ally interacts the same way with two people of the same generation initially and then adapts and learns based on more personalised preferences

But, initially the starting point created to personalization is important as it lowers the barrier of interaction between users and Ally thus, making the interaction more well defined and regular.

It is also important to note that this model covers only the voice interaction part of Ally. To be able to make medical predictions, which is the more technical aspect of the product, large databases of medical knowledge is required and it needs to be created in unison with doctors and other medical practitioners. This is beyond the scope of this thesis, as here, the focus is on the user interaction aspect.

Summary and conclusions

Deep Learning

Deep learning is a subset of Machine Learning, that uses Neural Networks with multiple hidden layers.

Features

The voice and phrase characteristics from sections 2.3.1 and 2.2.1 are refined to create features, that will act as the input to the training algorith and framework

Traning a deep learning network

In order to let a machine "learn" about the qualities of voices and phrases, they are trained with by creating networks for training.

1. A databse of voices and phrases are manually classified to segregate the database off different features.
2. These are then fed into the deep learning algorithm for the algorithm to learn the attributes that make the voices and phrases have different characteristics.

A developed networks are thus created.

Wavenet

Wavenet is a speech synthesis technique created by DeepMind to synthesize speech by using neural networks. Wavenet has been rated the closest to human voice based compared to other TTS methods.

Speech synthesis model

In a speech synthesis model, words are send from a database to the text analysis where the nature of text is analysed and sent to the linguistic analysis where the phasing, intonation and duration is processed to create a phoneme which is then sent to the waveform generator, where speech is produced as an output

A Deep Learning Framework to create the a personalised interaction between Ally and users

With all the elements put together, a deep learning framework for personalised interaction between users and Ally is created.
This framework will be a foundaton to create interactions between Ally and people of different ages. This is because the framework caters to each generation's value, thereby reducing the barrier of interaction and making the user feel more comfortable.
This initial interaction is vital for people to start using Ally

04
Reflection

This is the final chapter of this report. It starts with reflecting on the concept that was created in Chapter 3 and will go on to discuss the feasibility of the framework. The chapter then proceeds to the reflection on the research, hypothesis. Lastly, the limitations of the research, recommendations for the future and personal reflections are discussed.

Reflecting on the concept

When Ally was first conceptualized, the concept was very abstract, even though there was an embodiment prototype. While the physicality of a product might play an important role in the usability of the product, Ally's core function is conceptualized to talk, listen and interact. Creating a verbal or voice based interaction is an abstract concept. It is hard to associate with affordances, unlike with physical objects. The framework created in Chapter 3 makes the voice interaction a more tangible and quantifiable attribute. It has set the baseline and constraints to create interactions between different users and a medical device.

The results of the research conducted to understand how different users interact with Ally and with respect to the voice characteristics, were learning towards the creation of this framework. The results of the research fit like a glove for the framework, especially considering the fact that feature engineering is very hard thing to do and requires lots of analysis and data extraction. Although there are a lot of common features across the three generations, it is the probability in which they occur that make their interactions more personalised. For example, Gen X prefer a voice to be friendly, just like the other two generations, but they do not want that to be the most telling characteristic in a voice. They like the voice to be more pragmatic. Thus, these fractions in which each feature is fed into the framework is what causes the personalization. In this project, the features have been manually extracted. Usually it is done by engineers who analyse data. Creating features from the user research means the end users have directly been involved in the process of the design project, thus giving it a more personal and emotional value, as opposed to features extracted through data. These features are more humane.

This aspect of talking to a device about one's health is relatively new and it is good to have guidelines before creating such a product because the barrier of interaction might be high and people might be unwilling to confront to it. This framework paves way for that.

Feasibility of concept

While this framework cannot be tested in this project, it is quite relevant to people who are into deep learning and feasible. This framework can be used by deep learning experts to create a fully functional working AI, with guidelines that will help in reaching out to a large group of people across different generations. DeepMind is already well into creating AI research to help patients get from test to treatment as quickly as possible. They've created tools to send notifications to doctors if patents' health deteriorates. This strengthens the fact this framework can be used to create personalised voice based on the user, using WaveNet. Feasibility wise, it is a very plausible model. The complexity of elements in the framework need to be taken into consideration, as these are time consuming factors.

Reflecting on the research

In terms of research, typical to the Design for Interaction master, as the first step, we zoomed out of the current context because Ally, without it's VUI, is an inanimate object. Technically, it is an product that can talk, but physically speaking it isn't an object that has sentience. So it was interesting to see if people connected with animate objects and spoke to them. It turned out that they did, and there were a lot of characteristics that lead then to it. Be it the participants assigning or building a personality over the object, of the physicality of the object eliciting a certain attribute. The results from this acted as a baseline from which the research ideas were built upon.

The generative session conducted with all three generations was one of the first opportunities to get in touch with people of different ages and observe practically, how each generations had distinct traits. It helped in gaining an understanding as to who this product/interaction is being designed for.

Creating the ***VUI questionnaire*** was one of the most challenging parts of this project as it had to be constantly iterated to make it user friendly and at the same time get useful results. A standardised test was not considered as an option for this section as the goal was not to identify the most suitable voice, but rather, what the characteristics different voices evoked. At that point, the outcome of the project was still unknown, but it was clear that a voice was not going to be designed. The project has consistently focussed on trying to humanise Ally and lower the barrier for users, since affective speech creates a positive experience for the users.

Testing interactions between Ally and users was the closest we got to understanding how people might interact with Ally. Although the Wizard of Oz technique was used in the user test and there was no real AI present, the participants believed that there was an actual product. At no point did they realise that it was being controlled. It was interesting to see how easily some people got comfortable with Ally while it took longer for others. Largely, Baby Boomers and Millennials humanised Ally, especially because it was recorded with an actual human's voice and wasn't synthetic. This is what the framework also intends to achieve. Thus the experiment was conducted effectively, enough for participants to actually believe Ally existed as a product.

Proving the hypothesis

After literature research, it was concluded that the focus should be on different generations as each generation seemed to have different opinions and values. Based on the research, a hypothesis that each generation's interaction was influenced by their values and what they see in products. Although there are some places where the preferences overlap, like for example Baby Boomers' and Millennials' approach to Ally in terms of interaction was somewhat similar as opposed to Gen X, who had a very goal oriented approach (page 59). It is true that there are some distinct characteristics that each generation has and it influences their interaction with Ally.

Limitations of the research

More literature research would have helped in exploring more qualities that could have been incorporated in the voice and more features could have been created. Another limitation is the lack of standardised metrics to evaluate the results. Most of the research was qualitative and although there were some quantifiable data, they were not significant. A more methodological research approach that is leading in terms of the results would have created more convincing insights. With respect to the final concept, it would have been nice to have a visualization of an interaction other than the framework, like an interaction vision. A visual representation to understand the interaction might have helped the readers of this report.

Reflecting on new learnings

The conceptualization chapter was quite interesting from a designer's perspective. It was more about learning about Deep Learning. Artificial Intelligence and the likes have always generic at first sight for non-experts. For a non-expert, it is easy to assume that a machine automatically becomes intelligent, but the amount of work that goes into creating this machine with intelligence is something that we often don't realise. It was interesting to understand feature extraction. The mathematics behind neural networks is quite exciting, although it is not applicable in this project. It is intriguing to learn how a machine is able to learn things and patterns that we as humans cannot imagine. The fact the outcome of this concept is a feasible framework is something that is inspiring.

Recommendations for the future

Validating the framework:
It is important to validate the framework that has been created in this project. Although it isn't within the scope of this project, validating the framework would help prove the reduction in barrier of usage between users and a medical device.

Testing with a larger group:
Since this project was carried out by an individual, the datasets are small. Conducting the user tests, particularly the voice user tests with a larger dataset might help in refining the results, thereby creating a richer set of features. The probabilities of features created using a larger dataset will be more accurate.

Influence of context in the interaction between users and Ally:
This project focussed solely on the verbal interaction aspect. However, the usage of a product is also influenced by the surroundings and context. Hence, further research on the context of use is important.

Embodiment
Finally, the physicality of the product needs to be discussed. More research and testing on how Ally should look, and if it needs to be a product that is handheld needs to be done.

Privacy Policy:
Data privacy is extremely important for a product like Ally. Hence, there need to be stricter enforcement of rules to protect user data. More research on this is compulsory, as this is a subject concern the ethics of the community.

Personal reflection on project process

The excitement to work on this project has remained consistent through course of 7 months. The initial month involved a lot of "learning by doing" and it was fun to explore a new topic. The duration in which the user tests were created, iterated upon and tested felt extremely realistic. It felt like working on an official project, since they also usually come under strict time constraints. It was nice to experience to practice what it is like to work under strict time constraints, which is how it should be. Managing this project in the stipulated time span feels positive. The final concept, i.e the creation of a framework has also been a very positive experience. Althought the outcome of this project was different from what was personally expected at the beginning, it has been a very satisfactory one, nonetheless.

References

References

- Abelin, Å., & Allwood, J. (2000). Cross linguistic interpretation of emotional prosody. In ISCA Tutorial and Research Workshop (ITRW) on Speech and Emotion.
- Allen, J., Hunnicutt, M. S., Klatt, D. H., Armstrong, R. C., & Pisoni, D. B. (1987). From text to speech: The MITalk system. Cambridge University Press.
- Allen, J., Hunnicutt, M. S., Klatt, D. H., Armstrong, R. C., & Pisoni, D. B. (1987). From text to speech: The MITalk system. Cambridge University Press.
- Arning, K., & Ziefle, M. (2009, November). Different perspectives on technology acceptance: The role of technology type and age. In Symposium of the Austrian HCI and Usability Engineering Group (pp. 20-41). Springer, Berlin, Heidelberg.
- Baby boomers. (2018, January 17). In Wikipedia, The Free Encyclopedia. Retrieved 13:18, January 19, 2018, from https://en.wikipedia.org/w/index.php?title=Baby_boomers&oldid=820954895
- Bella, M. & Hanington, B., 2012. Universal Methods of Design, Beverly, MA: Rockport Publishers. P204
- Berger, Charles R. (2008). "Interpersonal communication". In Wolfgang Donsbach. The International Encyclopedia of Communication. New York, New York: Wiley-Blackwell. pp. 3671–3682. ISBN 978-1-4051-3199-5
- Buller, M. K., & Buller, D. B. (1987). Physicians' communication style and patient satisfaction. Journal of health and social behavior, 375-388
- C.Nass, Y.Moon, and N.Green, Are computers gender-neutral? Gender-stereotypic responses to computers with voices, Journal of Applied Social Psychology 27, no.10 (1997):864-876.
- Campbell N. (2008) Expressive/Affective Speech Synthesis. In: Benesty J., Sondhi M.M., Huang Y.A. (eds) Springer Handbook of Speech Processing. Springer Handbooks. Springer, Berlin, Heidelberg
- Cantor, N., & Mischel, W. (1979). Prototypes in Person Perception1. In Advances in experimental social psychology (Vol. 12, pp. 3-52). Academic Press.
- Coleman, A comparison of the contributions of two voice quality characteristics to perception of maleness and femaleness in the voice; R.O.Colman, Male and female voice quality and its relationship to vowel formant frequencies, Journal of speech and Hearing Research 14 (1971): 565-577; Gunzburger, Bresser, and Keurs, Voice identification of prepubertal boys and girls by slightly sighted and visually handicapped subjects (via Wired for Speech)
- Collecteren met je mobiel. (n.d.). Retrieved March 27, 2018, from https://www.hartstichting.nl/home
- Cooper-Wright, M (2012, November). Prototyping and Boundary Objects [Blog post] Retried from https://medium.com/@matt_speaks/prototyping-and-boundary-objects-b469b63d5115 [Accessed 27 Mar. 2018].
- D.Biber, Variation across speech and writing (Cambridge, Cambridge University Press, 1988) (via Wired for Speech)
- De Lemos, J. A., Drazner, M. H., Omland, T., Ayers, C. R., Khera, A., Rohatgi, A., ... & McGuire, D. K. (2010). Association of troponin T detected with a highly sensitive assay and cardiac structure and mortality risk in the general population. Jama, 304(22), 2503-2512.
- DeepMind. (2018). WaveNet: A Generative Model for Raw Audio | DeepMind. [online] Available at: https://deepmind.com/blog/wavenet-generative-model-raw-audio/ [Accessed 27 Mar. 2018].
- Dix, A. (2009). Human-computer interaction. In Encyclopedia of database systems (pp. 1327-1331). Springer US.
- Flanagan, J. L., & Golden, R. M. (1966). Phase vocoder. Bell Labs Technical Journal, 45(9), 1493-1509.
- Gaul, S., & Ziefle, M. (2009, November). Smart home technologies: Insights into generation-specific acceptance motives. In Symposium of the Austrian HCI and Usability Engineering Group (pp. 312-332). Springer, Berlin, Heidelber
- Generation X. (2018, January 15). In Wikipedia, The Free Encyclopedia. Retrieved 13:17, January 19, 2018, from https://en.wikipedia.org/w/index.php?
- Goetz, J., Kiesler, S., & Powers, A. (2003, October). Matching robot appearance and behavior to tasks to improve human-robot cooperation. In Robot and Human Interactive Communication, 2003. Proceedings. ROMAN 2003. The 12th IEEE International Workshop on (pp. 55-60). Ieee
- Hanington, B., & Martin, B. (2012). Universal methods of design: 100 ways to research complex problems, develop innovative ideas, and design effective solutions. Rockport Publishers.
- Korsch, B. M., & Negrete, V. F. (1972). Doctor-patient communication. Scientific American, 227(2), 66-75.
- Kurzweil, R. (1990). The Age of Intelligent Machines, MIT Press
- L.W Harmon, J.-I.Hansen, F.Borgen, and A.Hammer, Strong Interest Inventory: Applications and technical guide (Palo Alto, Calif: Consulting Psychologist Press, 1985).
- Lewis, J. R. (2016). Standardized Questionnaires for Voice Interaction Design. Voice Interaction Design, 1(1).
- Markowitz, J. A. (2016). Speech and Language for Acceptance of Social Robots: An Overview
- Mavridis, N. (2015). A review of verbal and non-verbal human–robot interactive communication. Robotics and Autonomous Systems, 63, 22-35.
- McLeod, S. (2018). Social Identity Theory | Simply Psychology. [online] Simplypsychology.org. Available at: https://www.simplypsychology.org/social-identity-theory.html [Accessed 27 Mar. 2018].
- Millennials. (2018, January 15). In Wikipedia, The Free Encyclopedia. Retrieved 13:18, January 19, 2018, from https://en.wikipedia.org/w/index.php?title=Millennials&oldid=820517538
- Morris, M. G., & Venkatesh, V. (2000). Age differences in technology adoption decisions: Implications for a changing work force. Personnel psychology, 53(2), 375-403.
- Ng, Andrew (2017), Neural Networks and Deep Learning - Coursera, Retrieved from https://www.coursera.org/learn/neural-networks-deep-learning/lecture/eAE2G/what-is-a-neural-network, Accessed on dd/mm/yyyy
- Norton, R. W. (1978). Foundation of a communicator style construct. Human Communication Research, 4(2), 99-112.
- Ommen, O., Janssen, C., Neugebauer, E., Bouillon, B., Rehm, K., Rangger, C., ... & Pfaff, H. (2008). Trust, social support and patient type—Associations between patients perceived trust, supportive communication and patients preferences in regard to paternalism, clarification and participation of severely injured patients. Patient Education and Counseling, 73(2), 196-204.
- Pilcher, Jane (September 1994). "Mannheim's Sociology of Generations: An undervalued legacy" (PDF). British Journal of Sociology. 45 (3): 481–495. doi:10.2307/591659. JSTOR 591659. Retrieved 10 October 2012.
- Prosody. (n.d.) In Merriam-Webster’s collegiate dictionary. Retrieved from http://www.merriam-webster.com/dictionary/prosody
- Rana el Kaliouby (Nov–Dec 2017). "We Need Computers with Empathy". Technology Review. 120 (6). p. 8.
- Reeves and Nass, The media equation;

D.Voelker, The effect of image size and voice volume on the evaluation of represented faces, unpublished doctoral dissertation, Stanford University, Stanford, Calif., 1994.

- Rich, E. and Knight, K. (1991). Artificial Intelligence (second edition). McGraw-Hill.
- Russell, S. J., & Norvig, P. (2016). Artificial intelligence: a modern approach. Malaysia; Pearson Education Limited,.
- Sanders, L., & Stappers, P. J. (2012). Convivial design toolbox: Generative research for the front end of design. BIS.
- Stewart, O. T., & Blanchard, H. E. (2008). Linguistics and psycholinguistics in IVR design. In Human factors and voice interactive systems (pp. 81-115). Springer, Boston, MA.
- Suter, E., Arndt, J., Arthur, N., Parboosingh, J., Taylor, E., & Deutschlander, S. (2009). Role understanding and effective communication as core competencies for collaborative practice. Journal of interprofessional care, 23(1), 41-51.
- Taylor, P. & Keeter, S. (Eds.) (24 February 2010). "The Millennials. Confident, Connected. Open to Change". p. 5.
- Wei, Y., & Zhang, Q. (2012). Common Waveform Analysis: a new and practical generalization of Fourier analysis (Vol. 9). Springer Science & Business Media.
- Wikipedia contributors. (2017, October 7). Feature (machine learning). In Wikipedia, The Free Encyclopedia. Retrieved 16:20, March 27, 2018, from https://en.wikipedia.org/w/index.php?title=Feature_(machine_learning)&oldid=804252725
- Wikipedia contributors. (2018, March 13). Intonation (linguistics). In Wikipedia, The Free Encyclopedia. Retrieved 16:39, March 27, 2018, from https://en.wikipedia.org/w/index.php?title=Intonation_(linguistics)&oldid=830202291
- Wikipedia contributors. (2018, March 27). Phoneme. In Wikipedia, The Free Encyclopedia. Retrieved 16:40, March 27, 2018, from https://en.wikipedia.org/w/index.php?title=Phoneme&oldid=832646038
- Wikipedia contributors. (2018, March 7). Duration (music). In Wikipedia, The Free Encyclopedia. Retrieved 16:39, March 27, 2018, from https://en.wikipedia.org/w/index.php?title=Duration_(music)&oldid=829225258
- Wmfc.org. (2018). [online] Available at: http://www.wmfc.org/uploads/GenerationalDifferencesChart.pdf [Accessed 27 Mar. 2018].
- Zimmerman, J., Forlizzi, J., & Evenson, S. (2007). Research through design as a method for interaction design research in HCI. In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 493-502). ACM.

Appendix

Owing to the sizes of the report and Appendix, the Appendix is not attached with this report. It has been printed out separately. A digital version is available in the TU Delft Repository.