The User Value of Speech Recognition at Home

Albert Johansson and Maria Blomberg

DIVISION OF INNOVATION ENGINEERING | DEPARTMENT OF DESIGN SCIENCES
FACULTY OF ENGINEERING LTH | LUND UNIVERSITY
2018

MASTER THESIS

The User Value of Speech Recognition at Home

An Explorative Study of the User Value Created by Speech Recognition for the IKEA Customer

Albert Johansson and Maria Blomberg

LUND UNIVERSITY

The User Value of Speech Recognition at Home

An Explorative Study of the User Value Created by Speech Recognition for the IKEA Customer

Published by
Department of Design Sciences
Faculty of Engineering LTH, Lund University
P.O. Box 118, SE-221 00 Lund, Sweden

Subject: Innovation Engineering (INTM01)
Division: Division of Innovation Engineering
Supervisor: Lars Bengtsson
Co-supervisor: Emil Åkesson
Examiner: Malin Olander Roese

Abstract

Humans have interacted using speech for thousands of years. Since the middle of last century, humans have been able to interact with computers using speech. Recent investments in speech recognition have resulted in the technology making its way into the homes of mainstream consumers. However, little is known about the user value of this target group using speech recognition in the home environment. The user value is relevant to IKEA for determining how, or if, the technology can deliver to their vision of *creating a better everyday life for the many people*.

The purpose of this thesis is to explore the user value for the broader audience using speech recognition in the home environment. Further, the aim is to understand if, and how, speech recognition is relevant to IKEA. To reach the purpose, an exploratory, mixed method design, combining qualitative and quantitative strategies has been used. The primary approach of this study is a market research approach for identifying user value called means-end chain. Derived values are triangulated with feedback from experts and compared to existing literature. Aspects of implementation, such as product design and business have been excluded from the thesis.

Four user values of speech recognition in the home environment are concluded; *facilitate daily life*, *everyday efficiency*, *comfort* and *increased calmness*. All four values refer to tasks complementing, not replacing, already existing interfaces. Moreover, a scepticism towards the new technology of speech recognition is identified.

Speech recognition is also concluded to, entirely or partly, fulfil the IKEA product development criteria for smart products, namely; *convenience*, *easy to understand*, *clear use case* and *solving a user need*. Therefore, speech recognition can contribute to the IKEA vision of *creating a better life for the many people*.

Keywords: Speech Recognition, Means-End Chain, User Value, Smart Home, Voice Control

Sammanfattning

Människan har interagerat genom att använda sin röst i årtusenden. I mitten av förra seklet började datorer förstå människans röst. På senare år har stora investeringar för röststyrning i hemmet gjort att teknologin börjat nå en bred konsumentmarknad. Användarvärdet för röststyrning i hemmiljö för denna målgrupp är varken välkänt eller väldokumenterat. Användarvärdet är intressant för IKEA för att avgöra hur, eller om, teknologin kan bidra till deras vision om att *en bättre vardag för de många människorna*.

Syftet med denna uppsats är att utforska användarvärdet av röststyrning i IKEA-kundens hemmiljö. Utifrån detta avgörs om röststyrning är relevant för IKEA. Syftet uppnås med en explorativ ansats som utgår från en flerstegsmetod vilken kombinerar kvalitativa och kvantitativa data. Uppsatsens primära ansats utgörs av en marknadsanalysmetod för att förstå användarvärde som kallas means-end chain. De framtagna värdena trianguleras med feedback från experter och jämförs med befintlig litteratur. Aspekter som har med implementering att göra, såsom marknadsstrategi, affärsstrategi och funktionalitet ingår inte i denna uppsats.

Fyra användarvärden identifieras för röststyrning i hemmiljö; *underlätta vardagen*, *vardagseffektivitet*, *bekvämlighet* och *ökad avkoppling*. Samtliga värden syftar på aktiviteter som kompletterar, inte ersätter, redan existerande gränssnitt. Det framkommer även en viss skepsis mot den nya teknologi som röststyrning representerar.

Utöver detta dras slutsatsen att röststyrning, helt eller delvis, uppfyller IKEAs kriterier för utveckling av smarta produkter, nämligen; *bekvämlighet*, *lätt att förstå*, *tydligt användningsfall*, och *lösa ett användarproblem*. Därmed anses röststyrning kunna bidra till syftet *att skapa en bättre vardag för de många människorna*.

Nyckelord: Röststyrning, means-end chain, användarvärde, det smarta hemmet

Acknowledgments

This master's thesis has been performed as the closing part of the Master of Science in Engineering, Industrial Engineering and Management at the Faculty of Engineering at Lund University. During the work, we have been guided by several people.

First, we would like to express our sincerest gratitude to Rebecca Töreman at IKEA Home Smart for the open-hearted and sincere support along the way. We could truly not have asked for a better client supervisor. Also, a big thanks to Björn Vedmo at IKEA Lighting, for the nicest welcome at IKEA Malmö. We would also like to thank the rest of IKEA Home Smart for giving the opportunity of doing this thesis, despite the short notice.

Further, we want to thank the following people for contributing with their expertise to our research; Hiroshi Ishiguro at Osaka University, David House at KTH, Nils Lenke at Nuance Communications, Ekaterina Kruchinina at Nuance Communications and Dimosthenis Kontogiorgos at KTH.

Lastly, special thanks are given to both of our supervisors at Lund University, Lars Bengtsson, and Emil Åkesson. By providing us with constructive feedback, new ideas and happy moments we have learned a lot while having an inspiring and allowing work process.

Malmö, January 2018

Albert Johansson

Maria Blomberg

Table of contents

List of acronyms and abbreviations

AI	artificial intelligence
HVM	hierarchical value map (used in MEC analysis)
IT	information technology
MEC	means-end chain
RQ	research question

1 Introduction

This chapter provides a general understanding and purpose of the thesis. First, in the background, the setting where the thesis takes place is described. Further, the problem statement, purpose and research questions are presented. Followed by delimitations and the disposition of the report.

1.1 Background

Speech is fundamental in human communication. Interaction with voice has been practiced and developed for thousands of years. Since the middle of last century, humans have been able to interact with computers using speech. Artificial intelligence (AI) controlled by speech used to be a futuristic prediction introduced by the sentient computer HAL 9000 in Stanley Kubrick's movie *2001: A Space Odyssey* from 1968. Now, 50 years later, people use speech to control smart phones, light bulbs and appliances.

Using speech recognition at home has become increasingly popular. In 2016, Echo, a speech recognition system developed by the American company Amazon, was the most selling speaker in the United States (voicebot, 2017). According to Amazon, the idea of the device is to create *frictionless purchasing*, using speech recognition to stimulate their main business goals (Rhordi, 2016). Amazon is being followed by Google, Apple and other technological giants developing their own speech recognition systems and devices.

The emerging technology is a result of intense research. However, the literature found on speech recognition tend to focus on technological aspects. Little has been written about the potential user value. Previous research is often focused on specific target groups with certain needs, such as elderly and disabled (Horstmann, 2010; Portet et al., 2011; Cordasco et al., 2014; Arriany and Musbah, 2016).

Meanwhile, life at home has changed (Inter IKEA Systems, 2015; Inter IKEA Systems, 2017). As an example; the definition of a home has broadened and the need for privacy has increased (Inter IKEA Systems, 2015). Additionally, the home environment has been affected by technological change. Smart speakers, televisions and appliances are starting to find their way to where people live. Together, these devices are part of the smart home, a home that connects its

technology to respond to the needs of its occupants (Aldrich, 2003).

As a major provider of home environments, IKEA, the largest furniture retailer in the world (Statista, Inc., 2017), have released smart lighting in 2017 (Inter IKEA Systems B.V., 2017b). This after admittedly being slow on adopting to new business opportunities opened by technological development (Hammersley, 2016; Watts, 2015). However, the initiative has been praised in tech media for its ease of use, simplicity, and low price (Schoutsen, 2017).

The smart lighting system was released by IKEA Home Smart, the department developing smart products. They recognise speech recognition as an interesting opportunity for future products. However, the main challenge is to understand the user value[1] of speech recognition and its relevance to IKEA. For a smart product to be relevant, it needs to deliver to their vision of *creating a better everyday life for the many people*. (Töreman, 2017e)

1.2 Problem statement

Even though the technology of speech recognition has been around for a long time, its position in the home environment for the broad consumer market is a relatively recent phenomenon. Big technological companies, like Google, Apple and Amazon invest aggressively in the technology (Block, 2017). What they consider being the user value of speech recognition is unknown.

IKEA, the partner in this study, is showing interest in exploring what values speech recognition can create for their consumers. Previous research has explored the user value of speech recognition by focusing on limited targets groups, such as elderly and disabled. The technology has proved satisfying (Horstmann, 2010), making everyday life easier (Portet et al., 2011), providing convenience and increased efficiency (Arriany and Musbah, 2016; Cordasco et al., 2014).

However, the IKEA customers represent a broader audience and the user values created by speech recognition for that target group has not yet been studied. Thus, it is of interest to explore whether speech recognition can deliver user value and contribute to the IKEA vision of *creating a better everyday life for the many people*.

[1] User value is defined in 3.3.1 as *desirable end states of existence from the user perspective*.

1.3 Purpose

The purpose of this thesis is to explore the user value of speech recognition in the home environment for the broader audience. This is done in order to better understand the relevance of speech recognition in the home environment for the consumer market in general and for IKEA in particular. This will provide a foundation on the user perspective for speech recognition at IKEA.

1.4 Research questions

To reach the purpose, two research questions are introduced. Answering these two questions provide for the purpose.

(RQ1) What user value could be generated from speech recognition in the home environment for the broader audience?

(RQ2) Judging from a user perspective, how is speech recognition in the home environment relevant to IKEA?

1.5 Delimitations

This thesis does not cover the implementation aspect of speech recognition. Implementation entails areas such as product design, market and business strategy and functionality. Instead, recommendations are presented with a more overarching approach focusing on user value. Furthermore, this project is part of a master thesis determined by several conditions, one of them being a time limitation of 20 weeks.

1.6 Disposition of report

This section presents the disposition of the thesis, for structural purpose and ease of understanding for the reader.

Chapter 1 Introduction

This chapter gives the reader an introduction by explaining the scene where the thesis takes place. It also consists of the problem statement, purpose and research questions. As well as delimitations and this disposition of the report.

Chapter 2 Methodology

In this chapter, the reader is given an explanation of the methodology. It introduces the research strategy and design. The chapter also presents the methodology of means-end chain and the other data collection and analysis methods. The working process is described and the chapter ends with a discussion of the research quality.

Chapter 3 Theory

In chapter 3 the theoretical framework is introduced. This is done by describing and defining speech recognition and user value. The chapter also presents the historical context and explains the characteristics of the user value of speech recognition.

Chapter 4 IKEA

In this chapter, the focus is on IKEA and their IKEA Home Smart department. The product development criteria IKEA are described. The chapter also explains IKEA's mind-set towards speech recognition, as well as the technological aspects of the life at home for the many people.

Chapter 5 Means-end chain analysis

This chapter presents the analysis of the primary research approach, means-end chain. First, the results of the survey are presented. Followed by the coding procedure to structure that data. The chapter also introduces the graphical presentation of the results from the means-end chain approach, i.e. presents the derived user values of speech recognition at home and its linkages.

Chapter 6 Analysis

This chapter provides the reader with the analytical dimension. The means-end chain result is triangulated by expert reviews and the relevance for IKEA is analysed.

Chapter 7 Discussion

In this chapter, the results and analysis are discussed. The definition of speech recognition is examined and the derived user values are discussed. This is followed by a discussion from the IKEA perspective and by recommendations for IKEA. The chapter also examines the validity of the means-end chain approach. Lastly, the academic contributions and proposals for further research are

presented.

Chapter 8 Conclusions

This chapter summarizes the main findings in the report and provides the reader concluding remarks.

2 Methodology

This chapter presents the methodology being used in this research. It explains an overview of the research strategy followed by research design. Furthermore, the data collection methods are motivated and followed by an explanation of the data analysis. Moreover, the work process of the different data collection methods is described. Lastly, the quality of the chosen methodology is discussed.

2.1 Research strategy

This master thesis conducts a *mixed methods* research, i.e. a research that combines the traditional strategies quantitative and qualitative research (Bryman and Bell, 2015; Teddlie and Tashakkori, 2009). A mixed method strategy rejects the either/or choice between qualitative or quantitative data and uses information with both words and numbers (Teddlie and Tashakkori, 2009). It copes with the weaknesses and empowers the strengths of the traditional strategies (Bryman and Bell, 2015). A mixed method was chosen from a bottom-up approach. First, a suitable approach to understand the user value of speech recognition was identified. This approach was a mixed method in itself, involving both qualitative and quantitative data. The method was seen as the primary approach and resulted in user values. To further understand those user values, qualitative interviews seemed appropriate. Hence, the decision of going for a mixed method strategy. However, using a mixed method research strategy lays great responsibility in the design of the research, therefore effort was made when designing the research in next section.

Furthermore, this research takes an *exploratory* approach. Exploratory research seeks to understand a phenomenon with little previous research. This enables basic knowledge and settings to develop in further research (Neuman, 2003; Lekvall and Wahlbin, 2011). Also, an exploratory approach typically asks questions of what and rarely conducts definitive answers (Neuman, 2003). For this thesis, no definitive answers are needed. Instead, the thesis aspires to understand more about the user value of speech recognition as that field is relatively unexplored in previous research. Therefore, an exploratory approach was considered appropriate for this thesis. This demands of the researcher to be creative, flexible and open-

minded to all sources of information (Neuman, 2003).

Lastly, the scope and perspective have been decided upon in collaboration with IKEA and their needs (Blomkvist and Hallin, 2015). Combining the exploratory approach and the cooperative nature, the project has changed purpose and problem statement throughout the process.

2.2 Research design

The research design creates a framework for data collection and analysis (Bryman and Bell, 2015). Using a mixed methods approach means that the embedded design has either the qualitative or quantitative research as its primary approach but also uses the other approach to enhance the initial approach (Bryman and Bell, 2015). In Figure 2.1 a graphical scheme of the research design in the thesis is presented. This section explains the intended steps, starting with the primary approach.

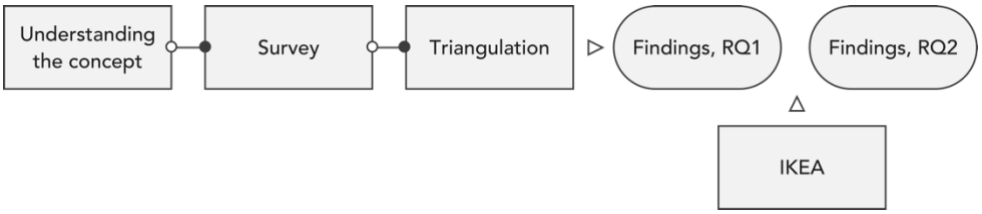

Figure 2.1 A graphical presentation of the research design.

The primary approach of this research is based on a specific cross-sectional, or survey, approach that combines qualitative research, with quantitative. In cross-sectional research, data is typically collected by a questionnaire on multiple cases at one point in time (Bryman and Bell, 2015; Lekvall and Wahlbin, 2011). Also in survey research, the collected data is either quantifiable or quantitative, which is examined to draw conclusions from the comparison of individuals in the collected data (Bryman and Bell, 2015; Lekvall and Wahlbin, 2011).

For this research, the aim is to collect data from multiple IKEA customers, quantify the data and find patterns in the responses. Hence, the cross-sectional approach was considered suitable. Further, cross-sectional design is less costly, making it appropriate for this thesis having limited time (Neuman, 2003).

The primary approach was enhanced by complementing approaches. Here, it consists of one previous and one subsequent step.

First, the previous step was *understanding the concept*. This step aims to perform a literature review, to explore the current state of speech recognition in research, thus to support the development of the primary approach.

Second, the subsequent step is to cross-check the results from the primary approach. This procedure is a strength in mixed methods research since it provides the opportunity to collect results using one method and cross-check those results with another method (Bryman and Bell, 2015). This is the process of triangulation, it aims to generate perspective, control and validity to the results (Bryman and Bell, 2015). Critics mean that triangulation can lead to separate findings from different approaches (Bryman and Bell, 2015). In this research, this is dealt with by considering the results from the survey approach as primary, meaning the findings are primarily based on those results and critically reviewed later on.

The triangulation makes it possible to answer RQ1 and present the findings of the first part of this thesis. These findings are compared with and related to the results from the IKEA perspective. This is resulting in the findings regarding RQ2, which is the last step of the research design.

2.3 Data collection

Several different data collection methods are applied to support the research questions, strategy and design. These are: literature review, the means-end chain (MEC) approach, qualitative interviews and organisational document gathering. In this section, those methods are described and motivated.

2.3.1 Literature review

Two different literature reviews are presented in this thesis. The first is used to understand the concept of speech recognition. The second is used to compare and relate the result from MEC with previous research on the user value of speech recognition.

For both of the literature reviews, the traditional, or narrative approach is used. This means that the researcher critically describes and assess what is already known about a certain research area by using previous research (Jesson et al., 2012). The aim is to get a broad perspective of that area, develop ideas and identify research gaps (Bryman and Bell, 2015). Further, the traditional literature review is less time consuming than a structural review and is both exploratory and creative (Bryman and Bell, 2015; Jesson et al., 2012). The characteristics of the traditional approach fit well with the research question, strategy, design and limitations described above.

Critics of the traditional, narrative, review argue that the absence of systematic protocol means that the reader cannot judge the completeness of the review (Jesson et al., 2012). Therefore, the procedures of the literature review are carefully described to enable better judgement of the review. Furthermore, critics

claim that the traditional review can be affected by unfair selection of sources (Jesson et al., 2012). This is dealt with by aiming at an unbiased selection of sources and to critically comparing and contrasting those sources.

2.3.2 Means-end chain approach

The primary method for data collection in this research is the means-end chain approach. MEC is a specific type of cross-sectional research that combines qualitative and quantitative research. It aims to understand the important values that consumers assign to a specific product or service. The approach was developed as a marketing research method in the 1980's and has been widely adopted ever since (Gutman, 1982; Walker and Olson, 1991; B. Klenosky, 2002; de Souza Leão and Mello Benício de, 2007; Jung and Kang, 2010; Jung, 2014; Deng and Christodoulidou, 2015).

More specifically, MEC provides a conceptual tool that aims to explain the linkages between attributes in products (*means*), the consequences of the attributes, and the personal values (*ends*) the consequences lead to (Gutman, 1982). In other words, the user perception is hierarchically structured in three levels, *attributes*, *consequences* and *values* (Figure 2.2). Attributes describe the physical properties, or functions, of a product. Consequences are the benefits achieved by the attributes. Lastly, values are the highly abstract motivation explaining the user behaviour (Jung and Kang, 2010).

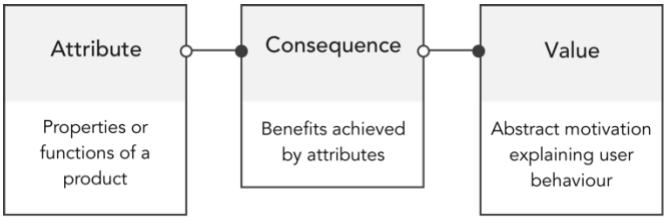

Figure 2.2 The abstractions levels in the MEC approach.

Furthermore, the typical data gathering method for MEC is the laddering procedure, performed by asking three questions. The attribute question (1) (*What attributes makes that product attractive for you?*), consequence question (2) (*Why is the attribute desirable to you?*) and the value question (3) (*Why is that important to you?*) (Jung and Kang, 2010). These questions are often modified to fit the purpose of the study (Jung and Kang, 2010; Jung, 2014; Deng and Christodoulidou, 2015). The adaptation and modification of the MEC approach are described when explaining the work process further on.

During the last decade, MEC has been successfully used in the context of information technology (IT) (Jung and Kang, 2010; Jung, 2014; Deng and Christodoulidou, 2015). MEC also gives an improved understanding of relevant consumer needs. It is well suited for an exploratory study, such as this one (Costa

et al., 2004). The fact that this thesis aims to explore user value of speech recognition in the home environment, a part of IT and the advantages of the method described above provide arguments for MEC to be an appropriate approach for this study.

Critics of the MEC approach claim that the method does not have a clear distinction between attributes, consequences and values. Critics also state that parts of the method are complex and subjective (Costa et al., 2004). Therefore, the working process of the MEC approach is carefully described to enable better judgement of the results.

2.3.3 Qualitative interviews

This study used qualitative interviews in two ways. First, IKEA staff were interviewed to provide insight needed to answer RQ2. Second, experts in speech recognition were interviewed about the results from the MEC analysis.

Qualitative interviewing is a broad term that describes a flexible research method where the researcher can ask several different types of questions (Lekvall and Wahlbin, 2011; Bryman and Bell, 2015). This provides for a deep and extensive understanding of the studied phenomenon and the possibility to discover new dimensions (Lekvall and Wahlbin, 2011; Bryman and Bell, 2015; Blomkvist and Hallin, 2015). For this thesis, both the concept of speech recognition and its relativeness for IKEA requires a rich contextual understanding. Hence the approach of qualitative interviews is found suitable.

Advantages of qualitative interviews are that it emphasises on the interviewees own perspectives, rather than on the researchers. It gives an insight of what the interviewee recognises as relevant since he or she can elaborate freely, which allows rich and detailed answers. Yet, an interview does not have value in itself, it only has a meaning when compared and related to other collected data and interviews. Hence, multiple interviews are conducted and compared to other data. (Bryman and Bell, 2015)

However, qualitative interviews are time-consuming, both when conducting the interviews and the subsequent work when managing the data (Bryman and Bell, 2015; Lekvall and Wahlbin, 2011). Therefore, the number of interviews are limited to fit the timeframe.

Furthermore, two different techniques of qualitative interviewing are used in this thesis, unstructured and semi-structured interviews. Unstructured interviews can be compared to an ordinary conversation, where the respondents answer freely to just a few predetermined questions, which generates a genuine understanding of a phenomenon (Bryman and Bell, 2015; Lekvall and Wahlbin, 2011). This type of interview is suitable when gaining general understanding of the IKEA perspective. Consequently, used when interviewing IKEA staff.

Semi-structured interviews are loosely structured, with an interview guide as a basis. This is appropriate when the research has a fairly clear focus (Bryman and Bell, 2015; Blomkvist and Hallin, 2015). Therefore, semi-structured interviews are suitable for this research when triangulating the MEC results. Hence used when interviewing experts.

2.3.4 Document gathering

Other than qualitative interviews, insights about IKEA were acquired from organizational documents. Organizational documents provide valuable background information about an organization and its different aspects (Bryman and Bell, 2015). Typically, these documents are authentic and meaningful, however, they can have issues with credibility and representativeness. Therefore, in this thesis, the documents are critically reviewed and only used in relation to other sources when drawing conclusions.

2.4 Data analysis

Data analysis in mixed method research combines quantitative and qualitative data analysis strategies (Teddlie and Tashakkori, 2009). In this research, it means that the quantifiable data collected through the MEC approach is analysed through the MEC analysis. Thereafter the data is triangulated with a qualitative analysis of interviews, document gathering and literature review.

2.4.1 MEC analysis

The MEC analysis corresponds to three steps, which are briefly presented in Figure 2.3. First, classifying the collected responses into categories. Second, the linkages between the categories are quantified in an implication matrix (Reynolds and Gutman, 1988). Last, the linkages in the implication matrix are used to generate a hierarchical value map (HVM). This is where the dominant connections are graphically presented in a tree diagram. In the HVM the connections are referred to as chains and consist typically of one or multiple attributes, consequences and values (Reynolds and Gutman, 1988; Jung and Kang, 2010).

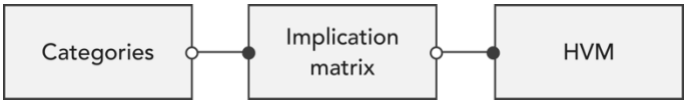

Figure 2.3 Graphical presentation of the MEC analysis.

In all steps of the MEC the level of abstraction from attribute to consequence to value are dealt with. Where attributes represent the lowest abstraction,

consequences in the middle and values the highest abstraction (Reynolds and Gutman, 1988).

2.4.2 Triangulation

The results from the MEC analysis are analysed by examining the collected answers from the expert interviews. This makes the triangulation of the research and is part of the data analysis.

2.5 Work process

This section explains how the different research methods were conducted.

2.5.1 Process

The connection between the elements conducted in this thesis is presented in Figure 2.4. First, the literature review of speech recognition gives an understanding of the foundation. Then, the MEC approach and analysis gives the primary results. Third, the results are triangulated with expert interviews and compared to the insights from a literature study of user value. The first three steps conclude the findings to RQ1: *What user value could be generated from speech recognition in the home environment for the broader audience?* These findings are then compared and related to the IKEA perspective, which in turn gives the findings to answer RQ2: *Judging from a user perspective, how is speech recognition in the home environment relevant to IKEA?*

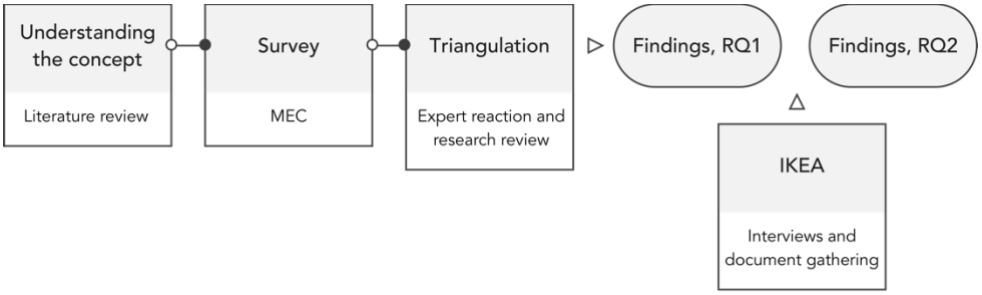

Figure 2.4 Graphical presentation of data collection methods with respect to the research design.

2.5.2 Literature review

Two narrative literature reviews were performed. One to understand what term to

use and how speech recognition can be defined. The other to learn about previous research on the user value of speech recognition.

The material used in both reviews was found by searching for keywords in different databases. The terms searched for were; speech recognition, voice command, voice control, voice recognition and automatic speech recognition. Also, the same terms were searched for while adding *user value* as a keyword (e.g. speech recognition user value, voice command user value etc.). The search was performed in three complementing databases. First, LUBsearch, Lund University shared database consisting of articles and other literature. Second, Google Scholar, a search engine of scholarly literature. Third, the Swedish national library service Libris.se. Subsequently, more relevant literature was found through reference lists in previously found articles, 53 articles and 6 books were found.

For the literature review to find and define the term speech recognition, 32 articles and 4 books were used. The criterion for an article to be part of the literature review was to describe human speech being understood and, or, registered by an artificial system, e.g. a computer.

In the literature review about the user value of speech recognition, articles in the primary search and articles regarding the MEC were used. In the end, 10 articles and 1 book were used in in this review.

2.5.3 MEC approach

This section describes the entire process of the MEC approach, including the choice of respondents, formulating questions and type of data collection.

2.5.3.1 Laddering technique with pen and paper survey

The original method of the laddering technique, suggested by Reynolds and Gutman (1988), recommends 60 to 75 minutes semi-structured in-depth interviews. While the interviewing technique provides a good understanding of the respondent, it is time-consuming. Walker and Olson (1991) conducted a revised version, where a questionnaire replaces the in-depth interview. They have been followed by others and is often referred to as the pen and paper version (Jung and Kang, 2010; Jung, 2014). With the pen and paper version, respondents are feeling more pressure to answer, because the respondents themselves choose when to finish (Botschen and Hemetsberger, 1998). On the other hand, this pressure might cause a made-up answer, in order to satisfy the interviewer. (Walker and Olson, 1991)

The pen and paper version was found suitable for this thesis, primarily because of time limitations. Moreover, the survey was physical since writing on a piece of paper was seen as less of a barrier than typing on a computer (Töreman, 2017a).

For the laddering analysis 30 to 50 usable responses are sufficient (Reynolds and

Gutman, 1988; de Souza Leão and Mello Benício de, 2007). However, according to Reynolds and Gutman, one quarter of the answers is normally not informative enough to be used (1988). For this study, the number of usable responses were decided at a minimum of 30. Hence, 40 or more responses had to be collected, including the insufficient quarter.

2.5.3.2 The process of creating the survey

In order to gain relevant responses, the survey was developed in an iterative process inspired by the multiple step process for survey research proposed by Bryman and Bell (2015). A total of 7 iterations were made including feedback from 13 different test persons. Test persons were picked with consideration of the final target audience, e.g. IKEA customers. The iterations were continuously documented.

2.5.3.3 Respondents and selection

The respondents sought for in this study were IKEA customers. To get hold of those respondents, a face-to-face selection at an IKEA store was performed, i.e. customers passing were asked to take part in the survey. For the sake of geographical convenience, the store in Malmö, Sweden was chosen. This sampling approach is called convenience sampling and is commonly used in exploratory research (Lekvall and Wahlbin, 2011).

The probability of not finding respondents perfectly representing the entire population, i.e. all current and potential IKEA customers, is high because of four reasons. First, the sample is limited to a specific geographic location. Second, some people entering the store are more likely to take part in the survey than others. Third, the sample is likely to differ depending on the time of day. Finally, there is a probability of only approaching people looking as if they are open to being interviewed, which of course is a subjective observation (Lekvall and Wahlbin, 2011). This was dealt with by the authors trying to approach people at random, no matter of their perceived willingness to answer. However, if a person did not wish to participate, this was immediately respected according to ethics. In order to vary the demography of the population, data were gathered at two different timespans, morning and afternoon. No compensation was given to any of the respondents for taking part in the survey.

2.5.3.4 Questions

The questions asked in the survey were designed to answer RQ1. As described earlier the original laddering questions were modified to fit the purpose of this study. The modifications were tested and reviewed in the iterative process explained above. The term *attribute* was replaced with *function* to reflect the more practical approach of speech recognition.

The laddering questions were organized in the typical hierarchical manner. With the attribute question, *A: What would you (or someone else in your home) use*

voice control[2] *for in your home?* as the first question. Followed by the consequence question *C: Why do you (or someone else in your home) want to use voice control in the manner you described in the previous question? Please explain.* Last the value question *V: Regarding the reasons you provided in the question above, why are those reasons important to you (or someone else in your home)? Please explain.* The survey can be found in Appendix A.1.

Table 2.1 Laddering questions in the survey on the three different levels of abstraction.

	Statement level	***Question***
1	A: Attribute	What would you (or someone else in your home) use voice control for in your home?
2	C: Consequence	Why do you (or someone else in your home) want to use voice control in the manner you described in the previous question? Please explain
3	V: Value	Regarding the reasons, you provided in the question above, why are those reasons important to you (or someone else in your home)? Please explain.

2.5.3.5 Survey

The MEC part of the survey, explained above, was complemented with questions asking about demographic information such as age, gender and living situation. People were also asked how many smart products they had at home and their habit of using voice control. Thus, the respondents could be analysed and complemented with concrete data about the current level of technological acceptance.

To provide the respondent with a credible rationale for the study they are participating in, the survey had an introductory statement, as suggested by Bryman and Bell (2015). The introductory statement briefly explains the purpose of the study and why the research is important. It also informs the respondent that the survey is voluntary and that all submitted answers are entirely anonymous. Lastly, for the sake of transparency contact details of the authors were listed.

In order to help the respondents with ideas, an inspirational poster (Appendix A.2) was provided. The inspirational poster shows examples of how speech recognition could be used in the home environment based on the categories of IKEA's division of rooms on their website, e.g. bedroom, living room, kitchen, dining room, children's room, bathroom, home office, hallway and laundry (Inter IKEA

[2] At the time of submitting the survey, the literature review had not been completed. Therefore, *voice control* was used instead of *speech recognition*.

Systems B.V., 2017c).

The authors had previously agreed on what to say (Appendix A.3) when introducing the survey to the respondents. If the respondent did not know what to answer, there were also prompts that could be used. In order for the respondent to answer as honestly as possible, respondents were told there was no right or wrong answer, as recommended by Reynolds and Gutman (1988). The survey was provided in both Swedish and English.

2.5.3.6 Coding

Coding is the first analytical step in the MEC approach, followed by implication matrix and HVM. It is about turning qualitative data into quantitative through coding (Reynolds and Gutman, 1988). This is carried out in multiple steps described below. But before the coding could start the responses were transferred from paper to computer and a manual audit decided which responses to eradicate. Responses were cleaned out due to not being informative enough, e.g. they did not answer any question or gave answers that did not make any sense. The coding procedure is summarized in Table 2.2 and elaborated further down in this section.

Table 2.2 Coding procedure in the MEC approach.

	Participant	***Action***	***Outcome***
First step	First coder	Open coding. Sub-categories were derived from the answers.	Coded data Set of sub-categories
Second step	Both coders	Sub-categories were discussed, clarified and merged.	Set of reconsidered sub-categories
Third step	Both coders	Independently recoding using the elaborated set of sub-categories. Verifying the categorization agreement. Reconciling disagreements through discussion between the two coders	 71% agreements Set of 29 sub-categories
Fourth step	Both coders	Merging sub-categories into categories Translation from Swedish into English	Final set of 22 categories in English

The first step involves the first coder following an open coding procedure. This is a procedure where categories are emerged from the studied data rather than predetermined. It consists of breaking down, examining, comparing, contextualizing and categorizing the data, which in turn leads to grouped categories (Bryman and Bell, 2015). As proposed by Jung (2014), the coder tried to use the respondents words and phrases, to make it easier and accurate when continuing the process of coding.

As a second step, the two coders discussed the elaborated set of sub-categories, clarified some of them by exemplifying according to their corresponding answers and merging some of the sub-categories.

Based on the new elaborated set of sub-categories, the coders classified the data again, independently. The categorization agreement was verified by calculating the ratio of agreements, which was 71 percent. To solve inter-rater disagreements, those were discussed by the two coders and they reconciled. (Deng and Christodoulidou, 2015; Jung and Kang, 2010)

The final step of the coding procedure was to merge sub-categories into categories for more distinctive terms and smoother handling further on in the analysis, as performed by Jung and Kang (2010). Lastly, the categories and parent categories were carefully translated into English, trying to retain the original meaning of the words and phrases from the respondents' answers in Swedish.

To illustrate the results of coding the answers, an example is presented in Table 2.3.

Table 2.3 Example of answers and how the coding was performed.

Question	***Q:1***	***Q:2***	***Q:3***
Answer	Turn on or off the lights when I'm sitting on the sofa.	To avoid standing up.	Because I'm lazy.
Sub-category	Lighting	Change habits	Laziness
Category	Lighting	Change habits	Comfort

The four-step coding procedure resulted in an elaborated set of categories used to further analyse the data using the MEC approach.

2.5.3.7 Implication matrix

The second step in the MEC analysis is to create an implication matrix. A matrix with the size of X*X with X representing the total number of categories.

This was performed by comparing the linkages between the categories in the answers given by the respondents. Responses to the three questions generated a ladder. For instance, if a respondent answer category X on the first question, Y on the second question and Z on the third question, this generates a ladder of X → Y → Z. In a ladder the direct linkages are found between the linkages that are directly connected. In this example X → Y represent one direct linkage and Y → Z another. The indirect linkages are found among those categories that have at least one other category between them, in this example X → Z.

The total number direct and indirect linkages were noted and summarized in the implication matrix. Were the direct linkages are represented by an integer to the

left of the decimal and the indirect linkages are represented by an integer to the right of the decimal. I.e. a number of 4.02 means 4 direct linkages and 2 indirect linkages.

2.5.3.8 Hierarchical value map

The implication matrix described above works as a basis for designing the HVM. This is the third and last step in the MEC analysis. The ladder linkages are on the aggregated level termed *chains* (Reynolds and Gutman, 1988).

An HVM is gradually built by considering the linkages between categories (Reynolds and Gutman, 1988). I.e. A → B and B → C, create a chain with the categories A–B–C. To decide which categories that should be a part of the HVM a cutoff value was used. The reason for using a cutoff value is to illustrate the bigger picture and not include every single linkage. The value is determined by calculating the percentage of active cells and the number of active linkages based on different cutoff values in the implication matrix, as recommended by Jung (2014).

Thereafter the method explained by Reynolds and Gutman (1988) was used. First, the first row was examined to find a cell with a number equal or higher than the cutoff value. The linkage between the first row and the row belonging to that cell was identified. Then, the row of that cell was examined in a similar manner. Then the same method was used on that row and so forth. This led to a created chain of for example A–B–C–D.

Multiple chains emerged from the implication matrix and the different linkages were examined to decide which categories to identify as values and which to be seen as consequences in the HVM. As explained by Reynolds and Gutman (1988), this required some ingenuity. When the different chains had been designed and the categories abstraction level had been decided upon, the chains were merged, creating the final HVM.

The HVM makes the result for the primary approach of this thesis. In following sections, the work process for complementary data collection methods is described.

2.5.4 Qualitative interviews

In this thesis, the qualitative interviews were conducted for two different purposes. First to get a contextual understanding of the IKEA perspective. Second, to triangulate the findings from the MEC analysis.

2.5.4.1 Contextual interviews (unstructured)

Interviews were conducted with the supervisor at IKEA and the responsible manager of IKEA Home Smart. The interviews focused on creating a context of

the IKEA perspective. They were held by telephone or in person and were all recorded. All contextual interviews were unstructured, which gave the interviewee a possibility to speak freely about what he or she felt was relevant for the subject. The interview guides can be found in Appendix C.

To analyse the interviews, word-by-word transcripts were created. By transcribing the interviews, a more thorough examination of responses could be made. Thorough transcription allowed repeated examinations of material and transparency by enabling public scrutiny (Bryman and Bell, 2015).

2.5.4.2 Expert interviews (semi-structured)

The expert interviews aim to triangulate the collected data and results from the MEC analysis.

Ten people, considered experts, were asked to participate in the triangulation. The selection process consisted of searching for professionals within the academy and in the business world. Six of them chose to participate, two from KTH Royal Institute of Technology in Stockholm, one from Graduate School of Engineering Science at Osaka University in Japan and two from Nuance Communications – world leader in speech software (Nuance Communications, 2017). The supervisor at IKEA was also considered an expert and hence interviewed as both an IKEA party and expert. All experts were given the opportunity to proofread their input for ethical reasons as suggested by Bryman and Bell (2015).

The interviews were preluded with the MEC results in a summarized research paper sent by email. The authors developed a questionnaire to use in a semi-structured interview, presented in Appendix D. The interviews began with a brief explanation of the MEC results, followed by questions regarding the results. The interviewees were also asked to tell their general opinion on the user value of speech recognition. The interviews were held over telephone and lasted for 10-15 minutes. Except for one, all interviews were recorded and transcribed. For the interview that was not recorded, extensive notes were taken during the interview, complemented by proofreading and correction immediately after the interview.

2.5.5 Document gathering

As part of the research design, data were collected directly from IKEA. Two different types of data were collected: presentational material and the IKEA Life at Home Report.

The presentational material was reviewed and used for contextual understanding as a complement to the IKEA interviews. The material consisted of two slideshows with presenter's notes, used as an external introduction to IKEA Home Smart.

The IKEA Life at Home Report is an annual report funded by IKEA to understand

the way people live at home (Inter IKEA Systems, 2017). Since the first report from 2014, different themes have been explored including relationships, morning routines and cooking. The most recent report, Beating the Battles from 2017, involves many aspects of the technological home, thus is relevant for this study. The report is based on statistics from a survey with over 21 000 respondents in 22 countries. Different data is used, including qualitative deep interviews, global surveys and social media interaction (Inter IKEA Systems, 2017).

2.6 Quality of research

Typically, quality of a research is evaluated based on the criteria reliability, replication and validity (Bryman and Bell, 2015). Guba and Lincoln (1994) describe alternative criteria for evaluation, where the primary criteria are trustworthiness and authenticity. To increase the trustworthiness of this study one technique from trustworthiness is derived – triangulation, as described when introducing the research strategy. The technique of triangulation is elaborated below. Triangulation is common in mixed methods research strategy and works as a complement to the typical criteria, validity, reliability and replication, all discussed below.

2.6.1 Validity

Validity refers to the issue of whether or not a measure of a concept really measures that concept (Bryman and Bell, 2015). LeCompte and Goetz (1982) divide validity into internal and external validity. Where internal validity deals with if the researchers' observations match the developed theory, and external validity refers to what degree the conclusions can be generalized.

This study focuses on the unique phenomena and aims to understand its complexity. Typically this makes the internal validity a strength since the theory and observations are developed in symbiosis (Bryman and Bell, 2015). The generalization of the study might suffer since it is concentrated around a specific subject and small samples. However, the results can be used by others in the field of speech recognition and not just by the faculty and IKEA. Thus, the results are fairly generalizable.

2.6.2 Triangulation

Triangulation works as a technique to establish credibility, which parallels internal validity (Bryman and Bell, 2015). This means that more than one source of information has been used to describe the same concept. This generates perspective, control and possibly validity of the results.

In this thesis the process of triangulation aims to increase the validity and involves the following methods:

- The MEC approach and analysis, which results in derived attributes, consequences and values and a graphical representation of those, an HVM.
- Semi-structured interviews with experts from the academy, business world an IKEA. They described their view and gave comments on the MEC results.

2.6.3 Reliability

Reliability of research refers to the consistency of a measure of a concept. It deals with the question of the repeatability of the study. In other words, if the measure in the study was performed again with a gap in time, there will be little variation of the results collected (Bryman and Bell, 2015; Lekvall and Wahlbin, 2011). For this study, there are risks of reliability with the MEC approach as well as the interviews.

The MEC approach is, as explained earlier, a widely accepted method when studying customers perceived value of products or services. In this research, there are risks of reliability when collecting the data, and when coding the data. If, when doing the examination once again, it generates varied results, the measure could be considered to be unreliable. However, the reasons could be changes in external factors or the respondents' personal circumstances (Bryman and Bell, 2015). Therefore, the distribution of respondents is thoroughly documented and aimed to be heterogeneous. Further, there is a risk with reliability when doing translations of data into categories, as performed when coding answers in MEC. This is called inter-rater reliability, to mitigate this risk the method of coding was planned in detail and evaluated with the quantifiable measure inter-rater agreement.

The issue of reliability for interviews are the fact that the interviewees might be subjective or affected by the current context. Other interviewees might give another perspective on the same prospect. To reduce these risks interviewees were hand-picked from different areas of expertise and from both the academy and the enterprise.

2.6.4 Replication

The concept of reliability is closely related to the criterion of replicability. This is a criterion that copes with the ability to recreate or replicate the research to audit the findings of a study (Bryman and Bell, 2015). This is uncommon in qualitative research but highly valued in quantitative research. Since this study combines both strategies, although with, a focus on the qualitative strategy, the issue of

replication is aimed to be dealt with. This is done by documenting the MEC approach as detailed as possible.

3 Theory

The theoretical framework is presented in this chapter. First, a literature review on speech recognition is described resulting in a definition of the concept. Second, the history and usage of speech recognition today is presented. Finally, the current state of the user value of speech recognition in the literature is presented along with a definition of user value.

3.1 Definition of speech recognition

Before studying the user value of speech recognition later on in this chapter, the phenomena will be appropriately defined. No uniting term or definition of speech recognition has been found in the academic literature. This section aims to provide a theoretical foundation to define speech recognition.

3.1.1 Term

A literature review was done to understand what term should be used, and how it should be defined. The criterion for an article to be part of the literature review was that it should describe the concept of human speech being understood and, or, registered by an artificial system, e.g. a computer.

Derived from the literature, the terms: *speech recognition*, *voice control*, *voice recognition*, *speech command controlling*, *voice interface* and *smart home automation system*, including their variations, were all used to describe the concept and are presented in Table 3.1 . The three most cited terms; *speech recognition*, *voice control* and *voice recognition*, are examined below, starting with the third on the list.

Voice recognition, the third most cited term, is claimed by Homayoon Beigi (2011), as synonymous to *speaker recognition* but has been "mistakenly applied to speech recognition" (Beigi, 2011, p. 3). Speaker recognition is the technology of recognizing who is speaking rather than what is being said. (Beigi, 2011)

Table 3.1 Summary of different wording explaining the phenomena used in literature.

Concept	*Sources*	*Citations*
Speech Recognition, Speech Recognition System, Automatic Speech Recognition	Rabiner and Juang, 1993; Kamm and Helander, 1997; Jurafsky and Martin, 2000; Newman, 2000; Anusuya and Katti, 2009; Meisel, 2010; Homayoon Beigi, 2011; Gemmeke et al., 2013	20 570
Voice control, Voice command	Obaid et al., 2014; Kim and Lee 2016; Simpson and Levine, 2002; Busatlic et al., 2017; Baig et al., 2012	212
Voice recognition, Voice recognition systems, Voice command recognition	Beigi, 2011; Mittal et al., 2015; Bala et al., 2010; Arriany and Musbah, 2016	73
Speech command controlling	Zhang et al., 2008	33
Voice interface, Voice based user interface, Voice user interface	Cordasco et al., 2014; Vacher, et al., 2015; Soda, 2012	29
Smart Home Automation System	Mittal et al., 2015	10

The second to most cited term was *voice control* (including variations). However, articles using this term focused on applied technology. Assuming the main contribution was other than theoretic, less caution would have been taken when naming the concept. Also, the number of citations were significantly lower than the number of citations for the most commonly cited term.

Speech recognition (including variations) was the term with the articles most commonly cited, with 20 570 citations compared to 212 for *voice control*. Two of the most cited journals using speech recognition are amongst the earliest published; Rabiner and Juang (1993) with 10 897 citations and Jurafsky and Martin (2000) with 9 078 citations. Since they have been published for a longer time it is likely that it increases the number of citations. However, the big difference between first- and second place allows confidence to determine the suitable name of the phenomenon. The most appropriate term is *speech recognition*, which includes the variations speech recognition system and automatic speech recognition.

The term *speech recognition system* will be used in this thesis to describe a product rather than the technology itself. E.g. Google Home is a speech recognition system using speech recognition. A speech recognition system can

also include smart products connected to the interface, e.g. a fridge or TV. The term automatic speech recognition will not be used in this thesis.

3.1.2 Definition

Identifying the term of the phenomenon is not enough. Thus, this section seeks to answer what speech recognition refers to. The criterion for an article to be part of this step is the same as when identifying the term. A comparison of different descriptions and definitions of speech recognition is reviewed below. The different descriptions can be found in Table 3.2.

No unifying definition of speech recognition could be found. Additionally, only a few publications presented a definition rather than a description. Instead, the phenomenon was most commonly described in text or by using a figure. Further, the different descriptions do not necessarily explain the same phenomenon. As an example, Gemmeke et al. (2013) describe automatic speech recognition by naming different products. Meisel (2010) on the other hand, claims speech recognition not to be a product but rather a technology.

Similarities between the different descriptions can also be found. First, all definitions directly or indirectly describe speech recognition as a technology. Further, Gemmeke et al. (2013), Newman (2000), Beigi (2011), Jurafsky & Martin (2000) and Rabiner and Juang (1993) describe speech recognition not only as a technology but as a *computer* technology.

All descriptions are meeting the requirement of *human speech being understood and, or, registered by an artificial system, e.g. a computer*. However, following descriptions lack the relevance or credibility to make a satisfactory definition. The description from Rabiner & Juang (1993) and Meisel (2010) is an interpretation and therefore less viable. Describing different types of products, as done by Gemmeke (2013), is not relevant to the definition of this thesis. Nor is a description of minimal requirements for understanding human language, as done by Jurafsky & Martin (2000). The description by Beigi (2011) is a side note in a book on a different topic (speaker recognition) and therefore not regarded viable.

Helander and Kamm (1997) use the words: "*Automatic speech recognition (ASR) is a technology for communicating with a computer using spoken words or phrases*" (Candace Kamm & Martin Helander, 1997, p. 103). Since speech recognition is more commonly used for automatic speech recognition and there is no clear distinction between them, automatic speech recognition is replaced with just speech recognition. To describe the different usages of the term speech recognition and enable other terms to be used, the Newman (2000) description of speech recognition being an *umbrella term* is interlaced to the definition by Helander and Kamm.

After those alterations, the definition of speech recognition for this thesis is:

> *Speech recognition is an umbrella term for a technology communicating with a computer using spoken words or phrases.*

Table 3.2 Descriptions of speech recognition as found in the literature.

Description	***Source***	***Citations***
Human speech is processed and understood. The received information can trigger a task and, or, prompt the user with speech. *(interpretation of model describing speech recognition)*	Rabiner and Juang, 1993	10 897
Minimally, such an agent would have to be capable of interacting with humans via language, which includes understanding humans via speech recognition and natural language understanding (and of course lip-reading), and of communicating with humans via natural language generation and speech synthesis. *(about speech recognition)*	Jurafsky and Martin, 2000	9 078
In a speech recognition application, it is not the voice of the individual which is being recognized, but the contents of his/her speech.	Homayoon Beigi, 2011	204
Automatic speech recognition (ASR) is a technology for communicating with a computer using spoken words or phrases.	Kamm and Helander, 1997	27
These days, Automatic Speech Recognition (ASR) is firmly rooted in everyday life, with ample examples such as talking to your iPhone using Siri, vocal interfaces for home automation or directing your navigation device by voice while driving.	Gemmeke et al., 2013	23
Speech recognition is an umbrella term that covers a number of different approaches to creating software that enables computers to recognize natural human speech.	Newman, 2000	13
Speech recognition is a technology that can be used in products, not a product itself. (publication not providing a description of speech recognition but mentioning examples of e.g. mobile phones being able to understand human speech)	Meisel, 2010	6

3.2 Speech recognition in context

With speech recognition defined, it is of value to describe the context, both historical, spatial and conceptual. In this section the historical context is presented. This is followed by a description of the smart home, for spatial and conceptual

context.

3.2.1 Historical overview

The earliest form of speech recognition appeared after the World War II with machines producing "print-ready" dictation (Shagoury, 2010; Jurafsky and Martin, 2000). In 1952, Bell Labs built a machine recognising 10 different digits from a single speaker; further developed by MIT allowing 10 words to be recognised by any speaker (Shagoury, 2010).

In the 1960s, some of the fundamental ideas of speech recognition were published (Rabiner and Juang, 1993). In the 80's the US defence research constructed a system with a 1000-word recogniser (Rabiner and Juang, 1993). A commercial interest was taken in the machines. In 1984 Dragon System, now Nuance, released the first commercial application of speech recognition – an advanced device for dictation.

The first consumer-affordable systems, with a vocabulary of over 20 0000 words, were introduced in 1997 by Dragon System (Shagoury, 2010). In the years to come, speech recognition made its first steps towards becoming more mainstream. One example is speech integration for Microsoft Word, released in 2002 (Bhuiyan, 2016). The development of the cloud further enabled speech recognition to develop. A breakthrough for the mainstream consumer market was the integration of Siri to the Apple operating system in 2011 (Rhordi, 2016). In June 2015, Siri had more than 1 billion requests per week through speech (Bhuiyan, 2016).

In the last couple of years, speech recognition systems are finding the way into mainstream consumer homes (McKinsey&Company, 2016). Newly introduced products such as Google Home, Amazon Echo and Apple HomePod are consumer goods. Amazon Echo was the fastest selling speaker in 2016 and now owns about 25% of the market for speakers (Bhuiyan, 2016).

3.2.2 Speech recognition and the smart home

The smart home has been around for several decades, the idea is to engage a number of different connected devices to respond to the needs of its occupants (Aldrich, 2003). An easy and intuitive interface is needed and speech recognition is a promising candidate (Soda, 2012; McKinsey&Company, 2016; Fleury et al., 2011). Speech is seen as especially promising for comfort in the smart home, even though it is unexplored compared to other interfaces such as the visual (Fleury et al., 2011). When designing for the smart home, a user-centred mind-set is needed. Portet et al. (2011) claim several studies show that smart home applications are only successful if including the user in the design process.

In the future, the smart home is predicted to be automated. The data will provide

information to make the system smart and attend to the needs of the user automatically. (Aldrich, 2003)

3.3 User value and speech recognition

To understand the user value of speech recognition at home, the concept of user value is defined in this section. Further, the current state of the user value of speech recognition in the literature is described. Finally, design challenges are presented.

3.3.1 Definition of user value

In modern times, the term value is often referred to as the economical worth of an object (Friedman et al., 2013). This thesis studies value from a different perspective; the subjective experience of an individual. Also, this study uses the MEC approach to identify the value. Because of those two reasons the definition of value presented by the inventor of MEC, Jonathan Gutman: "desirable end states of existence" (1982, p. 60) is found to be suitable. However, this thesis is focusing on *user* value. Hence, adding the user perspective to the definition becomes:

> *Desirable end states of existence from the user perspective.*

3.3.2 The natural system

Speech is the fundamental way humans communicate. Through 200 000 years of evolution, humans have become voice activated. That skill has enabled societies to grow and become complex structures (Nass and Brave, 2005). This natural skill learned from birth is why speech interfaces are characterized as being more natural than other types of input devices (Kamm and Helander, 1997). Further, speech recognition allows the user to move around freely, without engaging hands and eyes, unlike visual interfaces. (Simon and Paper, 2007)

An ideal speech recognition system is known in the literature as a natural speech system. Such system would not only perfectly transcribe what is being said but also understand and generate appropriate responses or actions back (Kamm and Helander, 1997). Despite not having reached this level of perfection, the human brain rarely makes distinctions between speaking to a machine or another human. This also applies to very poor speech understanding with low-quality speech production (Nass and Brave, 2005). Another advantage is speed, already in 1989, speech recognition was claimed to be faster than typed input (Martin, 1989).

When being used as human-human interaction, too much might be expected from

the speech recognition system. Therefore, Shneiderman (2000) claims human-human relationships are not a good inspiration when designing user interfaces. Instead, the system should be designed by adjusting to computer limitations. Jennifer Lai (2000), claims the key to a successful speech recognition system is focusing on the everyday problem of the user – not the new technology.

Either way, managing expectation is important, with a word recognition of 95 percent accuracy, which is considered high, one out of twenty statements will be misunderstood (Nass and Brave, 2005).

3.3.3 Elderly – the early adopters?

Previous research in speech recognition has mainly explored specific target groups, often being elderly and disabled. From those group, the research result is positive. One study integrated speech recognition to the everyday life of elderly. Life was made easier for the users as well as the caretakers and loved ones (Portet et al., 2011). In another experimental study involving elderlies, Cordasco (2014) says speech recognition was greatly appreciated due to simplicity. Speech recognition provided the elderly with a better everyday use of technology products such as mobile phones, tablets and computers. Likewise, analysing multiple studies, Horstmann (2010) suggests people with disabilities are satisfied with the use of speech recognition.

Arriany and Musbah (2016) agree on speech recognition providing vital assistance to people with disabilities and other vulnerable citizens. While for people without disabilities, speech recognition works as a secondary tool. By this, speech recognition could be more inclusive than visual interfaces to interact with the surrounding environment. (Arriany and Musbah, 2016)

Furthermore, one study found that 95 % of persons in that study would continue to use the voice-controlled system even if it sometimes was wrong in interpreting commands (Portet et al., 2011). However, another study had interviewees expressing fear about a system not recognising what is said (Koskela and Väänänen-Vainio-Mattila, 2004).

3.3.4 Design challenges

Being a social interface, designing speech recognition demands a new way of thinking compared to when designing a visual interface. As one example, Simon and Paper (2007) stress the importance of social norms. This was concluded from a study having speech recognition as an operating system on a naval fleet. The biggest barrier to adoption was found to be social norms. Simon and Paper (2007) further suggest that once a critical mass of adopters for the technology is reached, users will have a hard time refusing or delaying their adoption process.

Understanding the information of the spoken text is not enough. Our brain has a liberal definition of speech including tones, pronunciation and non-vocal communication such as facial expressions. This challenge demands advanced technology involving artificial intelligence and coordination between different systems (Nass and Brave, 2005).

Further, by analysing existing research in speech recognition, Nass and Brave (2005) presents a fundamental guideline when designing speech recognition systems; the more similar to the user, the better. People want to talk to a person alike themselves including gender, personality type and tonality.

3.4 Summary of theory

To sum up, two of the terms used in research questions are defined. Speech recognition is defined as *an umbrella term for a technology communicating with a computer using spoken words or phrases*. User value is defined as *desirable end states of existence from the user perspective*.

Speech recognition has been around since the 1950s and has developed rapidly for the broader audience the last years, with the introduction of speech recognition systems for the home such as Google Home.

One basic characteristic of speech recognition, among others, is the natural way of communicating. Previous studies have shown that speech recognition benefits elderly and people with disabilities.

Also, the most efficient system of social interfaces, such as speech recognition, is the one that is mimicking the personality of the user.

4 IKEA

This chapter focuses on IKEA, and their development into the smart home. Firstly, the smart home department at IKEA is described along with what is interpreted as their criteria for user value. Secondly, the IKEA mindset towards speech recognition is presented. Lastly, the impact of technology on the life at home is described.

4.1 IKEA and the smart home

IKEA is a global furniture retailer, founded in Sweden, during the 1940s. Since the first IKEA furniture catalogue in 1951, the company has grown into a global organisation with 400 stores in 47 countries (Block, 2015). Today, IKEA is considered the largest furniture retailer in the world (Statista, Inc., 2017).

The IKEA vision is to *create a better everyday life for the many people*. This is done by their business idea "To offer a wide range of well-designed, functional home furnishing products at prices so low that as many people as possible will be able to afford them" (Inter IKEA Systems B.V., 2017a).

With technology playing a bigger part in everyday life for the many people, IKEA started a smart home initiative in 2012. The initiative was named IKEA Home Smart and integrated into the lighting department. Block (2017), the business leader of IKEA Home Smart, describes the initiative growing out of curiosity and having an entrepreneurial spirit. The initial idea was to explore technology in the home environment and how IKEA could be part of the ongoing technological change. (Block, 2017)

As of now, a range of wireless charging pads and a smart lighting system has been launched (Töreman, 2017d). Other projects are in the pipeline. The ambition of IKEA Home Smart is not only to make smart products. The vision is to influence the entire IKEA with smart, digital solutions, to help the company move forward on the digital agenda (Block, 2017).

4.1.1 Criteria for user value at IKEA Home Smart

As part of IKEA, IKEA Home Smart follows the IKEA vision and way of

working. To further understand how two staff members at the IKEA Home Smart department were interviewed along with analysing internal material. Access to internal material has been limited due to confidentiality.

When developing a product, the user is always the main priority. Rebecca Töreman (2017f) states the goal is to create user value, but the term user value is not talked about in itself. Instead, there are several aspects of product development being discussed. If these are met, the product is assumed to generate user value (Töreman, 2017e). In this thesis, these aspects are interpreted as product development criteria.

To understand the criteria and prioritise amongst them, the interviews were analysed. It was noted how many times a word, or a similar meaning to that word, was mentioned. Criteria and the number of times mentioned are presented in Table 4.1. The most common criteria were *convenience* (12 mentions) followed by *easy to understand* (8 mentions), *solving a user need* (8 mentions), *clear use case* (5 mentions) and *affordability* (4 mentions). *Beautiful home*, *wellbeing*, *peace of mind* and *accessibility* were mentioned twice or less and therefore not further elaborated. Also, *affordability* is excluded from further analysis. Affordability is related to implementation, which is not included in this thesis.

Table 4.1 Criteria for user value as perceived from interviews.

Name	Number of times mentioned in correspondence
Convenience	**12**
Easy to understand	**8**
Solving a user need	**8**
Clear use case	**5**
Affordability	4
Beautiful home	2
Wellbeing	2
Peace of mind	1
Accessibility	1

The most mentioned criterion was *convenience*. By convenience, IKEA means having an easier life and being able to do tasks more effectively. Töreman (2017b) exemplifies by having your lights turn on when waking up. Convenience generated by speech recognition could be translated into controlling the lights without leaving the sofa. The purpose of convenience is to complete tasks quicker. That way, the user is able to spend time on what really matters, such as spending time with your family. (Töreman, 2017b)

Furthermore, the product should be *easy to understand*. Unlike many tech players on the market, IKEA wishes to provide smart products for the broad audience, regardless of tech skills. Therefore, the product should easily be understood.

Already in the store, the user should understand the functionality and potential benefits of the product (Töreman, 2017d). One example is manifested in the smart lighting system. Unlike other systems on the market, the colours are curated and limited. Also, the app controlling the light is made with a less-is-more approach (Töreman, 2017d).

Another criterion was for the product to *solve a user need*. In the context of technology, IKEA is not interested in producing gadgets. For IKEA, it does not matter whether the solution is analogue, mechanical or digital. Preferably, products should be integrated with the physical home, as was done with the wireless charging and the smart lighting. IKEA wants to keep focusing on the physical home. (Töreman, 2017d; Block, 2017)

Finally, the product should have a *clear use case*, meaning real user problems are being solved by the product. As a consequence, IKEA has not entered the field of smart appliances. This is because IKEA Home Smart has not yet seen a strong use case for the technology available today. (Töreman, 2017c)

As of today, no systematic tools are used to evaluate and quantify user value or the product development criteria above. However, feedback and ideas are given in dialogue with the different product departments within the company, e.g. living room, to learn about their challenges and needs (Block, 2017). Also, personas and user testing amongst staff are used as soft tools (Töreman, 2017d).

4.2 Speech recognition at IKEA

Speech recognition is an aspiring candidate as the primary interface of the smart home. Investments by major technological players have already made a breakthrough in the consumer market. Those factors combined make speech recognition interesting for IKEA (Töreman, 2017d; Block, 2017). Potential users also seem interested, 25% of respondents in the Life at Home Report claim they are excited to use voice-controlled technology to enable social interaction in their home (Inter IKEA Systems, 2017).

A step towards speech recognition was taken on the first of November 2017, when the smart lighting system at IKEA was integrated to speech recognition systems by Amazon and Apple. As of now, the feedback has been overwhelmingly positive. The negative complaints have been from Google users, still waiting for a similar integration due next year. Feedback from retailers has not yet been received and it is still too early to draw any conclusions according to Töreman (2017f).

Furthermore, IKEA is not and nor aspires to be a company developing the newest technology. Instead, the IKEA Home Smart approach is to integrate and enable other systems with their products, as was done with the smart lighting system (Block, 2015). Hence, IKEA is dependent on adopting to technology available on

the market. In practice, that means being perceptive to what the key players are doing (Block, 2017). If entering the market, IKEA will presumably affect it. As an example, the IKEA wireless charging technology was one of the aspects taking into consideration when Apple decided to integrate wireless charging. (Block, 2017)

For IKEA to further develop speech recognition, Block (2017) thinks the technology needs to be perfected. The IKEA customer has short patience and is unforgiving with a dysfunctional system. He recognises challenges such as entering a credit card and changing the system settings by using voice. He is interested in seeing whether technological companies will go all-in on voice. A possible scenario could be a combination of a visual interface and speech recognition. (Block, 2017)

4.3 Life at home with technology

IKEA has produced a Life at Home Report since 2014 to learn about the present meaning of *better everyday life for the many people*. Aspects relevant to this thesis are brought up below.

4.3.1 Technology, the ambivalent aid

Life at home has benefited from the use of new technology. Loneliness can be reduced by connecting the world outside the walls of the house. Time can also be spent together through streaming music or watching TV. Dull, daily activities can be handed over, such as vacuum cleaning and keeping a schedule. Owners of smart homes appreciated being able to remotely control their home. One example mentioned was being able to check in on pets and follow up on deliveries while being away. (Inter IKEA Systems, 2017)

On the other hand, 27% of people think they spend less time with their partners because of technology. Also, 17% feel guilty about spending too much time on social media when at home. One third of people in the study put a time limit on how much they use screens at home. The same proportion of people think it is hard to find the right balance of using technology at home. (Inter IKEA Systems, 2017)

Despite being a global problem, the attitude towards new technology differs with country. In Sweden, only 10-29% feel excited to bring future tech into their homes, compared to Russia where 66% agreed. (Inter IKEA Systems, 2017)

The ambivalent feelings towards technology are true also coming to safety. On the one hand, smart products often increase the feeling of safety in our home by giving a *peace of mind* when leaving the home and helping to remember routines. On the

other hand, privacy might be experienced as intruded. The latter seems to vary with country, people from Germany and Russia are significantly more worried than their counterparts in the US and India. (Inter IKEA Systems, 2017)

4.4 Summary of IKEA

In conclusion, IKEA entered the smart home market in 2012 with their IKEA Home Smart product range. The vision of *creating a better life for the many people* is realised in a set of product development criteria. A product should be *convenient, easy to understand, solve a user need* and have a *clear use case*. IKEA's mindset towards speech recognition is positive but awaiting. A first step in the area of speech recognition was taken in November 2017, when the smart lighting system was integrated into speech recognition systems by Amazon and Apple.

Further, technology has both positive and negative impacts on our life at home. On the one hand it helps to create a better life, on the other hand it can be a barrier to growing human relationships.

5 Means-end chain analysis

This chapter performs the analysis of the MEC approach, which is the first step in the analytical foundation to answer the purpose. First, it presents the results from the MEC survey, followed by the identified categories in the coding procedure. Furthermore, the implication matrix is introduced which acts as the framework when designing the HVM at last.

5.1 MEC data

The first step in the MEC analysis is to collect and structure the data. A total number of 64 answers were collected, with 10 being disqualified for further analysis. Those answers that were cleaned out were incomplete, absent or meaningless. Thus, a total number of 54 answers could be analysed.

As presented in Table 5.1, those 54 respondents were heterogeneous in gender, with 57.4% female and 42.6% male. Coming to age, the groups 24-35 years and 65 years or older were most frequent, with 22.2% and 33.3% respectively. For other age groups, the distribution was fairly heterogeneous. Furthermore, the majority of the respondents were living with spouse or partner, representing 59.3%. The second most frequent living situation was living alone, 24.1% and the third was with children, 18.5%.

A complete table of the demographics is presented in Appendix B, Table B.1. Data in that table show respondents answering *yes* or *maybe*, 50.9% and 37.7% respectively, to the question of having an interest in using more speech control at home. The table also presents 40.7% of the participants having a smart TV/media player and roughly 20% have smart lighting or smart security camera at home. Lastly, none of the respondents had a speech recognition system in their home.

In Appendix B, Table B.2 demographics from disqualified answers are presented. Out of the 10 disqualified answers, 2 were completely empty and 8 had given complete answers on the demographic part of the survey, but incomplete answers on the laddering questions. Those respondents were heterogeneous in age and women were overrepresented. The distribution between different living situations was fairly consistent with the qualified answers, i.e. 60% living with spouse or partner, 20% living alone and 20% living with children.

Table 5.1 Demographic information of the 54 usable participants. Details to be found in Appendix B, Table B.1.

		Frequency	*Percent*
Age (years)	18 or below	1	1.9
	18-24	7	13.0
	25-34	12	22.2
	35-44	4	7.4
	45-54	5	9.3
	55-64	7	13.0
	65 or older	18	33.3
Gender	Female	31	57.4
	Male	23	42.6
Living situation	Alone	13	24.1
	With relative(s)	3	5.6
	With spouse/partner(s)	32	59.3
	With friend(s)	1	1.9
	With child(ren)	10	18.5
	Other	1	1.9
Using voice control at home today	Yes	9	17.0
	No	44	83.0

5.2 Coding

After collecting and structuring the data, the laddering questions from the 54 qualified answers were coded. The coding resulted in eleven attributes, six consequences and five values. The process of making the distinction between attribute, consequence and value, is described in next section. In this section the attributes and consequences are presented briefly and the values are presented in more detail.

First, the most frequent attributes were *lighting*, *entertainment*, *security equipment*, and *appliances*. Apart from the distinction explained when introducing the implication matrix in next section, the answers coded as attributes are straightforward to differentiate between the consequences and values since they refer to functions of speech recognition. All attributes, with descriptive examples and the total number of mentions, are presented Table 5.2.

Table 5.2 Attributes derived from respondent's answers on laddering questions with descriptive examples and the total number of mentions in the answers.

Attribute	*Example*	*Mentions*
Household	Tell the robotic vacuum cleaner to vacuum the kitchen floor or turn on the coffee machine.	9
Appliances	Set a timer on the oven or set the oven at 225 degrees.	12
Other & not specified electronics	Turn on or off my TV without pushing a button.	6
Find my phone	Tell me where my phone is if I can't find it.	4
Security equipment	Lock the doors when I'm in my bed.	15
Lighting	Switch off the lights when sitting on the sofa.	25
Indoor climate	Turn the heat up in the living room.	2
Time & information management	Wake me up at 7 o'clock tomorrow morning.	9
Communication	I want to be able to call on my children on the second floor when dinner is ready.	6
Information search	Find the closest pizzeria that is open at 9 pm.	5
Entertainment	Change TV channel without the remote control.	16

The categories for consequences are presented in Table 5.3. They are C1 (*simplify*), C2 (*physically engaged*), C3 (*change habits*), C4 (*security*), C5 (*fun & exciting*) and C6 (*saving energy*). The consequences differed in frequency with *simplify* being mentioned 27 times and *saving energy* being mentioned only once.

Table 5.3 Categories for consequences with examples and the total number of mentions.

Category	*Example*	*Mentions*
C1. Simplify	Find a restaurant with speech recognition because it would be a simple way of doing it.	27
C2. Physically engaged	Turn on the lights when my hands are full of groceries, coming home from the grocery store.	21
C3. Change habits	Turn off the oven etc., to avoid having to physically check the kitchen all the time.	14
C4. Security	Double check if the door is locked. Because of security aspects at home.	7
C5. Fun & exciting (removed)	Control washing machine, it can be a bit clever.	9
C6. Saving energy (removed)	Turn off the lights to reduce electricity.	1

The values derived from the data are C7 (*facilitate daily life)*, C8 (*everyday efficiency*), C9 (*comfort*), C10 (*increased calmness*) and C11 (*indifference*). Where the most frequent value far and away was *comfort*, mentioned a total of 20 times. All values are presented and explained in Table 5.4. Three values are derived from this study, while two are modified from previous research, this to build credibility and transparency to the derived values.

Table 5.4 The derived values with descriptive explanations and the total number of mentions.

Category	***Explanation***	***Mentions***
C7. Facilitate daily life	Live an easier life with speech recognition to help in daily tasks. (derived from the study)	9
C8. Everyday efficiency	The ability to complete a task faster than today. Be more productive in the daily life. (modified from Jung, 2014 and Deng and Christodoulidou, 2015)	7
C9. Comfort	Ability to feel relaxed and comfortable in the home environment. (modified from Jung 2014)	20
C10. Increased calmness	Ability to feel safe and secure in the home environment. Less concern and stress. (derived from the study)	8
C11. Indifference (removed)	Scepticism and incredulity towards speech recognition. (derived from this study)	7

The first value in the table, *facilitate daily life*, is achieved when the functions used in speech recognition facilitates the existence of the user. A typical answer from a respondent that was coded as *facilitate daily life* was:

> *I want to use speech recognition because it would be a smooth way to turn on the lights, send text messages if I am running late or to set a timer on the oven. It would be simple and facilitate in my everyday life.*

Moreover, *everyday efficiency* is modified from values in previous MEC research. First, Jung (2014) use *productive daily life* to explain a user value similar to this study, which is exemplified with *waste less time, manage schedule better, make better decisions and help concentration*. Second, Deng and Christodoulidou (2015) use *productivity in daily life* which is explained as *the ability to manage daily routines and schedules more efficiently*. The comparison with those two definitions and analysis of answers in this study results in the value of *everyday efficiency* which is perceived when enabling the user to complete tasks faster than today and to be more productive. An example of an answer from the survey, adjusted to formulate sentences is presented below:

> *I want to control lighting, radio, TV, oven or my computer because it is quick and effective. I can also do other things at the same time. It is great because I can do it in an automatic (or natural) way when it comes to my mind.*

Likewise, an alteration of *comfort* is also used in previous research. Jung (2014)

uses *sense of comfort* and exemplifies it as *emotional stability, feel connected and become comfortable*. In that study, it is described as reaching a certain psychological state by having positive social relations. However, in this study, the value is more strictly derived from *comfort* only, which by Kolcaba and Kolcaba (1991, p. 1302) is defined as "whatever makes life easy or pleasurable" and it is compatible with the self-indulgent goal of maximizing pleasure (Kolcaba and Kolcaba, 1991). Respondents in this study used terms as laziness, comfort or relaxation to describe their presumed value of speech recognition, hence *comfort* from Kolcaba and Kolcaba (1991) is most appropriate. A typical answer exemplifying *comfort* in this study is presented below:

> *I want to turn on or off the lights when I am sitting on the sofa, to avoid getting up. This is important to me because I am lazy.*

Another reoccurring motif was respondents feeling speech recognition could help them to reduce their stress or concern. Hence, the value of *increased calmness* was derived. To exemplify, a slightly polished answer from the respondents is displayed:

> *Lock the front door through speech recognition. It is easy to forget and it is important to me that the door is locked.*

The last value derived in this study, *indifference*, was given in answers with apparent scepticism and incredulity towards speech recognition. Answers coded as *indifference* could be stated as:

> *Speech recognition is probably great in some areas, but people need more human contact, not more technology.*

The categories presented above, derived from coding, are used in the implication matrix as the next step in the MEC analysis.

5.3 Implication matrix

The implication matrix reveals the relationship between the derived categories. I.e. the implication matrix depicts the number of times one category leads to another. It is displayed in Table 5.5 and the distinction between the categories is explained below.

There are three different types of categories in the implication matrix based on their linkages. These different types are used to decide the categories abstraction level, i.e. making the distinction between attribute, consequence and value. This process might seem complex and requires ingenuity, as described by Reynolds and Gutman (1988).

The first type points only towards other categories, these are on the lowest abstraction level and therefore attributes. As mentioned earlier, they refer to functions of speech recognition. For the sake of ease and convenience when interpreting the implication matrix, they are only shown in the rows, and not the columns.

The second type has multiple linkages both from and to itself, that means that they are in the middle of two other abstraction levels and therefore represents the consequences.

The categories on the highest abstraction level represent the values and do have linkages towards them but do not link further in the chain. This is the end of the chain and indicates the highest abstraction level, this makes the distinction between consequences and values pretty clear.

To summarise, attributes only have linkages *from* themselves, consequences have linkages both *from* and *towards* themselves, values only have linkages *towards* themselves. This understanding makes it possible to build the actual chains.

To exemplify the usage of the implication matrix in Table 5.5: *lighting* point directly towards C1 (*simplify*) a total number of 9 times and indirectly a total number of 4 times in the collected responses. Hence the number of 9.04 where the row of *lighting* crosses the column of C1. Furthermore, C1 (*simplify*) is linked to C7 (*facilitate daily life*) directly a total number of 4 times and 0 times indirectly in, hence the number of 4.

The relations between the categories in the implication matrix are the foundation when designing the HVM, it enables to build chains of the collected data. This is the next and last step in the MEC analysis.

Table 5.5 Implication matrix.

Categories	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11
Household	3.03	1.01	1	1.01			1	0.01	0.01	0.02	
Appliances	3.02	4.02	1	1.01	1.01		2.01	1.02	2	1	1.01
Other & not specified electronics	3.03	2					0.01	1.01		1	
Find my phone	1	2	1	1.01	1		1			0.02	
Security equipment	5.01	4.01	3	3.03	1.01		0.02	0.02	3.02	0.03	0.02
Lighting	9.04	7	6	3.02	1.01	0.01	2.03	1.02	7.07	1.04	0.02
Indoor climate	2.01								0.01		
Time & information management	2	2.01	2.01	0.01	1		3.01	1.01	2.01	0.01	2.01
Communication	5.01	0.02	2.01				1.01		1		
Information search	3	0.01	1.01		0.01		1.01	1.01	1.01		0.01
Entertainment	8.02	5.02	2				0.04	1.02	4.03	0.02	0.02
C1. Simplify		3	1	1			4		4	2	1
C2. Physically engaged	3			2			2	2	2	1	1
C3. Change habits	1	1		1		1		2	4	3	
C4. Security										2	
C5. Fun & exciting (removed)		1.01		1						1	0.01
C6. Saving energy (removed)	1							2			
C7. Facilitate daily life									1		1
C8. Everyday efficiency											
C9. Comfort		1			1		1			1	2
C10. Increased calmness									1		
C11. Indifference (removed)	1				1			1		1	

5.4 Hierarchical value map

The HVM is built up by connecting the chains that are formed by interpreting the implication matrix. To decide which cells in the implication matrix to include in the HVM, a cutoff value is used. The appropriate cutoff value is chosen by evaluating several cutoff values and choosing the one appearing most stable and informative (Reynolds and Gutman, 1988).

As performed by Jung (2014), the cutoff value is determined by calculating the percentage of active cells and the number of active linkages based on different cutoff values in the implication matrix, as presented in Table 5.6. The selected linkages should correspond to more than two-thirds of the total linkages (Jung, 2014). To satisfy those recommendations, the cutoff level 2 was chosen, i.e. only the relations with 2 or more linkages were included in the HVM. That corresponds to 38.7% of the active cells and 74.1% of the total linkages, which also shows a certain level of variance (Jung, 2014).

Table 5.6 Cutoff values on different levels. Level two was chosen and is highlighted in bold.

Cutoff level	Number of active cells in the implication matrix	Percentage of active cells at or above cutoff value (%)	Number of active linkages in implication matrix	Percentage of active linkages at or above cutoff value (%)
1	102	82.3	211	99.8
2	**48**	**38.7**	**156**	**74.1**
3	26	21.0	112	53.2
4	14	11.3	76	36.1
5	8	6.5	52	24.7

The cutoff value of two excludes two consequences, C5 (*fun & exciting*) and C6 (*saving energy*), since these two do not have linkages above a level of two.

The category of C11 *(indifference*) did nevertheless have linkages pointing to itself at or above the cutoff level. Both *time and information management* and C9 (*comfort*) link to *indifference* at a level above two. However, the answers classified as *indifference* were more about a general indifference to new technology. The answers were, for example; “There is really no reason for using speech recognition in the manner I described in the previous question”. Since that category of answers does not appear to be important in the MEC analysis it is omitted from the HVM. This was considered as an example of being ingenious when designing the HVM as recommended by Reynolds and Gutman (1988). However, the *indifference* approach by some respondents is an important aspect to be discussed later on.

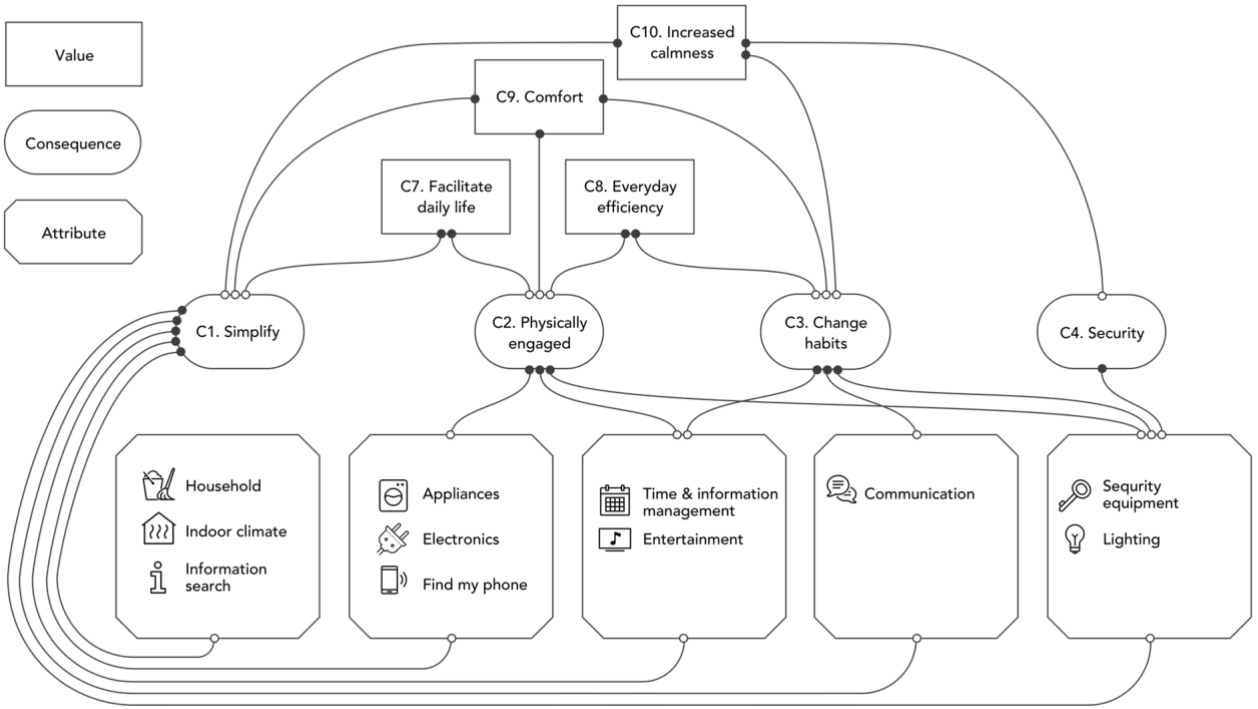

Figure 5.1 Hierarchical value map, a graphical presentation of the hierarchical relations between attributes, consequences and values derived in the study.

The HVM in Figure 5.1 shows the graphical summary of the hierarchical relations between attributes, consequences and values. The most important aspects of the HVM is explained below.

First, it is apparent that the attributes of *security equipment* and *lighting* are the most widespread attributes. I.e. they align with several different aspects of the perceived consequences.

Second, every attribute leads to the consequence of *simplify*. It is fairly obvious in the study, that users find an interest in simplifying their life at home. The consequence of *security* was the consequence with least linkages, and therefore the most distinguishable.

Third, the four values included in the HVM, *facilitate daily life*, *everyday efficiency*, *comfort* and *increased calmness* do all have multiple linkages. But *comfort* and *increased calmness* have most linkages.

Last, to exemplify the chain in the HVM three separate chains are presented. First, household → simplify → comfort. Second, time and information management → change habits → everyday efficiency. Third, security equipment → security → increased calmness.

5.5 Summary of MEC analysis

In summary, the MEC analysis is the foundation to explore how speech recognition can generate value in the home environment. The MEC analysis explains collected data and presents the derived categories in abstraction levels of attribute – consequence – value. The derived and used values are *facilitate daily life*, *everyday efficiency*, *comfort* and *increased calmness*. The linkages between categories are summarised in an implication matrix, which in turn acts as the foundation for designing the HVM, where the connections between categories are graphically presented. *Security equipment* and *lighting* are the most widespread attributes and *simplify* is the only consequence linked from all attributes. The most distinguishable consequence is *security*. The results from the MEC analysis are further used when analysing the phenomena of speech recognition at home when building the analytic framework.

6 Analysis

In this chapter, the empirics from the IKEA exploration and the MEC approach are analysed. First, the MEC analysis is examined by experts in the field and by IKEA. Second, the relevance for IKEA is analysed.

The analysis is accomplished to elaborate the analytical foundation of the study to understand the purpose of the thesis. It will be used to answer the research questions:

- (RQ1) What user value could be generated from speech recognition in the home environment for the broader audience?
- (RQ2) Judging from a user perspective, how is speech recognition in the home environment relevant to IKEA?

6.1 Experts opinions

In the previous chapter, user values generated from speech recognition were investigated by doing a survey using the MEC approach. As previously described those results were triangulated by experts who were asked about their view on the results. This section presents the expert's audit of the MEC to triangulate the result. In most cases, experts were asked to comment the specific values. If time, the experts were also asked about the consequences and attributes.

6.1.1 Values, consequences and attributes

The values previously identified in the MEC approach are *facilitate daily life*, *increased calmness*, *comfort* and *everyday efficiency*. These are all analysed by experts below.

First, speech recognition is able to generate the value of *facilitate daily life* (House, 2017; Ishiguro, 2017; Kontogiorgos, 2017). The technology is easy to learn, especially compared to visual interfaces (Kontogiorgos, 2017; House, 2017). Ishiguro, a prominent visionary within the field and professor at Osaka University thinks the speech recognition systems of today, such as Google Home and Amazon Echo, are primarily helping people to be more productive. Based on his world-leading work with humanoids (robots that look like humans), he predicts

future systems will evolve daily life beyond productivity. Better technology making robots more human could, amongst other things, help people be more creative.

David House (2017), professor in speech recognition at the Royal Institute of Technology in Stockholm, claims *increased calmness* has to do with the simplicity of the interface combined with the ability to perform productivity tasks. A speech recognition system is less distracting than its visual counterpart when it comes to certain tasks. The tasks performed by the system save time and energy. (House, 2017)

Regarding *comfort*, Ekaterina Kruchinina (2017), a research manager in natural language understanding at Nuance Communications, claims that speech recognition can avoid a lot of effort when doing information searches. She also expresses uncertainty towards the value of comfort and its linkages to the consequences.

The value of *everyday efficiency* was questioned as the system might not be good enough yet. House (2017) thinks people can get frustrated by using speech recognition with everyday tasks. When writing an email, for example, the system is likely to get it wrong and end up demanding more effort than writing in a visual interface. Not delivering to the expectations of users has caused problems for speech recognition (Kontogiorgos, 2017; House, 2017).

Coming to consequences, *security* is recognised as something that could be generated from a speech recognition system (Kruchinina, 2017; Ishiguro, 2017). One reason being the user can control things further away, such as the locking mechanism of the door, in a simple way (Kruchinina, 2017). Also, having a system helping with anything from emergency calls to simple tasks is likely to generate a feeling of safety and security (Ishiguro, 2017). Ishiguro (2017) adds a system generating more security also means less privacy.

Kruchinina (2017) agrees with the attribute of *entertainment* generating user value. She claims speech is well suited for easy access to an information system such as a television, which could otherwise be complicated. The same logic applies to *security equipment*. Coming to *appliances*, the situation is different. Smart appliances as of today are often limited to simple commands such as turning on and off. Therefore, the benefit of using speech as opposed to a visual system is not that great. In addition, people need convincing to invest in the appliances, which might take time. In that case, an integration would be a nice-have rather than something to rely on (Kruchinina, 2017).

To summarise, none of the values was clearly disregarded by the experts. On the other hand, new perspectives were given. These perspectives are further considered in the discussion.

6.1.2 Design challenges

The experts were not only asked to give comments on the MEC results but also to give their opinion on speech recognition in general. Hence some design challenges were identified and described below.

The market for speech recognition at home is still at its cradle. However, what is done today have an impact on the speech recognition of tomorrow. Nils Lenke (2017), research director at Nuance Communications, claims the perception of speech recognition has been heavily influenced by Siri, released 6-7 years ago. Similarly, the impact of Alexa and Google Home today should not be underestimated. Those systems dictate what people think about speech recognition at home. One way people are influenced is through advertisement. The first thing people try on their speech recognition system is what they have seen in the ads. This does not only illustrate the influence of examples. It also reviles there is a need to teach people what to do with the systems. (Lenke, 2017)

In addition, there might be other unexpected design related aces for the user. Navigating a menu is awkward in speech recognition and better suited for a visual system. One way of design is for the system to provide suggestions on functions to try out (Lenke, 2017).

Future development will entail a humanisation of the system (House, 2017; Ishiguro, 2017). A more natural communication can be used including facial expressions (Ishiguro, 2017). It is also likely for the home to become more automated. In such case, speech recognition is likely to become a suitable interface (House, 2017).

6.2 Relevance for IKEA

This section aspires to elaborate on RQ2 *Judging from a user perspective, how is speech recognition in the home environment relevant to IKEA?* This is done by comparing the MEC results with collected information about IKEA.

6.2.1 The relevance of speech recognition according to IKEA criteria

As described previously, for a product to be relevant for IKEA Home Smart, it should be *convenient, easy to understand, solving a customer need* and *having a clear use case*. Those criteria are compared to the findings from the MEC approach below and summarised in Table 6.1.

6.2.1.1 Convenience

Convenience from an IKEA perspective is a merge between the value of

facilitating daily life and *everyday efficiency*. Together, they represent having an easier life and being able to do things more effectively. Examples given by IKEA to exemplify convenience were all mentioned in the survey.

In addition, convenience is in many ways related to the consequence *simplify*. Simplify was the most popular answer in the MEC survey, linking to all of the 11 attributes. Similarly, convenience was the most mentioned criterion in interviews with IKEA. Additionally, the ability of technology is also described in the Life at Home Report, as making our lives better and bringing convenience (Inter IKEA Systems, 2017).

Summarising, speech recognition can make the life of an IKEA customer in the home environment more convenient.

6.2.1.2 Easy to understand

There is no direct value or metric in either the survey or the HVM telling if the respondents thought speech recognition was easy to understand. Still, there are some indications. One-third of respondents were 65 years old or older, which is more than the national average of about 20% (Heggeman, 1999). Despite overrepresentation by one demographic group, not being regarded as the most tech-savvy, 84% of the respondents were able to give examples and write about how they would use speech recognition. Further, when approaching people to take the survey, they were asked about whether they knew what speech recognition was. Surprisingly, most people knew what it was. Those who were not sure knew what was talked about when bringing up Siri, the speech recognition assistant used in Apple products.

Adding to that, the theory has presented speech being natural and fundamentally human. This way, one could argue, speech is impossible not to understand. It is in our DNA. Humans are evolved to easily understand speech.

However, none of the respondents had tried a speech recognition system at home. Therefore, it is hard to determine if the system is easy to understand. Both literature and by experts mention a discrepancy between expectations and reality. To better understand whether the system is easy to understand, it would be useful to have users trying the product. Therefore, it is likely that speech recognition is easy to understand for the IKEA customer in the home environment but more research is needed.

6.2.1.3 Clear use case

The first of the three open-ended questions in the survey asked *What would you (or someone else in your home) use voice control for in your home?*, a question giving an answer similar to describing a use case. The answers received were typically presented as a use case, e.g. “turning the lights on when having my hands full”. It seemed respondents were able to understand and relate to possible use cases. The number of attributes in the HVM was equal to consequences and values

added together (11 attributes, 11 values and consequences). A possible reason is people finding it easier to name a use case rather than a consequence or value.

When supporting respondents to answer, many explained their use case in a more elaborate way than what was done in the survey. Having people write their answers possibly limited the description of a use case. This will be discussed later in the report. Qualitative interviews are likely to provide more insights into the use case.

Summarising, judging by the many use cases given by the respondents, speech recognition has a clear use case for the IKEA customer in the home environment.

6.2.1.4 Solving a user need

As described above, users can imagine when to use the speech recognition (clear use case) and thinks it would make life easier for them (convenience). These are two factors supporting speech recognition solving a customer need. However, many functions of speech recognition require other smart products. E.g. monitoring the light with speech demands to have smart lighting. Some of those smart products could be unnecessary gadgets while some could be very relevant. Therefore, speech recognition is only partly able to solve a user need in the home environment. It comes down to the products connected to the interface.

Table 6.1 Criteria for user value for IKEA and relevance for speech recognition.

Name	***True for speech recognition used by the IKEA customer in the home environment?***
Convenience	Yes
Easy to understand	Probably
Clear use case	Yes
Solving a use need	Partly

6.2.2 Safety

While all values found though MEC being mentioned by IKEA, safety (consequence) was not. *Safety*, in this case, is the feeling of being safe or in control of your home. If developing a speech recognition system, this aspect has relevance. Technology enabling the feeling of safety was confirmed by both the Life at Home Report and experts (Inter IKEA Systems, 2017; Ishiguro, 2017).

However, the Life at Home Report mentions privacy as an issue connecting to safety (Inter IKEA Systems, 2017). With more safety, there is also less privacy, adds the expert Ishiguro (2017). Privacy is something sought after in the home environment (Inter IKEA Systems, 2015; Inter IKEA Systems, 2017).

6.2.3 Response on MEC analysis from IKEA

When presented with the result, Rebecca Töreman (2017f), product developer at IKEA Home Smart and the IKEA supervisor for this thesis reacted by saying "It's exactly what's in our business plan" (a business plan which content is not known by the authors due to confidentiality). The business plan mentioned is for IKEA Home Smart as a unit, not speech recognition itself. "This is an indicator that we could be on the right track", she added.

The use cases, central to the product development at IKEA are described as attributes. Even though attributes are presented and explained in the HVM, entire user stories with use cases could not be provided from the data. This would be needed when taking speech recognition further at IKEA Home Smart, says Töreman (2017f).

6.3 Summary of analysis

In summary, the analysis aims to build the foundation to answer the two research questions, RQ1 and RQ2.

The expert interviews triangulate the results from the MEC analysis and analysed the validity of the values. None of the values were disregarded but some of them problematized. These new perspectives are brought into further consideration in the discussion.

To answer RQ2, the relevance for IKEA is analysed. It was concluded speech recognition for the IKEA customer in the home environment could generate both convenience and a clear use case. Speech recognition was also concluded as likely to be easy to understand but having users evaluate a real interaction is needed. Coming to solving a user need, speech recognition could partly be helpful, depending on the products connected to the interface.

Following, the idea that speech recognition making people feel safe was presented as something to be regarded by IKEA. Finally, the results from the MEC analysis are confirmed by IKEA's business plan.

7 Discussion

In this chapter, the results and analysis are discussed. First, the definition of speech recognition is examined. Then, the derived values from the MEC approach are discussed, followed by a discussion about the IKEA perspective and recommendations for IKEA. Further, the validity of the MEC approach is discussed and the academic contributions are presented. Lastly, proposals for further research are elaborated.

7.1 Answers to research questions

First of all, short answers to the two research questions are given. The theory, results and analysis build the foundation to answer these research questions. The answers are further discussed in this chapter.

(RQ1) What user value could be generated from speech recognition in the home environment for the broader audience?

This study concluded the results that the user value generated by speech recognition in the home environment is:

- Facilitate daily life
- Everyday efficiency
- Comfort
- Increased calmness

(RQ2) Judging from a user perspective, how is speech recognition in the home environment relevant to IKEA?

Speech recognition in the home environment is likely to provide user value for the IKEA customer. Also, speech recognition fulfils or partly fulfils, the product development criteria for smart products at IKEA.

7.2 The definition of speech recognition

When elaborating the theory around the definition of the phenomenon, a literature review was done concluding *speech recognition* is the most accurate term to describe the phenomenon of a computer registering and understanding speech. Later, speech recognition was defined as *an umbrella term for a technology communicating with a computer using spoken words or phrases*.

Speech recognition was by far the most frequent term used to explain the phenomenon in the examined literature. The number of citations of articles using the term *speech recognition* far outnumbered articles using other terms, such as *voice command* and *voice recognition*. Therefore, speech recognition was considered appropriate to use to describe the phenomenon in this thesis. A definition for speech recognition was created by fusing together two separate definitions.

Even if the definition provides a correct description of speech recognition as used in this thesis, it is not enough to describe the entire scope of speech recognition in the smart home. Speech recognition is often representing an entire system of smart home products. E.g., turning on or off the lights, the most popular attribute in this research, cannot be done only by a computer system communicating using spoken words or phrases. To control it, the system also needs a connected smart lightbulb, a communication gateway and maybe more. However, formulating a completely new definition would require a thesis in itself, studying the concept more deeply. Such a comprehensive study or review has not yet been written. In this thesis, the derived definition worked on existing definitions, as intended.

Furthermore, the literature study was done because different definitions and concepts are being used to describe the phenomenon. The definition of speech recognition in this study was good enough to serve the purpose of this thesis. However, a more comprehensive study is needed. The 34 articles and 5 books included in the study were primarily picked because of their relevance for this particular research topic. A comprehensive literature review should include material chosen with the specific purpose of that literature review. Not, as in this case, material focusing on the user value of speech recognition in the home environment.

7.3 User value of speech recognition

The MEC analysis concluded four user values of speech recognition used in the HVM, namely *facilitate daily life*, *everyday efficiency*, *comfort* and *increased calmness*. During the research, the value of *indifference* was also identified. In this section, all values will be discussed in relation to previous research and expert

interviews. It is aimed to discuss the answer to (RQ1) *What user value could be generated from speech recognition in the home environment for the broader audience?*

7.3.1 Facilitate daily life

First, the value of *facilitate daily life*, defined as *living an easier life with speech recognition to help in daily tasks* is discussed. The value is recognized by the experts, some thinking it is the main user value of the speech recognition systems existing today. The systems of tomorrow, however, will be able to transcend from setting a timer to enable creativity.

Previous research confirms the value for specific target groups. Elderly from several studies think speech recognition makes life easier and simplifies the use of technology. The same goes for disabled people. Looking at a broader audience, research suggests similar benefits, however less crucial.

Another aspect of facilitating daily life is the fact that the user benefits of speech recognition when not using their hands. This is confirmed by the connection of the consequence *physically engaged* with the value *facilitate daily life*, which is linked several times.

7.3.2 Everyday efficiency

The second value, *everyday efficiency* is explained as *the ability to complete a task faster than today and be more productive in the daily life*. One aspect of efficiency is speed. Research has suggested speech recognition to be faster than typed input since a long time. In expert interviews, productivity was mentioned as the number one benefit of the speech recognition systems as of today.

On the critical side, the value of *everyday efficiency* is questioned due to system inaccuracy causing frustration. However, the example referred to was writing long emails; a task more relevant to the workplace rather than the home environment.

7.3.3 Comfort

Thirdly, the value of *comfort*, explained as the *ability to feel relaxed and comfortable in the home environment* can be put in relation to previous studies which brings up concepts of comfort or relaxation. Derived from experts, this is especially true in systems handling a lot of information, such as a smart TV. Also, uncertainty about the consequences linking to comfort was expressed. A pilot study using speech recognition to help people with disabilities confirms comfort to be a likely benefit of speech recognition in the home environment.

Comfort was the most frequent value in the responses. However, experts express a hesitation concerning this value. One reason can be that the value is somewhat hard to understand by those deep into the technology of speech recognition. Another reason can be that respondents answer what they think in advance will be useful with speech recognition, without knowing that much of the experienced value as the experts know.

The opposition explained above does highlight the need for further examining this value, but also pinpoints that people feel the need for comfort because of their own laziness or need for relaxation.

7.3.4 Increased calmness

The fourth value, *increased calmness*, is described as the *ability to feel safe and secure in the home environment together with less concern and stress*. Previous research confirms security to be a promising field for speech recognition at home.

According to experts, speech recognition can empower the feeling of safety for two reasons; being in control of something outside your immediate reach and having something always attending to your needs, being able to help. However, more security, or the value of *increased calmness*, also means less privacy.

7.3.5 Indifference

All four values discussed above are a part of the HVM previously constructed. The value of *indifference* was, however, removed from the MEC analysis since answers connected to *indifference* did not appear important in the analysis. But no matter what, it is an important aspect to discuss.

Social norms play an important role into the adoption of speech recognition. Since speech recognition at home is a recent phenomenon, the technology acceptance and adoption are important. The research concludes that reaching a critical mass is an important factor for creating social norms. The critical mass is not at this stage clearly defined for speech recognition at home.

It is known that speech recognition in the home environment is not yet widely adopted and that it in many eyes is seen as something odd and unreliable. Several of the respondents in the study even mean that it is strange talking to a machine. Hence it is fairly clear that speech recognition still is several steps from being considered a social norm. The lack of social norm is probably the reason for the perceived indifference towards speech recognition at home. But, while the technology gets more adopted and in due time reaches a critical mass, the value of indifference will probably be less of an issue. Although, it shall be noted that it may affect speech recognition systems and implementers should pay attention to if

a critical mass is not reached. If not, it will be difficult for users to adopt to the technology, since it will still be seen as something odd and outside the social norm.

7.3.6 Summary of user value of speech recognition

Summarising, both experts and previous research mostly confirm the derived values in this study. It is, of course, not black and white, there are complications with all four values but *facilitate daily life* seem to be the one that almost everyone experiences as most likely. The key to a successful implementation of speech recognition in people's everyday life is that it reaches the critical mass to become a social norm and thereby bridge the value of indifference.

The discussion can later be compared to the discussion of next section, where the IKEA perspective is examined.

7.4 The IKEA perspective

This section discusses the answer to (RQ2) *Judging from a user perspective, how is speech recognition in the home environment relevant for IKEA?* by interpreting results from chapter 6 and 4. Further, the results are both justified and critically evaluated.

To start off, speech recognition in the home environment is relevant for IKEA. The results from the MEC analysis and the triangulation by experts were aligned and complemented with existing information from the Life at Home Report. These, results were compared with the IKEA Home Smart criteria for user value. Together, the different perspectives illustrate how user benefits of speech recognition are relevant for IKEA.

Perhaps, the major strength of the system's ability of being natural. Since an early age, humans around the world have practised a spoken interface. The technology is in our DNA and has enabled humans to build societies. That way, the technology is both convenient, easy to understand and suitable for the broad range of IKEA customers. The current ambivalent feeling towards technology described by the Life at Home Report could be improved with speech recognition as a social interface. The feeling of spending too much time with technology instead of nurturing relationships at home could be improved by adopting a social interface instead of a visual interface.

Speech recognition is not only to be seen as a product. Also, since IKEA Home Smart is not only a product range, it is a driving force of an ongoing project of digitalising IKEA with an ambition of moving into other product categories.

Speech recognition could enable that ambition and be seen as a tool to make digitalisation happen.

Going forward, the current view of IKEA Home Smart is generally aligned with the findings of this thesis. If speech recognition was to be considered further than what is done today, the current ideas of IKEA Home Smart could be kept with the addition of thinking about speech being able to generate safety. Björn Block (2017) stated speech recognition should be close to perfection to be adopted by IKEA. However, research suggests some of the benefits of speech recognition can be experienced even if the technology is of poor quality. And the demand exists, one-fourth of the respondents in the Life at Home Report states they are excited about speech recognition at home. Moreover, a perfect speech recognition system is a long way ahead. Speech recognition with a perfect understanding would not only interpret the meaning of words. It would understand the meaning behind tonalities, language styles, facial expressions and body movements. That would require data power and technologies existing in sci-fi movies at best.

Regardless whether IKEA goes deeper into speech recognition or not, the development is going forward. If adopting to speech recognition, this is likely to affect many products. Aside from the integrations and possibilities with the technology itself, the data generated could be used. It could create an understanding of life at home many times better than extensive research such as the Life at Home Report used in this study. In the long run, data would be fundamental to enable home automation as predicted by experts.

One of the challenges with speech recognition pointed out by all experts, is expectations. The expectations are rooted in human-human conversation but can be altered. Siri and ads for new speech recognition systems are heavily influencing the view of speech technology. If IKEA decides on moving forward with speech recognition, the first step could have a long-term effect. Therefore, it could be important to diverge from the use case centred product development approach of today. As was presented from the MEC analysis, a use case, or attribute, could lead to many different experiences. Complementing with a systematic product development approach could raise important points. Imagine speech recognition being preferred over a visual interface: *How will that affect the use of common spaces? Will less time be spent on social media? What will happen to the experience of making dinner?* Nevertheless, the use case perspective should still be kept, research even says successful smart home products cannot be done without designing for user needs.

Stepping into the future, there are exciting times ahead. Experts claim speech recognition will develop into understanding underlying messages, the tone of voice, gestures and even facial expressions. Unlike visual technology, a perfected speech recognition system is not that hard to imagine; it would be similar to human conversation.

To concretize the discussion about the user value of speech recognition and the

IKEA perspective, the next section provides some recommendations for IKEA on how to proceed.

7.5 Recommendations for IKEA

IKEA is recommended to go forward with speech recognition as it is relevant to their business and current strategy. The technology does not have to be perfected to generate user value. Rather, speech recognition is able to generate user value relevant to IKEA as of now.

According to the findings, the IKEA Home Smart product development criteria are aligned with the user value of speech recognition. Additionally, speech recognition is able to generate safety at home, something not being mentioned by IKEA Home Smart. However, safety is not to be considered a product development criterion but rather a product development area. If generating the user value of safety, privacy should be kept a priority.

7.5.1 Recommended points of action

- **Inform and educate**

  For IKEA to understand and possibly adopt to this new technology, relevant staff should be educated. Speech recognition is relevant for the entire home. Therefore, a seminar with people involving many different departments could be interesting.

- **Go deeper**

  Understanding what people would like to do with a speech recognition system is not enough. Instead, understanding the way it is being used is fundamental for further development.

- **Find your spot**

  There are many active players in the speech recognition market. Nevertheless, IKEA has the opportunity to find an offer not existing on the market, just as was done with the smart lighting system.

  Further, recommending speech recognition as a technology for IKEA does not necessarily mean developing the system itself. One scenario could be IKEA products communicating with other speech recognition devices as a part of the smart home.

7.6 Means-end chain discussed

The MEC approach is the primary methodology in this thesis, used to explore (RQ1) *What user value could be generated from speech recognition in the home environment for the broader audience?* In this section, five aspects of the validity of the MEC approach in the context of this study are discussed.

First of all, as described earlier 64 answers were collected in the pen and paper study, a loss of 10 was recognized because of severe shortcomings and absent answers. Hence 54 answers could be analysed. According to Reynolds & Gutman (1988), one-fourth of the answers contain only one ladder and three-fourths of the respondents contain two or three ladders. In this study 23 of 54 respondents could give an answer of three ladders, and 23 of 54 respondents gave answers containing two ladders. That is a total of 46 answers, roughly 85%, containing two or three ladders, this is somewhat higher than the typical level stated by Reynolds and Gutman (1988). Reasonably, the higher level of valid answers can be because of the inspirational poster, or by an allowing coding process. Despite the helping poster and the coding process, it is reasonable to state that the amount of answers provides sufficient support for the derived values.

Secondly, in contrast to previous MEC research (Jung and Kang, 2010; Jung, 2014; Deng and Christodoulidou, 2015) this study examines an IT phenomenon that has not been utilised by the respondents. A superior majority of the respondents, 86%, does not use speech recognition in their home environment, but almost everyone could understand the phenomenon when explained to them. This means that this study measures what people think they want. This was handled by the triangulation of the MEC results by experts and the comparison with previous research, which found those results credible. Hence, this study claims MEC analysis can explore future products as long as the product can be explained and is easy to understand for the users. However, to confirm this ability the results should be followed up when or if the speech recognition in the home environment has been more widely accepted.

Thirdly, the information provided the respondents may not be without complications. The inspirational poster (Appendix A.2), might have had an influence on the answers. Respondents were able to pick a use case from the poster without wanting that function. When examining all use cases given in the answers 12% were very similar (e.g. find my phone), 51% somewhat similar (e.g. turning on/off the light) and 37% different (e.g. opening the gate) to an example in the inspirational poster. It is hard to determine the impact of the inspirational poster, it was perceived some respondents read through the entire information whilst many only had a glance. Nils Lenke (2017) says people are receptive to suggestion when it comes to speech recognition. The first thing people try on their new speech recognition system is what they have seen at others do. On the other hand, it is reasonable to claim that the nudging in this pen and paper study is fully

comparable with a standard deep interview that takes 60 to 75 minutes. This because in a deep interview the conversation is carried forward by natural discussion and guiding questions.

Lastly, there are limitations with the data gathering in the MEC approach. The respondents were limited in both numbers and demographics. Data was collected at an IKEA store in the south of Sweden for one day, thus the respondent population was limited. The survey was handed out in the lighting department, this might have affected the answers to reach the most common attribute as lighting. The collected data was limited in such way, that all respondents, except for one, were Swedish.

7.7 Academic contribution

The academic contribution of this thesis, apart from the purpose, is twofold. First, the research contributes to a novel perspective on the term and definition of speech recognition. Second, this thesis claims that a technology not currently being used by the respondents can be subject to the MEC approach.

7.8 Further research

The concept of user value of speech recognition is still fairly new in the academy and further research is needed. Those aspects are discussed below.

Firstly, an academic contribution of this study is to introduce a digested definition of speech recognition. Further research can implement that definition in order to understand the widely used concept of speech recognition. Since this study has a limited scope of speech recognition there is room for alterations of the definition to fit other specific areas of research. There is also room for comparing the definition to other terms, such as voice control and voice command, to derive a more novel definition.

Secondly, it would be interesting to complement the study by using the MEC approach on people having experience with speech recognition at home. This could be done as of now if carefully selecting respondents using the technology at home. Outside the smart home, the MEC approach could evaluate the existing speech recognition services, such as Siri and Google Voice. An interesting MEC method for this is used by Jung (2014), to study the user values of smartphones.

Thirdly, variations of the MEC study is likely to provide results complementing this study. Although the pen and paper approach was perceived as adequate, it was noticed that the respondents were more detailed in speech than their written

answers. Thus, it is reasonable to claim that in-depth interviews could give more truthful and useful answers. Interviews, in turn, could lead to more distinguishable ladders and consequently more developed chains to use in the HVM. To more deeply explore the attributes and consequences of speech recognition this study could be complemented by deeper in-depth laddering interviews proposed by Reynolds and Gutman (1988). Further, it would be interesting to perform a study with an adequate representation of the *broader audience*. This could mean performing the study in different countries and having a larger number of respondents.

Lastly, the implementation of speech recognition requires an understanding of the technology and market. This thesis does not explore the implementation of speech recognition. To provide a more specific recommendation for IKEA, a study investigating the attributes given in this thesis could be analysed and further developed.

8 Conclusions

This study was carried out with the MEC approach as the primary approach. It was complemented with literature reviews and qualitative interviews. The study yielded two main findings in relation to the purpose of exploring the user value of speech recognition in the home environment for the broader audience.

Firstly, the most imperative finding: the user values enabled by speech recognition was concluded to be *facilitate daily life*, *everyday efficiency*, *comfort* and *increased calmness*. *Facilitate daily* life refers to the user's ability of living an easier life with speech recognition as relevant aid in daily tasks. *Everyday efficiency* indicates the user's ability to complete a task faster than today and to be more productive in the daily life. *Comfort* refers to the ability to feel relaxed and comfortable in the home environment. *Increased calmness* refers to the ability to feel safe and secure in the home environment and feel less concern or stress.

A denominating factor of all these values is that they are making existing attributes easier compared to the current interface. Therefore, speech recognition is seen as an interface not replacing, but complementing or improving current interfaces. Hence, speech recognition represents an evolution rather than a revolution in the home environment for the broader audience. The evolution is, so far, received with indifference by several users. To overcome that obstacle the technology must reach a critical mass of adopters and thereby get socially accepted.

Secondly, speech recognition is found to entirely or partially fulfil the product development criteria for smart products at IKEA, namely: *convenience*, *easy to understand, clear use case* and *solving a user need*. Thereby, speech recognition is likely to be a relevant product for IKEA in addition to generating user value for the IKEA customer. This user value could be generated with existing technology, the system does not have to reach perfection.

Hence, from what has been found, there are several arguments for speech recognition enabling to the IKEA vision of *creating a better life for the many people*. Therefore, IKEA is recommended to further explore speech recognition.

References

Aldrich, F.K. 2003. Smart Homes: Past, Present and Future *In*: *Harper Richard (ed.) Inside the Smart Home*. Springer-Verlag, pp. 17–39.

Anusuya, M.. and Katti, S.. 2009. Speech Recognition by Machine: A Review. *(IJCSIS) International Journal of Computer Science and Information Security*. **6**(3),pp.181–205.

Arriany, A. and Musbah, M. 2016. Applying Voice Recognition Technology for Smart Home Networks *In*: *2016 International Conference on Engineering MIS (ICEMIS)*. Agadir, pp. 1–6.

B. Klenosky, D. 2002. The "Pull" of Tourism Destinations: A Means-End Investigation. *Journal of Travel Research*. **40**(4),pp.385–395.

Baig, F., Beg, S. and Khan, M.F. 2012. Controlling Home Appliances Remotely through Voice Command. *International Journal of Computer Applications*. **48**(17),pp.975 – 888.

Bala, A., Kumar, A. and Birla, N. 2010. Voice command recognition system based on MFCC and DTW. *International Journal of Engineering Science and Technology*. **2**(12),pp.7335–7342.

Beigi, H. 2011. *Fundamentals of Speaker Recognition* 1st ed. New York: Springer US.

Bhuiyan, J. 2016. Mary Meeker: Voice and image searches are going to make up at least half of all searches by 2020. *Recode*. [Online]. [Accessed 6 September 2017]. Available from: https://www.recode.net/2016/6/1/11829844/mary-meeker-voice-image-searches-2020.

Block, B. 2015. Internet of things the IKEA way [PowerPoint slides], Inter IKEA Systems B.V. 2015, Älmhult, Sweden.

Block, B. 2017. Telephone interview with Björn Block, Business Leader IKEA Home Smart (2017-10-10).

Blomkvist, P. and Hallin, A. 2015. *Method for engineering students. Degree project using the 4-phase Model* 1:1. Lund, Sweden: Studentlitteratur AB.

Botschen, G. and Hemetsberger, A. 1998. Diagnosing means-end structures to determine the degree of potential marketing program standardization.

Journal of Business Research. **42**(2),pp.151–159.

Bryman, A. and Bell, E. 2015. *Business Research Methods* 4th ed. Oxford University Press, USA.

Busatlic, B., Dogru, N., Lera, I. and Sukic, E. 2017. Smart Homes with Voice Activated Systems for Disabled People. *TEM Journal*. **6**(1),pp.103–107.

Cordasco, G., Esposito, M., Masucci, F., Riviello, M.T., Esposito, A., Cholletz, G., Schlögl, S., Milhoratz, P. and Pelosi, G. 2014. Assessing Voice User Interfaces: The vassist system prototype *In*: *2014 5th IEEE Conference on Cognitive Infocommunications (CogInfoCom)*. Vietri sul Mare, pp. 91–96.

Costa, A. de A., Dekker, M. and Jongen, W. 2004. An overview of means-end theory: potential application in consumer-oriented food product design. *Trends in Food Science & Technology*. **15**(7),pp.403–415.

Deng, X.N. and Christodoulidou, N. 2015. Understanding User Values of Wearable Computing *In*: *2015 Thirty Sixth International Conference on Information Systems*. Fort Worth.

Fleury, A., Vacher, M., Noury, N. and Portet, F. 2011. Development of Audio Sensing Technology for Ambient Assisted Living: Applications and Challenges. *International Journal of E-Health and Medical Communications*. **2**(1),pp.35–54.

Friedman, B., Kahn, P. and Borning, A. 2013. Value Sensitive Design and Information Systems *In*: *Forthcoming in P. Zhang & D. Galletta (Eds.) Human-Computer Interaction in Management Information Systems: Foundations*. NY: M.E. Sharpe.

Gemmeke, J., Ons, B., Tessema, N., Van hamme, H., van de Loo, J., De Pauw, G., Daelemans, W., Huyghe, J., Derboven, J., Vuegen, L., Van Den Broeck, B., Karsmakers, P. and Vanrumste, B. 2013. Self-taught assistive vocal interfaces: An overview of the ALADIN project *In*: *Proceedings Interspeech 2013*. Lyon, France: ISCA, pp. 2038–2043.

Guba, E.G. and Lincoln, Y.S. 1994. Competing paradigms in qualitative research. *Handbook of qualitative research*. **2**(163–194),p.105.

Gutman, J. 1982. A Means-End Chain Model Based on Consumer Categorization Processes. *Journal of Marketing*. **Vol. 46**(2),pp.60–72.

Hammersley, J. 2016. Ikea – are they struggling to build their e-commerce execution? [Accessed 7 September 2017]. Available from: https://www.linkedin.com/pulse/ikea-struggling-build-e-commerce-execution-james-hammersley.

Heggeman, H. 1999. Andelen äldre fortsätter öka. *VälfärdsBulletinen, Statistiska centralbyrån (SCB)*. (1),p.29.

Horstmann, K.H. 2010. User Performance With Speech Recognition: A Literature Review. *Assistive Technology*. **13**(2),pp.116–130.

House, D. 2017. Expert interview with David House (2017-11-14), Professor, Dept. of Speech, Music and Hearing, School of Computer Science and Communication, KTH (Royal Institute of Technology), Stockholm, Sweden.

Inter IKEA Systems B.V. 2017a. About the IKEA group. *Welcome inside our company*. [Online]. [Accessed 12 October 2017]. Available from: http://www.ikea.com/ms/en_US/this-is-ikea/company-information/index.html.

Inter IKEA Systems B.V. 2017b. Control IKEA Home Smart products with Amazon Alexa, Google Assistant and Apple's Home app. *IKEA newsroom*. [Online]. [Accessed 10 November 2017]. Available from: http://newsroom.inter.ikea.com/news/control-ikea-home-smart-products-with-amazon-alexa--google-assistant-and-apple-s-home-app/s/b81aae67-67bd-4dd8-bfd7-3b4ef4f48888.

Inter IKEA Systems B.V. 2017c. Furniture & Home Furnishings | IKEA United Kingdom. [Accessed 23 November 2017]. Available from: http://www.ikea.com/gb/en/.

Inter IKEA Systems, B.V. 2015. *Life at Home Report #3 What makes a home* [Online]. [Accessed 13 October 2017]. Available from: https://ikeamuseum.com/media/filer_public/9a/94/9a94647a-d356-4a6b-9803-3fc6a4ca0034/ikea_life_at_home_report_3.pdf.

Inter IKEA Systems, B.V. 2017. *Life at Home Report #4 Beating the Battles* [Online]. IKEA. [Accessed 8 November 2017]. Available from: http://lifeathome.ikea.com/home/images/LIFE_AT_HOME_2017.pdf.

Ishiguro, H. 2017. Expert interview with Hiroshi Ishiguro (2017-11-14), Director and professor at the Intelligent Robotics Laboratory, Department of Systems Innovation, Graduate School of Engineering Science, Osaka University, Japan.

Jesson, J.K., Matheson, L. and Lacey, F.M. 2012. *Doing Your Literature Review – Traditional and Systematic Techniques*. London: SAGE Publications Inc.

Jung, Y. 2014. What a smartphone is to me: understanding user values in using smartphones. *Information Systems Journal*. **24**(4),pp.299–321.

Jung, Y. and Kang, H. 2010. User goals in social virtual worlds: a means-end chain approach. *Computers in Human Behavior*. **26**(2),pp.218–225.

Jurafsky, D. and Martin, J. 2000. *Speech and Language Processing* Draft of September 28, 1999. Englewood Cliffs, New Jersey 07632: Prentice-Hall.

Kamm, C. and Helander, M. 1997. Design issues for interfaces using voice input *In*: *Design Issues for Interfaces using Voice Input*. Handbook of Human-Computer Interaction. Elsevier Science B.V, pp. 1043–1059.

Kim, D. and Lee, H. 2016. Effects of user experience on user resistance to change to the voice user interface of an in-vehicle infotainment system: Implications forplatform and standards competition. *International Journal of Information Management*. **36**(4),pp.653–667.

Kolcaba, K.Y. and Kolcaba, R.J. 1991. An analysis of the concept of comfort. *Journal of Advanced Nursing*. **16**(11),pp.1301–1310.

Kontogiorgos, D. 2017. Expert interview with Dimosthenis Kontogiorgos (2017-11-15), PhD Student in Social Robotics, Dept. of Speech, Music and Hearing, School of Computer Science and Communication, KTH (Royal Institute of Technology), Stockholm, Sweden.

Koskela, T. and Väänänen-Vainio-Mattila, K. 2004. Evolution towards smart home environments: empirical evaluation of three user interfaces. *Personal and Ubiquitous Computing*. **8**(3–4),pp.234–240.

Kruchinina, E. 2017. Expert interview with Ekaterina Kruchinina (2017-11-17), Research Manager, Natural Language Understanding Department, Nuance Communications.

Lai, J. 2000. Conversational Interfaces. *Communications of the ACM*. **43**(9),pp.24–27.

LeCompte, M.D. and Goetz, J.P. 1982. Problems of Reliability and Validity in Ethnographic Research. *Review of Educational Research*. **52**(1),pp.31–60.

Lekvall, P. and Wahlbin, C. 2011. *Information för marknadsföringsbeslut* 4th ed. Studentlitteratur AB.

Lenke, N. 2017. Expert interview with Nils Lenke (2017-11-15), Senior Director, Corporate Research, Nuance Communications.

Martin, G. 1989. The utility of speech input in user-computer interfaces. *International Journal of Man-Machine Studies*. **30**(4),pp.355–375.

McKinsey&Company 2016. There's No Place Like [a Connected] Home. *Connected Homes*. [Online]. [Accessed 27 October 2017]. Available from: https://www.mckinsey.com/spContent/connected_homes/index.html.

Meisel, W. 2010. "Life on-the-Go": The Role of Speech Technology in Mobile Applications *In*: *Amy Neustein (ed.) Advances in Speech Recognition*. Springer, pp. 3–18.

Mittal, Y., Toshniwal, Paridhi, Sharma, Sonal, Singhal, Deepika, Gupta, Ruchi and Mittal, V. K. 2015. A voice-controlled multi-functional Smart Home Automation System *In*: *2015 Annual IEEE India Conference (INDICON)*.

New Dehli, pp. 1–6.

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

Neuman, W.L. 2003. *Social Research Methods. Qualitative and Quantitative Approaches* 5th ed. USA: Pearson Education, Inc.

Newman, D. 2000. *Talk to your computer, speech recognition made easy*. Waveside Publishing.

Nuance Communications 2017. Fast Facts | Nuance. [Accessed 3 January 2018]. Available from: https://www.nuance.com/about-us/fast-facts.html.

Obaid, T., Rashed, H., Abu El Nour, A., Rehan, M., Muhammad Saleh, M. and Tarique, M. 2014. Zigbee based voice controlled wireless smart home system. *International Journal of Wireless & Mobile Networks (IJWMN)*. **6**(1),pp.47–59.

Portet, F., Vacher, M., Golanski, C., Roux, C. and Meillon, B. 2011. Design and evaluation of a smart home voice interface for the elderly: acceptability and objection aspects. *Personal and Ubiquitous Computing*. **17**(1),pp.127–144.

Rabiner, L. and Juang, B.-H. 1993. *Fundamentals of Speech Recognition*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc.

Reynolds, T.J. and Gutman, J. 1988. Laddering Theory, Method, Analysis and Interpretation. *Journal of Advertising Research*. **28**(1),pp.11–31.

Rhordi, M. 2016. Has voice control finally started speaking our language? *The Guardian*. [Online]. [Accessed 1 September 2017]. Available from: https://www.theguardian.com/technology/2016/dec/04/voice-control-amazon-echo-digital.

Schoutsen, P. 2017. IKEA Trådfri: Internet of Things done right. *Home Assistant*. [Online]. [Accessed 8 September 2017]. Available from: https://home-assistant.io/blog/2017/04/17/ikea-tradfri-internet-of-things-done-right/.

Shagoury, J. 2010. Dr. "Multi-Task": Using Speech to Build Up Electronic Medical Records While Caring for Patients *In*: *Amy Neustein (ed.) Advances in Speech Recognition*. Springer, pp. 247–274.

Simon, S.J. and Paper, D. 2007. User Acceptance of Voice Recognition Technology: *Journal of Organizational and End User Computing*. **19**(1),pp.24–50.

Simpson, R. and Levine, S. 2002. Voice Control of a Powered Wheelchair. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. **10**(2),pp.122–125.

Soda, S. 2012. Implementing Virtual Agent as an Interface for Smart Home Voice Control *In*: *2012 19th Asia-Pacific Software Engineering Conference*. Hong Kong, pp. 342–345.

de Souza Leão, A.L. and Mello Benício de, S. 2007. The Means-End Approach to Understanding Customer Values of a On-Line Newspaper. *BAR-Brazilian Administration Review*. **4**(1),pp.1–20.

Statista, Inc. 2017. IKEA - Statistics & Facts. *Statista*. [Online]. [Accessed 13 October 2017]. Available from: https://www.statista.com/topics/1961/ikea/.

Teddlie, C. and Tashakkori, A. 2009. *Foundations of mixed methods research: integrating quantitative and qualitative approaches in the social and behavioral sciences*. Los Angeles, California: SAGE.

Töreman, R. 2017a. E-mail correspondence with Rebecca Töreman (2017-10-17), product developer at IKEA Home Smart. Subject: Feedback on survey.

Töreman, R. 2017b. E-mail correspondence with Rebecca Töreman (2017-12-17), product developer at IKEA Home Smart. Subject: Selected parts as draft.

Töreman, R. 2017c. E-mail correspondence with Rebecca Töreman (2017-12-23), product developer at IKEA Home Smart. Subject: Feedback on the chapter about IKEA.

Töreman, R. 2017d. Interview with Rebecca Töreman (2017-09-25), product developer at IKEA Home Smart. Subject: Understanding IKEA Home Smart.

Töreman, R. 2017e. Interview with Rebecca Töreman (2017-10-05), product developer at IKEA Home Smart. Subject: Deep dive in IKEA Home Smart.

Töreman, R. 2017f. Telephone interview with Rebecca Töreman (2017-11-09), product developer at IKEA Home Smart. Subject: Feedback on results and update on TRÅDFRI and speech recognition.

Vacher, M., Lecouteux, B., Romero, J.S., Ajili, M., Portet, F. and Rossato, S. 2015. Speech and Speaker Recognition for Home Automation: Preliminary Results *In*: *2015 International Conference on Speech Technology and Human-Computer Dialogue (SpeD)*. Bucharest, pp. 1–10.

voicebot 2017. Amazon Echo & Alexa Stats. [Accessed 5 September 2017]. Available from: https://www.voicebot.ai/amazon-echo-alexa-stats/.

Walker, B. and Olson, J. 1991. Means-End Chains: Connecting Products With Self. *Journal of Business Research*. **22**(2),pp.111–118.

Watts, N. 2015. IKEA And Out Of The Flat Pack Thinking. [Accessed 7 September 2017]. Available from: http://www.ogilvy.com/topics/tech-

innovation/ikea-and-out-of-the-flat-pack-thinking/.

Zhang, M., Lv, X. and Li, H. 2008. Robot Control Based on Voice Command *In*: *2008 IEEE International Conference on Automation and Logistics*. Qingdao, China, pp. 2490–2494.

Appendix A Survey material

This appendix describes how the survey was carried out by presenting what was presented to the respondents.

Notes on the survey and inspirational poster

- Survey was available in both Swedish and English
- All respondents, except for one, used the Swedish version
- At the time of submitting the survey, the literature review had not been completed. Therefore, voice control was used instead of speech recognition.

A.1 Survey

Survey about voice control and the smart home

To learn more about how you use or could use smart products and voice control.

The survey takes about 5 minutes and is entirely voluntary.

Albert Johansson and Maria Blomberg are responsible for this study.

sve10ajo@student.lu.se
+46 - 767 - 19 78 71

- The result may be used as input for future product development at IKEA.
- The purpose is to understand how voice control can generate customer value.
- Your submitted answers are entirely anonymous and will only be used in connection to this study.

This survey is a part of a Master Thesis at Lund University (LTH) in collaboration with IKEA.

IKEA

LUND UNIVERSITY

please, turn →

1. Do you (or someone else in your home) use voice control today?

☐ a) Yes ☐ b) No ☐ c) Don't know

2. *If No/Don't know*: **What would you (or someone else in your home) use voice control for in your home?**
If Yes: **What do you (or someone else in your home) preferably use voice control for in your home?**

3. Why do you (or someone else in your home) want to use voice control in the manner you described in previous question? Please explain.

4. Regarding the reasons you provided in the question above, why are those reasons important to you (or someone else in your home)? Please explain.

5. Do you (or someone else in your home) have an interest in using more voice control in your home?

☐ a) Yes ☐ b) No ☐ c) Maybe ☐ d) Don't know

6. How old are you?

☐ a) Under 18 years ☐ b) 18-24 years ☐ c) 25-34 years ☐ d) 35-44 years ☐ e) 45-54 years ☐ f) 55-64 years ☐ g) 65 years or older ☐ h) Don't want to answer

7. I primarily identify my gender as…

☐ a) Female ☐ b) Male ☐ c) Don't want to answer ☐ d) Other:_____

8. Which of the following are applicable to your living situation?

Check all that apply

☐ a) Live alone ☐ b) Live with relative(s) ☐ c) Live with spouse /partner(s) ☐ d) Live with friend(s) ☐ e) Live with child(ren) ☐ f) Other:_____

9. Which of the following best describe how you use voice control today? (both inside and outside the home).

Examples: Siri, Google Voice Assistant or Cortana on your smartphone, smart TV or other.

☐ a) Every day ☐ b) Every week ☐ c) Every month ☐ d) More seldom/never

please, turn →

10. Which of the following smart home products exist in your home today?

Check all that apply. Please note, products presented in the images are only examples of products within that category.

☐ a) Smart doorlock

☐ b) Smart lighting

☐ c) Smart socket

☐ d) Security camera

☐ e) Smart sensor

☐ f) Smart thermostat

☐ g) Smart TV och media player

☐ h) Smart speaker and surroud sound

☐ i) Robotic vacumcleaner/ Robotic lawn mower

☐ j) Smart appliences

☐ k) Voice command device

☐ l) None

☐ m) Other:_____

Thanks for your participation

A.2 Inspirational poster

A.3 Prompts

In this section, the unstructured interview guide for the MEC survey is presented. Conversations were generally made in Swedish.

Introduction

Interviewees: "Hi, we are two master students from Lund University. Would you like to take part in a survey about speech recognition? It takes about five minutes and is completely anonymous."

If people showed interest: Interviewer: "Do you know about speech recognition?"

If yes: Interviewer: "This survey is to understand how speech recognition could be used by IKEA, do you want to have input on future product development at IKEA?"

If no: Interviewer: "Speech recognition is a computer understanding what you say. Do you know Siri? Siri is speech recognition."

Introduction for respondents taking the survey:

Interviewer: "This is the survey, it takes about five minutes to answer. Keep in mind, no answer is wrong or right, just answer what comes to mind. If you want inspiration on what to answer, have a look at the inspirational poster. Don't hesitate to ask questions."

Helping questions (if respondent had a hard time answering, derived from previous MEC research)

- Negative laddering: Why would you not want speech recognition?
- Differences by occasion, placing the respondents in a personal context: Have you been in a situation when you were unable to interact with things the way you use to?
- Evoking situational context: Imagine your Livingroom, what would you normally do in that room? Could speech recognition help do that activity?
- Postulating the absence of an object or a state of being: If you are doing the dishes and want to change songs on Spotify, how would you do that without speech recognition.
- Age-regression contrast probe: Is there a difference on how you monitored your TV or radio 10 years ago, compared to today?
- 3rd person probe: Why do you think my mother use speech recognition when using her smart phone?
- Redirecting techniques: Silence or communication check.

Appendix B Demographics

In this appendix, the compiled demographic information from the respondents of the MEC survey is presented.

In Table B.1 all demographic information from the MEC data gathering is presented. It is a detailed version of Table 5.1.

Table B.1 Demographic information of participants in the MEC data gathering.

		Frequency	***Per cent***
Age (years)	18 or below	1	1.9
	18-24	7	13.0
	25-34	12	22.2
	35-44	4	7.4
	45-54	5	9.3
	55-64	7	13.0
	65 or older	18	33.3
	No answer	0	0.0
Gender	Female	31	57.4
	Male	23	42.6
	Other	0	0.0
	No answer	0	0.0
Living situation (multiple choice)	Alone	13	24.1
	With relative(s)	3	5.6
	With spouse/partner(s)	32	59.3
	With friend(s)	1	1.9
	With child(ren)	10	18.5
	Other	1	1.9
Using voice control at	Yes	9	17.0

home today [3]	No	44	83.0
	Don't know	0	0.0
Interest in using voice control	Yes	27	50.9
	No	4	7.5
	Maybe	20	37.7
	Don't know	2	3.8
Usage of voice control	Every day	6	11.5
	Every week	9	17.3
	Every month	4	7.7
	Seldom or never	33	63.5
Smart home products in your home (multiple choice)	Door lock	3	5.6
	Lighting	11	20.4
	Socket	3	5.6
	Security camera	10	18.5
	Sensor	1	1.9
	Thermostat	2	3.7
	TV or media player	22	40.7
	Speaker and surround sound	6	11.1
	Robotic vacuum cleaner or lawn mower	2	3.7
	Appliances	1	1.9
	Voice command device	0	0.0
	None	14	25.9
	Other	4	7.4

For clarity and transparency, the demographic information of the respondents who provided answers which were disqualified is presented in Table B.2

Table B.2 Demographic information of participants with disqualified answers in the MEC data gathering.

		Frequency	***Per cent***
Age (years)	18 or below	0	0.0
	18-24	1	12.5
	25-34	1	12.5

[3] Excluded from research due to inadequate question. When half of the respondents had answered, the authors realised the question could be interpreted as using any kind of speech recognition at home - including smart phones. The question was intended to ask whether respondents used a speech recognition device specifically designed for the home environment. The remaining respondents received the first question together with a handwritten note about the smart phone not being included.

	35-44	1	12.5
	45-54	2	25.0
	55-64	1	12.5
	65 or older	2	25.0
	No answer	0	0.0
Gender	Female	6	75.0
	Male	2	25.0
	Other	0	0.0
	No answer	0	0.0
Living situation (multiple choice)	Alone	2	20.0
	With relative(s)	0	0.0
	With spouse/partner(s)	6	60.0
	With friend(s)	0	0.0
	With child(ren)	2	20.0
	Other	0	0.0
Using voice control at home today	Yes	0	0.0
	No	10	100.0
	Don’t know	0	0.0
Interest in using voice control	Yes	0	0.0
	No	3	37.5
	Maybe	3	37.5
	Don’t know	2	25.0
Usage of voice control	Every day	0	0.0
	Every week	1	14.3
	Every month	0	0.0
	Seldom or never	6	85.7
Smart home products in your home (multiple choice)	Door lock	0	0.0
	Lighting	2	20.0
	Socket	1	10.0
	Security camera	0	0.0
	Sensor	0	0.0
	Thermostat	0	0.0
	TV or media player	2	20.0
	Speaker and surround sound	1	10.0
	Robotic vacuum cleaner or lawn mower	0	0.0
	Appliances	0	0.0
	Voice command device	0	0.0
	None	3	30.0
	Other	0	0.0

Appendix C IKEA interview guides

C.1 Interview with Rebecca Töreman (2017-09-25)

Type: Unstructured interview

Role: Product developer at IKEA Home Smart.

Subject: Understanding IKEA Home Smart

Language: Swedish

Referred to as: Töreman, 2017d

Understanding the foundation of IKEA Home Smart with questions such as:

- What does IKEA Home Smart do?
- How does IKEA Home Smart work?
- How does IKEA Home Smart think when developing new projects?
- What does the IKEA Home Smart strategy or business plan look like?
- What are the guiding principles of IKEA Home Smart?

Getting to know more about possible scopes of the thesis by asking questions such as:

- What relevant data is available?
- What in the area of speech recognition is the most interesting for you to write about?
- Would it be possible to do user testing?

C.2 Interview with Rebecca Töreman (2017-10-05)

Type: Unstructured interview

Role: Product developer at IKEA Home Smart.

Subject: Deep dive in IKEA Home Smart

Language: Swedish

Referred to as: Töreman, 2017e

Going through expectations with questions such as:

- What kind of insights would you wish that we present on the day of presentation?
- How would you prefer the material to be delivered?
- What do you prefer to get out of this?

Going through the limitations and possibilities of doing an academic thesis

Understanding how IKEA works with users with questions such as:

- How do you define user value?

C.3 Interview with Björn Block (2017-10-10)

Type: Unstructured telephone interview

Role: Business Leader IKEA Home Smart

Language: Swedish

Referred to as: Block, 2017

Introduction

> *Presenting ourselves and the project. Presenting the purpose of the interview which was to get another perspective on IKEA Home Smart.*

IKEA Home Smart

- How did the IKEA Home Smart initiative start?
- What are you doing to understand user value at IKEA Home Smart?
- Have you identified any success factors from earlier product development? If yes, what are they?

- What do you think about the future for IKEA Home Smart?

Speech recognition

- What would make speech recognition more relevant to IKEA?
- What is your view on speech recognition in the years to come?
- How do you think about speech recognition in a longer perspective?

End

- Is there anything you would like to add?

Informing about the possibility to get in touch with us via email. Thanking for the interview

C.4 Interview with Rebecca Töreman (2017-11-09)

Type: Semi-structured telephone interview (expert interview)

Role: Product developer at IKEA Home Smart.

Subject: Feedback on results and update on TRÅDFRI and speech recognition

Language: Swedish

Referred to as: Töreman, 2017f

Introduction

Update on the project. Informing about the expert interviews. Informing about the main purpose of this interview is getting a reaction from IKEA on the results.

- Do you want to proof read the material about IKEA?

Result

The HVM is described and showed in a pdf being sent out prior to the interview.

- Do you have any spontaneous thoughts on the results?
- Does this result reflect the impression of the IKEA customer today?

- Is there something surprising?
- Is this information relevant to IKEA?

Other questions

- How is user value defined at IKEA Home Smart?
- How were the personas created? Have they in any way been made with speech recognition in mind?

TRÅDFRI and speech recognition – getting to know the feedback from the recently released speech recognition integration between TRÅDFRI and speech recognition systems provided by e.g. Amazon and Apple.

- How did the initial release turn out?
- What feedback have you received?
- Is there more feedback coming?

C.5 E-mail correspondence with Rebecca Töreman (2017-10-17)

Type: E-mail

Role: Product developer at IKEA Home Smart.

Subject: Feedback on survey

Language: Swedish

Referred to as: Töreman, 2017a

Asking for feedback on an online draft survey. At the time of writing, the survey can be found on the following link: https://voicesurvey.typeform.com/to/xOxWPw

C.6 E-mail correspondence with Rebecca Töreman (2017-12-17)

Type: E-mail

Role: Product developer at IKEA Home Smart.

Subject: Question about convenience

Language: Swedish

Referred to as: Töreman, 2017b

What does convenience mean for IKEA Home Smart? What does convenience mean from a user perspective? [4]

C.7 E-mail correspondence with Rebecca Töreman (2017-12-23)

Type: E-mail

Role: Product developer at IKEA Home Smart.

Subject: Feedback on the chapter about IKEA

Language: Swedish

Referred to as: Töreman, 2017c

Feedback

A draft of chapter 4 was sent for fact checking.

- Are the facts true?
- Do you agree with the description that is given?
- Do you think there is relevant information missing in the text?
- Are there any other comments you would like to make?

Feedback was given by commenting the pdf document and replying to the email.

[4] Answers from this question were excluded when counting the number of mentions in the product development criteria in Chapter 4. Asking a specific question about one criterion was considered to increase the number of mentions in a non-credible way.

Appendix D Expert interview guide

All experts had personal interview guides with variations in personal questions. Documents are available upon request. This section presents the general interview guide.

Experts

- David House – Professor, Dept. of Speech, Music and Hearing, School of Computer Science and Communication, KTH (Royal Institute of Technology), Stockholm, Sweden
- Hiroshi Ishiguro – Director and professor at the Intelligent Robotics Laboratory, Department of Systems Innovation, Graduate School of Engineering Science, Osaka University, Japan
- Dimosthenis Kontogiorgos – PhD Student in Social Robotics, Dept. of Speech, Music and Hearing, School of Computer Science and Communication, KTH (Royal Institute of Technology), Stockholm, Sweden
- Ekatarina Kruchinina – Research Manager, Natural Language Understanding Department, Nuance Communications
- Nils Lenke – Senior Director, Corporate Research, Nuance Communications

Introduction

The expert was thanked for participating in the study and informed of the length being 10 minutes. The expert was also asked if giving consent to the interview being recorded.

Before asking questions, the expert was asked if he/she had read the summary being send out prior to the conversation. Depending on how much the expert knew about the project, a description was given.

General questions

- Do you have any general reaction to our results?
- Going through our four values one-by-one, what do you think of them? Is there something you recognise or something unfamiliar? *Facilitate daily life*, *everyday efficiency*, *comfort* and *increased calmness*.

- What do you believe is the user value of speech recognition in the home?
- In your experience with speech; does the expected user value match the experienced user value?

Closing statement

The expert was thanked again for participating in the study. They were asked if they wanted to proof read the material from the interview being used in the thesis. If any further question or comments, they were asked to send an email.

Appendix E Time plan

This appendix presents the planned project plan as a timeline of the performed activities.

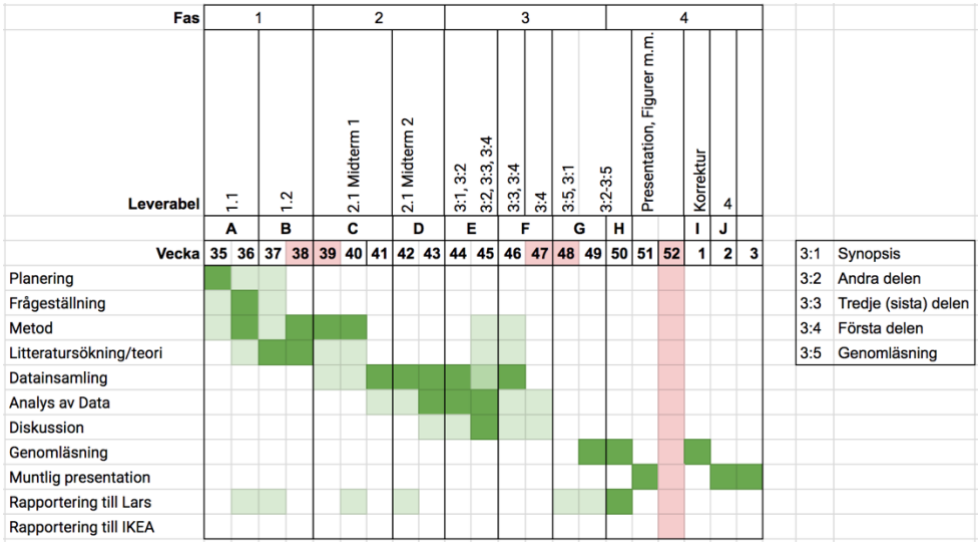

Figure C.1 Time plan of the work process, presented in Swedish.