What are the possibilities of integrating task based speech recognition into work processes?

Vilka möjligheter finns det till att integrera arbetsuppgiftorienterad röststyrning i arbetsprocesser?

ROBERT ANDERSSON
TAB PARR

Bachelor of Informatics Thesis

Abstract

Speech recognition (SR) technologies are on the rise and the use of SR in warehouses is the focus for the thesis. The research question is "*What are the possibilities of integrating task based speech recognition into work processes?*" The study collects qualitative data through an ethnographic field study, interviews and observations for assessing the requirements, and thereafter produces a prototype solution based on the ethnographic findings. In this thesis the authors get a glimpse through interviews and hands on trial of a SR system in a warehouse environment, conducting a field study inspired by ethnography. This is used to obtain a deeper knowledge of how SR is used in the warehouse environment. The knowledge and collected data is thematically analyzed to find relationships in the data and to help develop the requirements for a mobile SR solution. An inventory of speech recognition platforms, which is used to compare and evaluate alternative solutions. These alternatives provide a better understanding of how SR can be integrated into mobile devices and platforms to support a voice enabled business process. The results show that there are possibilities for the integration of SR into existing work processes but there are also barriers.

Table 1 list of terminologies that are used throughout the essay

Acoustic Model	AM
Artificial Neural Networks	ANN
Automatic Speech Recognition	ASR
Business Activity Monitoring	BAM
Business Process Modeling	BPM
Business Process Modeling Notation	BPMN
Business Process Simulation	BPS
Dynamic time warping	DTW
Hidden Markov Model	HMM
Human Computer Interaction	HCI
Input and output	IO
Linear Predictive Coding	LPC
Linguistic Model	LM
Natural User Interface	NUI
Service Oriented Architecture	SOA
Soft Systems Methodology	SSM
Speech Recognition	SR
Speech Synthesis	SS
Speech User Interface	SUI
Voice Control	VC
Voice Directed Warehousing	VDW
Voice Extensible Markup Language	VXML
Voice Recognition	VR
World Wide Web Consortium	W3C

Table of Contents

1. Introduction

Google has been quoted as saying that SR is the future for web search and has invested in SR technology. One such example is its browser Google Chrome with SR integration. Apple ® as mentioned above is one of the pioneers and visionaries of SR. Apple ® have invested in SIRI which is a tool for conversational speech recognition. With giants such as Google and Apple ® investing in SR there seems to be a definite future for voice recognition (PC Mag, 2013). One could imagine potential future areas of use whereas a fail-safe, and easy to use, main app talks to and can control other apps on mobile phones, so that also users with low vision can use the voice features for accessing critical mobile applications such as mobile banking. By increasing accessibility through the Universal Design principles it allows for broader accessibility and mobility. E.g. "Speech recognition is used in deaf telephony, such as voicemail to text." (Varun, 2012).

The main principle is that "over 94% of all companies are looking to enhance or acquire more dynamic and interactive warehouse processes. Across a broad spectrum of options"... "Including voice technologies." (Heaney, 2011, p.1).

"Many opt to replace or upgrade the system they have in place to move from batched and paper-based operations to dynamic real-time event processing. The commercial systems that exist today are robust in their capabilities to handle high volume real-time interactions with RF, mobility and high speed tracking/ confirmation, and can support task interleaving." (Heaney, 2011, p. 13).

Speech recognition (SR) tools are commonly used in many fields, and are today widely available in many devices. The scope of this thesis focuses on Speech Recognition (SR) and its use in the warehouse environment. Mobile phones are one example of devices that SR is becoming increasingly available on, with applications like Siri, available on the Apple ® iOS platform. Within the field of Informatics there have been many articles written about SR but there is a lack of articles about SR integration into existing work process and there is room for further studies in this area. There is also a trend to move existing work processes to mobile solutions. The existing SR solutions are largely embedded solutions that are costly to implement. In order to implement SR technology into mobile devices, the thesis looks at existing SR technology that is used in warehouses. The thesis also covers a hands-on trial of SR technology currently being used in a warehouse environment. This is in order to gain a better understanding of SR technology (Bocij, 2008; Orvinder, 2009).

There is a lot of potential for voice. One could see implementations of voice control technology, in e.g. fast food restaurants such as McDonalds. Where voice control works in noisy environments Akbarinia, Valdez Medrano & Zamani (2011) and can be used to improve the efficiency by working hands free by implementing a SUI instead of a GUI. The employees at the counter take orders by repeating the customers' orders in a microphone "one cheeseburger" and it would instantly be displayed in the kitchen where the burgers are made. There are plenty of existing improvements to incorporate into a SR system. E.g. the Acoustic Model can be "trained" by feeding it with data gathered from the noisy environment, in order to overcome the noisy environments. This was also the case at Dagab and is presented in the results section. On top of this Huang, Acero & Hon (2001) points to several methods that attained good results

and allows for efficient SR even in noisy environments. The software and hardware combined can provide for the best quality result in a SR system screen the payment order can be displayed on the employees' counter. However one problem was that of noise distortion and according to Akbarinia, Valdez Medrano & Zamani (2011) there are plenty of existing improvements to incorporate into a SR system. E.g. the Acoustic Model can be "trained" by feeding it with data gathered from the noisy environment, in order to overcome the noisy environments. These features were also present at Dagab and is presented in the results section. On top of this Huang, Acero & Hon (2001) points to several methods that attained good results and allows for efficient SR even in noisy environments. The software and hardware combined can provide for the best quality result in a SR system.

1.2 Purpose

The purpose of the thesis is to collect qualitative data in order to identify and present requirements for integrating SR into work processes in a warehouse environment. The thesis provides analysis and design and prototypes of a solution producing Low-fidelity and High-fidelity prototypes that are based on the collected data and technical requirements. This may provide the basis for future studies towards the creation of a technical artifact or Proof of Concept. The research question for the thesis is "*What are the possibilities of integrating task based speech recognition into work processes?*" To answer this question the thesis will rely on ethnographic methods, case studies and interviews to find an answer.

1.3 The context of the study

This section gives a summary of the relation and context of interviewees. There are two companies in focus: Novacura and Dagab. Dagab was chosen to study the phenomenon of task based speech recognition, through an ethnographically inspired field study in order to study an existing SR system (that is embedded). Also to produce a prototype. Novacura represents the owner of the system, Novacura Flow, which is looking at integrating SR into their Flow process. Novacura has existing customers who use their Flow system, IFÖ and Roxtec which represent the Flow users in the Warehouse environment. These could potentially be voice enabled in the future.

The interviews with the participants in the project[1] were undertaken and then presented in Table 2 (section 3.1) that provides an overview of the nine interviews that were carried out. The two companies of focus are presented in more detail in sections 4.1 (Novacura AB) and 4.2 (Dagab AB).

2. Theory

The thesis covers many areas such as Business Process Modeling (BPM), human computer interaction (HCI) and Speech Recognition Technology. The following theory section will provide relevant research and a brief history of the technology, the major terms and how this relates to

---

[1] The term project and thesis are used interchangeably throughout the thesis.

the thesis. For the reader that would like more information about speech recognition and Speech User Interface (SUI) the authors recommend a book titled Practical Speech User Interface Design written by Lewis (2010).

<u>Note</u> that this thesis uses many acronyms which may be confusing to the reader but Table 1 "list of terminologies" on p.3 of this thesis includes many of the terminology found in this thesis.

2.1 Speech Recognition (SR) A Historical Background

Speech is an example of an Active input mode which allows the user to give explicit commands to the system (Sears & Jacko, 2007). Speech Recognition (SR) has been around since the beginning of the 1950s and the term SR is used in this thesis as an umbrella term for both Voice Recognition (VR) and automatic speech recognition (ASR). SR is defined as the ability of a device or program to understand and translate the spoken words into machine readable text (Juang & Rabiner, 2005). It has its origins in machine generated speech, also known as speech synthesis (SS). SS is the precursor to SR and is the technology which later led to the improvements and focus on SR technology (Juang & Rabiner, 2005). Speech synthesis technology provided the foundations for the SR technology that was later developed to recognize spoken words and convert these into machine readable text. In Japan several Laboratories took on the challenge of speech recognition and in 1960 further discoveries such as Linear Predictive Coding (LPC) and pattern recognition, helped to lay the foundations for further advancements. In 1980 the research shifted from pattern recognition to using statically modeled frameworks to better recognize speech with the Hidden Markov Model (HMM) (Rabiner, 1989) which became the preferred model. Advances in Artificial Neural Networks (ANN) also led to better SR methods of pattern recognition gave rise to Dynamic Time Warping (DTW) and parallel distributed processing, which later is refined and integrated with HMM to create further breakthrough in SR.

IBM and AT&T represented two main approaches to SR. IBM choose to focus on a single user approach with their development of dictation software. AT&T focused on a speaker independent approach that was not user specific but allowed for greater range of users and dialects. AT&T used this approach in telephone call centers to handle call routing processes, based on voice recognition. This helped create the libraries with large-vocabulary speech recognition. These companies helped to developed and refine SR for commercial use; another such company is Dragon systems that developed "Naturally Speaking", software that could recognize natural human speech. The accuracy of SR software was low, around 10%, up until 1990, making the word error rate also high. This was improved as many private companies developed new applications and after the 1990s the Word error rate was improved. Many of the errors were based largely on factors such as background noise, hardware pickup such as poor quality microphone, the algorithms and pattern recognition models such as HMM.

The two main areas of speech are conversation and task oriented speech. Conversation oriented SR is more complex as humans do not always follow grammatically correct standards in a conversation and there is a wider verity and a larger range of word that are used in a conversation. Task oriented SR is more specific in the vocabulary used with a smaller verity of words that are used to complete a given task. The vocabulary is not large as it is specifically based on the words needed to accomplish a given task (Lewis, 2010) and task generated speech is the area of focus for this thesis. In task based SR the user is restricted to a small defined list of possible voice inputs in order to be able to go to the next step in the task an

therefore minimizing the risk for errors. This is needed in a business use of SR technology because it needs to be reliable and correct most of the time. As the technology improved, the speed of speech recognition also improved. This opened the market for speech recognition to move into the portable device arena. This was made possible because of decreases in the device size coupled with increases in speed and processing power. The vocabulary or syntactic model is found in a Linguistic Model (LM) and is used as a reference to the voice profiled or sound waves in an Acoustic Model (AM) the combination of the LM and AM are used for adapting to the LM in order to provide improved SR, for creating a better match to the users' voice.

The SR system can iteratively improve the word error rate by doing several passes of the same word to capture errors then substitute the error which adapts the LM. An example of this is provided in the Vocollect ®, Inc. patent number 8255219 how word error rates can be reduced using substitution. "One type of speech recognition error is a substitution, in which the speech recognition system's hypothesis replaces a word that is in the reference transcription with an incorrect word. For example, if the system recognizes "1-5-3" in response to the user's input speech "1-2-3", the system made one substitution, substituting the `5` for the `2`" (Patent Genius, 2011). The described process is included in a process called Voice Mapping and is explained in greater detail in the article by Rentzos, et al. (2005).
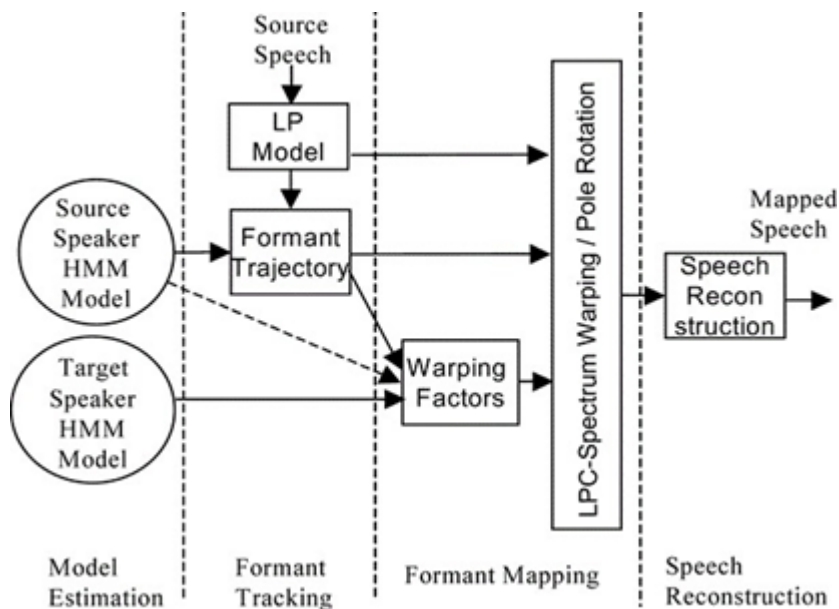
Figure 1. The voice mapping process is reproduced from (Parametric Formant Modeling and transformation in voice conversion cited in Rentzos, et al., (2005) Figure 1. p 229)

Some of the most important features of the technology that the thesis has encountered for improving the accuracy of the technology are voice mapping or templating. This together with a noise canceling microphone provides quality speech recognition; these are covered in more detail in the results section. Voice directed warehousing (VDW) is an umbrella term, for SR technologies that are used in a warehouse environment, such as Pick-by-voice, Voice packing and Pick to voice theses terms are all gathered under one term, SR, in the following sections.

2.2 Business Process Modeling (BPM)

Business Process Modeling (BPM) and Business Process Management Notation (BPMN) are modeling tools that use abstraction to show how the organizations business processes look like. This is done by drawing continuous processes with BPMN using standardized icons that describe the work processes as shown in figure 2. BPM makes it easier for decision making in an organization by simplifying the constant process of business improvements. This process allows the organization to easily identify, visualize and improve processes with BPMN and allows for better communication with other organizations and simplifies coordination between companies.

Business Process Modeling Notation (BPMN) is a widely used tool for Business Process Modeling. There are a number of tools used in process design: process maps, Business Process Simulation, Business Activity Monitoring (BAM), and Service Oriented Architecture (SOA). Often companies are in need of an improved BPM as BPM provides methods to define automate and or improve work processes. This translates into improved efficiency, productivity and performance management.
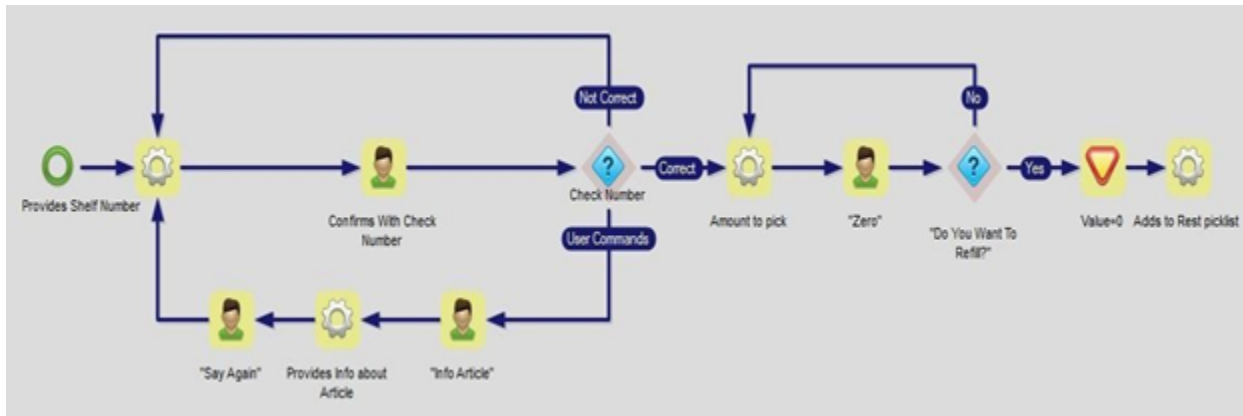

Figure 2 shows an example of a BPM that is modeled in Novacuras Flow Designer

BPM is a tool used for both high and low level of abstraction by visualization of a work process and its components. Business Process Simulation is a modeling technique. The improvement approach is done through Business Process Reengineering, workflow systems and the use of performance management technologies such as Business Activity Monitoring (BAM), and the use of SOA approach to resources (Bocij, 2008). Many companies want to improve their processes as they could potentially bottleneck performance. Therefore a process map helps to understand a business process; also it shows the interrelationships between activities in a process. The roles in the process become clear. And in larger projects it may be necessary to vary the level of detail. Business process management paves the way for the organization to empower and focus its workers on the processes and therefore on the main goals of the organization. One can bring consensus and collaboration in the work process. This can be used as a first step in using Business Process Simulation (BPS) (Bocij, 2008)on the process map.

Business Process Simulation(Bocij, 2008) is a tool that allows you to monitor the performance of the business over an extended time period; this can also be done quickly and in a number of

different scenarios. Process Improvement and BPM is central to Novacuras Business Model. They create new business processes or take existing ones and find improvements. BPMN shows how the organization works in an abstract way by visualizing how the business processes look like. This is done by drawing continuous processes with BPMN using standardized icons that describe the work processes. It facilitates decision making in an organization by using the collected data and helps an organization to work towards constantly improve their work processes. This process allows the organization to easily identify, visualize and improve processes with BPMN and allows better communication with other organizations and simplifies coordination between companies.

2.3 Human computer interaction (HCI)

In this section the key theories that are used for the design of the prototype are described in brief. It is not in the scope of the thesis to describe the following theories and terms in depth but however to present a brief description of HCI, Speech User Interface (SUI), Soft Systems Methodology (SSM) and Prototyping.

Human-Computer Interaction (HCI) is a multidisciplinary field that uses theories and research from several fields to create a large knowledge base of design theories and patterns. HCI looks at human users and how they interact with a system with the aim of improving this interaction (Sears & Jacko, 2007). As HCI is a very broad field the thesis selectively uses the design theories and patterns that are relevant for the thesis.

Looking at the design aspect of the SR system there are some principles that could be highlighted, and features that are of interest in a low-fi prototype. In addition there are theories that are of interest for the thesis. They are the design principles presented by Preece, Rogers & Sharp (2011) that deal with Human Computer Interaction (HCI). Principles such as communicating with the system using gestures, touch or speech termed Natural User Interface (NUI) and is considered too broad for the thesis. The design patterns for Speech User Interface (SUI) which use speech to communicate with the system are more relevant for the thesis. IBM (2006) and Lewis (2010) have both produced a guide for SUI design based on VoiceXML (VXML) which is a web standard for voice based web browsing. SUI is based on speech and not based on the standard Graphical User Interface (GUI). SUI extends beyond the limits that a graphical environment in that the user uses speech or SR to communicate with a system and allows the user to respond through voice commands and speech inputs.

Prototyping is defined by Preece, Rogers & Sharp (2011 p.390) as "one manifestation of a design that allows stakeholders to interact with it and explore its suitability; it is limited in that a prototype will usually emphasize one set of product characteristics and de-emphasize others." Prototypes are a communication tool that allows the designer to physically show and test their conceptual ideas. The two types of prototypes that this thesis uses are Low-Fidelity and High-Fidelity prototyping. The main differences are Low-Fidelity prototypes are produced in the beginning of the design phase and are largely paper based whereas High-Fidelity prototypes are produced after the Low-Fidelity has been used and tested and are software based prototypes.

The difference between SUI and other SR systems such as Siri from Apple ® is that Siri uses conversional SR which is a form of continuous SR and works with large vocabularies. SUI on the other hand uses task based SR and has a very limited vocabulary and only responds to specific words or short phrases. SUI is also called voice user interface (IBM, 2006; Lewis, 2010). Also Kjeldskov (2013) provides guidelines from HCI for mobile computing design one of which is the context. Where and how the system is used, is an important factor. Avgerou (2001) also highlighted the importance of context when designing and implementing a system. The themes from the analysis find the relationships which the design should include and which context the mobile device is to be used in, for example the device is to be used indoors in a warehouse which is a noisy, dusty environment. Noise cancelling microphone will be one of the hardware requirements though there is software that can deal with this issue and the dust can also be a problem. The context of the warehouse has other requirements that of speed and accuracy and the design of the warehouse is also an important factor (Heaney, 2011) as seen in the ethnographically inspired study. It also showed that space was a factor and how users move in the space should be considered as the space can be better planned. The warehouse can be designed to make voice packing easier with the heavier items first in the picking order; this allowed the picker to better pack the picked items. The context the device is used in also requires that the design should be robust but affordable and the T5 is a robust device but is costly. The T5 device costs around 30 000 Swedish kronor each. The headset costs around 1000 Swedish kronor. The Key to this solution is finding the right commercially available hardware and accessories like a Noise-Cancelling Wireless Bluetooth Hands free Headset with USB Docking, a charging cradle and a robust android device would cut the costs for the required hardware. The cost can be cut to roughly 3000 to 4000 Swedish kronor (1/10 of the T5) for each unit.

Soft Systems Methodology (SSM) is a design method that was developed in the early 1960 by Peter Checkland. SSM is used to develop models through abstractions of reality by using the systems rules and principles. It allows a structured thinking about the real world and allows for the conceptual thinking about the real world. This merger allows the SSM models to present a wider view of a situation making the models both descriptive and normative. SSM deals with complexity by constraining thinking in order to expand thinking. SSM begins by redefining the problem as a situation and looks at the situation in an unstructured way. This helps to develop a model of the situation with the complexity and relationship to the situation. The thesis will use one of the SSM modeling tools by creating a CATWOE model for the system and its situation.

The CATWOE model is a business analyst tool. The CATWOE (Checkland, 2000) is a soft systems methodology that is used "To build a model of a concept of a complex purposeful activity for use in a study." and provides a framework to complement the Template analysis as CATWOE focuses on organizational issues and provide an outline for the system being studied. It is based on the collected data and helps to identify the key question on what the SR system is trying to achieve, what are the problem areas, what are the external factors for the solution and consider the impact on both the business and people involved.
The central theme in the CATWOE is T for transformation process, it is used to provide a wider view of the situation that is being studied (Checkland, 2000).
The letters in CATWOE stand for:

C for Customers who (or what) benefits or falls victim from this transformation.
A for Actors the one/s that carries out the transformation to these customers.
T for Transformation process of the "input" being transformed into "output".
W for Weltanschauung (world view) what makes this transformation meaningful.

O for Owner of the "system" with ultimate power and can cause it not to exist.

E for Environment constraints that influences but does not control the system.

The model helps the thesis to produce a prototype from Dagabs situation and in order to produce an artifact for Novacura.

3. Method

This section contains an explanation of the data collection methods and how this data is analyzed. In this section a presentation is made of the ethnographically inspired method and a brief description of how it is used within the project.

The following data collection techniques were used, interviews, observations, ethnographically inspired field study(in which observations are included), and literature studies. Observation consists of observing others and participant listening, during which the ethnographically inspired field study emphasizes the whole concept of going to Dagab and getting into their work and to work with their voice system as users would do. The project undertook both participant listening and participant observation and got into the roles of employees.

3.1 Interviews

The questions were created with inspirational ideas from Bocij (2008) based on the "Who What, When, Where?" method and from the first set of themes in the template analysis as well as input from Novacura. The use of triangulation is also applied to the interviews. Interviews are conversations with a purpose (Bocij, 2008), in order to collect valuable data. The interviews with Novacura have mainly covered the Flow system and one of the aims was to document the process of designing a Flow [2] process using the Flow Designer application, the project has done this through interviewing an experienced Flow designer at NovaCura.

The Interviews with Dagab cover two supervisors and employees using an embedded SR system. The interviews were conducted using unstructured and semi structured questions. The semi structured interview consisted of a mix of open and closed questions (Bocij, 2008). The interview questions started by capturing the role and a regular work day and then probed to find the problems. The interviews were also used to cross reference the data from the observations. COOP is also an example of existing SR system in a warehouse environment and is used to cross reference the data from Dagab.

The data collection stage consists of a total of nine interviews and consisted of one unstructured and eight semi structured interviews. These interviews were conducted with the users, managers, flow designers, system owners, and embedded system providers (Vocollect ®) to gain several perspectives. All the interviews have been recorded and selectively transcribed. The interviews have been analyzed by a thematic analysis method called Template Analysis (King, 2004) which is provided as themes in the Results section. This is to categorize different findings.

---

[2] Flow is a BPM tool for the streamlining of a work process or a business process improvement, where Novacura streamline the interface and minimize the steps required for the user to complete a task.
2

Table 2. A list of the nine Interviews with the company name and interview number.

NovaCura	Dagab	Vocollect ®	COOP
Interview #1 with Flow Designer 2013-04-22	Interview #2 with Supervisor 1 2013-04-15	Telephone Interview #3 with Vocollect ® partner, CDC software Sweden 2013-04-25	Telephone Interview #4 with SR system Project leader 2013-04-25
Telephone Interview #1 with Manager 1 2013-04-23	Interview #3 with Specialist user 2013-04-26		
Telephone Interview #2 with Manager 2 2013-04-23	Interview #4 with a Voice Systems Coordinator 2013-04-26		
	Interview #5 with SR User 1 2013-04-27		

NOTE: For ethical details the interviews for this thesis can be provided upon request.

3.2 Ethnographically Inspired Field Study

Ethnography is defined by Helander, Landauer & Prabhu (1997, p.1435) as "a method that comes from anthropology" and "it is the work of describing the culture".
When observing in a rich environment to study "it is necessary to filter and focus".
"Participant listening is an important technique employed by ethnographers." Forsey (2010, p.1); Crang & Cook (2007) define ethnography as 'participant observation plus any other appropriate methods'.

Having conducted a field study at Dagab the aim was, to "Capture a deeper knowledge of how their daily use is formed through interaction with the organization's culture." (Patel & Davidson, 2003, p. 23). The field study included hands on trial of the Vocollect SR system in order to experience and use the leading SR system and to take the role of the employees or worker to gain a glimpse of a working day for this group of users. In addition employees were observed without interference, however primarily the field study presented us the opportunity to get a hands on experience as a supervisor would guide us through the process and show us how they work with the voice system. In ethnographic work "You gather what is available, what is

'ordinary', what it is the people do, say, how they work" "Data gathering is opportunistic in you gather what you can and make the most of it." (Preece, Rogers & Sharp, 2011, pp.252-255).

The thesis adapts some of the guidelines from Crabtree (2003, p.53) to be observant of e.g. activity and job descriptions of everyone involved, what they say that they do, and what they actually are doing. And with that criteria in mind try to explore and think about the environment and to see if there could be brought an improvement in the working environment and the layout of it. These findings are later presented in the Results section.

The thesis focuses on the use of reliable sources on research done in the area. Firstly it looks at research papers, theses, course literature and, scientific articles (Nutter, 2010). Also it is important to consider the fact that the designer has a role that is central and critical to the design (Löwgren & Stolterman, 2004). Therefore the aim was for the project to be as prepared as possible before conducting the field study. For this purpose, a framework was created to work from by observing acts, activities, objects, actors, events (Löwgren & Stolterman, 2004). Furthermore we got into the role of working there and got hands on experience with the technology, and accustom to their way of working with the technology. Half a day the project was given the chance to interact with employees and interview two different types of supervisors with important roles at Dagab. The project participants spent two hours picking with voice and packing roughly two pallets each. The observation we undertook at Dagab consisted both of observing employees use the system, but the participants of this project also got the chance to use the system and to pick by voice.

Advantages of ethnography are that it captures the culture of the setting that you want to bring an artifact into. The project has collected valuable data that will be used as the base for improvements. Observation conducted in the work environment can yield how the people actually interact within objects in a space and this can be cross referenced with the interview data. This can capture requirements, social dynamics, positive and negative aspects of the technology. These aspects can be used to provide possible improvements and are important aspects of further study.

In this particular field study, one had to pay more attention to the auditory memory more than the visual aspects. The field study provided the opportunity to confirm what had been said in interviews and to view the actual work process and compare this to the user's own concept of the work process. The combined effects of the ethnographic field study and the two interviews at Dagab were cross referenced, or Triangulated to see what had been said and what actually happened in reality, of knowledge that is based on experience and observation and cannot be documented easily and is tacit knowledge. Tacit knowledge can be found easier using ethnography. For example the regular users of the voice picking system at Dagab automatically use the fastest speed, furthermore they might take several items in one go to avoid returning to the forklift more than once. Through ethnography tacit knowledge becomes explicit but this can have the effect of the data losing its richness in the transformation process (Nonaka, 1994). The ethnographic method together with a literature study of the research articles, industry reports and other relevant information helped to identify contributing theoretical foundations as well as identify any previous research and techniques that had been studied, in order to bring about a prototype(section that can show and include the relevant requirements.

Ethics is an important aspect of the project where the observer has to consider many aspects of the thesis such as it is not ethical to time workers as they are at work. They may not take kindly to being measured for performance on an individual level. The researcher cannot foresee the effects of the information presented in the thesis and must be wary not to point the finger at anyone nor name any names. In this vain all video or audio recordings will remain anonymous and will only be available for the authors and for the duration of the thesis. The data sources are only to be kept for the duration of the project and they will all be deleted after the project has ended. All participants remain anonymous. According to Berger & Ludwig (2007), ethics is a key part of a qualitative research project as this helps to promote professional behavior and build trust.

Literature studies have helped to organize and highlight the categories and the previous research on the topic of voice control which helped to define the domain and scope for the thesis. According to Patel & Davidson (2003, p.34) the ethnographic field study is used to capture "a deeper understanding of how everyday knowledge is shaped by the interaction of the current culture" and the organization. This helps us to identify the contexts present within the work environment that the artifact is used in. In a warehouse it can be very busy; therefore it is very important with a clear intent before performing such a field study (Löwgren & Stolterman, 2004).

Observations are carried out in order to detect inefficiencies in the existing way of working with the SR system. This can be used to compare different systems' inefficiencies. When observing as a participant the tacit knowledge can be interpreted through experience, as this knowledge is not written nor realized. The dangers of observation analysis is the "Hawthorne effect" (Bocij, 2008) where it was noted that people behaved differently when under observation. Regarding observations Yin (2003) comments on the possibilities of direct observations with a field study to discover tacit knowledge.

A Task-Based Speech Recognition system is expected to provide a positive Return-on-Investment as described in the case studies and by Berger & Ludwig (2007). According to Berger & Ludwigs (2007) article, many of the effects of SR are positive, such as increased efficiency and a decrease in errors. For adopters of this technology the attraction is reduced costs and the need for a fast Return on investment (ROI) of the specific SR system. In most cases the ROI should less than one year (Heaney & Pezza, 2010; Voxware, 2011; Miller, 2004).

3.3 Template Analysis

The interviews and case studies are analyzed using the qualitative data analysis method developed by King (2004) and called Template analysis which is a thematic analysis that is structured as a hierarchy of themes. The first step is to provide a template or framework of coded themes by organizing the textual data into a hierarchal structure with a group of main themes that include sub themes which may have an interrelated to other sub themes or to other main themes. The textual data in the form of transcripts, notes and diary are extracted from the semi structured and unstructured interviews, field study observations, discussions, audio and video files create a large amount unstructured data that is to be analyzed to create a template. In this way Template analysis reduces the data into a manageable size which is relevant for the study. Template analysis is not strict in that the themes are flexible and can be expanded reduced or edited as the study progresses. In Template analysis there is no distinction between

the descriptive and interpretive sections in an interview. King (2004) believes this to be a false dichotomy in that the researcher is being interpretive when describing events in the interview. Template analysis allows a dynamic relation between the descriptive and interpretive and allows some themes to be more descriptive and other themes may be more interpretive than others.

The key focus in Template analysis is the context and relationships the themes have to the research question. The project found case studies that were relevant for task based speech recognition in a warehouse environment. These were used to create a template analysis on themes together with a mind map. The case studies consist of three industry based reports which are: "American Eagle Outfitters implement voice" (Zebracom, 2007), "Voice inside the food bank warehouse" (Voxware, 2012) and "Från grosshandel till Sveriges största centrallager" (Intelligentlogistik, 2008). These provide the first preliminary set of guideline themes for the project. The themes are grouped and used in the interviews in section 4.4.

1) Main theme of Work processes.
    Sub themes of picking accuracy, inventory accuracy, employee productivity, training time are reduced. The relation and context of the different groups that are chosen to participate in the interviews

2) Main theme of Work environment.
    Sub themes of worker Safety, job satisfaction, ergonomics and customer satisfaction.

3) Main theme of Technology.
    Sub themes of requires specialized SR software, WMS, ERP, SR accuracy, SR speed. Requires specialized hardware, durable, noise canceling microphone, mobile or wearable computing device.

King (2004) noted that Template analysis aids in providing insights into observed behavior and the meanings that underlay it. Template analysis also helps to compare perspectives from the involved groups and their experience of the context under study. It focuses on the communication between the involved groups, the user and the system. This provides a template for the factors influencing participants communication and behavior when using SR. Template analysis is however not suitable for a small study with only one or two small data sources. Templating is not done once but is an iterative process that begins with a preliminary set of starting themes that are flexible and are changed as needed under the length of the study. It is important to note that templating can remove too much of the contextual basis of the original data. One must be wary not to just create a list of themes but how these themes are related to the research question being addressed. For this purpose a mind map has been used to distinguish themes and help to disperse the themes, hinder the linear listing of themes and also to help identify the themes with the largest number of relations or the more prominent ones.

3.4 A selected inventory

Appendix A includes a table with an inventory of SR solutions. The list includes some of the off the shelf solutions and some SR development toolkits. The table was created both as a request

from Novacura and as a tool for an overview of the available mobile SR platforms. This helps to identify the integration possibilities for a voice enabled Flow. The source for the Appendix A is Avios (2013). It displays the different possible voice platforms starting with embedded systems and going through the Android, Windows Mobile and iOS platforms. The table shows an overview of embedded and non-embedded solutions. It gives an overview of providers. The embedded providers used at Dagab and Coop (Vocollect ®) and also the available development tools that could be used for integration into the prototype of this project. The main distinction that should be made between different voice platform providers is that of Conversational and task based SR. Appendix A is designed for the purpose of gaining better understanding of the voice platforms that are available and could potentially be used to integrate the Novacura Flow application. It is not a complete list of all the available SR tools and off the shelf SR systems.

4. Results

The project made use of the Ethnographic method in order to tackle a real-world problem needing a solution. During the study qualitative data was collected from interviews, observations and the field study. The results were furthermore based on the interviews, a field study, project meetings, textual analysis, and the literature study.

4.1 Novacura AB

Novacura does not have a speech recognition module but is currently looking at producing such a solution (to integrate with their Flow). Process Improvement and BPM is central to Novacuras Business Model. Novacura works with existing processes to find improvements. Business Process Management is used to identify, design, improve, and analyze an organization's business processes. BPM has its strength in that there are tools that can be used systematically to identify, visualize and improve their processes (Orvinder, 2009). BPM is used by Novacura as an aid in decision making and to help an organization to improve their work processes. This allows Novacura and the client organization to easily identify, visualize and improve processes with BPM. The software that allows this is Novacura Flow 5 and Novacura Flow Designer.

The thesis has gained inspiration and ideas through the access to software and examples of Novacura's Flow models with the aim of the thesis, to design a Voice enabled Flow process that includes important requirements in the work processes found in the warehouse study. The thesis first came in contact with Novacura during an initial meeting when some general requirements were established that the mobile application is required to work with both hands free in the warehouse and that the solution should be a mobile solution. In the process, the project also aims to produce an inventory of available SR solutions. Novacura has an existing Android application that runs the Novacura Flow 5 client. Therefore Novacura seeks an improvement of their Flow application in the form of SR integration into the existing mobile application. The current application is controlled with a touchscreen interface.

The meetings with Novacura have also resulted in information that during the project, there will be access to the source code for the existing application, which is already integrated into the Flow client. A module to the existing application is meant to be used in a warehouse and it will basically be performing text to speech and interpret vice-versa speech-to-text. The information required is generated from Novacuras Flow Client, via the Flow5 mobile application. The Flow Client interacts with an ERP application that provides access to information databases, and this

will be further detailed. There are three requirements from NovaCura that were defined for the SR integration.

Figure 4 shows how the platforms and solutions are connected together. On the far left is the mobile application by Novacura, a Flow app for a range of mobile devices. It is also for this app that the thesis aims to find out the requirements for a voice module. The puzzle piece in the middle which is the Flow client is the same API or system, as the puzzle piece to the left. The difference is that to the left is the mobile version of the Flow client. The Flow client works with the business system, in this case the IFS application. It is the Flow 5 mobile application which the thesis focuses on in our investigation. What are the opportunities and how voice control can be integrated into it?
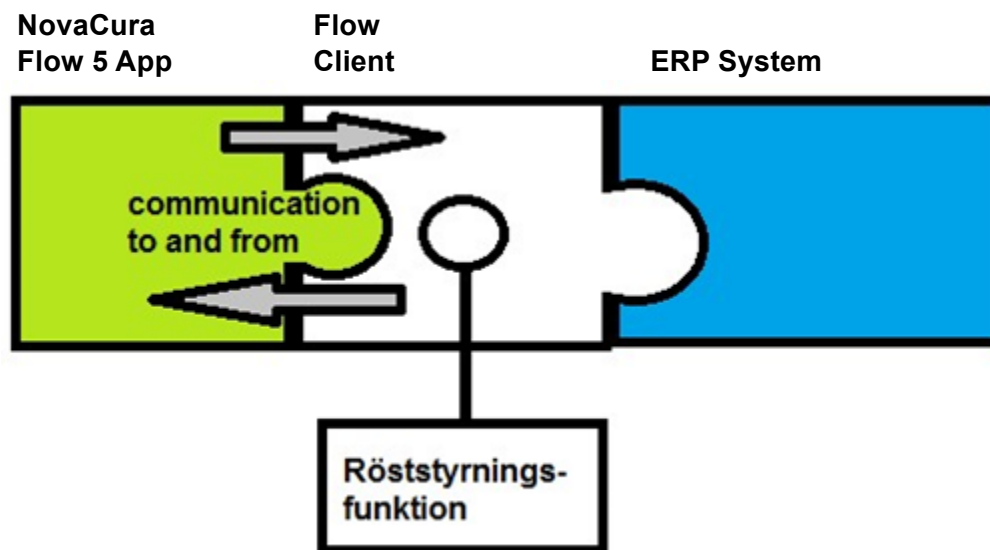
NovaCura Flow 5 App **Flow Client** **ERP System**

communication to and from

Röststyrnings-funktion

Figure 4 shows how the SR module can be integrated with NovaCura Flow client
Novacura does not have an existing SR module. But are interested in the future development of a SR module for their Flow Client.

4.2 Dagab AB

Dagab is a large warehouse distributor that delivers a range of groceries to Hemköp, Willys, Willys Hemma, Tempo and some smaller supermarkets. Dagab is a large warehouse which is used as an example of a warehouse where pick-by-voice has been successfully implemented. They use an embedded system, a one purpose solution. They are potentially interested in an integrated SR module-app such as Novacura aims to produce. The warehouse workers wear headsets and a mobile computer to receive commands from the voice control system e.g. which shelf the user should go to next. Once the user arrives at the shelf a feedback and verification number is requested by the system and is verbally given as "A12 done". The next step is to receive the number of items to be collected for the batch order. Then the cycle repeats and a new item and shelf location are given to the user. In order to collect data, during observation, about how it is possible to implement voice control, you should look at the way a warehouse worker goes through i.e. picking and how workers use the SR system. The advantage of this

kind of observation is that it does not require much from those observed (Patel & Davidson, 2011).

4.2.1 Ethnographic Field Study, observations and interviews at Dagab

Figure 5 the Dagab warehouse. The picture was taken on the day of the Dagab ethnographically inspired field study, 2013.

The project undertook half a day of an ethnographically inspired field study at the Dagab warehouse. The aim was to get hands on experience with the SR system and see how it works. This also provides a real world example of how the users interact with the SR system. Partly also to confirm what has been said in the interviews and case studies. This provides a valuable opportunity to get a firsthand experience of the SR technology. The Ethnographic Field Study resulted in 10 videos 63 pictures, a copy of the voice profiling commands sheet and a daily report sheet with all the orders that are handled under the work shift. There have also been made separate voice recordings of the interview and the observations as well as written notes of these.

4.2.2 Voice Hardware for Voice Picking at Dagab

Below is a list of the SR hardware that was used in the warehouse. The hardware that is presented is an off the shelf solution from Vocollect ®. This Hardware is required for the SR to work successfully with the WMS and VDW systems are presented in greater detail in section 4.4.3 under the sub theme of Hardware.

Figure 6 the Talkman T5 hardware showing the four function keys. The picture was taken at the Dagab field study 2013

The central device for workers at Dagab is the Vocollect ® Talkman T5 see (Figure 6). The device is a mobile computing solution that has a SUI (Lewis, 2010) and no screen or keyboard. Instead the Talkman T5 uses the headset and microphone as the Input and output (IO) device so workers can work with their hands and eyes free from distraction. The T5 is a robust device this including other facts is confirmed in the interviews in the results section. The battery is designed to last a full shift (Vitech, 2012).

4.2.3 Multimodal Devices
Refers to the user's communication which can occur through two or more combined modes of input for example: speech, touch and scanner. Speech is used to interact with the system in this case of the Talkman series T5, a device is a multimodal device if other input devices or accessories can be attached to it such as a ring scanner, used in warehouses to scan barcodes. The T5 is multimodal but lacks a screen for visual display of information. The T5 can be modified by adding accessories such as scanners through the accessories port.

4.2.4 Headsets
The SR20 headset is a lightweight single cup headset that rests on one ear and is designed for use in normal noise environments see (Figure 7). SR20 has noise-cancelling for better speech recognition headsets. The headset is a hidden cost as the headset and the wire input connection for the headset need to be replaced because of damage or wear and tear. The headsets are personalized meaning that each user has one headset that is their personal work tool.

Figure 7 SR20 headset the picture was taken at the Dagab field study 2013

4.2.5 Accessories
Vocollect ® T5 - 5 Bay Battery Charger is the accessory that is needed to charge many batteries at a time see (Figure 8).

Figure 8 charge station (Vitech, 2012)

4.2.6 SR Software

The SR technology Vocollect ® Voice Solutions was introduced in 2010 at Dagab. The selected SR system used is Vocollect ® Voice ®. The software was developed by a company called Vocollect ® which was founded 1987 and with over twenty years of experience in SR technology in the warehouse environment. The software allows for Voice-Enabled Workflows that works with the Vocollect ® Voice software solution and is based on the VoiceCatalyst architecture (see Appendix A). Vocollect ® also provide many case studies and industry reports that highlight mobility, ease of use and hands free computing (Vocollect ®, 2013). The SR software has increased in efficiency and allowed SR to work efficiently even in noisy environments (Huang, Acero & Hon, 2001). Companies have found additional usage areas for SR within the warehouses environment. The SR system can be used for picking orders, replenishment of articles, to monitor orders and total performance in real time. Other software used at Dagab is the software from Vocollect called Plock Monitor which is used in the warehouse to monitor, supervise and send messages to workers when an order deadline is due.

It is important to understand that SR has come a long way and that it has potential, for many companies like Dagab. The greatest potential is an application that enables this SR technology on an android device instead of an embedded system. Further this can potentially reduce the costs for companies investing in SR technology for business use.

4.2.7 Observations at Dagab

 The initiation process to configure the SR system takes 30 minutes for a new user. However this is only done once; the first time a new user initializes a new voice profile. The project includes an improved prototype (Low-fi) where parts of the login process have been simplified and made more efficient and faster by enabling touch. To map a voice profile the user had to follow the procedure of the system verbalizing a list of words. Thereby the user was required to repeat the words in this process at least three times for each word in the voice profiling sheet (Figure 9). An example of this was that the system said "Adam" first, then the user said "Adam" into the microphone, the system said "Adam" the second time, then the user said "Adam" the second time, the system said "Adam" the third time, and then the user said "Adam" the third time. This was repeated for each word in the list (Figure 9). The process is called voice profiling

at Dagab and is also known technically as voice mapping which is the Vocabulary that the system makes use of. Mapping is described in more detail by Rentzos, et al. (2005).
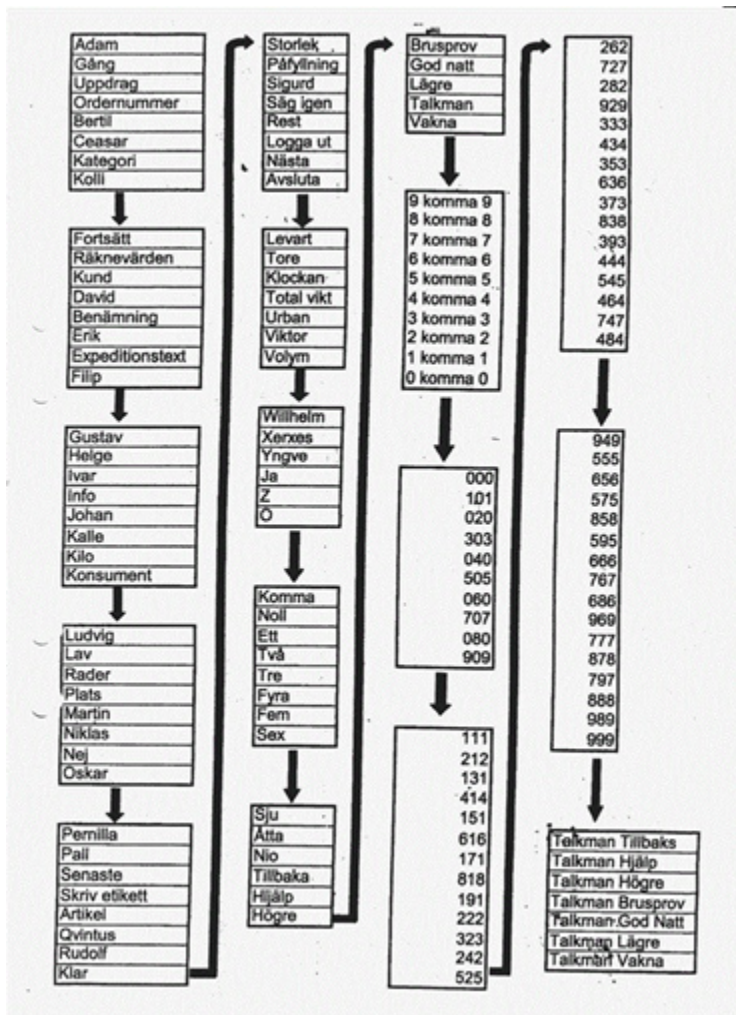
Figure 9 voice profiling process sheet that is used for voice profiling. The picture was taken on the day of the Dagab field study 2013.

After creating the voice profile one had to log in with the user account. We were assigned Chalmers one and Chalmers two, and the voice profiles were saved to these accounts. To use the T5 one had to scroll with buttons on the T5 to find these among the other employee names. Also one had to give the employee id number. One had to confirm the warehouse area that was chosen to work in. This was done by saying a code. Then the T5 was paused. When the user started up the next time the startup had been paused and the start was much faster. The next step was to write down the order detail for the order, the order number. This startup process was still paper based, for the daily report. It is a process that is repeated daily and could be replaced by a speech or digital version and included in the login process. After a 35-40 minute initialization and configuration a new user could start working with the system.

The system provides the user with information about the customer, the pallet type, the package, number of rows. However before starting it asked for the order number. When the user had filled in this information into the daily report of orders, he or she could get started and it would provide the first shelf to pick. The positive aspects were for instance that work could be done with both

hands free, the possibility to rotate, look around easily and lift things without the distractions of pen and paper. The only limitation that was encountered was that as we had just started with the voice picking, the voice would say the next shelf to go to. This turned out to be harder with auditory memory as a new user would not be accustomed to the speed of the computer voice, even at the lower speeds. However it was pointed out this is something that each user becomes accustomed to. Likewise the process of packing the items correctly on the pallet and in the warehouse, is knowledge that is gained through experience and tacit knowledge. The instructor commented that he usually uses it in the fastest speed on the T5.

Figure 10 Example of the warehouse's aisles the picture was taken at the Dagab field study 2013

One smart feature was the design of the aisles (Figure 10) labeled alphabetically from A to T where the heavier items were located first in aisle A as these would be picked and packed first. However the aisles are all one way only and this meant that the user had to drive further to get to the next row of aisles as one could not entre the aisle from the opposite direction.

Figure 11 Vocollect ®s Talkman T5 and headset. The picture was taken at the Dagab field study 2013.
Figure 11 is a picture of the T5 unit and headset with microphone used at Dagab.

Figure 12 is a picture of a shelf label note. The arrow shows the shelf below is the one below to be picked from. The picture was taken at the Dagab field study 2013.

As an example a pick process begins with the system saying the aisle name as shown in Figure 12. So the system would say "Adam" for the aisle named A, then "one five Adam" as the shelf number (15A). Then the user would drive the forklift to aisle A, locate shelf 15A and once there a check number is required to confirm that the user is at the correct shelf. The check number is found under the shelf number, the worker provides this check number as a feedback, in order to trigger the next command. The user provides the feedback by saying the check number to the system upon arriving at the shelf number. To verify the example "check 06" the user can just say six but in the case where the check number is twelve, the user must vocalize each number separately saying "one, two" not "twelve". Once the user has confirmed the checking number then the system would say the number of items to pick from that shelf. Then the system says, "One", which is the number of items on that shelf to pick, the user confirms this by vocalizing the number of items that have been picked by saying, "One", to confirm that the user had picked one item from that shelf.

The user can obtain information on an item to be picked by saying "Info benämning". Thereby the system responds with the type of article or product to be picked. An error that was encountered was that the system could mistakenly hear other conversations in the background as a confirmation, whereas the user had not yet confirmed that the article had been picked. However it should be noted that we worked together in a group and this is uncommon practice in a warehouse. Someone else in the group said"one" or something similar, so when the picker said one the system said the checking number was wrong. The user had to find the last shelf that the user received by saying the command "Info last" and the system would say the last shelf that was given to the user. So sometimes the voice system was sensitive to voice input from others when working in a group or it could be that "One" was sometimes picked up from the background noise.

An example of positive use of voice pick is that if a shelf needs refilling it can be done relatively smooth (compared to paper). If the user has to pick three items and the user replies two. The system asks "You said two, but you have to pick three do you want to refill?" The user answers "Yes" and within a few minutes the shelf is refilled by other users with appropriate fork lifts for the task. This provides a real time flow of work processes and goods, as both refilling and picking are combined and can be monitored in real time. It was observed that a user could bypass the one item at a time process. He took an item and said "1 " and went on to the next

aisle, as he had prior experience of the same customer and had gotten accustomed to the placement of items and also of the standard order.

4.3 CATWOE

CATWOE is explained in the theory section 2.3. it describes Dagab's use of SR technologies. What the business is trying to achieve. It relates back to the ethnographically inspired study results from Dagab. Its relation, context and feasibility to our project with Novacura. It helps to outline the impact of our proposed idea. Following is an initial CATWOE analysis of Dagab.

C = Customers

Warehouse workers, better interaction with SR. Increased accuracy and decreases errors leading to fewer returns. This leads to increases customer satisfaction. The users can be both for and against the SR system but the company gains benefits from the SR technology. The solution is proven to provide a quick return on investment.

A = Actors

The warehouse managers. They carry out transformation in order to make the system more accurate and perform less errors. In addition to allow them to work hands free giving better ergonomics. It Improves the focus, safety, and takes the eyes off any screen or paper. They might react negatively towards the limited ability to plan ahead. As with voice you can perform only one step ahead.

T = Transformation process

Input is from touch to voice. Output is voice. The output is interpreted by the application and responds accordingly. There is feedback to the user. The user verifies at every step which triggers the next step. It works step by step verifying each step to minimize errors.

W = Weltanschauung (World View)

The voice system allows for real time performance monitoring, and overview of total picks. For example attendance is monitored, and also no pick lists have to be given out manually. On the other hand the impact is that of efficiency that you cannot give two pick lists to one worker. The real problem is that of warehouse picking accuracy. And a digitalization/ monitoring enabling of the picking in real time.

O = Owner

Vocollect are the owners. However in the future Novacura could be the owner. As they aim to create a module for a solution to the problem.

E = Environmental constraints

Health issues like a sore throat impact the speaking ability and enhance performance negatively of the system. Background noise has to be dealt with as embedded systems like the Vocollect T5 have. The SR will not work with workers that have hearing impairments. The hardware needs

to be, like the T5, very robust. Many of these technical solutions are covered by patents and laws of copyright and ownership of some of the fundamental quality SR technologies and methods.

4.4 Thematic Analysis

In the previous section a broader CATWOE model was presented of the problem domain. In the following section the interview results are presented. They are categorized by themes. This provides for a more focused approach. As pointed out earlier in section 3.3 a mind map was used to create the most important themes covering the key aspects discovered in the interviews, as the interviews as whole are too lengthy to cover in detail. The themes are used to group, validate and to capture the requirements which are needed for the prototypes. The results have been analyzed using the Template analysis that identifies the central themes of interest. Also to find the relationships between the themes, and how they are related to the research question (King, 2004).

4.4.1 The central theme of technology

Under the central theme of technology, there are the sub themes of software, hardware, patents and mobility. In this section the results of the thematic analysis of the technology is used to help group themes of the SR system requirements. Including the right hardware in order to attain good speech recognition results. The process of collecting data and narratives helped to disclose that hardware was an issue, as it could be a potential source of work based conflicts between management and workers.

4.4.2 The sub theme of Software

The software of focus in this theme is Vocollect ® that specializes in SR technology. The Novacura software: Flow Designer and Novacura Flow 5 Client are connected to BPM.

The Novacura Flow Designer(interviewee) explained how the Flow process is created. It is done by using existing or known work processes which are then visualized by the Flow Designer(software tool) together with the customer. The Flow Designer is a good tool for designing the detailed steps in a flow process but lacks the high level abstraction. The Flow process integrates different windows or views from many different systems and integrates them into one view or window that the user can work more efficiently with. The designer can help improve business processes but they cannot break any of the business rules in the ERP system. A flow process doesn't override a business process but it aims to streamline it.

The Vocollect ® software and hardware are packaged and sold kits and integrate as one system. At Dagab the selected SR system is Vocollect ® Voice ®. The software allows for Voice Enabled Workflows that work with the Vocollect ® Voice software solution. The SR software has increased in efficiency and allowed SR to work efficiently even in noisy environments. Companies have found additional areas of use for SR within warehouse environments. The SR system can be used for picking orders, the replenishment of articles, to monitor orders and total performance in real time. Other software used at Dagab is the software from Vocollect called Plock Monitor which is used in the warehouse to monitor, supervise and send messages to workers when an order deadline is due. At Dagab they chose the SR software to improve accuracy.

"The biggest reason is that we don't have many errors. Accuracy is a lot higher (now). Every pallet we make has a specific member. When we print the waybill. We can see each page and what is on the pallets, a summary of all the orders."
(Interview #2 with Supervisor 1)

4.4.3 The sub theme of Hardware

The importance of hardware is not the focus for this thesis. However it is apparent that there are work-based problems that arise as a side effect of the limitations in the hardware such as battery life. In terms of hardware the T5 is an embedded device and has a limited memory (64 MB) and the processing power is based on a 416 MHz PXA270 CPU. As the T5 is an embedded mobile device, it has only one specific purpose, as a voice picking device so the memory or the processing power are not limiting factors. The output method is audio and is delivered through a headset with one headphone and microphone. The input method is speech through the microphone. The T5 has also 4 buttons, one for menu, one for pause and play, plus and minus are used as both volume control when in the menu, and when working these are used for adjusting the speech speed**.** The hardware within the T5 is patented by the company and the device is reliant on proprietary software (Vitechgroup, 2009).

The battery issue was highlighted after the interview with Supervisor 1 where the theme of battery life was noted and it was a recurring theme in several other interviews.

The problem with battery life is easily overcome by replacing the battery with a charged battery and does not disrupt the work flow, but it delays it because the user is required to change the battery and probably to drive the forklift back to the T5 battery charging station to look for a charged battery.

The problem with the batteries arise from the unstructured process of placing the batteries in the charge stations. Some of the batteries remained uncharged the next day due to laziness or other factors but the batteries were not placed correctly in the charge station.

> "The T5, its battery is not the best. So we try to come up with a solution to store the T5 and the batteries. But because people tend not to put it away in the charger correctly, they don't push it down and it's not charged the next day."
> (Interview #2 with Supervisor 1)

The project sees cost savings and time savings by having a worker or supervisor who can perform a routine check, and to make sure that the batteries are charging at the end of every shift to check that the T5s are correctly docked in the battery charging station. Supervisor 1 also added that if the batteries lasted for less than 3 hours they were replaced, this is seen as an unexpected or hidden cost with the Vocollect ® T5 system.

> "I am not sure how many batteries that we have ordered in the last 3 years and then have to throw away afterwards but it's quite a few."

> "I would say it is the biggest problem with the voice pick for us. In our other warehouse in Stockholm I think they have them (T5) downstairs and give them out with 2 specific batteries with the T5 and I don't think they have the same problem as us."

(Interview #2 with Supervisor 1)

4.4.4 The sub theme of Patents

According to the interview with Vocollect ® and their website it is clear that a number of patents protect their technology. Partly because they sell hardware (Patent Genius, 2013) kits with software. An example of one patent that protects the fail-safeness and functions for error reduction below in (Figure 13).
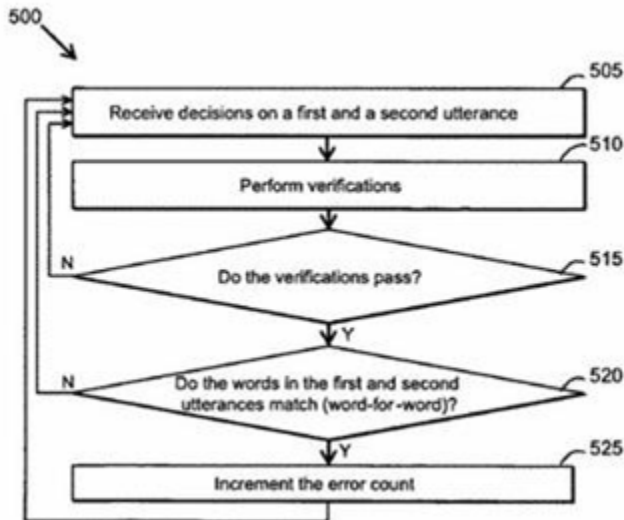

Figure 13 Example of a patent from Vocollect ® (Patent Genius, 2013)

Upon having contacted Vocollect ® partner in Sweden and explained in an interview that explained the importance of the Voice Mapping and Templating features of Vocollect ®s software. If for instance a user starts to pronounce a word differently, with a sore throat or slower or faster, then the software can be reconfigured to adapt to this new pronunciation of a word. You can retrain certain commands so that it recalibrates to the user. Basically necessary for professional use in noisy environments, are protected in forty three patents (Patent Genius, 2013) of which some cover hardware, how the headset is designed, also patents exist for the messaging system.

4.4.5 The sub theme of hardware Mobility

The hardware that has enabled for greater mobility at Dagab is the T5. The T5 is a mobile computing device that has an embedded Microsoft® Windows® CE 5.0 OS and the T5 is the Vocollect ® device that the Vocollect ® voice recognition software runs on. The device is constructed as a robust, wearable computing device that has WIFI and Bluetooth capabilities which runs on rechargeable batteries. They are designed to last at least a shift of 8 hours.

Mobility here describes the hardware that enables features of mobility e.g. Wi-Fi, and Bluetooth connectivity to communicate with the Vocollect ® system. There are design patterns for mobile devices that are discussed in the following theme of mobility in use. Mobility requires such attributes, that the device is limited in size, weight, performance and power (battery size). Other mobile computing devices are the A500, T5 and T2 (Vitechgroup, 2013).

4.4.6 The central theme of people and how they use the technology

The second central theme is how people use these technologies and this is also broken down into sub themes. The field study, interviews and observations, participant observation and other documents are used to show how workers use these technologies in their daily work environment. The themes are used to group and highlight the data to find relationships.

4.4.7 The sub theme of Ergonomics

Previously Dagab used scanners which were much less ergonomic than voice picking. The workers complained of physical side effects of carrying or using scanners. Ergonomics in this case is about the hardware, the T5 which is a mobile computing device. Ergonomics is also about how the users uses the T5 device and in which context the device is used. How the hardware allows the workers to use the technology to communicate with the system while on the move. The hardware allows for the user to have both hands free and their eyes free to focus on the work and while they can interact with the picking system through speech and speech synthesis. There are design patterns for mobile devices that include mobility that are applicable from SUI (Lewis, 2010). The physical size the placement and weight of the device are more ergonomic solutions for the user and allows them to move about quickly and efficiently in their daily work.

> "It was more accurate and it was more ergonomic, they could work with both hands."
> (Interview #2 with Supervisor 1)

In interview with the project leader at Coop stated they also use Vocollect ® and have tested a range of devices like the T2, T5 and the newer A500 which are all mobile computing devices that are from Vocollect ®. They implemented the SR system in 2003 because of ergonomic and the previous hardware was heavy, slow and uncomfortable for use over a longer period of time.

4.4.8 The sub theme of Auditory Memory

The memory is largely visual and textual when using Novacura Flow and Flow Designer and at Dagab the use of SR system requires the users to rely on auditory memory instead, which requires getting accustomed to. The Flow client limits memory to one step at a time, and by doing this the software limits errors.

In the with Supervisor 1 he highlighted that the SR system only allows for one pick command at a time and the memory that the user needs is only one step at a time, which simplifies the work process and reduces errors. In the Interview with the Voice Systems Coordinator at the Dagab warehouse noted that with paper picking the paper list gives a good overall view and ability to

plan ahead several upcoming nearby items, on the pick order, but the users visual memory was need to be greater as the picker had to remember more. The workers are not limited to one step at a time but have the freedom to choose, for example to hop over one or more picks and come back to them at a later stage.

In the Supervisor 1 and the Specialist user at the Dagab warehouse explained that the voice commands and feedback (check nr, 1) are found under each shelf address and helps to create limits and restrict the work flow to one step at a time. This limits also the number of errors that the user can make while also limiting the memory that is required by the user to perform the work.

4.4.9 The sub theme of Flexibility

It was noted that flexibility was a deciding factor in the Novacura interviews. Flexibility in this context is the ability of the artifact to adapt to the existing work processes or existing systems within the client company.

> "The other company had every work process pre-defined. Whereas Novacura let them dynamically adapt the Flow process to the existing processes within the company. They sought improvements for the warehousing packing, and picking where they use handheld devices."
>
> "The problem with the other company (Flow provider) was that they had predefined how an order receiving process would look like."
> "-With Novacura we could ourselves decide how our goods-receiving process would look like, so it's two different ways of doing it just"
> (Telephone Interview #1 with Manager 1)

Novacura allowed for flexibility by adapting the flow process to the existing work processes this is cross referenced with the Flow Designer where he explained the process of creating a flow process.

> "Normally they have a problem or they have something they cannot do or something that they need to improve. Defining process is quite heavy as the designer requires a broad knowledge of many different of a business, aspects such as ERP systems, business rules that need to be used, detailed knowledge of work processes and this helps the Flow designer to make configuring the flow process easier. The flow integrates different windows or views from many different systems and integrates them into one view or window that the user can work more effectively with. The designer can help improve business processes but they cannot break any of the business rules in the ERP system. A flow process doesn't override a business process but it aims to streamline it."
> (Interview #1 Flow Designer)

The Vocollect ® system is flexible and can be adapted to the user's voice by creating a voice profile. The voice profile can be calibrated to changes in the user's voice or the commands the user wants to use. The SR system could be used for many other tasks other than just picking. This was the main area where a computerized voice pick system showed the greatest flexibility

as compared to the paper based pick list. The SR system could be used for a real time overview of the picking activity. The SR system can also be used to send voice messages to one or more users.

> "And here you can see the people that are picking in real time and what their order is." (Interview #2 with Supervisor 1)

4.4.10 The sub theme of Speed and increased accuracy

The main reason that Dagab chose the Vocollect ® system was to increase accuracy and thus reduce costly and timely product returns and increase customer satisfaction.

> "No I don't know about the system. But I know a bit about why they chose to go to voice pick; the biggest reason is that we don't have as many pick errors. Accuracy is a lot higher (now). Every pallet we create has a specific number. When we print out the waybill (with the order number). We can see each page and exactly what is on all the pallets, you get a summary of all the orders." (Voice is accurate because the picker can only see from pallet to pallet. You don't see the whole order as is the case with a paper pick list. With voice pick you go from one pick to the next, one step at a time).
> (Interview #2 with Supervisor 1)

Supervisor 1 in the interview stated that accuracy and working with two hands were benefits and reasons for implementing the voice pick system. Supervisor 1 also added that any improvements in pick speed had not been observed. Dagab also had an accord based wage with a minimum pick amount for each worker. The system coordinator however, said that both accuracy and pick speed have improved after the voice based picking system was introduced, replacing the paper based pick list. The improvements in pick speed and accuracy have also been noted in the literature studies and in the interview with Coop.

Speed and increased accuracy has reduce costly returns and increased customer satisfaction as accuracy has improved the pick error rate from 3 in 1000 to 0.6 in 1000 picks when going from paper based pick list to voice picking.

The accuracy can be affected by background noise as highlighted by the example of how voice profiles were created when the system was first implemented, the workers chose to create their voice profiles in areas that were quite where there was little background noise. The accuracy was poor in the first attempts at voice profiling. This was because of the environment as it did not resemble actual background noise in the work area. To gain better accuracy the voice profiles were created by the users in the noisy working environment where it would later be used.

Set of design issues that are important for accuracy and speed when using voice picking is voice templating where commands are mapped to a defined set of commands (vocabulary) for the system that can be used in the system (Figure 9) In the ethnographically inspired field study the project took the chance to test the actual voice picking technology and the T5 mobile computing devices. The field study showed that voice picking was an easy technology to learn and after voice profiling the user can start picking with voice. An issue was the requirement of accustomization: for the user to adapt to the speed of the voice commands and to train the auditory memory to make voice pick efficient.

4.4.11 The sub theme of Ease of Use

The design should be easy to learn, and easy to use. This was indicated in the interviews by several sources an can be revised as simplicity with restriction. The device should be simple to use and with the term error restriction is defined as creating a level of restriction so the user can efficiently do the tasks but is limited in that the task is simplified to the degree that many of the errors are designed out. The system only gives one command at a time and requires feedback by the user to advance to the next step restricting the user. The vocabulary is restricted to reduce SR errors, in the vocabulary there should be, the right words and enough words to do the work required (Lewis, 2010). The Vocollect ® T5 has only four buttons that are used for navigation and are used to pause, exit, increase, decrease voice speed or to increase, decrease voice volume. The system is simple to use as the user is given only one command at a time. A voice profile has to be created before the user can start using the voice pick system which takes less than 30 minutes. The transport organizer also said that the system is easy to learn because after half an hour the workers can start working.

4.4.12 The sub theme of Criticism of Voice

The critical aspects of the SR system that have emerged from the interviews and case study of Coop Broterminalen and interviews and observation at Dagab. Some of these are summarized and presented here.

A case study by Computer Sweden (2009) explains how some users at COOP experience that the work day has become more monotonous because they only repeat the same numbers all day and feel like robots. Others see the clear advantages of having your hands-free. The hardware is important as the headsets have to be very efficient at noise canceling. Initially the system was tested with inexpensive headsets which turned out to be impossible to use in terms of quality. After the users had created their own profile and they had configured each user seldom there were any problems with the Speech Recognition technology itself. There have been some troubles with dialect as the number 28 for example is pronounced differently in southern Sweden, and in Västerås 28 has to be pronounced like two numbers, two and eight. In Skåne, southern Sweden there are no such problems, tvau, and autta. But in Västerås the numbers glide together, tvåååtta. Also the system is able to understand several languages which can be useful in a warehouse.

The voice pick allows the user to pick one item at a time, although this can be bypassed as seen in section Observation at Dagab. The T5 lacks an overview of the pick list. In this sense the SR system is limiting the user to plan ahead.

> "You cannot plan the ahead of time with your work. With voice you only get one command at a time."
> (Interview #2 with Supervisor 1)

With paper pick list the user had more freedom to plan the pick process but this also produced more pick errors and lead to pallets being returned to the warehouse.

> "With the paper you can take maybe five different that are nearby.
> The move away from paper reduced the number of steps in the pick process like waiting in line to receive the pick list but the advantage of the paper pick list was that the order could be divided or combined as needed.
> "When we had paper it was a physical list for each order which was given out downstairs. So everybody had to line up. And if you had a big order and then you had a little order for the same store you could give them both to one employee. Say for example you had an order with 400 packages and then you had one with twelve then he could pick both orders. With voice you can't do that. You can have Combined Picking. But it's a little bit of problem because you have to use one pallet for the other order that is only twelve."
> (Interview #2 with Supervisor 1)

Supervisor 1 also explained the need for more voice commands there could for example be one for refilling as this is a little bit more time consuming. As it was illustrated in the example above some of the steps could be skipped to provide for an immediate voice command saying "Talkman Refill".

4.5 Requirements for the SR system

The requirements are primarily based on input from Novacura, Dagab, interviews, field study, template analysis and observations. The requirements are the building blocks for the prototype. Many of the requirements have derived directly from the themes and the reader can refer to the themes for a detailed description of each and how they are related. The requirements are:

1) Accuracy: the voice recognizer understands and gives the right answer and the answer by either using templating and or mapping the users' voice (not perceive voice input wrong from time to time. Systematically perceives it as what the user wants the input to be.

2) Performance & Speed: Performance, the system should scale to the number of users and be precise and accurate by using templating.

3) Ease of use. It should be a system that is easy to set up and use.

4) The solution should be mobile, which provides both benefits and limits such as the processing power of the mobile device is limited. Hardware is a factor that affects performance and accuracy. Hardware is important for the solution e.g. battery life and noise cancelling microphone. It is important to have durable connectors for the headset. Also the use a Bluetooth headset can increase hardware durability. Some benefits are the portability for the device but the thesis also takes into account the environment that the system is used under as it may be used indoors and outdoors while working with the mobile device. Others can be battery/power limits and so on. The platforms open for consideration are Android and Windows Mobile.

5) Robust. The device and system should be able to withstand the daily work environment e.g. Physical ware and tare.

6) The system should provide information and support e.g. monitor and send messages to users.

4.6 Prototype

Novacura provided the initial requirements that have been listed in section 4.5. The requirements have been redeveloped and some functions have been altered or removed throughout the process. The observations, field study, and the interviews help to define some but not all of the requirements that are needed for the design prototype. Some of the requirements are not known or visible to the users but have emerged from research or industry reports and can be found in the literature study and other collected data. With this in mind the design prototypes are presented here in the form of design proposals and mockups in Low-Fidelity and High-Fidelity prototypes.

The Theory section 2.3 about HCI should be had in mind as relevant theories are provided with guidelines from HCI for the following prototype section. Three factors affect the prototype: space, environmental context as well as costs-savings. The hardware can be combined with SR software that is designed for limiting errors by using restrictive simplicity (Lewis, 2010). These are two design patterns that have repeatedly mentioned in the interviews. An example of this is that the voice pick only allows one step at a time with simple commands that have restricted possibilities of verbal feedback that is expected from the user. These directly affect accuracy, ease of use and processing power required. The user is given only one simple command at a time and the expected response is also limited to a few simple answers. If the response is incorrect then the system will repeat the command or ask the user to repeat the response but the user will not be able to advance to next step without the correct response.

4.6.1 Low-Fidelity Prototyping

The layout of the hardware is described below and the prerequisite for this particular layout is a noise canceling Bluetooth headset either with mono sound or stereo this could be personal for every employee and additional to lend if lost. The mobiles phones should run Wi-Fi only and not need any 3G connection. Alternative two is that every user has it in the pocket with the noise cancelling headphones directly connected by cord. The mockup (Figure 14) has the benefit of cordless freedom and endless battery on the forklift. While having the unit in the cradle the pick list can be displayed for a hands free working environment. So the unit remains docked during use and the Bluetooth headset is worn by the user.
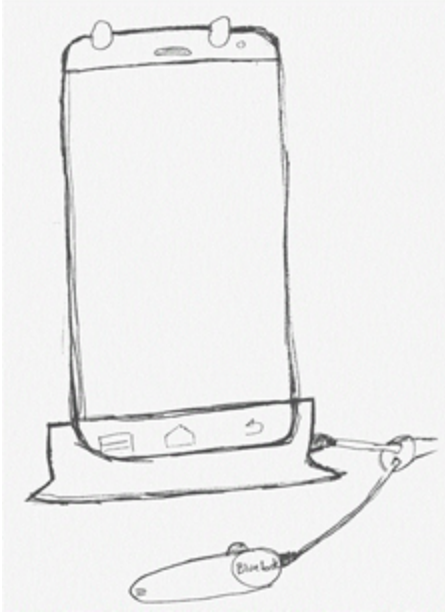
Figure 14 Design of a mobile Android solution (Hand sketch, 2013)

The two (Figures 14) and (Figures 15) show two possible solutions in the sketches visualize the different hardware components that are required in the design of a prototype. These different components are derived from the requirements from Novacura, observations, field studies and interviews. One possible solution in figure 14 is where each forklift has a charging cradle in which the Android phone sits in. This is to eliminate battery concerns, as this was highlighted in the ethnographic field study and in the hardware theme. And the Pick list is available for show at all times on the display as well as a wireless Bluetooth headset is used together with the mobile device for hands free working.
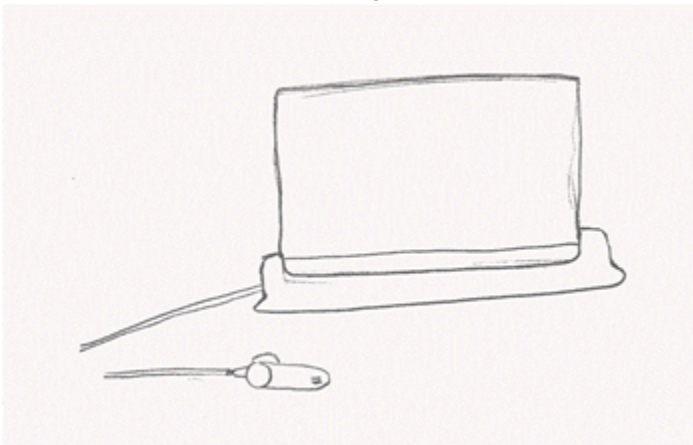

Figure 15 Design of a tablet Android solution (Hand sketch, 2013)

The second solution as seen in the sketch in (Figure 15) shows an Android tablet device e.g. a 7 inch android tablet, instead of an Android phone. The major difference in this solution is that the device lies down limited to using Wi-Fi for connectivity and the device is specifically used for voice picking in the warehouse. This provides a multimodal solution that can show the pick list on a touch screen. And gives the user the hands free through the Bluetooth headset. The pick list should be available even if the system goes down. This could be managed through not

having it interactive and is only displayed as a static screenshot. Also Novacuras system provides asynchronous communication allowing for work when not online. An example of noise canceling Bluetooth headsets used in Call Centers (Jabra, 2013). If opted for a wireless BT solution the headsets should perform at least10 hours talk time and be noise canceling and Bluetooth enabled. If it proves difficult with pairing and or charging the headsets another possible solution is to have a mobile phone simply, in the pocket, with a headset with a cord connected(from the headset to the device). This eliminates pairing and charging of Bluetooth headset concerns. However it brings back the battery concerns of the mobile phone, without the charging cradle. The prototype that the project has undertaken was drawn in IPhone Mockup tool[3] available online[4]. Following are sketches of a prototype. The prototype is based on primarily the requirements from Dagab and with design principles in mind. The project has led to a prototype according to the requirements from Dagab. The three figures show some of the functionality present in the prototype

Figure 16 Login design prototype (drawn in IPhone mockup, 2013)

In (Figure 16) describes the process of logging in to the system. The user uses touch to quickly enter the employee number and scrolls down with the touchscreen to select a profile which has been created prior on a PC. The PC profile also has details of which type of picking the user prefers for instance pick with pallet A or pallet B. And other profile details which is controlled via middleware software e.g. such provided by Vocollect ® to their customers.

---

[3] A mockup is a model for studying and testing a design.
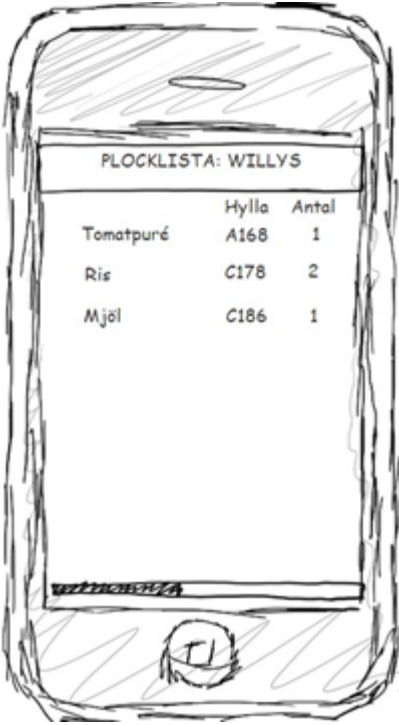[4] IPhone mockup tool available at. http://iphonemockup.lkmc.ch

Figure 17 pick list prototype (drawn in IPhone mockup, 2013)

The project aims to keep the prototype as simple and easy and intuitive as possible.

In Figure 17 the user has logged in and confirmed the order number and is ready to work. Originally the prototype included buttons above the pick list: three buttons plus (+) and (-) minus for the voice speed as workers need a way to regulate this. However thinking this process through it would make the users too dependent of the screen. It should be used mainly for an overview of the pick list. Therefore the above described commands + - pause should be voice controlled with a keyword said before e.g. "Android Pause" or "Android Faster". When paused, the worker can go to take a break. The navigation of the menu is performed by scrolling left and right between pick list and a list of available voice commands.

Figure 18 command list prototype (drawn in IPhone mockup, 2013)

Figure 18 displays a command list easily available to the user during work. By gesturing left or right the user can either display the pick list in the previous figure 18, or when the user has scrolled to the right in the menu bar the list of available voice commands is displayed.

From the low-fidelity prototypes the project can move to the next step and further develop the technical aspects by showing a hi-fidelity prototype. This Novacura voice flow could be constructed to support an android (or other) SR platform. In order to see which platforms are available for this voice flow that the project has designed in 4.6 is an appendix with technology that could support our high fi prototype.

4.6.2 High-Fidelity prototyping

The prototype was produced in the Novacuras Flow Designer tool. The High-Fidelity prototype (Figure 20 to 23) shows a prototype of a voice enabled Flow. The project has resulted in several different Flows which have been constructed for the purpose of illustrating how to integrate voice into the Flow process. Novacura Flow designer is the tool that has been used to model the Flow processes. Novacura. Flow Designer uses a set of tools for modeling; these are found to the left in the work area and are shown in (Figure 19). The symbols or tools that are used are User Task, Machine Task, Decision Task, Assertion Task, Termination Task, Contextual Information, Assignment, and Log Port. The main Symbols used in our prototype are the first three. The green circle is the start point for the flow process which follows in the direction of the arrows.
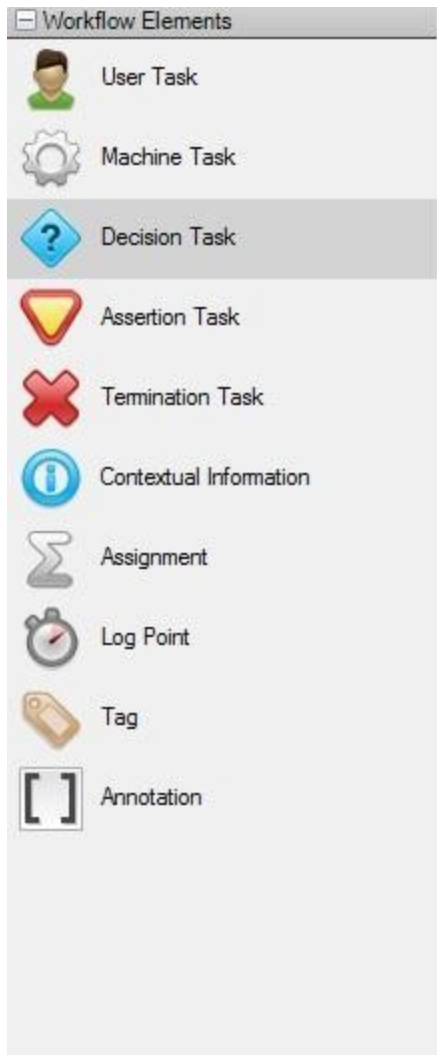
Figure 19 Novacura Flow designer tools (Novacura Flow Designer, 2013)

The Flow in (Figure 20, 21, and 22) is a prototype for a voice pick Flow process and (Figure 23) is a prototype for a voice activated refill process. The prototypes are all modeled in Novacura Flow designer as examples of HI-FI design prototypes for a SR Flow process. (Figure 20) describes the process of logging in both for registered and new users. It is important to note that the flow process in (Figure 20) takes place after scrolling to and choosing a voice profile and the user's voice has a voice profile that is connected to the user's login.
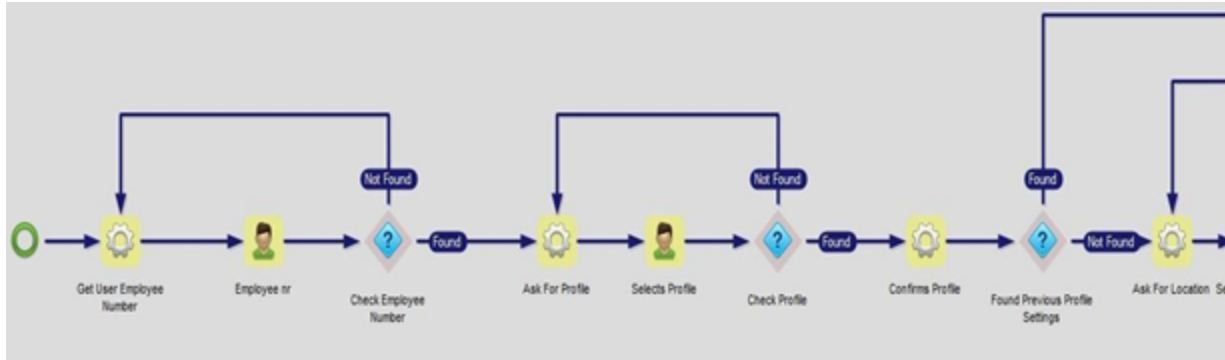
Figure 20 Novacura Flow designer login and pick process part a (modeled in Novacura Flow Designer 2013)

From start the process is that the user logs in with an employee number. If this is found in the system the profile is asked for. If not the system asks for the employee number again.
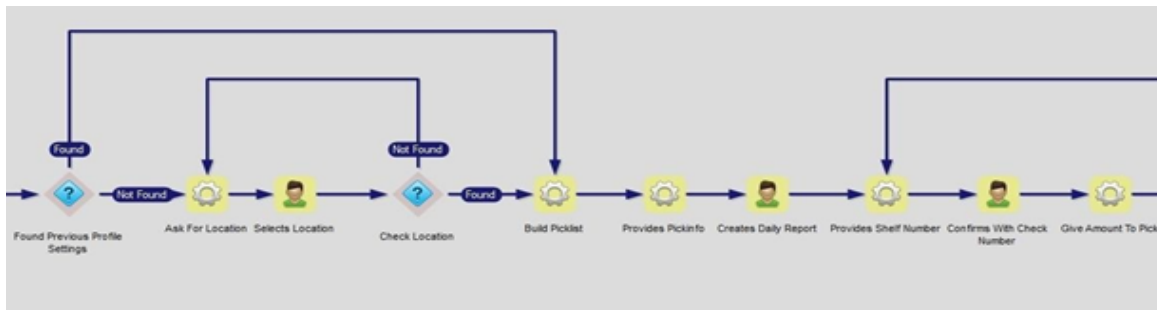

Figure 21 Novacura Flow designer login and pick process part B (modeled in Novacura Flow Designer, 2013)

If a Profile is used that has already been set up, there is no need to enter the location information. If the user is a new user the location will be entered (probably in the form of a code for each area).When the user is set up a pick list is built. And the system starts providing the user with a shelf to go to. And the user confirms its location throughout the process with the check number under the shelf info. The system provides info with how many to pick from the specific shelf.
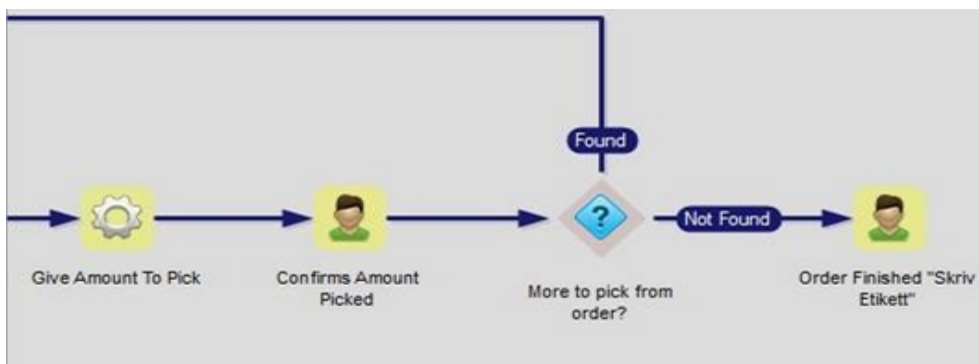

Figure 22 Novacura Flow designer login and pick process part C (Modeled in Novacura Flow Designer, 2013.)

In (Figure 22) the user confirms amount picked. And if there is more work to be done in the order new shelves are provided to go to. If else the order is complete and the user confirms by saying "print ticket."

(Figure 23) is another example of a flow diagram that describes error and demonstrates the fail safeness of the voice pick process. The user confirms a check number. If there is an error, the user can say any appropriate user command to find out where the error is. The user command say "Info Article" which asks the system to provide information about the article on the shelf too see if this correct shelf. The article is described by the system as 5Kg rice which is the shelf above. The user can also say the user command "Say again" to the last command from the system, in this case the system repeats the last information it provided, e.g. shelf number. Then confirms the check number to confirm that the user is at the right shelf. The flow diagram showcases some of the information that should be possible to request from the voice system in case of uncertainty or to check for errors using the available User Commands.
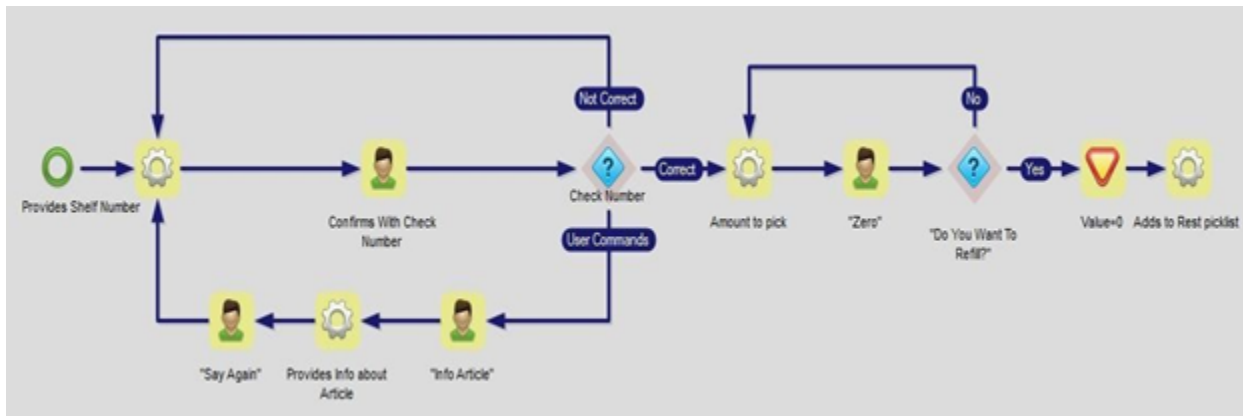

Figure 23 Novacura Flow designer refill process (modeled in Novacura Flow Designer, 2013)

5. Discussion

The project has conducted an ethnographic field study, interviews and design solutions, of speech recognition software that is used in the context of a large warehouse. It can be argued that there are benefits from a SR solution and even greater gains through integration of SR into an existing application. Moving from a touch screen interface to a SUI is the optimal solution but the field study questioned this rhetoric and puts forward that a well-balanced combination of the two can be used to achieve a better solution. With the SR system users can gain advantages such as working with both hands free, not bound to other input devices and it allows for greater mobility. This is a more natural way of communicating with the system through a natural user interface (NUI) (Wigdor& Wixon, 2011). This frees the hands and eyes but the user is then trapped in the audio realm to communicate with no other choice but speech, so a complement to SR is required. It is important to note that a touch screen should only be needed and activated at necessary steps in the work process in order not to distract the user from the task at hand. An example can be the daily report of orders which could be touch based when filling in the information. It could be a compliment to the speech based interaction.

5.1 Limitations

The scope of the thesis is on SR systems and solutions that are task based, and not on conversational SR. The section is based on theory outlined in section 2.1about Task based SR. Task based SR is chosen also because it resembles the BPM process, step by step. The environment or context the SR is used in is the warehouse. We were provided half a day of studies however this was not enough for a full scale ethnographic study therefore the thesis has focused on inspiration from ethnography in order to make the most of half a day that was provided by Dagab.

When it comes to the limitations of the technical platform the existing Flow client operates on is the Android platform. So preferably the solution should work with either the Android is the first priority and the other platforms, iOS or Windows Mobile, are of secondary importance. This proved to be a technical limitation buy focusing on one platform and as the majority of SR systems for VDW was embedded with the Windows CE platform. However the project was not limited to just one platform but was free to look into any relevant platform. Novacura provided many of the tools in the project to develop a Flow processes that could replace the existing pick processes with the use of the Novacuras toolsets for Design. The testing of the design results was limited and performed only a summary evaluation of the designs. This it was clear that more time was needed in this area, in order continue on to the next step, that of creating a proof of concept.

5.3 A Retrospective of the thesis

The project's initial scope was too wide and the ambition level too high, particularly the technical solution or proof of concept. This led to the undesired effect of the project becoming harder to manage and the project concluded with prototyping. In retrospect the project has been managed well and the technical aspect can be a subject for future studies in this area. The project diary also provided partial insights into problems like side tracking off into other related areas caused the project to lose focus of the central research question. This had the effect of the thesis becoming bloated with text, as the authors covered terminology and definitions that are not directly relevant to the problem. The design relied heavily on the collected data. The design patterns in HCI were expected to provide more, however many design patterns are largely developed for GUI and thus the thesis gained inspiration from SUI principles.

6 Conclusion

The study used ethnographically inspired methods to gain access into an organizational culture and discover tacit knowledge (Nonaka, 1994) how speech recognition systems are used in a warehouse environment. This tacit knowledge and the collected data were used to form the Low Fidelity and High-Fidelity prototypes.

The answer to the research question is provided in the following points of patents, hardware and added functionality. The answer to the question "What are the possibilities of integrating task

based speech recognition into work processes?" is that yes it is possible. However in the field study it can be concluded that preferably this would be done by a multimodal device, which allows the user to interact also with a touch screen. As it allows for improved interaction and options than a pure voice only (SUI) solution would. There are other areas that are a hinder such as the theme of patents is likely to hinder the final solution. One solution would be that one acquires the patents or one develops their own customized solution that does not infringe on these patents. The most realistic solution however is to make use of the patents that fulfill the requirements for the SR system and allows for better quality SR for industrial use. Voice Mapping and Templating are necessary for SR to be more accurate in a noisy environment. There are many patents (Patent Genius, 2013) which cover hardware e.g. headset and software patents like the messaging system. The costs can be reduced by using existing hardware solutions. The SR solution could be further improved upon as shown in the design of the prototype. Mainly with added functionality through a display of the active pick list and available voice commands. The SR system should also include a function that shows the user where the finished pallets should be placed; this would make use of the screen on the Android devices and provide further time savings for the workers.

The SR technology was only one piece in the puzzle for improving work processes as the ethnographically inspired field study and interviews highlighted. Although the SR technology did provide a fast ROI (Berger & Ludwig, 2007; Miller, 2004). Increased efficiency and increased accuracy was experienced for adopters of the technology (Miller, 2004). The design or layout of the warehouse can make the process of packing and picking more efficient and so the planning of the work space is also a factor to be considered (Heaney, 2011). The SR solution requires the right software and hardware to make SR work and described in more detail by Voxware (2012), in the white paper titled Optimizing Worker Performance. The combined effect of workflow improvements using BPM can be combined with the benefits of SR technology to create a better work environment that results in increased efficiency and accuracy. This can be achieved by integrating SR directly to the Flow process; Novacura could provide further incentives for companies wishing to implement SR technology into mobile devices in the warehouse environment (logistics, 2013). The thesis showed only some of the positive and negative aspects of SR while trying to capture requirements that could be used for future studies for integrating SR into an existing application. The thesis ends with a call for future studies, which are required, for the development of a functional prototype and a proof of concept for a multimodal SR system.

References

Akbarinia, A., Valdez Medrano, J., & Zamani, R. (2011). Speech Recognition for Noisy Environments-Feasibility of Voice Command in Construction Settings.

Avgerou, C. (2001). The significance of context in information systems and organizational change. Information systems journal, 11 (1). pp. 43-63.

Berger, S., M., Ludwig, T. D. (2007). Reducing warehouse employee errors using voice-assisted technology that provided immediate feedback. Journal of Organizational Behavior Management, 27 (1), 1-31.

Bocij, P. (2008). Business Information Systems: Technology, development and management. Ft Press.

Crabtree, A. (2003). Designing Collaborative Systems: A Practical Guide to Ethnography, London: Springer-Verlag.

Crang, M., & Cook, I. (2007). Doing ethnographies. SAGE Publications Limited.

Checkland, P. (2000). Soft systems methodology: a thirty year retrospective. Systems Research and Behavioral Science, 17, S11-S58.

Forsey, M. G. (2010). Ethnography as participant listening. Ethnography, 11(4), 558-572

Helander, M. G., Landauer, T. K., & Prabhu, P. V. (Eds.). (1997). Handbook of human-computer interaction. North Holland.

Huang, X., Acero, A., & Hon, H. W. (2001). Spoken language processing (Vol. 15). New Jersey: Prentice Hall PTR.

Juang, B. H., & Rabiner, L. R. (2005). Automatic speech recognition–a brief history of the technology development. Georgia Institute of Technology. Atlanta Rutgers University and the University of California. Santa Barbara.

King, N. (2004). Using templates in the thematic analysis of texts. Essential guide to qualitative methods in organizational research, 256-270.

Kjeldskov, J., (2013). Mobile Computing. In: Soegaard, Mads and Dam, Rikke Friis (eds.), The Encyclopedia of Human-Computer Interaction, 2nd Ed. Aarhus, Denmark: The Interaction Design Foundation. Available online at: <http://www.interaction-design.org/encyclopedia/mobile_computing.html> [Accessed 20 May 2013]

Lewis, J. R. (2010). Practical speech user interface design. CRC Press, Inc.

Löwgren & Stolterman (2004). Design of information technology: the material without qualities. University Press. ISBN: 9789144042039

Forsey, M. G. (2010). Ethnography as participant listening. Ethnography, 11(4), 558-572.

Miller, A. (2004). Order picking for the 21st Century. Unpublished white paper: Tompkins Associates, Raleigh, NC.

Nonaka, I. (1994). A Dynamic Theory of Organizational Knowledge Creation. Organization Science, 5(1), 14-37.

Orvinder, C. A (2009). A business process in execution: en investeringsbedömning av Business Process Execution Language med Business Process Management. Department of Computer Science, Lund University. (R)

Patel, R., & Davidson, B. (2003). Forskningsmetodikens grunder. Att planera, genomföra och rapportera en undersökning.

Preece, J. Rogers, Y. Sharp, H. (2011). Interaction Design: Beyond Human-Computer Interaction

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE, 77(2), 257-286.

Rentzos, D., Vaseghi, S., Yan, Q., & Ho, C. H. (2005). Parametric formant modelling and transformation in voice conversion. International Journal of Speech Technology, 8(3), 227-245.

Sears, A., & Jacko, J. A., (2007). The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications. CRC Press.

Varun, S. (2012). An Implementation of Voice Based Mapping

Wigdor, D., & Wixon, D. (2011). Brave NUI world: designing natural user interfaces for touch and gesture. Morgan Kaufmann.

Yin, R. K. (2008). Case study research: Design and methods (Vol. 5). SAGE Publications, Incorporated.

Internet based sources

AVIOS (2013). Tools for Developing Speech-enabled Mobile Applications. [Online] Available at: <http://avios.org/app_dev.htm> [Accessed 10 May 2013].

Jabra.com (2013). Noise Cancelling Headsets. [Online] Available at: <http://www.jabra.com/headsets-and-speakerphones/professional-products/contact-cente> [Accessed 10 May 2013].

logistics.about.com (2013). Smartphones and the Supply Chain. [Online] Available at: <http://logistics.about.com/od/supplychainsoftware/a/Smartphones-And-The-Supply-Chain.htm> [Accessed 20 May 2013].

Iphonemockup.lkmc.ch (2013). Online IPhone mockup tool. [Online] Available at: < http://iphonemockup.lkmc.ch/?id=GkbfCSGgDDAZNZgFmMGHbaddSaBCGTASQMAOeGRhK &style=pencil > [Accessed 18 May 2013].

Patent Genius.com (2012). Patent 8255219. [Online] Available at: <http://www.patentgenius.com/patent/8255219.html> [Accessed 10 May 2013].

Patent Genius.com (2013). Vocollect ® Patents. [Online] Available at: <http://www.patentgenius.com/assignee/Vocollect ®Inc.html> [Accessed 10 May 2013].

PC Mag.com (2013). Google Adds Speech Recognition to Chrome. [Online] Ziff Davis, Inc. Available at: <http://www.pcmag.com/article2/0,2817,2414277,00.asp> [Accessed 05 April 2013].

Vocollect.com (2013). Vocollect ® Voice Solutions. [Online] Available at: <http://www.intermec.com/Vocollect/solutions/solution-configurations-interfaces/index.aspx> [Accessed 18 May 2013]

Vitechgroup.com (2013). Voice Terminals: Vocollect ®'s Talkman Series. [Online] Available at: <http://www.vitechgroup.com/web/solutions/voice/voicehardware/Vocollect ®devices.asp> [Accessed 17 May 2013]

Wikipedia.com (2013). List of speech Recognition Software. [Online] Available at: <http://en.wikipedia.org/wiki/List_of_speech_recognition_software> [Accessed 10 April 2013]

Case Studies

Zebracom.net (2007). American Eagle Outfitters implement voice. [Online] Available at: <http://www.zebracom.net/article-MMH-cover-American-Eagle.pdf> [Accessed 15 April 2013]

Voxware.com (2012). Voice inside the food bank warehouse. [Online] Available at: <http://www.voxware.com/fileadmin/media/resources/case_studies/CaseStudy_Harvesters.pdf> [Accessed 16 April 2013]

Computersweden.idg.se (2009). Stadigs, J. "Jag har en dam i örat ". [Online] International Data Group. Available at: <http://computersweden.idg.se/2.2683/1.237184/jag-har-en-dam-i-orat-> [Accessed 16 April 2013]

Intelligentlogistik.se (2008). Från grosshandel till Sveriges största centrallager. [Online] Available at: <http://www.intelligentlogistik.se/files/IntLog_nr8_lag.pdf> [Accessed 06 April 2013]

Industry Articles and Papers

IBM (2006). Speech User Interface Guide. [Online] Available at:
<http://www.redbooks.ibm.com/redpapers/pdfs/redp4106.pdf>   [Accessed 10 April 2013]
Voxware (2012). OPTIMIZING WORKER PERFORMANCE: Voice-Directed Operations in the Warehouse. [Online] Available at:
<http://www.voxware.com/fileadmin/media/resources/white_papers/WhitePaper_OptimizingWorkerPerformance.pdf> [Accessed 20 April 2013]

Heaney, B., (2011). Voice Technology Meets the Ecommerce Distribution Challenge
<http://www.voxware.com/fileadmin/media/Voice_Pioneers/Aberdeen_report-_Voice_Users_Speak_Out.pdf> [Accessed 15 May 2013].

Heaney. B., & Pezza. S., (2010). A Sound Decision: Utilizing Voice Technology for Improved Productivity and Quality. Analyst Insight. [Online] Available at:
<http://www.spiritdatacapture.co.uk/papers/aberdeen-group-utilizing-voice-technology-forimproved-productivity.pdf>[Accessed 20 April 2013]

Vitechgroup (2009). Vocollect ® T5 Fact Sheet. [Online] Available at:
<http://www.vitechgroup.com/web/solutions/voice/voicehardware/Vocollect_Talkman_T5_Series.pdf>[Accessed 20 April 2013]

Voxware (2012). VOXWARE VOICE MANAGEMENT SUITE: Product overview. [Online]
Available at:
<http://www.voxware.com/fileadmin/media/resources/product_info/VMS_Product_Overview.pdf>
[Accessed 20 April 2013]

APPENDIX A Inventory of existing SR platforms

Inventory of existing SR platforms
The appendix was designed for the purpose to get overview of SR platforms and developer toolsets that could potentially be integrated in the Novacura Flow application. Table 1 in this appendix consists of a selection of SR systems platforms currently available and some toolsets for integration of SR and its development. It is a small sample of what is available. For more detailed description of other SR solution providers are listed in greater detail by Wikipedia (2013). The distinction that can be made to the toolsets in Appendix A is that Appendix A shows toolsets where tools can be bought in order to program with a voice platform and this is from a professional use for warehouse environments. Appendix A is only a selection of leading SR systems and a few toolsets that may be useful for future studies for the integration of SR into an application.

For the reader that wishes to gain a more detailed description the authors recommend reading the industry or research reports in the reference section of this appendix, in particular the report by Heaney (2012) and Heaney & Pezza (2010). Table 1 references some of the SR systems available from the larger AVIOS list of Tools for Developing Speech-enabled Mobile Applications available from the Avios (2005) linked reference.

Table 1 Inventory of existing SR platforms (Avios, 2005)

Platform & Company	Solution & hardware	System & its Advantages	System & its Drawbacks
Microsoft: Windows Windows CE Company **Intermec**: **Vocollect ®** SOURCE: http://www.intermec.com/Vocollect /index.aspx	VoiceCatalyst, Vocollect ® Voice-Applications, Vocollect ® VoiceConsol. **Hardware**: Vocollect ® Talkman Mobile Computing Devices and Voice-RecognitionOptimized Headsets	Hardware and software are from the same company. Off the shelf solution and allows for fast implementation. Robust hardware solution. Specialize in voice solutions Offers an Integrated Development Environment (IDE) software platform. VoiceArtisan enables customization and expansion capabilities of their Vocollect ® Voice Solution.	The solution is multimodal but has no inbuilt display. The customer is locked into the embedded hardware and software combination. Requires the use of expensive, proprietary accessories. Price for the voice device units are high. High price for modifications to the propriety software.
		VoiceArtisan is an open,	

		flexible, and extensible IDE that allows for customer defined solutions to adapt to changing business requirements. Vocollect ® supports scanners and displays for mobile computing devices from Intermec, Honeywell/LXE, and Motorola/Psion. Sorce:http://www.intermec.com/Vocollect/solutions/productsservices/integrateddevelopmentenviornment/index.aspx	
Windows Windows CE Company **Voxware** SOURCE: http://www.voxware.com/ourproduct/productoverview/	Voice Applications, Voice Toolset, Voice Architecture, Voxware Voice Management Suite (VMS), Enterprise Voice Architecture (EVA), VoxConnect, VoxStudio. **Hardware**: Any certified voice-capable devices. A list of these are available at SOURCE: http://www.voxware.com/fileadmin/media/resources/product_info/Voxware_Certs_Feb242012.pdf	Hardware independent solution and the customer is not locked into specific hardware. Specialized in voice solutions Voxware VMS is based on a unique three-level software architecture, with Voxware WMS at the top and supported by solutions that are based on configuration tools such as VoxStudio and VoxConnect are the middle layer At the base is the Enterprise Voice Architecture, a powerful software platform that makes Voxware VMS which creates an adaptable, portable, and scalable solution. Source:http://www.voxware.com/fileadmin/media/resources/product_info/VMS_Product_Overview.pdf	Require specialized training to create voice work processes in VoxStudio. The task of choosing the right mobile voice device requires specialized knowledge.
Windows Windows CE	Lydia® Voice-Suite **Hardware**: Voxter® Suite	Lydia®-PDA is a "pick by voice" with system world-wide support and one that works independent of the	The solution is not multimodal. The customer is locked

Company **topVox** SOURCE: http://www.topvox.com/lydiavoicesolutions.html		hardware. The hardware is standard PDAs from several manufacturers with a specialized SR software developed by topsystem for the voice picking.	into the hardware and software.
Windows Windows CE Company **Motorola** SOURCE: http://www.motorolasolutions.com/US-EN/Business+Product+and+Services/Mobile+Computers/Wearable+Computers	**Hardware**: WT4090 VOW Voice Only Wearable Terminal, RCH51 Rugged Cabled Headset	Large global organization Provide a large variety of hardware solutions including voice only or multimodal solutions.	Provides only the wearable computing device that can be used with the other SR systems like Voxware system.
Open source Company **Zanzibar OpenIVR** SOURCE: http://www.spokentech.org/index.html	Zanzibar OpenIVR Sphinx4 ASR engine FreeTTS speech synthesizer **Hardware**: None	Zanzibar OpenIVR is a complete, standards based, open source IVR. It has support for industry standards such as voiceXML. It can be integrated with a VOIP PBX (like Asterisk) using SIP and RTP VOIP standards You can use openIVR to voice enable your existing Asterisk PBX. You can deploy Asterisk and openIVR together to voice enable your web applications. You can deploy and run VoiceXML applications.	It is only a development tool for SR. Used mainly for dictation and continuous speech recognition used in many areas for dictation like the health care services Not used within warehouse voice peaking

		You can write and deploy speech applications in java using the MRCP4j library. Source: http://www.spokentech.org/ind ex.html	
Open source Company **CMU Sphinx** SOURCE: http://cmusphinx. sourceforge.net/ wiki/	Sphinx(1-4) CMUSphinx toolkit PocketSphinx **Hardware**: None	CMU Sphinx has many useful tools for SR for developers Flexible development toolset.	Development tool only for SR.
Android Company **NDEV** SOURCE: http://www.nuanc emobiledeveloper. com/public/index. php?task=howI tWorks	NDEV Mobile **Hardware**: None	NDEV Mobile provides mobile application developers the opportunity to integrate their Nuance's renowned and proven cloud-based Dragon Mobile speech recognition and text-to-speech capabilities. Developers access our speech services through a variety of platform SDKs as well as via a web services interface. Registration to the program is free - so don't wait and start speech-enabling app your today!	Not specific for voice pick Development tool only for SR.
Android Apple ® IOS Company **iSpeech** SOURCE: http://www.ispee ch.org/ DEMO: http://www.ispee	iSpeech **Hardware**: Android or iOS device	Speech-Enable Your Android® iSpeech Text to Speech (TTS) and Speech Recognition SDK for Android lets you Speechenable any Android App quickly and easily with iSpeech Cloud. The SDK has a small footprint and supports 27 TTS and ASR languages and 15 for free-form dictation voice recognition. Last updated: 03/20/2013.	Set of development tools only Requires developer or specialist knowledge to use. Not specific for voice pick

ch.org/text.to.sp eech		You can use iSpeech to convert text to speech, convert documents to speech, convert web content to speech, or convert blogs to speech. For personal use, iSpeech's simple online text to speech software tool supports over 20 languages. Developers can take advantage of iSpeech's free text to speech (TTS) SDKs for major mobile platforms. Human quality TTS is available for iPhone, Android, and BlackBerry. iSpeech SDKs are tested for rock-solid results with our iSpeech Cloud.	
Windows Phone 8 Company **Microsoft** **Tellme Speech** SOURCE: http://www.micros oft.com/enus/Tell me/business/appli cations.a spx	Windows Phone 8 Microsoft Tellme Speech **Hardware**: Windows Phone	The SR solution is part of a large global company. Microsoft provides Speech WinRT APIs to allow third party developers take advantage of the Microsoft Tellme Speech Platform. These APIs enable two primary types of speech functionality for third party applications: (1) integration into the global speech experience, and (2) rich in application "dialog" using speech recognition (with cloud based web search and short message dictation grammars) and speech synthesis (male and female voices) in multiple languages. The developer capabilities are directly integrated within the Windows Phone Developer Platform with appropriate extensions to	Not specific for voice pick

Table 1 provides a sample of the mobile platforms that is relevant to the thesis and the development tools can be used in future studies. In table 1 the leading platform is the windows CE-Windows mobile platform and the other platforms IOS, Android trying to catch up. Vocollect ® provide embedded systems and are one of the leaders in the area of SR systems in WMS

and provide a turnkey solutions for Windows CE platform, they use devices that are built specific for SR in noisy environments. One example is the Vocollect ® T5 which runs on the Windows CE platform. Windows CE is commonly used in embedded systems like the T5 as it is lightweight and requires very little processing power and yet is flexible. For the purpose of copyright the diagram of the Vocollect ® architecture is available for inspection from the referenced internet link at Intermec (2011) diagram of product architecture. The Vocollect ® VoiceApplications can be viewed from the referenced link at Intermec.com (2011) diagram voicetasks voiceapplications.

Voxware, have created a hardware-independent voice software product, The Voxware Voice Management Suite (VMS). VMS is a flexible solution where the customers is not locked into software that can only run on one device but has a Mobile Device Adaptability scheme which gives the customer can choose from a list of available devices which can then be integrated with a 30-step process for certifying voice-capable mobile computers. Voxware has implemented android integration which is one of the user requirements that are stated in the requirement section, in their existing Voice solution. The referenced diagram gives an overview of the Voxware solution and is available in the reference section at Voxware (2011).

Nuance Communications are leaders in dictation or continuous SR and can be used in conversional SR and for professional needs and is available on the following mobile platforms, Windows OS, Android and iOS. Nuance Communications is a company that provides solutions for a diverse range of SR domains such as Nuance Healthcare, for the health care industry. Nuance Dragon dictate is one of the leading dictation software solutions for speech to text.

One open source platform solution is the Zanzibar IVR which is based on dual server architecture with Java speech resources on a cairo MRCPv2 server. Zanzibar IVR has VXML in its architecture which is a W3 standard for voice browsing as shown in (Figure 1) the Zanzibar OpenIVR architecture.
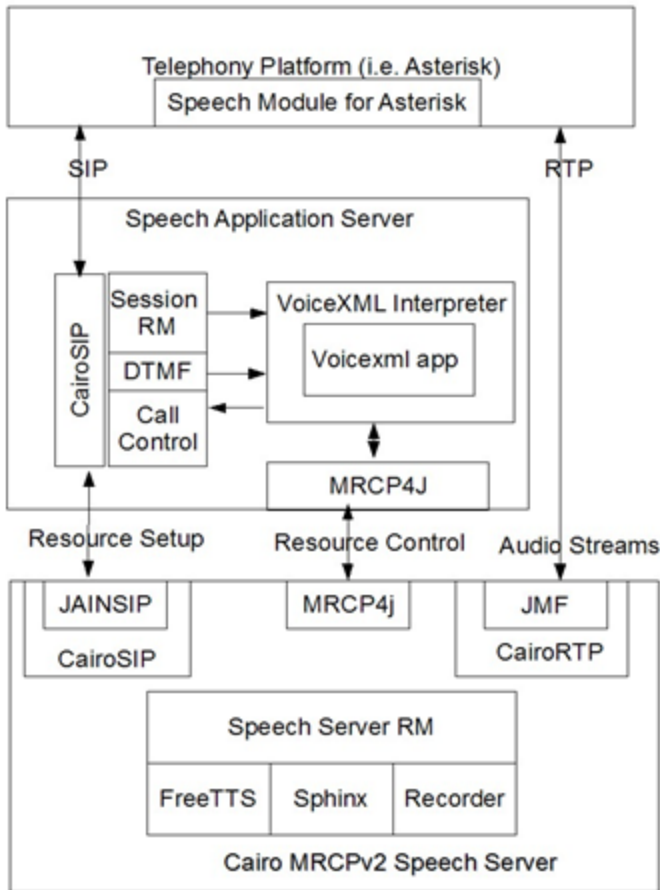
Figure 1 Zanzibar OpenIVR architecture (Zanzibar OpenIVR (2011) image source: http://www.spokentech.org/images/StackDiagramWithVoicexml.png)

From this inventory it the reader can choose the embedded SR systems that are good at what they do. Even though cost of each system is not included in Table 1 it give a short overview and outlines the possibilities for professional SR solutions in a warehouse environment. The challenge for future studies is to implement useful and professional SR technology to the standards of Vocollect ® with voice mapping and implement this into an existing application that can run several mobile devices. One issue is that of patents on the SR technology that is held by many of the leading actors and is a limiting factor for new actors wishing to enter the noisy environment of the VDW field.

Internet links for Appendix A

Avios.org (2005). Tools for Developing Speech-enabled Mobile Applications. [Online] Applied Voice Input Output Society (AVIOS). Available at:
<http://www.vitechgroup.com/web/solutions/voice/voicehardware/Vocollect ®devices.asp>
[Accessed 22 May 2013]

Motorolasolutions.com (2013). Wearable Computers. [Online] Motorola Solutions, Inc.
Available at: <hhttp://www.motorolasolutions.com/US-EN/Business+Product+and+Services/Mobile+Computers/Wearable+Computers> [Accessed 22 May 2013]

Intermec.com (2011). Diagram of product architecture. [Online] Vocollect ®, Intermec Inc. Available at: <http://www.intermec.com/Vocollect ®/public-files/diagrams/diagram-productarchitecture-overview-web.pdf> [Accessed 22 May 2013]

Intermec.com (2011). Diagram voicetasks voiceapplications. [Online] Vocollect ®, Intermec Inc. Available at: < http://www.intermec.com/Vocollect ®/public-files/diagrams/diagram-voicetasksvoiceapplications-web.pdf> [Accessed 22 May 2013]

Heaney, B., (2011). Voice Technology Meets the Ecommerce Distribution Challenge <http://www.voxware.com/fileadmin/media/Voice_Pioneers/
_Voice_Users_Speak_Out.pdf> [Accessed 15 May 2013].

Heaney. B., & Pezza. S., (2010). A Sound Decision: Utilizing Voice Technology for Improved Productivity and Quality. Analyst Insight. [Online] Available at:
<http://www.spiritdatacapture.co.uk/papers/aberdeen-group-utilizing-voice-technology-forimproved-productivity.pdf> [Accessed 20 April 2013]

Voxware.com (2011). VOXWARE VOICE MANAGEMENT SUITE Product overview. [Online] Vovware Inc. Available at: <http://www.voxware.com/fileadmin/media/resources
/product_info/VMS_Product_Overview.pdf> [Accessed 22 May 2013]

Wikipedia.com (2013). List of speech Recognition Software. [Online] Available at:
<http://en.wikipedia.org/wiki/List_of_speech_recognition_software> [Accessed 10 April 2013]

Zanzibar OpenIVR (2011). Zanzibar OpenIVR Project. [Online]
Available at: < http://www.spokentech.org/architecture.html > [Accessed 22 May 2013]