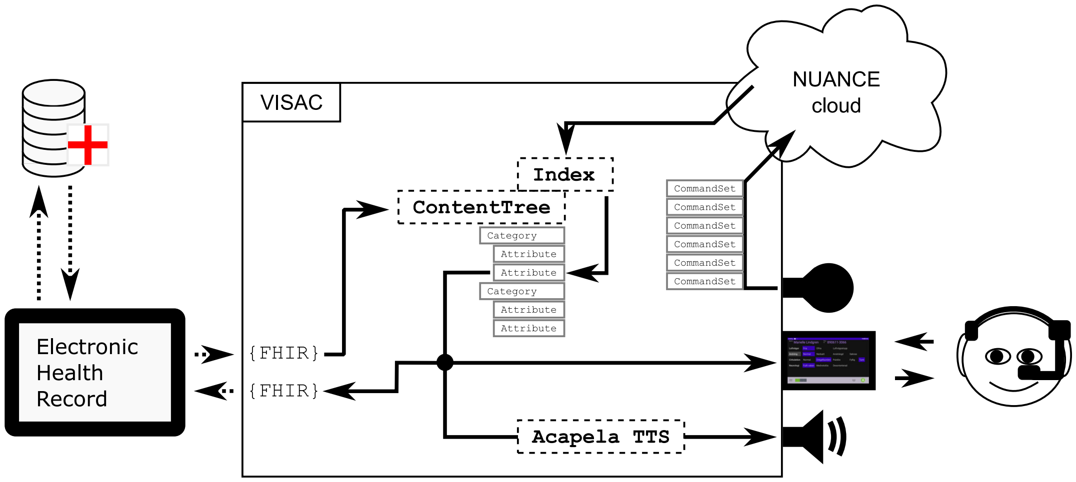
CHALMERS
UNIVERSITY OF TECHNOLOGY



# VISAC - Towards a Voice Interface for Swedish Ambulance Care

Master's thesis in Biomedical Engineering

ARTO KEMPPAINEN

DEPARTMENT OF ELECTRICAL ENGINEERING
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2022
www.chalmers.se

Master's thesis 2022

# VISAC - Towards a Voice Interface for Swedish Ambulance Care

ARTO KEMPPAINEN

Department of Electrical Engineering
*Division of Biomedical signals and systems*
Remote and Prehospital Digital Health
Chalmers University of Technology
Gothenburg, Sweden 2022

VISAC - Towards a Voice Interface for Swedish Ambulance Care

ARTO KEMPPAINEN

Main supervisor: Dr. Stefan Candefjord
Associate Professor, Department of Electrical Engineering
Chalmers University of Technology

Co-supervisor: M.Sc. Anna Bakidou
PhD Student, Department of Electrical Engineering
Chalmers University of Technology
and Faculty of Caring Science, Work Life and Social Welfare
University of Borås

Examiner: Prof. Bengt-Arne Sjöqvist
Professor of Practice emeritus, Department of Electrical Engineering
Chalmers University of Technology

Master's Thesis 2022
Department of Electrical Engineering
Division of Biomedical signals and systems
Remote and Prehospital Digital Health
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 (0)31 772 1000

Cover: An overview of the VISAC system. A more detailed description can be found in Chapter 4.

Gothenburg, Sweden 2022

# Abstract

**BACKGROUND** Ambulance Personnel (AP) work in a complex and dynamic environment and must make clinical decisions in a broad range of clinical cases. Patient information, as well as time, is often scarce and access to clinical expertise limited. Increasing demands on providing early clinical care and prepare hospitals with information on incoming patients, can be assisted with decision support and a more efficient documentation process. Commercially available devices for medical documentation and provision of decision support use point-and-click interfaces. Speech recognition and natural language processing can provide a hands free and natural way of quickly entering information into computer systems.

**OBJECTIVE** Leveraging commercially available hardware and software in constructing a pilot system to explore the current feasibility of a Voice Interface for Swedish Ambulance Care (VISAC).

**METHODS** Hardware and software components were explored and selected. A "primary survey" application was developed in Android Studio. Speech recognition capabilities were evaluated at noise levels between 30 and 93 dB(A) and usability was evaluated with AP at Örebro University Hospital by System Usability Scale questionnaire and interviewing.

**RESULTS** VISAC was constructed using the head mounted display HMT-1 (RealWear, Inc., United States) and the Dragon Medical SpeechKit (Nuance Communications, Inc., United States), with the application of performing an ambulance care primary survey. A command-based approach using a limited vocabulary was chosen. Testing at noise levels found inside an ambulance indicated recall of command intent of between 50 and 75 % , about 85 % in quiet conditions. AP user evaluation highlighted deficiencies in rejection of external speech, need for a natural and extensive vocabulary, as well as need for a method of determining whether speech is directed towards VISAC.

**CONCLUSION** A voice-only pilot system for in-field recording of electronic health record data was constructed. The available clinical vocabulary needs to be expanded to allow realistic evaluation in primary survey scenarios – preferably in concordance with a coding such as SNOMED CT. Tests indicate robustness to noise to at least 65 dB(A). Rejection of external sources of speech and a method of determining when VISAC is spoken to is needed for feasibility in clinical practice.

Keywords: Natural language interface, Head mounted display, Emergency medical services, Clinical decision support systems, Voice user interface, Augmented reality, Swedish ambulance care.

# Acknowledgements

I would like to express my deepest thanks to Dr. Stefan Candefjord, Prof. Bengt-Arne Sjöqvist and M.Sc. Anna Bakidou who have been my incessantly supportive supervisors through this thesis work. They have shared their valuable suggestions and ideas and have always met my more or less informed ramblings with positivity and interest. Anna has given many good suggestions on the structure and presentation of the thesis, which I am thankful for.

I would like to offer my very great appreciation to specialist nurse in ambulance, Erik Höglund at Örebro University Hospital ambulance department. I am immensely grateful for his help with visits to the ambulance department, letting me ride along in the ambulance, and gathering participants for the usability tests. Erik has graciously shared his effort and knowledge and has met me with nothing but enthusiasm and charity throughout the thesis work.

I am also very grateful to other ambulance personnel at Örebro University Hospital ambulance department who have lent me their invaluable perspective in testing VISAC and shared their experiences and expertise.

My special thanks are extended to Nuance for bestowing me with a license for their SpeechKit, and to the kind people working there: Torbjörn Hurtig who assisted in navigating products and services offered, Marcus Agehall who helped me getting started with his technical support and Loredana Cerrato who has offered solutions and suggestions whenever we have spoken.

I wish to acknowledge the help provided by Anders Stöt at Omilon who dispensed generously of his long experience on Swedish speech recognition software.

I must also thank Acapela Group for their kindness in lending me a software license for Acapela TTS.

Finally, I wish to thank my lovely wife, Ida, for her unwavering support and encouragement throughout this process.

Arto Kemppainen, Örebro, March 2022

# Contents

# 1

# Introduction

The context that prehospital emergency health care professionals work in is characterized by ongoing decision making in a dynamic and complex environment [1, 2]. Flexibility in planning and preparation is necessary [3], as well as a broad medical knowledge base and alertness to change [2, 4]. Rapid response to the patient, appropriate life supportive care, and prompt transport to appropriate definitive care are paramount to patient survivability [5].

As ambulance personnel (AP) are oftentimes the first point of contact between a patient and clinical care, information available at dispatch can be both limited and misleading: The report given by the one calling for help might be lacking [2] and a higher priority is often set by the emergency medical dispatch centre than what is assessed on site [6]. Urgency and time constraint, combined with the dynamic and uncertain work environment, may be contributing factors to clinical errors [7, 8]. Flawed decision making by emergency medical providers is thought to greatly add to patient harm [9]. Study of critical incident reporting has shown deficiencies in staff knowledge and experience to be the major cause for critical incidents [8].

The information that AP gather in field, is used both to provide the patient with immediate care, as well as to decide on the next care instance and allow them to prepare for the incoming patient. Demands on AP competence has been, and is, increasing both in Sweden [10, 11, 12] and elsewhere [10, 7] as the need for early clinical decision making is emerging in acute conditions such as sepsis and stroke. As a means of structuring the work process – in order to promote patient safety – clinical guidelines and protocols are widely used in prehospital care [13, 14]. Poor adherence to such guidelines has been noted in studies in various types of clinical cases [15, 4]. The physical format of guidelines has been shown as an obstacle to explicit use in assessment of patients [16].

Decision support tools may improve diagnostic accuracy as well as decrease the time to definitive care [17]. Computer based decision support tools are being employed in healthcare and have shown varying effect on different aspects of clinical performance [18]. Increased compliance with guidelines has been reported with use of computer based decision support tools [15, 19].

Research has been conducted to identify features of clinical decision support systems important to clinical practice improvements. Decision support being provided automatically as part of clinician workflow, at the time and location of the decision and in the form of a recommendation, without need for additional data entry, have been found to be notable correlative factors [20]. Providing AP with computerised decision support tools able to adapt to patient data ensures individualised treatment while simultaneously standardizing clinical decision making [21], plausibly increasing patient safety.

For each patient encounter, an Electronic Health Record (EHR), is filled out in which the AP records observations and may recommend further treatment. EHRs allow for sharing of clinical patient information within the care chain [22]. Filling out EHRs has been shown to be time demanding [23] and is a distraction in the interaction between AP and the patient. In critical situations, AP often keep information only in their memory or write this down on available surfaces such as their surgical gloves, for later recording in the EHR. This conduct is seemingly necessary but detrimental to information completeness [24].

Commercial prehospital EHR solutions exist. Among these are the MobiMed healthcare solution (Ortivus AB, Sweden) and the medicalpad (WEINMANN Emergency Medical Technology GmbH, Germany). For military use, the Mobile Computing Capability (Joint Operational Medicine Information Systems, United States) has also been developed. These systems operate on handheld screens via point-and-click interfaces, or when applicable, by acquiring data from attached sensors. To date, 15 of Sweden's 21 regions use the MobiMed healthcare solution [25]. Among these are Region Örebro län. MobiMed is compatible with the HL7 FHIR interoperability standard. More on this in section 2.4.3.2.

The utilization of digital speech recognition tools for medical dictation, as a complement to medical amanuenses, is increasing in Sweden [26, 27], indicating the existence of sophisticated speech recognition tools capable of Swedish medical terminology. The research team "Remote and Prehospital Digital Health" at the department of Electrical Engineering at Chalmers University have identified voice interfaces to be momentous components for building real time digital decision support tools. This thesis is an exploration into that research area.

## 1.1 State of the Art

Computer systems for on-site recording of EHRs in prehospital care situations have been proposed several times over the last decades [24, 28, 29, 30]. A theme in the designs has been to distribute the system between a server that handles heavy computations, such as the speech processing, and one or multiple light weight client devices that present a form for the user to fill out using voice or point-and-click. The clients have been carried, strapped to the user's arm, or mounted to the users head.

Head Mounted Devices (HMDs), computers fitted to the head of the user with a screen in front of one or both eyes, saw a new era with the launch of Google Glass (Google LLC, United States) in 2013-2014. The Google Glass was small enough to be worn over long periods of time, had support for voice commands, featured a 5 MP/ 720 p camera and was internet connected. Applications for the medical field were promptly proposed [31]. Today the Google Glass is the most widely used HMD [32] and has been utilized in dozens of pilot studies for use cases in clinical care [33, 32]. Often specifically in operating theaters for displaying information or gathering images and video [33, 34]. Use in prehospital care has been explored for triage, remote consultation and displaying of information, with noted drawbacks of short battery life and unstable internet connection [35, 36].

The HoloLens 2 (Microsoft Corporation, United States) is a holographic mixed reality HMD that allows the user to interact with 3D holograms and thus provides a novel user experience. Microsoft advertises the device for use in clinical settings as well as for industrial remote assistance. The device has seen clinical use cases in surgery [37, 38], clinical education and remote clinical assistance [39], among others.

The HMT-1 (RealWear, inc., United States) is a fully voice controlled HMD primarily aimed towards industrial use. The Norwegian based company Jodapro AS is developing and testing a remote guidance solution for Norwegian prehospital care, at Innlandet Hospital in Norway since 2019 [40, 41].

Voice User Interfaces (VUIs) have become more prevalent in consumer products in the last decades. Notable examples are virtual assistants such as Siri for Apple operating systems, the multi-platform Google Assistant and the Amazon Alexa, as well as home automation products such as Google Home and Amazon Echo. Among these some are voice-only while others employ multiple modes of user input. Research specifically on design of VUIs for use in prehospital care is – to the authors knowledge – scarce. Examples of designs can be found in the larger body of pilot and case studies, that exist [33, 42, 29, 43, 35, 30].

Among studies of speech recognition tools for clinical documentation between 1990 and 2018, nearly half used products by Nuance Communications [44]. Nuance Communications and 3M M*Modal are the only providers of commercial speech recognition software with support for Swedish clinical terminology – to the author's knowledge – to date. Both provide software development kits: 3M M*Modal only for Windows platforms and Nuance Communications for Windows and Android platforms. Nuance Communications software has been reported in one study to achieve satisfaction among a majority (77 %) of clinical users, despite frequent need for manual corrections [45], and has shown promising recognition results for dictation (accuracy $>$ 90 %) in a project similar to VISAC [29]. The software is designed to be used for free text dictation but also supports listening for predefined commands.

## 1.2 Objective

The objective of this thesis is to explore the feasibility of a voice interface for Swedish prehospital settings, based on contemporary, commercially available technology components. This question is investigated by selection of system hardware and software components and subsequent design and construction of a demonstrator product which implements a “primary survey”, and on which performance tests are to be conducted.

## 1.3 Demarcations

To assess existing research in the field and to grasp what technology is available, a state-of-the-art survey is to be done.

The choice of hardware platform is limited to the RealWear HMT-1 and the Microsoft HoloLens 2 as the project has access to these HMDs and they can thus be evaluated first-hand. They are to be evaluated, and a choice made between them. Alternative hardware platforms will not be considered for use in the project, but will be considered in the literature research as this project is intended as a pilot project. If a better choice of hardware platform exists, that is of interest to the research group.

Software-wise, a speech recognition software and a text-to-speech software for the Swedish language are to be chosen for use in the system. These are needed for converting human speech to digital text, and the reverse. As the system is intended for use in a Swedish context, support for the Swedish language is central.

As for software development and testing of the finished system, an ABCDE “primary survey” is to be implemented and the resulting system evaluated on speech recognition accuracy in a noisy environment. The “primary survey” is widely used and is interesting for this pilot study as a simple model for filling out patient records. The intended use case – in ambulance care – demands functionality in noisy environments. Such testing is thus important to conduct.

# 2

# Theory

To grasp the development process and end result, knowledge of certain concepts in medicine, linguistics, language processing, programming and usability testing are helpful. This chapter is intended to give the reader an introduction to these disparate concepts.

## 2.1 ABCDE Primary Survey

Prehospital Trauma Life Support (PHTLS) is a program of best practice principles for prehospital trauma care, developed since 1981 by the National Association of Emergency Medical Technicians. The American born program has reached international recognition and is used in dozens of countries [46], including Sweden. The "primary survey" is used for VISAC at this early development stage as it is well known and relatively easy to implement and is thus a good example of VISAC's intended functionality. A production version of VISAC is intended to support all relevant ambulance protocols and surveys.

Providing prompt care for trauma patients lowers mortality, thus it is important for Emergency Medical Services (EMS) professionals to quickly get the patient to correct definitive care [5]. On the scene, ensuring perfusion and oxygenation, to maintain energy production in vital organs, is paramount [5].

The primary survey is a tool for EMS professionals to quickly assess the condition of a patient according to life critical parameters. These being, in increasing order of severity: free airways, proper ventilation and oxygenation, haemorrhage control, perfusion and neurological function [5]. In training, the survey can be performed step-wise as a structured protocol, but in practice observation and questions such as *What has happened to you?* gives the EMS professional information about several of the parameters. A patient that answers has respiratory function and airways free enough to allow speech, as well as cerebral perfusion enough to make themselves understood [5].

The mnemonic ABCDE is often used for the categories of the primary survey. According to PHTLS [47] and treatment guidelines for Swedish ambulance care [14] the categories and short descriptions are as follows:

**A - Airway management and cervical spine stabilisation** Open and clear airways without risk for obstruction are essential for keeping the patient oxygenated. The airways might need clearing of bodily substances like blood and mucus, or from external objects. The possibility of injury to the cervical spine needs to be kept in mind.

**B - Breathing (Ventilation)** To ensure oxygenation of the patient, oxygen needs to be delivered to the lungs. Either by the patient's own ventilation, or by assistance.

**C - Circulation (Haemorrhage and Perfusion)** Bleeding, internal or external, might demand treatment. Abnormalities in the pulse of the patient are assessed, as well as color, temperature and moisture of the skin.

**D - Disability** Altered level of consciousness or motor capabilities can alert the AP to problems with cerebral oxygenation, injuries to the central nervous system, substance overdose or medical conditions like diabetes or cardiac arrest.

**E - Exposure/Environment** During examination of the patient's body to find injuries, body heat should be preserved.

## 2.2 Ambulance Noise Environment

AP spend time in the ambulance on their way to and from a patient scene. On scene the work is typically conducted in the patient's home [48], but can also be outdoors or in a different indoors location. The working environments are diverse and can thus present a great variability in sources of noise. In for example traffic accidents, surrounding traffic as well as power sources driving hydraulic tools used in cutting open cars, can produce loud noise. Differences in ambulance vehicles and their equipment – between and within countries – naturally results in different noise levels inside the ambulance. In research from Poland and Denmark mean average background noise is reported below or at 80 dB(A) inside the ambulance [49, 50]. Much louder peaks, reaching 123 dB(C), as well as individual trips averaging up to 94 dB(A) have been reported in one Danish study [50]. Another study reports a German average of 85 dB(A) and a Latvian average of 87 dB(A) [51]. An important source of noise inside a vehicle is vibrations from contact between tyres and ground. The speed affects the noise levels inside the ambulance [51]. According to Swedish regulations, 80 dB(A) is the upper allowed limit for constant exposure during an 8 hour shift, without providing hearing protection [52].

The measurements dB(A) and dB(C) denote A-weighted and C-weighted dB measurements, respectively. A- and C-weightings are defined in IEC 61672-1 and used in Swedish and European noise legislation [53]. A-weighting models the human ear's response to sound, giving a lower weight to energy content of frequencies that the human ear registers less well [53]. C-weighting gives equal weight to energy content of frequencies within normal human hearing, with a sharp roll-off outside of this range [53].

## 2.3 Speech and Language Processing

Making computers respond to human speech in a meaningful way is a complex task involving both extraction of information from spoken language into a digital representation, and subsequent action on that received information. Text is a widely used digital representation for spoken language. Automatic transcription of speech into text is called Speech Recognition (SR).

### 2.3.1 Speech Recognition Technology

The rise of deep learning has been revolutionizing to SR [54]. Machine learning requires large annotated data sets for training, as the systems are fundamentally trained to recognize specified patterns. Different languages need separate training and within languages, different groups of people differ in for example dialect and vocabulary. These factors have to be considered for a well functioning SR system.

Long Short-Term Memory (LSTM) is a machine learning architecture that is well suited for sequentially correlated data, such as speech, where the context of one word depends on surrounding words [55].

Handling background noise has long been [56], and still is [57], a hindrance to effective SR. Variations in dialects and handling multiple simultaneous speakers [58] are other current major challenges in the field [57].

### 2.3.2 Natural Language Processing

In order to make any sense of a string of text, the contained lingual information needs to be extracted. Natural Language Processing (NLP) encompasses techniques for understanding, as well as producing human language, using computers [59].

Certain tasks demands pre-processing of the used texts. A customary pre-processing step for normalizing text in NLP is removing of stopwords: common words that carry little information [60]. Removing such words will not affect the intended message in a significant way, but will decrease the number of possible word combinations.

In the field of linguistics a lemma is the canonical form of a word [61], the one typically found in a lexicon. Lemmatizing words is a common way of normalizing text by discarding morphological information. Morphemes are the smallest meaningful units of language [62]. For illustration, the word “unrepentive” can be broken into the morphemes “un-repent-ive”. “Repent” is here the root, with a prefix “un” and a suffix “ive”. A similar effect to lemmatizing can be accomplished using stemming, that is, reducing inflected words to a stem by removing suffixes. Pre-processing text by removing stop words and lemmatizing or stemming achieves identity between e.g.

the phrases “the airways are cleared” and “airway clear”.

## 2.4 The Software Platform

The RealWear HMT-1 is a voice controlled Android device running Android 10 (Google LLC, United States). The device implements the concept “Say What You See”, adding an interface layer on top of the Android operating system, called WearHF [63]. The interface layer consists of graphical overlays numbering each selectable item on screen, as well as a set of voice commands for selection of these items.

### 2.4.1 Android Architectural Components

User applications for the Android operating system are typically built using the programming languages Java or Kotlin. For building applications for Android, a set of Application Programming Interfaces (APIs) are available. These APIs provide building blocks for the application and allow access to core operating system components and services [64].

Android applications are built around the concept of what is called an “`Activity`”, in order to facilitate applications invoking specific functionality of other applications [65]. This allows e.g. a web browser to invoke an email client’s “new mail”-dialog directly without other parts of the client such – as the inbox – being invoked [65]. An application needs at least one `Activity` and can be composed of several.

The `Activity` has a “lifecycle”. From creation to termination certain stages of the lifecycle are accessible to the programmer. These are important for initiating the application before displaying it to the user (via the lifecycle method `onStart()`), deciding what happens when other `Activities` are given focus, and when returning back from them (lifecycle methods `onPause()`, `onResume()`) and cleaning up services when the user leaves an application (lifecycle method `onStop()`) [65].

When building Graphical User Interfaces (GUIs) for Android, the `View` and `ViewGroup` elements are used [66]. By combining different `View` elements which are given attributes – like dimensions, color and text – and collecting them in `ViewGroups`, complex GUI elements can be formed. Among predefined such elements are `TextView` and `Table`, providing text fields and tables respectively. The more complex element `RecyclerView` is used for creating lists with dynamic content and has an API for handling the information that is to be shown. `View` elements can be declared in a layout-file using XML or be defined and created programmatically at runtime. Attributes of the created `View` elements can also be programmatically altered.

The Android framework directs UI controller lifecycles, dictating what `Activity` is displayed as a result of both user actions and device events [67]. If an `Activity` is destroyed and recreated, information will be reloaded from sources, possibly resulting in loss of information entered by the user. Such information thus needs to be saved in a lifecycle aware manner. This can be done using a `ViewModel`. The `ViewModel` object is accessible during the entirety of the `Activity` lifecycle; from it's creation to it finishing [67]. Declaring methods and data in the `ViewModel` rather than in the `Activity` class, also provides a means of source code task separation [67].

### 2.4.2 JSON Standard

JavaScript Object Notation (JSON) is a language-independent data interchange syntax standard also called ECMA-404 [68]. JSON is easy for humans to read and write and has gained widespread use. The standard "provides a simple notation for expressing collections of key/value pairs" [68] called *objects*, with keys being strings and values being one of *object, array, number, string*, `true`, `false`, or `null` [68]. Objects are delimited by curly brackets and arrays by square brackets. An example of a JSON object is shown in Figure 2.1.

```
{
  "object with an object as value": {
    "object with a string as value": "a string",
    "for multiple key/value pairs": "separate them by a comma",
    "objects and arrays nest well" : [
        {
            "using an array as a value is OK" : ["one", 23, true]
        },
        {
            "key" : "value",
            "key" : "value"
        }
    ]
  }
}
```

**Figure 2.1:** An example of a nested JSON object.

JSON objects can be interchanged between systems using plain text which is parsed on reception, according to the specific needs and data types of the receiving system. Thus these need not be known by the transmitting system.

### 2.4.3 Electronic Health Records

The emergence of Electronic Health Records (EHR) makes swift interchange of patient data between healthcare providers possible. Patients can also be offered access to their health records via the Internet [69]. Lack of widespread adoption to available standards is a hindrance to interoperability between healthcare information systems [70].

#### 2.4.3.1 SNOMED CT

Systematized Nomenclature of Medicine Clinical Terms (SNOMED CT) is a collection of clinical terms organized in hierarchies with an “is a”-relation from children to parents within the hierarchies [71]. There exists 19 hierarchies which include “clinical finding”, “procedure” and “body structure” as well as the possibility to describe relations between concepts [71]. The concepts defined in the hierarchies are each represented by a code. An example of a body structure code relevant for VISAC is “89187006 | Airway structure (body structure) |”. The numeral code is accompanied by a human readable description. For each concept code, synonymous descriptions exist, both within and between languages. Code 89187006 has the English synonyms “Airway structure”, “Airway” and “Respiratory airway” as well as Swedish synonyms “luftvägar, struktur” and “luftvägar”.

#### 2.4.3.2 FHIR Interoperability Standard

Fast Healthcare Interoperability Resources (FHIR) is an interoperability standard for description of data formats called “resources” as well as a specification for the exchange of these resources, developed and maintained by the non-profit standards development organization Health Level Seven (HL7) [72]. The standard has predefined resources and allows for creating custom resources with simple or complex data types. Resources can, according to the standard, be represented in different ways, among which are as JSON objects.

According to Release 4 of HL7 FHIR, the resource `Observation` can be used for “Measurements and simple assertions made about a patient, device or other subject.” [73]. A minimal example of an `Observation` is shown in Figure 2.2 [73]. The “status”-key of an `Observation` can take as value a valid `ObservationStatus`, e.g. “preliminary”, “final” or “amended” [74]. The “code”-key needs a `CodableConcept` to be specified as value. The `CodableConcept` can be in the form of a plain text description and/or an external clinical term coding according to clinical coding standards such as Logical Observation Identifiers Names and Codes (LOINC) or SNOMED CT. Such an example of a `CodableConcept` is shown in Figure 2.3.

```
{
  "resourceType" : "Observation",
  "status" : "preliminary",
  "code" : { CodableConcept },
  "valueString" : "value"
}
```

**Figure 2.2:** A minimal example of the resource `Observation`, with a string as value. The “code”-key needs the `CodableConcept` to be specified for completeness.

```
{
    "coding": [
        {
            "system" : "http://snomed.info/sct",
            "code" : "89187006 | luftvägar, struktur |"
        }, {
            "system" : "http://loinc.org",
            "version" : "2.40",
            "code" : "69046-1",
            "display" : "Airway status"
        },
        "text" : "Luftvägar"
    ]
}
```

**Figure 2.3:** A FHIR `CodableConcept` describing the object of observation in three different ways: Using SNOMED CT code "89187006 | luftvägar, struktur |", LOINC code "69046-1": "Airway status", and using a plain text description.

## 2.5 System Usability Scale

The System Usability Scale (SUS) is a widely used [75] 10 item standardized questionnaire. It was first developed in 1986 as a simple and low-cost means to assess perceived usability of industrial system user interfaces [76]. The questionnaire has since seen several variations and has been used for evaluating user interfaces of many sorts, ranging from pagers to web applications [77]. Research assessing the quality of the SUS questionnaire has found it robust and valuable on a broad spectrum of user interfaces [77, 78].

The process is the following [76]: For each item the user is asked to grade their agreement with the given statement on a scale from 1 to 5 without thinking for too long on each question. If the user is unable to answer one item, they are to be instructed to fill in the middlemost alternative. The questionnaire is to be filled out by the user after testing the system but before further discussions about the system. The SUS score is subsequently calculated according to eq.2.1 where $S_{odd}$ denotes the score given for an odd-numbered statement and $S_{even}$ likewise the score given for an even-numbered statement. The maximum score is 100.

$$score_{SUS} = 2.5 * (\sum(S_{odd} - 1) + \sum(5 - S_{even})) \tag{2.1}$$

Using the SUS as an absolute scale for assessing the usability of an interface has been explored [77, 78]. For one such scale, a score of 70 is proposed as passable, with scores above 90 indicating a "truly superior product" [77].

## 2.6 Speech Recognition Tests

In testing of speech recognition software, comparison is often made between the input string which is spoken – the “gold standard” – and the resulting SR output – the “hypothesis”. The comparison can be made in several different ways. Common measurements include the Word Error Rate and the Word Information Lost, which are described below. A way of measuring recall of intended actions is also defined as well as a variant of the Word Information Lost-metric, modified with intention to account for quality of the recognized information.

### 2.6.1 Word Error Rate

Word Error Rate (WER) infers word-wise errors made between a gold standard and a hypothesis. It is a common metric in measuring the performance of speech recognition software [79, 80]. The WER does not infer language understanding accuracy [81] but only describes a word-wise distance between two text strings [79]. The word-wise edit distance is the minimal sum of the operations deletion, insertion and substitution, needed in order to transform the hypothesis into the gold standard [82]. The WER is acquired by dividing this word-wise edit distance by the number of words in the gold standard, and can be calculated according to eq.2.2 [80], where D, I, S, respectively denotes the number of deletions, insertions and substitutions, with a total number of N words in the gold standard.

$$WER = \frac{D + I + S}{N} \tag{2.2}$$

It should be noted that the WER can be greater than 1. The measurement is not to be thought of as a true percentage [83] and cannot be used as a measure of absolute performance [83]. As an illustration, assume that the word “foobar” is mistakenly recognized as “foo bar”. This adds one insertion along with a substitution to the total. Mistakenly recognizing the word as “foodbar” would have resulted in only a substitution. The WER is said to not be D/I-symmetric [79]. This asymmetry can lead to WER-values $> 1$ if the proportion of such errors is large.

### 2.6.2 Action Recall

With multiple command phrases leading to the same program action being performed (i.e. selecting a specific item), incorrect recognition of a certain phrase can still result in the intended action being performed. For such a case, the WER would indicate error despite the intended action being achieved. To account for multiple command phrases achieving the same action, the measurement “Action Recall” ($A_{rec}$) was formed by the author.

Assuming all possible speech recognition results, $\mathfrak{T}$, and using a set $\mathfrak{P} \subset \mathfrak{T}$ of cardinality $|\mathfrak{P}| = n_P$ available phrases. Further assuming that each available phrase $u \in \mathfrak{P}$ has at least one program action $a \in \mathfrak{A}$ which is specified to be performed upon recognition, out of $|\mathfrak{A}| = n_A$ available program actions. Where $a(u)$ is a non-injective function $a : \mathfrak{P} \to \mathfrak{A}$. A vector $u^{gold\ standard}$ with elements $(u_1^{gold\ standard}, \ldots, u_{n_u}^{gold\ standard}) \in$

$\mathfrak{P}$ of $n_u$ uttered phrases results in some vector $r^{\,hypothesis}$ of $n_r$ speech recognition results with elements $(r_1^{hypothesis}, \ldots, r_{n_r}^{hypothesis}) \in \mathfrak{T}$. Given a vector $u^{(m)}$ of $\dim(u^{(m)}) = m$ uttered phrases, where each phrase $(u_1^{(m)}, \ldots, u_m^{(m)}) \in \mathfrak{P}$ lead to the same action $a(u_1^{(m)}) = \cdots = a(u_m^{(m)})$ , $a(u^{(m)}) \in \mathfrak{A}$ being performed. Misinterpreting one of these for any of the others will nevertheless give the intended result.

The A$_{rec}$ is then calculated according to eq.2.3 where $a$ is an iterator over the $n_A$ available actions of $\mathfrak{A}$. For each available action $a \in \mathfrak{A}$, $N_a = m$ is the number of unique phrases in $\mathfrak{P}$ that lead to program action $a$ being performed. $C_a = \dim(a(u^{\,gold\,standard}) \cap a(r^{\,hypothesis}))$ is the number of times an uttering of a phrase $u$ with intended action $a(u)$, resulted in action $a(u)$ being performed.

$$A_{rec} = \sum_{a=1}^{n_A} \frac{C_a}{N_a} \tag{2.3}$$

### 2.6.3 Word Information Lost

Word Information Lost (WIL) is an approximation of the more complex measurement Relative Information Lost (RIL) [79]. RIL can be calculated according to eq.2.4, where $H(Y|X)$ is the conditional entropy of the transmitted message $Y$ given that the received message $X$ is known. $H(Y)$ denotes the entropy of the transmitted message $Y$ [79].

$$RIL = \frac{H(Y|X)}{H(Y)} \tag{2.4}$$

The approximation WIL is based on counts of hits, substitutions, deletions and insertions rather than knowledge of the mutual information of a gold standard and its interpreted hypothesis [79]. Furthermore, in contrast to WER, WIL is D/I symmetric and gives a true percentage [83]. WIL can be calculated according to eq.2.5.

$$WIL = \frac{H^2}{(H+S+D)(H+S+I)} \tag{2.5}$$

#### 2.6.3.1 Word Information Lost with Simplified Words

In the use case of VISAC, all of the information contained in a given phrase is not equal. While WIL approximates the information lost between a gold standard and an hypothesis, this measurement is agnostic of the quality of the information contained, for example whether it results in the intended action or not.

The measurement WIL$_{simpl}$ denotes the Word Information Lost computed for a data set pre-processed by removing of stopwords and reducing each word to their word stem. This indicates how much of the information which is critical for conveying of the intended action, has been lost.

# 3

# Materials and Methods

The material for this thesis is, to a large extent, derived practically via programming and subsequent testing and discussion with intended users. Orientation in literature as well as manual pages for software and hardware have also been necessary. In this chapter the exploration and development are described together with the chosen hardware and software. As the HMT-1 was eventually chosen as hardware platform, this choice is assumed in several of the sections.

## 3.1 State of the Art Survey

In order to assess the current state of the art in the field of HMDs and VUIs in pre-hospital care, a literature survey was conducted using the Chalmers Library search tool as well as Google Scholar and Google Search. This survey aimed at finding prehospital applications for the hardware and software components chosen for this project. For all searches, sorting on "Relevance" was used and for the article searches, 40 of the highest ranked articles were regarded for each search term. The articles were screened by title and the abstracts of relevant articles were read. The article was then selected for reading or discarded.

When using Chalmers Library search, ebooks, books and patents were excluded in search results. As the state of the art was to be surveyed, only works published in the latest 10-year period were considered, that is, between 2011 and 2021. Search terms returning relevant articles are collected in Table 3.1 – 3.3. No relevant hits were found in any of the article searches for search terms: HMT-1, Dragon Medical prehospital, RealWear prehospital, Jodapro.

**Table 3.1:** Search terms, Chalmers Library. There may be some overlap between chosen articles for the different searches.

| Search term | Hits | Chosen |
|---|---|---|
| voice user interface | 95,783 | 8 |
| prehospital voice user interface | 845 | 8 |
| prehospital head mounted device | 804 | 0 |
| Human computer interaction prehospital | 1352 | 2 |

For Google Scholar only articles published 2016-2021 were considered.

**Table 3.2:** Search terms, Google Scholar.

| Search term | Hits | Chosen |
|---|---|---|
| prehospital voice user interface | 859 | 7 |
| RealWear prehospital | 3 | 1 |

For JMIR publications, searches for hmt-1 or realwear returned no interesting results.

**Table 3.3:** Search terms, JMIR publications.

| Search term | Hits | Chosen |
|---|---|---|
| head mounted display prehospital | 3592 | 7 |
| voice user interface prehospital | 10079 | 3 |

For Google Search two pages (ca. 20 hits) of the highest ranked hits were considered. Relevant web pages were browsed and relevant links from these were followed. A few web pages were collected and a few articles were also found in these searches. Due to the disparity in web page content and structure, it is hard to treat these searches systematically. Search terms included: prehospital voice user interface, head mounted display prehospital, RealWear prehospital, HMT-1, jodapro.

This survey provided a foundation for further literature investigation and was a starting point to branch out from. Interesting sources referenced in articles found in the survey were studied, and further searches based on key-words and concepts were conducted. Apart from this focused survey, articles and literature have been accumulated and studied during the remainder of the project. Such accumulated knowledge about the state of the art in the field has also been added to the description, which is presented in Section 1.1.

## 3.2 Study Visit

A day was spent visiting the ambulance department at Örebro University Hospital, accompanying a crew on their shift. The aim of the visit was to get first hand experience of the work environment of the AP. Notes were taken on activities that the crew perform, technical equipment they use and the general sound environment.

## 3.3 Target Device Selection

The hardware and software components were evaluated with a wider use case in mind than the subsequently designed application. This was because at the time of making a decision for hardware platform, the exact application was not decided on. The research team had access to the RealWear HMT-1 as well as the Microsoft HoloLens 2 and a choice was made between these devices. The Vuzix M400, M4000 devices were considered in literature as they were the closest market alternative in early fall of 2021. Google Glass was also considered in literature as this device has seen use in clinical applications.

### 3.3.1 Important Capabilities

With the wider goal of adding decision support capabilities to a voice controlled, HMD for use in Swedish ambulance care, important parameters for a suitable target device can be identified: It needs to be rugged to handle adverse weather conditions with a wide temperature operating range. It needs good voice control capability with noise suppression for use in noisy environments with several consecutive voice sources. The display interference on the user's field of view should be kept low and the display needs to function in differing lighting conditions. The device should not hinder the communication between AP and patient and should be comfortable to wear during physical activity.

### 3.3.2 Available Devices

The choice of target device was limited to HoloLens 2 and HMT-1 from early on. Other HMDs used in clinical and industrial applications were explored to get an understanding of what alternatives exist. The devices are described below.

#### 3.3.2.1 Microsoft HoloLens 2

The Microsoft HoloLens 2 is a holographic, mixed reality device running Windows 10 Holographic OS, adding a 3D desktop environment to the Windows 10 operating system. The display is comprised of two waveguides – one in front of each eye – which gives the user a 3D holographic experience by displaying separate 2D holograms before each eye. Parallax differences between the two images allows the user to experience the projected holograms as if they were present in three dimensions in the room. The device is primarily controlled by hand tracking with gestures but does have voice-control capabilities and can track the users gaze as a third mode of input [84].

Wearing the HoloLens 2, the user can place 3D objects or flat application windows around themselves in space. Application windows can be docked to a cylinder around the user so that they are located at the same relative position, as the user moves about. A depiction of the device being worn can be seen in Figure 3.1.

**Figure 3.1:** A person wearing the HoloLens 2 together with glasses.

The HoloLens 2 has cameras in front of the device which register the geometry of the room that the user is located in. The HoloLens 2 creates a room model using this data. The boundaries of this room model function as boundaries for where objects can be placed.

The 3D environment offered by the HoloLens 2 is a novel feature which could lend itself well to innovative applications. The applications are developed in real time 3D platforms such as Unity or the Unreal Engine [85]. For this project an appealing benefit would be the availability of practically cost free virtual screen surface for displaying information.

#### 3.3.2.2 RealWear HMT-1

The RealWear HMT-1 is a fully voice controlled Android 10 device with a viewfinder beneath either eye. The viewfinder is mounted to an arm which can be adjusted according to where the user wants to have the screen located, normally below the dominant eye. The relative size of the display is about that of a 7-inch display at one arms distance. A depiction of the device being worn can be seen in Figure 3.2.

The HMT-1 can be used with different straps for mounting on the head, and is PPE (Personal Protective Equipment) compatible, allowing for clipping onto for example hard helmets.

**Figure 3.2:** A person wearing the HMT-1 together with glasses.

The user navigates Android applications using the principle of "Say What You See". Menu alternatives are numbered and can be selected using the command *Select Item N*, or by saying the title of the menu alternative. This way, many Android applications can be navigated without any prior modifications. Specific commands can also be defined programmatically in the source code of an application using RealWears WearML application programming interface. RealWear reports accurate voice recognition even in 95 dB(A) of typical industrial noise [86].

#### 3.3.2.3 Other Device Alternatives

The Vuzix M400 and M4000, as well as the Google Glass are other interesting HMTs, all running the Android operating system. The three devices are similarly designed and are made to be worn mounted to a frame, such as glasses. They have a touchpad on the side and can be interacted with using voice. The M400 has an opaque display while the displays of the M4000 and Glass are translucent. All have cameras.

The Vuzix devices are IP67-certified [87, 88]: dust-proof and resistant to being submerged at 1m depth for 30 minutes. The Glass is IP53-certified [89]: resistant to limited dust ingress and water sprays. The M400 is safe for drops from 2 meters [88] and the M4000 from 1 meter [87]. The Vuzix devices have hot-swappable batteries and can be used with different types of batteries depending on battery life needs, 2-12 hours [87, 88].

### 3.3.3 Comparison of HoloLens 2 and HMT-1

Both the RealWear HMT-1 and the Microsoft HoloLens 2 are targeted towards and used in various industrial and medical applications, among which are video calls to remote experts. Both devices can be used while wearing prescription glasses. Table 3.4 shows a comparison of device specifications given by the manufacturers.

**Table 3.4:** Comparison of specifications for the two devices HoloLens 2 and HMT-1.

| | HoloLens 2 [84] | HMT-1 [86] |
|---|---|---|
| Camera | 8 MP with 1080p30 video | 16 MP with 1080p30 video, 4-axis optical image stabilization |
| Speakers | “built-in spatial sound” | mono speaker, 3.5mm audio jack |
| Microphone | 5 channel array | 4 microphones with active noise cancellation |
| Battery life | 2–3 hours | 8-10 hours, hot swappable |
| Weight | 566 g | 380 g |

#### 3.3.3.1 Robustness

The HMT-1 is IP66 and MIL-STD-810G certified and thus dust-tight, waterjet-proof, drop proof from 2 m onto concrete and can handle temperatures between -20 °C and 50 °C [90]. The Industrial edition of HoloLens 2 is certified for use in clean rooms according to the ISO Class 5.0 classification [84], but does not have any Ingress Protection (IP) certification and is thus not well suited for outdoor use. The device is not built to handle drops and bumps.

#### 3.3.3.2 Noise rejection

Both devices have noise cancellation using multiple microphones which are either located to pick up speech or background noise [91, 92]. RealWear reports their voice recognition working accurately at 95 dB(A) of industrial noise [90]. Microsoft have not released any such data.

#### 3.3.3.3 Display

When wearing the HoloLens 2, the user’s field of view is partly obstructed by components in the visor. The visor itself is semi-transparent and thus reduces the amount of light that reaches the eyes of the user. The experience is that the view is constantly obstructed, apart from when flipping the visor up. The waveguides do not cover the field of view of the user but are limited only to a frame. The user needs to turn their head rather than move their eyes to look at different holographic objects.

The display of the HoloLens 2 is sensitive to lighting conditions. The holograms are “diluted” by sunlight and other light sources and can become invisible on a sunny day. The possibility of changing the intensity of the holograms exists, but in sunny conditions the available range is not sufficient. Light phenomena like glares and colourful rays are visible in certain lighting conditions, further distracting the user and making the displayed contents unclear.

During use of the HMT-1, frequent small adjustments of the viewfinder are needed. Between standing up and bending over, the angle at which the screen is viewed is altered enough to require manual adjustment of the viewfinder to keep the display in view. The viewfinder uses optics to give a fixed focus of 1 meter. This helps in keeping external light from reaching the display – effectively increasing contrast – but also limits the angle at which the display can be observed. A light source to the side and behind the user can introduce glares in the display.

As the viewfinder is located in front of one eye only, the part of the field of view covered by the viewfinder can still be seen with the free eye, making the display appear to be transparent.

#### 3.3.3.4 Camera

The positioning of the cameras differ between the two devices. The camera of the HoloLens 2 centered between the eyes and captures accurately what the wearer is seeing, while that of the HMT-1 is positioned slightly to the side of the head.

## 3.4 Software Component Selection

In the process of evaluating the HMDs, speech recognition and text-to-speech software compatible with the platforms were also sought out.

### 3.4.1 Nuance Dragon Medical SpeechKit

Nuance is a market leader in speech recognition [93]. Information found on the internet regarding available products was often incomplete and even misguiding. Contact with the sales department at Nuance made it clear that the SpeechKit SDK was available both for Windows and Android. Nuance Dragon Medical SpeechKit was chosen as it was the only commercially available speech recognition software with support for Swedish medical terminology known to the author at the time of making a choice. It is also the only such software with an API for Android. 3M M*Modal Fluency offers only Windows support. Nuance provided a test license for this project.

### 3.4.2 Acapela TTS

The text-to-speech software needed to produce a naturally sounding Swedish voice and have an SDK for Android. The Acapela TTS (Acapela Group, Europe) software provides a natural sounding synthesized Swedish voice and Acapela provided a test license for the project. The Android operating system has native support for text-to-speech, but does not provide a Swedish voice.

## 3.5 Developing VISAC

VISAC was developed in an explorative manner as availability and functionality of software and hardware components was not known at project start. One objective was to choose such components and evaluate if they were fit for the purpose of constructing a voice driven interface for Swedish prehospital care.

During evaluation of the devices HoloLens 2 and HMT-1 – prior to making a choice between the devices – cross-platform development was tried using the cross-platform development platform Xamarin 2.0 (Xamarin, United States). Xamarin was used with the integrated development environment Visual Studio (Microsoft Corporation, United States) and code was written in C#. Simple applications using the standard libraries worked well, for deployment both to Android and Windows. Platform specificity was introduced when adding custom WearHF functionality for the HMT-1. Adding speech recognition and text-to-speech tools for the specific platforms would have led to more platform specificity, defeating the purpose of using Xamarin. This approach was thus abandoned.

After choice of HMT-1 (model Model T1100G) as target device, Android Studio (Google LLC, United States) was used for development. This is an integrated developer environment geared towards the Android operating system. Android applications are developed mainly using the programming languages Java and Kotlin. Java was used, as the author was familiar with this language.

### 3.5.1 Development Process Overview

The development was conducted in two main segments; one preparation segment where the focus was on deciding on platform and seeing to that hardware and software components functioned together and requisites were met for the application development in the subsequent segment. This second segment consisted of three “development loops” in which the main application development was performed and feedback from AP was gathered along the way.

For Android programming, the Android operating system provides a development paradigm of its own, and uses APIs for adding predefined software components. Development guides are available and were used in familiarization with the development paradigm and Android concepts, in order to gain an understanding of available functionality. This familiarization was done step-wise, incrementally introducing functionality that was needed for the primary survey-application while simultaneously learning how to apply new elements and solving conflicts between different system components. The application programming process was reverted several times, as new insights were gained about software design problems. The version control system git was used for easier reversion. Android development guides, SDK specific developer documentation, customer service, as well as `stackoverflow.com` was frequently inquired for help during development.

On acquiring the SDKs for SpeechKit and Acapela TTS, functionality was tested on the HMT-1. The HMT-1 being voice controlled posed integration problems with SpeechKit. A way of switching between the two modes of voice control is described in section 4.1.2.1.

A method of handling recognized user speech was needed. SpeechKit allows two “modes” of speech recognition: “dictation” and “commands”. When the user selects a text field, speech is dictated into the field. During dictation, SpeechKit also listens for pre-defined commands. Recognition of such a command takes priority, and the application executes the actions that are predefined for the given command. Use of commands was chosen for controlling VISAC as this allowed for quick mapping of spoken words into actions.

Certain functionality was assumed to be critical for constructing the final application and was achieved before introducing AP in the development process. This included integrating the SpeechKit and Acapela TTS with the HMT-1, and writing back-end code for generating menus from a JSON-specification, further described in Sections 4.1.4 and 4.1.5. Adding support for HL7 FHIR using the SDKs “HAPI FHIR” [94] or “Android FHIR” [95] was examined but abandoned due to lack of time, as external communication was not an integral part of the project.

In the second segment, the three development loops spanning three weeks each were defined for developing the final application after finishing the preparations described above. In Loop 1, an application exhibiting a minimal primary survey, was produced and the vocabulary for the application was defined. “Vocabulary” here refers to words used for menu categories and words used for selectable attributes for those categories. The vocabulary was selected by studying choices in other primary survey-applications. An ambulance nurse at Örebro University Hospital provided a list of common terms used in a primary survey situation, from which words were also selected. In Loop 2, the minimal application from Loop 1 was tested by AP and comments and suggestions were used for further development through Loop 3. Tests for evaluating the final application were planned and methods for generation of command phrases – described in Section 3.5.2.1 – were explored. In Loop 3, the VISAC system was finalized and tested with AP. The testing is described in Section 3.6.

### 3.5.2 Application Design

The application functionality was specified to be “filling out an ABCDE primary survey” as this is an example of a structured survey. For such a survey, the categories for which parameters are to be reported, are specified by the acronym ABCDE, described in section 2.1. Measurements to take for each of the categories and the contents of the categories needed to be decided. The choice was made to use a shallow tree structure with categories that have a number of selectable predefined attributes each, as this functionality was expected to be sufficient for a simple primary survey and for testing the speech recognition capabilities. For the contents of the survey to be easily exchangeable, a JSON declaration of the tree structure defining them was developed. The declaration was constructed as a naive FHIR resource in order to prepare for interoperability with FHIR-compatible systems. The

VISAC application was, in this way, designed for displaying of consecutive predefined surveys that the user can respond to.

The first version of the user interface, shown in Figure 3.3, used for showing the VISAC concept to AP, employed two `RecyclerView` elements, one for displaying categories, and one for displaying attributes in the currently selected category. After the first evaluation with AP, an alternative design which constantly displayed all categories with their available attributes was suggested. A larger vocabulary and the ability to select and deselect several attributes for each category was also requested.

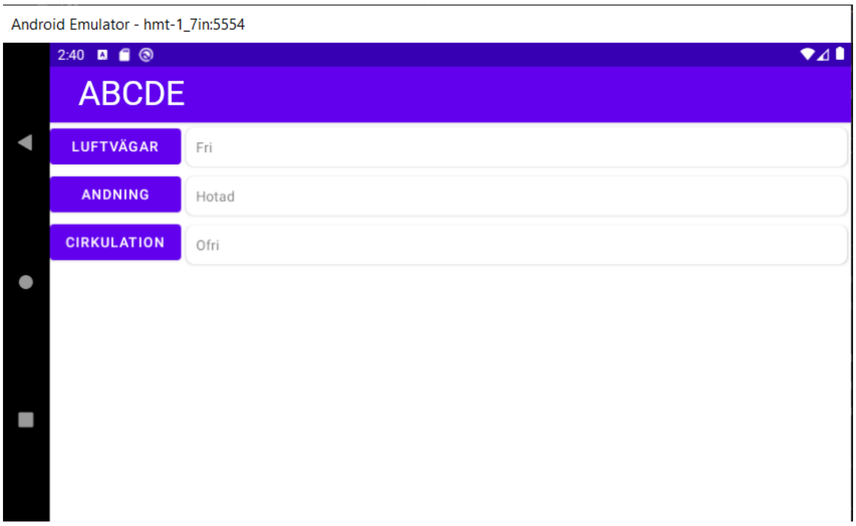


**Figure 3.3:** The first version of the visual user interface that was shown to AP, here run in an emulator. Only three categories are shown, with three selectable attributes. More categories and attributes were available at time of demonstration to AP.

#### 3.5.2.1 Generation of Commands

Command phrases were selected by finding common ways to combine the available category and attribute words into selection phrases. To exemplify, the attribute "'fria" in the category "luftvägar" was combined into phrases such as "fria luftvägar", "luftvägarna är fria" and "luftvägar fria". During Loop 1, AP suggested prepending the phrases "patienten har", "vi har" and "det är" for the command phrases. This was thus done. The generation of commands was done semi-automatically with use of Python for scripting as well as search-and-replace. A Python script generated phrases from word stems by appending relevant suffixes to these. The list of generated command phrases was then manually screened for incorrect phrases, which were removed. Complete sets of commands for some attributes followed the same structure and only needed replacement of one word between them, in which case, search-and-replace was used. An exhaustive list of selection command phrases that were generated for use can be found in Appendix A. For deselecting previously selected items, a separate set of commands was created for each attribute item. This set was constructed by prepending the three strings “avmarkera”, “avregistrera” and “ta bort” respectively, to each existing command.

An effort was made to find an automatic way of generating the commands, before choosing the method described above with Python scripting together with search-and-replace. The most promising among these was morphological analysis and generation. This was tried using the Python package UralicNLP [96] which has support for Swedish. Too many special cases would have needed handling for this approach to be implementable in the available project time frame because of how UralicNLP outputs data in morphological analysis. Other software projects and methods were considered but often did not support Swedish. At a later stage, it was found that Språkteknologigruppen at Kungliga Tekniska Högskolan provide an API for inflection of Swedish words [97]. This is not a complete solution but could prove useful for morphological generation.

## 3.6 Testing of VISAC

The HMT-1 was evaluated with respect to device ergonomics and physical factors as a part of choosing the target device. See Section 3.3 for this evaluation.

### 3.6.1 Recognition of Command Phrases

In order to evaluate the speech recognition capabilities of the VISAC system, tests of command phrase recognition were performed by a user speaking them to the VISAC system in different noise conditions. The noise levels chosen for the tests are enumerated in Table 3.5. The level 80 dB(A) was chosen as it has been reported in ambulances in Poland and Denmark, as well as this being a limit for allowed constant exposure during an 8 hour shift in Sweden. 95 dB(A) was chosen as RealWear reports HMT-1 rejecting noise to this level. The test rig could only produce 93 dB(A), so this level had to be used instead. After choice of 80 and 95 dB(A), the last level was calculated as $80 - (95 - 80) = 65$ dB(A), in order to get one more data point somewhere below the ambulance noise level of 80 dB(A). 30 dB(A) was the background noise measured in quiet conditions.

**Table 3.5:** Background noise levels chosen for command phrase recognition tests. 95 dB(A) could not achieved with the test rig used, thus 93 dB(A) was used instead.

| Description | dB(A) |
|---|---|
| room background noise level | 30 |
| $80 - (95 - 80)$ | 65 |
| level expected inside ambulance | 80 |
| HMT-1 reported noise rejection | 93 (95) |

### 3.6.1.1 Data Acquisition

A test rig was formed by connecting a Galaxy Note10 Lite (Samsung, South Korea) Android phone running a white noise generator application (Noice, version 1.3.3 [98]) to a set of stereo speakers of the brand Creative, model 265 (Creative Technology, Singapore). The speakers were placed 120 cm apart, facing each other, with the user's head centered in between them. Figure 3.4 describes the test setup graphically.

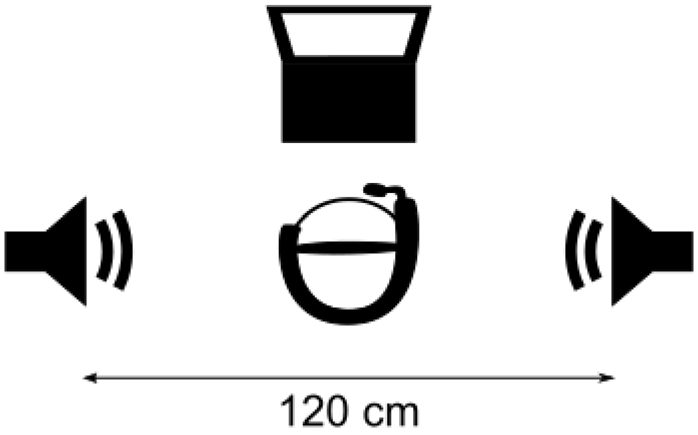


**Figure 3.4:** Test setup for noise tests. The user is sitting down between the loudspeakers, with a computer screen in front of them which displays the phrases to be spoken.

The sound level of the white noise was measured using the application "NIOSH Sound Level Meter" [99] (version 1.2.5) on a 2:nd generation iPhone SE (Apple, Inc., United States). Correct sound level was determined by holding the meter where the user's head would be located, while adjusting the sound level to the specified value. NIOSH Sound Level Meter is produced by the US agency *The National Institute for Occupational Safety and Health* and is tested and validated to be accurate to within ±2 dB(A) [100]. The app has been reported accurate within ±0.7 dB in the range 40 - 85 dB without calibration [101].

All 240 generated commands available to the user (excluding commands for deselection), were collected as gold standard in a list to be spoken in testing. Tests were conducted both for SpeechKit commands as well as SpeechKit dictation. For testing SpeechKit commands, the primary survey application was used. For testing SpeechKit dictation, a minimal application with a dictation user interface was constructed. The same sets of commands were enabled for both the applications. In order to simplify the process of testing, certain additional command phrases were added to VISAC. A command to handle registering of missed phrases (command phrase: *missade fras*) was available during testing of SpeechKit commands. A command phrase to print dictated text to the terminal (command phrase *skriv ut text*), was available during testing of SpeechKit dictation. Neither of these phrases were falsely registered when speaking a command phrase during testing.

Tests of SpeechKit commands were performed for all four noise levels while testing of SpeechKit dictation was performed only at levels 30 dB(A) and 80 dB(A). SpeechKit dictation was not to be used in the application but was tested to get a comparison of performance between the two modes and to see if use of free text dictation is viable.

As SpeechKit automatically trains a voice model during use, previously unused voice models were activated for the tests. Voice models were switched between using dictation or commands, but not between voice levels, as the number of available voice models too small to allow for this.

In testing, the user was sitting down with their head located in between the speakers. Each phrase was displayed on a screen in front of the user and phrases were read out loud in order, waiting a reasonable amount of time for the system to register a phrase. The waiting period between phrases was not timed. Experience on how long the system needs for a registration to complete was instead used for the waiting period. All registered phrases were printed to the system log. If a phrase was not registered, this was noted in the log. On completion, the information from the log was transferred to a text document and cleared from log information in order to only retain the registered phrases, separated by line breaks. The data was cleared from white spaces – except single spaces in between words – and all phrases were converted to lower case. Line-by-line alignment of the hypothesis to the gold-standard was done manually using the diff tool Meld [102] (version 3.20.4) when needed.

#### 3.6.1.2 Data Analysis

A Python script was written for the data analysis computations. For calculating WER and WIL the Python package jiwer [103] (version 2.3.0) was used. Stemming words and removing of stopwords for the $\text{WIL}_{simpl}$ calculation, was done using the Python package nltk [104] (version 3.7) with the Snowball stemmer [105] (version 2.2.0). Code for calculation of $\text{A}_{rec}$ according to eq.2.3 was written in Python. To achieve this, each phrase in the gold standard was manually labeled according to their intended action. From this mapping, labels were transferred to the hypothesis. Empty lines and phrases not contained in the gold standard were given an own label. This grouping of phrases by label according to intended action, allowed for checking whether registered phrases led to the action intended by the uttered phrase.

Figure 3.5 displays the Python code used to calculate $\text{A}_{rec}$ from the labeled data. The calculated parameter values were collected in Tables 4.1 and 4.2 and are shown plotted in Figures 4.9 and 4.10. A number of examples of misinterpretations made by the VISAC system were also collected in Table 4.3.

```python
    nbr_intended_actions_recognized = 0

    for i in range(len(reference)):
        if (reference_labeled[i] == hypothesis_labeled[i]):
            nbr_intended_actions_recognized += 1

    A_rec = nbr_intended_actions_recognized/len(reference)
```

**Figure 3.5:** Python code used to calculate $A_{rec}$ from data labeled according to intended action.

### 3.6.2 Usability Testing

The usability testing was performed during one day at the Örebro University Hospital Ambulance Department where six ambulance crew members participated in the testing. Participants performed the test procedure one at a time in a private room, together with the author. Each test session consisted of the following four phases:

**Familiarization with the device** The participants were given a sheet with instructions on basic usage and were then asked to familiarize themselves with the VISAC system. They could ask questions and get advise from the author during this familiarization phase. When they reported feeling comfortable using the system, the next phase was initiated.

**Performing of test cases** Four test cases, each briefly describing a patient contact, were provided with the instructions. These cases were now acted out, using the VISAC system to record clinical observations. The participants were asked to fill out the primary survey given the information provided for each case, according to their own experience. They were told that no observations were regarded as “correct” or “incorrect”. The author made observations on the process.

**Questionnaire: System Usability Scale** The participants were asked to fill out the System Usability Scale questionnaire after they had performed the four patient test cases. The participants were instructed to not think long before answering each question and also to fill in the middlemost value if they could not give an answer.

**Semi-structured interviews** In the last phase, semi-structured interviews were performed with all six participants. Interviews were conducted in Swedish and were recorded by taking notes. The interview questions were chosen with the goal of getting feedback on the user interface experience and use of the VISAC system in service. As a last part of the interviews, the participants were asked to try the VISAC system with TTS enabled. After filling out a few parameters, they were asked about their opinions of the system reading back the input and also if they would use this feature.

**Observation** Strategies for entering information and handling errors were observed while the participants performed the patient cases. Syntax used by AP in speaking command phrases was noted.

# 4

# Results

The VISAC system is described in this chapter in terms of design choices that have been made, together with results from the tests that have been performed on the system.

## 4.1 The VISAC System

The VISAC system was constructed using the head mounted display HMT-1 and the NUANCE Dragon Medical SpeechKit for speech recognition of Swedish medical terminology. A command-based approach using a limited vocabulary was employed. Acapela TTS was used for Swedish text-to-speech. The SpeechKit and Acapela Software Development Kits (SDKs) were installed according to instructions in documentation. An overview of the finished VISAC system is shown in Figure 4.1. The constituent parts will be further described below.

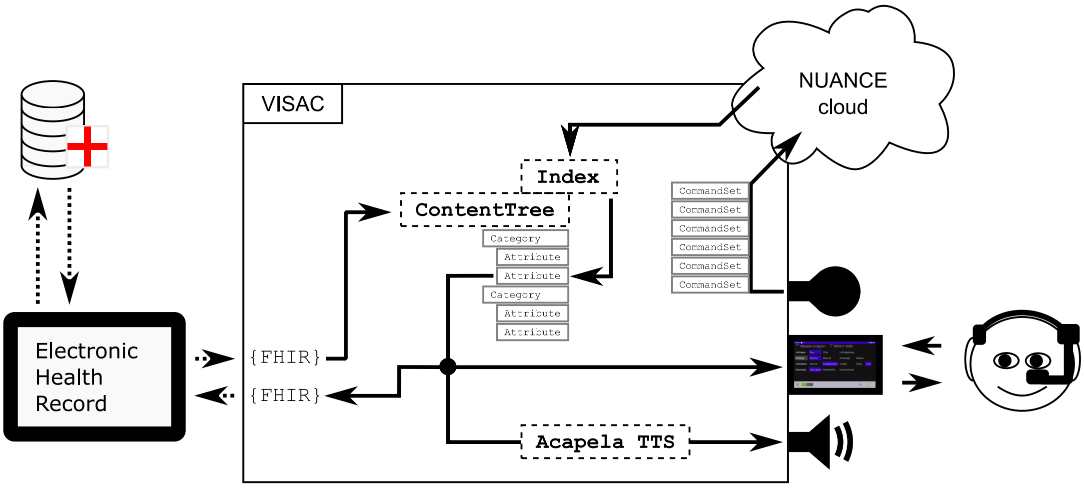


**Figure 4.1:** An overview of the VISAC system together with information inputs and outputs (dotted arrows not implemented).

### 4.1.1 Hardware platform

While the HoloLens 2 provides more screen surface, the HMT-1 provides a higher quality visual experience in terms of colour accuracy and screen flicker. The waveguide related visual distortions and the general obstruction of the field of view exhibited in the HoloLens 2 visor are not present in the HMT-1. The HMT-1 has a lower weight, a more robust build and a longer battery life, all three factors rendering it a more usable device for the application at hand. Both devices have noise cancellation. The HMT-1 being a fully voice controlled device, in contrast to the mainly gesture controlled HoloLens 2, is more in line with the intended use in this project. Neither of the devices have native support for Swedish dictation or Swedish text-to-speech.

### 4.1.2 Nuance Dragon Medical SpeechKit

The Nuance Dragon Medical SpeechKit for Android is delivered as an SDK in the form of a Java library. SpeechKit adds a control bar to the GUI (shown in Figure 4.2), which indicates status of the SpeechKit dictation and displays clickable items for control of SpeechKit functionality such as pausing dictation or viewing user help sections.

**Figure 4.2:** The SpeechKit control bar is added at the bottom of the speech enabled activity. The green meter to the left indicates that speech is currently registered. The color of the microphone icon to the right indicates whether the SpeechKit is listening for commands or not. Green indicates listening and red indicates not listening.

After installation, SpeechKit is automatically available if any text input field exists on screen and hidden when no text fields are available. Dictation is automatically paused after 20 seconds of silence, and then needs reactivation. Both deactivation and reactivation can be actuated manually as well as programmatically.

SpeechKit performs cloud based speech recognition, thus requiring an internet connection. Nuance specifies a need for approximately 10 kB/s in both directions to function adequately. The application registers with the cloud server by a user name, which is bound to a specific voice profile. These voice profiles are automatically trained during use of the service, adapting to the user's voice.

#### 4.1.2.1 Integration with the HMT-1

Integration of the “point-and-click” SpeechKit control bar, with the hands free voice controlled HMT-1, posed several problems. SpeechKit does allow for programmable voice commands but the recording of these commands is controlled via the SpeechKit control bar. The control bar in turn, is intended to be controlled via point-and-click, while the HMT-1 uses voice commands for all navigation. Because of this, switching was needed between the native HMT-1 voice control and the SpeechKit voice control, to avoid interference. Switching between these modes of voice control was done by implementing a voice command in each mode, which switches over to the other mode. The application is started in “SpeechKit-mode” and all application commands are implemented in this mode. “HMT-1-mode” is used for system-wide operations such as switching between applications. The only intended use of HMT-1-mode inside the application is to reactivate SpeechKit-mode when this has been paused.

The switching between RealWear-mode and HMT-1-mode is summarized in Figure 4.3. The HMT-1 device delegates control of the microphone using Android Intents. The Intent `ACTION_RELEASE_MIC` can be broadcast to force the HMT-1 to release the microphone, allowing other services to use it. Likewise, the Intent `ACTION_MIC_RELEASED` tells the HMT-1 to retake control of the microphone. These intents are broadcast in the SpeechKit callback methods `onRecordingStarted` and `onRecordingStopped`. In each mode a command is configured to allow switching over to the other mode. Android is also configured to handle gentle return to HMT-1-mode whenever the application is paused or closed. Without this configuration, closing the application in SpeechKit-mode would leave the user without voice control of the device, as the microphone is not controlled by HMT-1 commands.

As the application was to be controlled in SpeechKit-mode, the WearHF graphical overlay icons were deactivated in the application, to remove visual clutter. The soft keyboard is not used in the VISAC application and needed deactivation as it is otherwise displayed whenever a `TextView` is focused, taking up much of the screen space.

### 4.1.3 Information Handling

A `ViewModel` is created at application start and is used as a means for retaining and distributing application wide data during the application lifetime. In the `ViewModel` a `ContentTreeItem` is instantiated and data added from a locally stored JSON-description. The `ContentTreeItem` is a Java object and was designed as a representation of the state of selectable content in the application. An UML description of the `ContentTreeItem` is provided in Figure 4.4. The `ContentTreeItem` aggregates `MenuItem`s representing selectable categories, which in turn aggregate `DetailItem`s representing selectable attributes.

The `ContentTreeItem` is generated from a JSON object of the form described in Figure 4.5. Commands appertaining to the respective categories and attributes can be specified in the fields `commandPhrases` and `localCommandPhrases`. Commands in `commandPhrases` will be active during the entire lifetime of the application, while `localCommandPhrases` will be active only when the category is selected to which the given attribute belongs. This is necessary in cases where the same attribute exists in several categories.

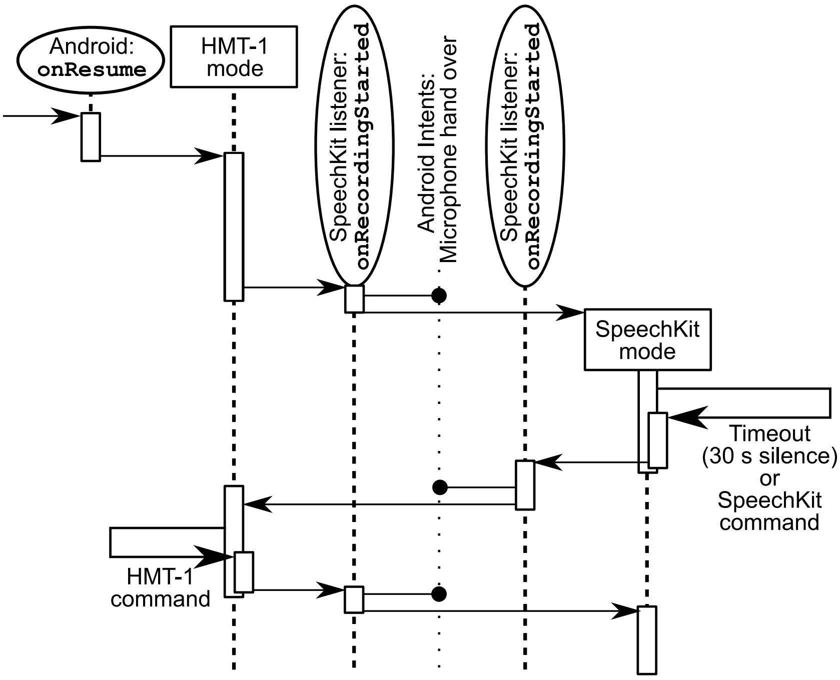


**Figure 4.3:** Sequence diagram describing switching between controlling the application in HMT-1-mode and SpeechKit-mode. The application will enter SpeechKit-mode at start and the user can then switch between modes. In HMT-1-mode, the application can also be closed, which is omitted from the diagram.

Each command in a `commandPhrases`- or a `localCommandPhrases`-listing will be collected in an `ItemCommandSet`. The `ItemCommandSet`s are then used to connect that collection of commands to the same intended action, e.g. selecting a specific available attribute. The `ItemCommandSet` is an adapter for the SpeechKit-object `CommandSet`. An UML description of the `ItemCommandSet` object is provided in Figure 4.6.

Commands for deselection are automatically formed when generating the `ContentTreeItem`. Each specified command is assigned three deselect commands by respectively prepending the phrases “avmarkera”, “avregistrera” and “ta bort” (meaning: “deselect”, “make unmarked” and “remove”) to the specified command phrase.

### 4.1.4 User Interface

The user interface has a voice component and a graphical component. The graphical component is exemplified in Figure 4.7 and is generated from the `ContentTreeItem` held in the application `ViewModel`. The graphical component of the user interface is only meant to give the user an overview of the current state of selections made, as

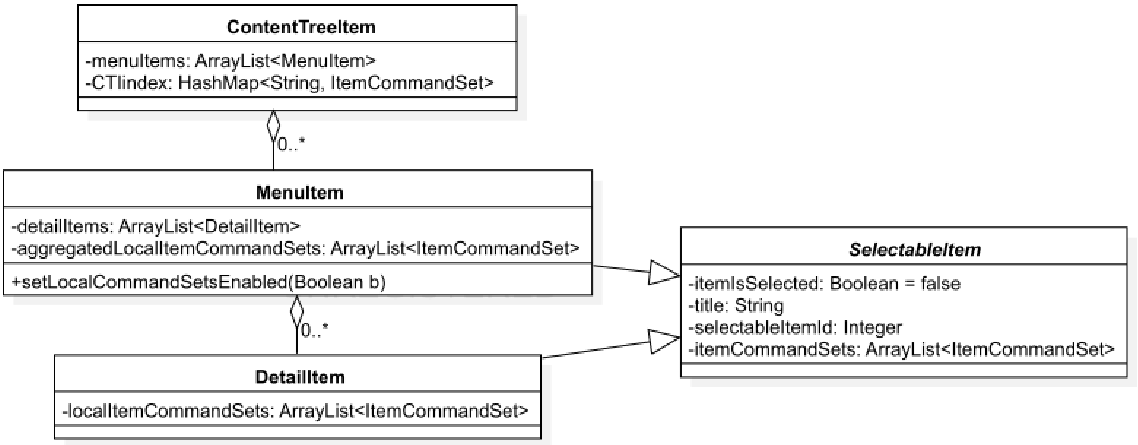


**Figure 4.4:** UML description of the `ContentTreeItem` object design. Get- and set-methods for private variables are omitted. "0..*" denotes aggregation of zero or more children.

```json
{
  "resourceType" : "ContentTree",
  "menuItems" : [{
    "title" : "Category 1",
    "commandPhrases" : ["", ""],
    "detailItems" : [{
      "title" : "Attribute 1:1",
      "commandPhrases" : ["", ""],
      "localCommandPhrases" : ["", ""]
    }, {
      "title" : "Attribute 1:2",
      "commandPhrases" : ["", ""],
      "localCommandPhrases" : ["", ""]
    }
    ]}, {
    "title" : "Category 2",
    "commandPhrases" : ["", ""],
    "detailItems" : [{
      "title" : "Attribute 2:1",
      "commandPhrases" : ["", ""],
      "localCommandPhrases" : ["", ""]
    }, {
      "title" : "Attribute 2:2",
      "commandPhrases" : ["", ""],
      "localCommandPhrases" : ["", ""]
    }
    ]}
  ]
}
```

**Figure 4.5:** JSON object description of a `ContentTreeItem` with two categories containing two attributes each. Command phrases are here indicated with empty strings for readability.

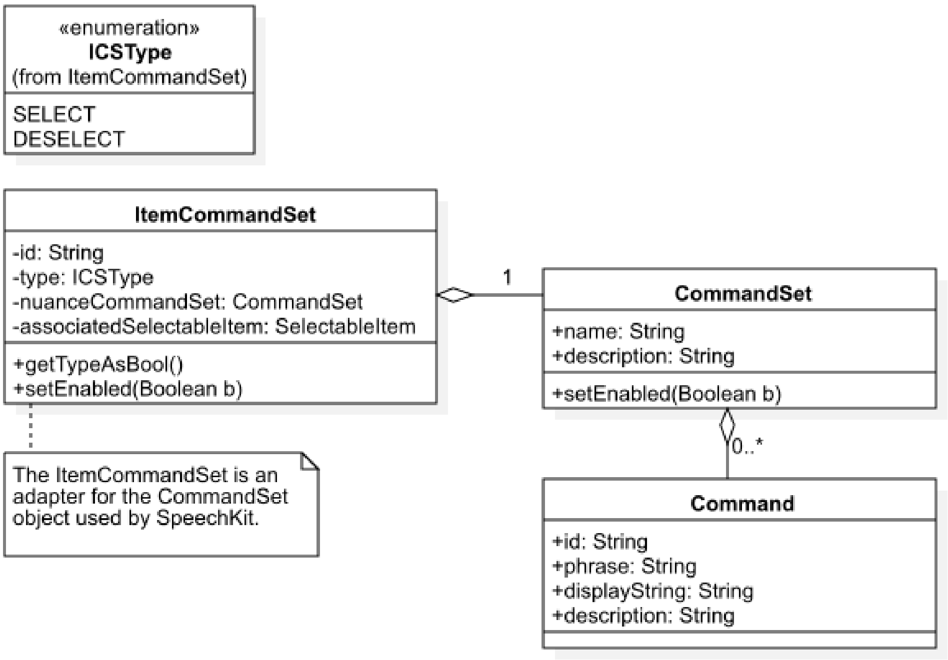


**Figure 4.6:** UML description of the `ItemCommandSet` adapter object. Get- and set-methods for private variables are omitted. For the SpeechKit classes `CommandSet` and `Command`, only relevant variables and methods are shown.

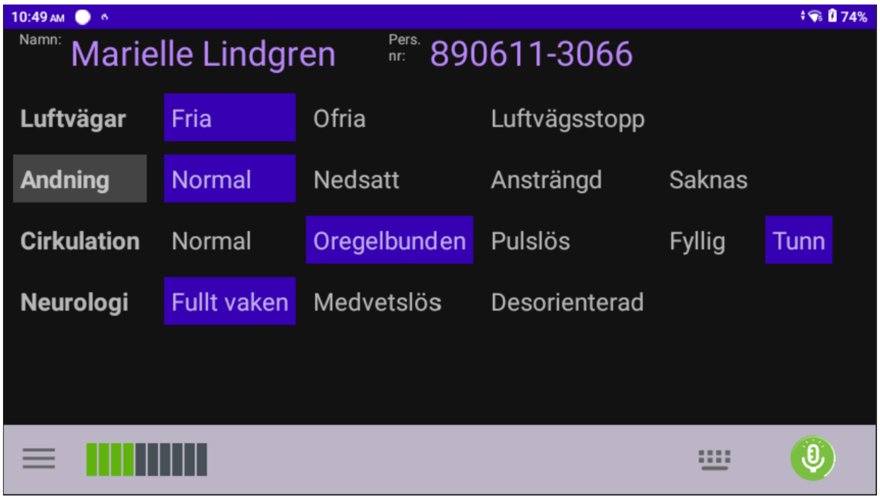
well as some additional information such as the name and personal number of the patient whose record is being edited.

Voice commands are used for selections and deselections. When a selection is made, this is recorded in the `ContentTreeItem` and the graphic user interface is updated together with a notification sound. The user can also switch mode to having the selected attribute be read out using TTS, instead of the notification sound being played.

### 4.1.5 Program Flow

At application start, the user is presented with a table specifying available categories and their respective attributes. The user can then utter phrases such as *det är stopp i luftvägarna* (the airways are occluded), *andningen är ansträngd* (the breathing is forced) or *patienten är fullt vaken* (the patient is fully awake) as well as simpler phrases akin to *cirkulation normal* (circulation ordinary). The complete list of available phrases can be found in Appendix A.

Output to and input from the user are local-only in the current implementation. The JSON specification of the `ContentTreeItem` as well as the patient information are stored on the device. In a production environment this would instead be loaded from a server, presenting the user with new content trees as selections are made and surveys finished. The user selecting an attribute currently results in a FHIR message (shown in Figure 4.8) being written to the terminal. After adding a FHIR payload


Namn: Marielle Lindgren

Pers. nr: 890611-3066

**Figure 4.7:** The VISAC graphical user interface. Selected category is marked with a gray background, selected attributes with a purple background.

to each selectable attribute, the respective payloads could be sent to a server on selecting an attribute, together with metadata such as a timestamp, SNOMED CT code, patient information and AP information.

```
{
  "resourceType" : "Observation",
  "status" : "preliminary",
  "code" : {(not implemented)},
  "subject": {
    "reference": "Patient/890611-3066",
    "display": "Marielle Lindgren"
  },
  "issued": "2022-03-12T16:54:12.458",
  "performer": [
    {
      "reference": "Practitioner/f005",
      "display": "First Last"
    }
  ],
  "valueString" : "Oregelbunden"
}
```

**Figure 4.8:** A mock-up of a FHIR message that could be sent to a server when selecting an attribute. The “code”-value could be e.g. a SNOMED CT code describing the category of the selected attribute.

## 4.2 Test Outcomes

Testing of the VISAC system was two-fold. Speech recognition capabilities were evaluated at noise levels between 30 and 93 dB(A) and usability was evaluated with

AP at Örebro University Hospital by questionnaire (System Usability Scale) and interviewing. The results are described below.

### 4.2.1 Recognition of Command Phrases

Speech recognition test results are presented below for each noise level, both when using recognition of commands and dictation. Table 4.1 presents values calculated for the complete data sets, 240 phrases (737 words in total, 48 unique words). Not every uttering of a phrase was recognized by the VISAC system, thus empty responses are registered for some of the phrases. The number of such missed phrases are noted in the column "empty responses". Table 4.2 presents values for the same parameters, calculated on filtered data sets, where phrases that returned empty responses have been removed prior to the calculations. The number of remaining phrases is noted in the column "non-empty responses". Figures 4.9 and 4.10 present the calculated values as plots. Because the parameters WER, WIL and $\text{WIL}_{simpl}$ quantify errors, a small value will denote a small error-percentage, meaning better performance. Keep in mind that $\text{A}_{rec}$ quantifies recall, meaning that a higher value will instead mean better performance.

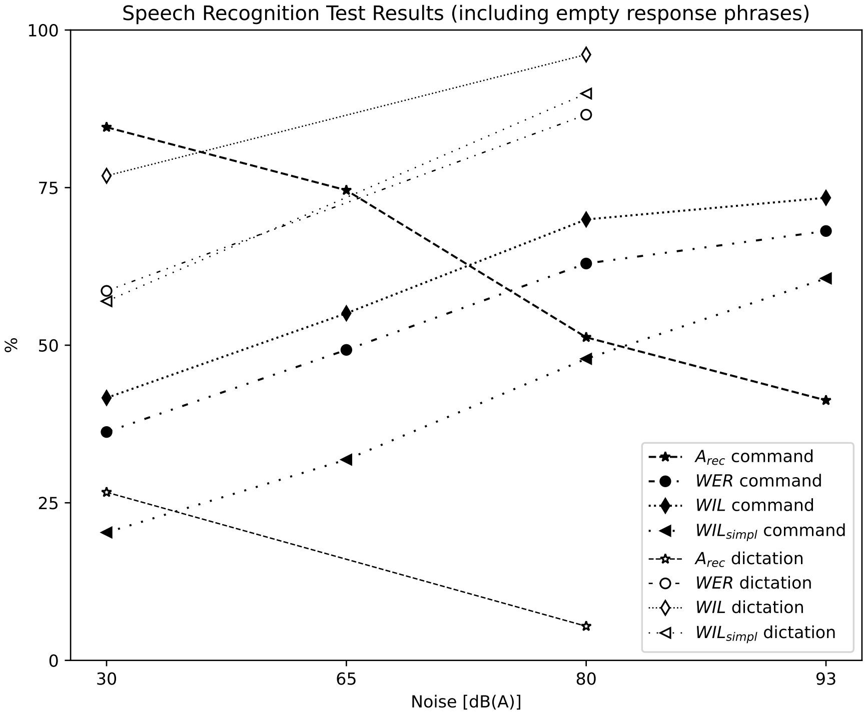


**Figure 4.9:** Speech recognition when including all phrases that were spoken but not recognized, and thus resulted in no response. Filled markers denote tests using SpeechKit commands and hollow markers tests using SpeechKit dictation. Note that only for $\text{A}_{rec}$, higher reported percentage means better performance.

**Table 4.1:** Speech recognition when including all phrases that were spoken but not recognized, and thus resulted in no response.

| | dB(A) | $A_{rec}$ | WER | WIL | $WIL_{simpl}$ | empty responses (of 240) |
|---|---|---|---|---|---|---|
| Using SpeechKit commands | 30 | 0.85 | 0.36 | 0.42 | 0.20 | 33 (13.8 %) |
| | 65 | 0.75 | 0.49 | 0.55 | 0.32 | 55 (22.9 %) |
| | 80 | 0.51 | 0.63 | 0.70 | 0.48 | 66 (27.5 %) |
| | 93 | 0.41 | 0.68 | 0.73 | 0.61 | 104 (43.3 %) |
| Using SpeechKit dictation | 30 | 0.27 | 0.59 | 0.77 | 0.57 | 0 (0.0 %) |
| | 80 | 0.05 | 0.87 | 0.96 | 0.90 | 36 (15.0 %) |

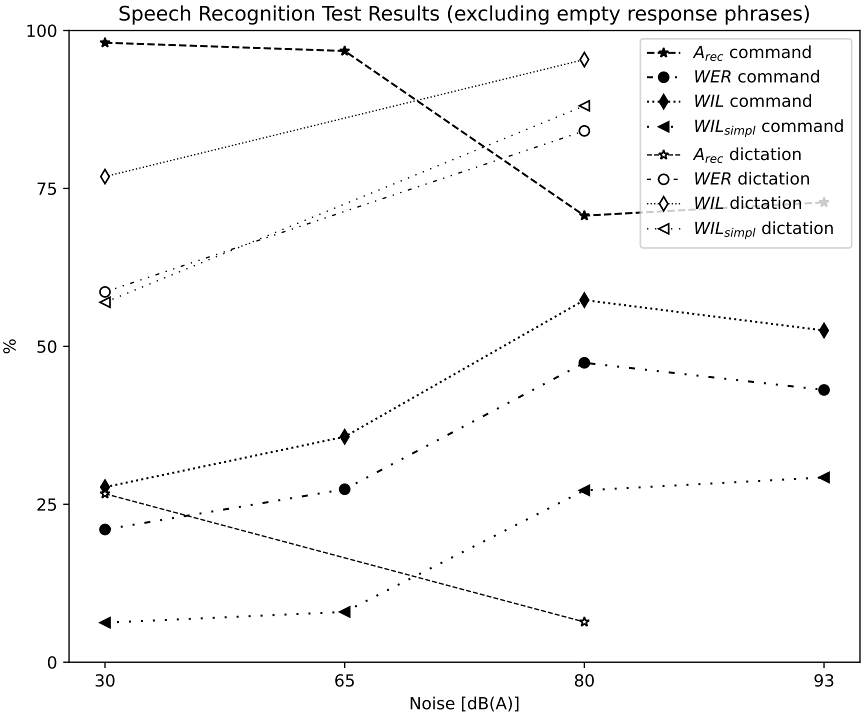


**Figure 4.10:** Speech recognition when excluding all phrases that were spoken but not recognized, and thus resulted in no response. Filled markers denote tests using SpeechKit commands and hollow markers tests using SpeechKit dictation. Note that only for $A_{rec}$, higher reported percentage means better performance.

**Table 4.2:** Speech recognition when excluding all phrases that were spoken but not recognized, and thus resulted in no response.

| | dB(A) | $A_{rec}$ | WER | WIL | $\text{WIL}_{simpl}$ | non-empty responses (of 240) |
|---|---|---|---|---|---|---|
| Using SpeechKit commands | 30 | 0.98 | 0.21 | 0.28 | 0.06 | 207 (86.3 %) |
| | 65 | 0.97 | 0.27 | 0.36 | 0.08 | 185 (77.1 %) |
| | 80 | 0.71 | 0.47 | 0.57 | 0.27 | 174 (72.5 %) |
| | 93 | 0.73 | 0.43 | 0.53 | 0.29 | 136 (56.7 %) |
| Using SpeechKit dictation | 30 | 0.27 | 0.59 | 0.77 | 0.57 | 240 (100 %) |
| | 80 | 0.06 | 0.84 | 0.95 | 0.88 | 204 (85.0 %) |

Some interesting and typical speech recognition errors from the noise test sessions are tabulated in Table 4.3. Results from all four noise levels are shown for commands and dictation respectively. Common errors include missing complete phrases and missing part of a phrase. Note that for commands, misinterpretations will result in blank fields if the misinterpretation does not exist in the set of available commands. For dictation, a misinterpretation will never be blank, as long as VISAC can hear the user speaking. This is also the reason as to why the percentage of empty responses in Table 4.1 is much lower for dictation than commands, and why some dictation results are deviating considerably from the gold standard.

**Table 4.3:** Registered phrases for all test cases when uttering reference phrases. Chosen phrases are selected to show errors of interest or typical errors found in the data set.

| Gold standard | Commands | | | | Dictation | |
|---|---|---|---|---|---|---|
| | 30 dB(A) | 65 dB(A) | 80 dB(A) | 93 dB(A) | 30 dB(A) | 80 dB(A) |
| vi har fri luftväg | | | ofri | | friluftsdag | jag friluftsdag |
| luftväg fri | luftväg fri | luftväg fri | luftväg fri | luftväg fri | luftfärd fri | i |
| ofri | ofri | ofri | | fri | ofri | ofri |
| vi har ofria luftvägar | | | luftvägar | luftvägar | ofria luftvägar | luftvägar |
| patienten har ofri luftväg | patienten har ofri luftväg | | patienten har ofri luftväg | | patienten har friluftsdag | patienterna har |
| vi har inte fri luftväg | | | inte fri | | det är inte friluftsdag | jag inte friluftsdag |
| luftvägen inte fri | luftvägarna inte fria | luftvägen inte fri | luftväg | luftvägen | luftvägar inte fri | det verkar inte bli |
| det är ej fria luftvägar | | | ej fri | ej fria | det är ju fria luftvägar | det är friluftsdag |
| luftvägar är ej fria | luftvägar ej fria | luftvägar ej fria | luftvägar ej fria | luftvägar ej fria | luftvägar i fyra | luftvägarna lite nya |
| luftvägen ej fri | luftvägen ej fri | luftvägen ej fri | luftvägarna | luftvägen ej fri | luftvägarna i frid | i |
| vi har stopp i luftväg | | | luftväg | neurologi | stopp i luftvägar | det luktar |
| det är stopp i luftvägarna | | andning | luftvägar | luftvägarna | det stopp i luftvägarna | stopp i luftvägarna |
| patientens andning är nedsatt | patientens andning är nedsatt | patientens andning är nedsatt | andningen nedsatt | | presenter samlingen nedsatt | patientens andning nästa |
| ingen andning | ingen andning | ingen andning | ingen andning | ingen andning | ingen aning | ingen aning |
| det är pulslös cirkulation | cirkulation | | | | det kostar cirkulation | twitter sticka |
| cirkulationen är pulslös | cirkulationen pulslös | cirkulationen pulslös | | | situationer / | situationen |
| vi har trycklös cirkulation | trycklös | trycklös | trycklös | | det cykla cykla gå | + |
| det är trycklös cirkulation | trycklös | trycklös | patientens luftväg fri | luftväg | uttryckliga cirkulation | det ryktas situation |
| fyllig | fyllig | fyllig | fyllig | | fyllig | i |
| cirkulation tunn | cirkulation tunn | cirkulation tunn | cirkulation | | situation tum | situation |
| vi har en medvetslös patient | medvetslös patient | medvetslös patient | medvetslös patient | medvetslös patient | jag är medvetslös patient | jag är medvetslös patient |
| desorienterad | desorienterad | desorienterad | desorienterad | desorienterad | desorienterad | |

### 4.2.2 Usability Testing

Tests were performed together with six ambulance crew members at Örebro University Hospital Ambulance Department during one evening shift. Responses have been degendered and shuffled for anonymization.

#### 4.2.2.1 The System Usability Scale

The SUS scores given by the six participants had a mean of 63.75 points and a median of 66.25 points out of 100. The individual scores are given in Table 4.4.

**Table 4.4:** SUS scores given by the six participants in the system testing session. Mean: 63.75, Median: 66.25

| participant | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| SUS score | 67.5 | 77.5 | 32.5 | 65.0 | 97.5 | 42.5 |

#### 4.2.2.2 Semi-structured interviews

The content of the interviews is recorded below. The notes that were taken have been translated and transformed for presentation by the author. The topmost response corresponds to the same respondent, and so on, for all questions.

**How natural did you find using VISAC?**

Very natural. The way of speaking is akin to what is used in training, where the ABCDE-procedure is used.

Completely natural.

Not very. Never used a system like this before. Would like better comfort and more experience with the device.

Not at all. It felt cumbersome. Would work in a calm situation but not in an hectic one. That would be stressful. Would want the device to be lighter and more as a part of the user, not as something one is wearing. Would also want more experience, to know what commands are available and how they should be spoken.

Not very natural as it is the first time. After a weeks practice one would know the application flow, what words work and what don't as well as how to act when something goes wrong.

Not very natural wearing something that is attached to oneself. Would rather have a separate device to speak to, e.g. mounted in the car. It gets in the way between oneself and the patient.

### What problems did you run into during use?

The vocabulary was lacking. Solved by choosing the closest option.

When you move around, the screen needs adjustments. The device falsely recognizes speech from surrounding speakers as commands.

The device did not recognise what was said. Changing the order of words did help.

Hard to see the screen. The available words and phrases were not natural. Normally one speaks more like “bedömer att luftvägen är fri”.

It was hard figuring out in what order to say certain words.

Not enough alternatives existed. Had to chose something “close enough”.

### Did you notice any changes in your behaviour while using VISAC?

No.

Was held up and could not advance because alternatives were lacking. Had to figure out what to choose.

The difference to using the MobiMed device is that you say the words out loud instead of pointing and clicking.

VISAC forces one to adhere to a work flow. Structuring the information into the ABCDE categories is normally done in hindsight. VISAC also forces more verbosity between the crew members. Generally a lot of what is going on is inferred from the situation by each member without being explicitly stated. Would like more alternatives to achieve better precision in clinical description.

Tried not to look at the screen, but still continued doing this. This, and figuring out what to say to be understood by the system, took focus off of the task.

A bit uncomfortable to work in a manner that is not familiar.

### How would VISAC affect your everyday work?

The speech recognition may not work well with a lot going on in the ambulance. Multiple people speaking and noise from the vehicle. In a trauma case, with a time-critical patient, a system like this will not be used. In a case like that, only vital parameters will be taken with the MobiMed system. VISAC is a great system for the common 80 % of cases, but not for emergency cases.

Less to write when in the ambulance. Free hands, MobiMed requires both hands for use. Be able to dictate the journal while both crew members are working with the patient. In emergency cases the crew are still to this day taking notes of vital parameters on their gloves.

VISAC would make filling out patient journals much easier as it documents while work is being conducted. This way things might be retained that would else be forgotten, if there is no time to take notes during an encounter.

VISAC would be helpful up until the point where miscommunication happens. Having to solve problems that emerge with VISAC takes focus off of the work at hand.

If the speech recognition works well, it would be helpful. Writing by hand is not that hard anyway. Good support when one wants to capture the gist of the situation now and fill in details later. Needs to be quick and reliable.

**During a patient encounter, what benefits and detriments do you see with VISAC?**

The patient might find it odd speaking to a device, which is a detriment. Some react negatively to the use of a mobile phone or the MobiMed system.

Free hands is a benefit. A detriment is that focus is split between VISAC and the patient, who might wonder whom one is speaking to. The device could seem odd to the patient.

This version of VISAC would be very straining to use. The device itself would need constant adjustments and might fall off when crouching or bending over, which is done regularly.

Especially older patients could be uncomfortable with the device.

Frees up hands but is in the way between the caregiver and the patient, who might wonder who one is speaking to.

The patients might find the device odd, which is a con. On the pro side, VISAC could make working with the patient journal easier and provide the emergency department with information at an earlier stage.

**How could that which is shown on the screen be made clearer?**

More alternatives. A backup touch-based system if VISAC does not function.

It is clear enough as is. Maybe a different color for the selected category alternative.

It was pretty clear once the display was properly positioned.

It is pretty clear as is.

Maybe higher contrast.

It is clear.

**What are your thoughts on VISAC repeating selected attributes?**

This function is superfluous as one can see what was entered on the screen. The patient record will be checked before signing it, anyway. This feature exists in the on-car systems but nobody uses it.

It is good to get a second check that the correct information was entered. This would be useful in some environments. In busy environments it could take focus off of the task, but also be helpful for clarity.

The function provides a “closed loop”. It assures the speaker that the device performed the correct action. The function might be good to be able to turn off, depending on the situation. E.g. it could be distracting in a critical situation. Could also be distracting if one listens to VISAC and misses what the patient is saying.

If the screen is there, the function is not needed, but some sort of indication is necessary. It is probably a question of getting used to.

The short audio signal is enough, as one is able to simply glance at the screen to see what was filled in. Would not use the repetition function.

Might be useful but would not use as it would take time to listen to the repetition. Also doesn’t like that the repetition makes sound.

**Observations made by the author during the test session**

Attempt to use command “<given name> har fri luftväg”. Solves unrecognized command by switching “<attribute> <category>” to “<category> <attribute>”.

One participant who methodically puts short waits in between phrases has good outcome on recognition while another participant who is very verbose and acts the patient encounter out has bad outcome. Finally reverts to using only one-word phrases, which then work well.

For one participant a confirming “mmh” while conversing with the author is repeatedly registered as “neuro”.

At one point VISAC registers its own repetition of an entry as a command, and thus gets stuck for a few loops.

Most participants conform to using commands as the words are laid out on the screen and speak in the manner “<category> <attribute>”. This is especially true after the system fails to register their uttered phrases. (Whether the phrases they failed in using are supposed to work or not.)

The device frequently registers conversation between participants and the author as commands. This is true both when VISAC is equipped and when it is lying on the table by itself.

# 5

# Discussion

The construction of the VISAC system had few constraints in the early phase and the end product is one outcome among many possible solutions; the result of successive decisions and several pivots. Below are a few suggestions, pointers and comments that might prove helpful for someone aiming to refine VISAC or a similar system. Interesting or odd results are also discussed.

## 5.1 Hardware Components

At the beginning of this project, decisions needed to be made on both which target system to use for development, as well as what software components to use for adding speech recognition capabilities and text-to-speech (TTS). As it was not known to the author what software was available and what software and hardware combinations were compatible, information about this was studied for making a choice. The outcome is presented in Sections 3.3 and 3.4. The choice of HMT-1 over HoloLens 2 is certainly correct for a production system, as the HoloLens 2, being too fragile, would not be able to cope with the working conditions of AP. For research purposes it is still an interesting platform and has seen some research on use as a heads-up-display in military medical situations [43]. The 3D environment allows for displaying a lot of information, but the visibility of the information is not great: holograms have low resolution, low contrast and become dimmed in brightly lit surroundings.

The HMT-1 does have drawbacks as well. About half of AP trying the device on found it hard to get a good view of the display through the viewfinder and to position the viewfinder correctly. The author found that this becomes easier with experience. Still, manual adjustments were needed during use, often demanding both hands. The user bending their neck or changing their pose when squatting down from a standing position is enough for adjustments to be needed. This could be a serious problem in a live scenario and renders the system less hands free. Several different straps for mounting the HMT-1 to the user's head are available from the producer. Only one such (the overhead strap) was available during this project. Other straps may perform better in holding the device steady. A difference in how well the device remained in position was also noted between users with shorter or longer hair, as long hair tends to be more slippery than for example buzz cuts.

In the interviews, AP were dubious that the device would interfere in the contact with patients and that patients would find the device odd and confusing, when AP speak to a third party whom the patient cannot see. This might become less of a problem as voice user interfaces become more prevalent. Certainly this design aspect needs attention as the well-being of the patient is paramount. A less visually accentuated design might make the interaction with VISAC passable as a phone call, which should be recognizable and not disturbing to most Swedish citizens.

In a scenario where natural language can be used to communicate with VISAC, or where the available commands are well known to the user, having a display might not be of importance for using the system. As the VISAC hardware is fundamentally an Android device with a good noise canceling microphone – in the mentioned scenario – the use of a headset with microphone and earpiece would provide the benefit of being lighter, less conspicuous and better secured to the user. A screen-based Android device, such as a phone or tablet, could be used if only occasional visual feedback was needed. This would also add a second mode of input via the touch screen and allow for a free choice of microphone to use.

Using a contact microphone, such as a throat microphone, might help in mitigating noise as well as help in rejection of speech from external speakers. Using a throat microphone alters the recorded sound considerably, but not beyond recognition to a human. Speech recognition for this scenario would probably need training on a data set specific for throat microphones. The filtering effect of throat microphones has been modeled as a low pass filter with a cutoff frequency of 300 Hz [106]. Performing tests with a throat microphone would be interesting to do in future studies.

## 5.2 Primary Survey Application

The early process of familiarization with the Android APIs shaped the direction of the project and the design of the first assessment application. This might have been detrimental to the end result and involving AP in the early design could have led to a more purposeful and user centered design.

A naive attempt of using the HMT-1 native support for English commands for use with Swedish was done early on in evaluation. This approach has serious limitations as the pronunciation differs severely between the languages. Adding a Swedish word and pronouncing it “with an American accent” functioned relatively well but is not fit for a production system as the user should be able to speak in their natural voice.

The choice of using SpeechKit commands rather than SpeechKit dictation was made because a method of directing program actions depending on spoken commands was needed. Commands provided this functionality, while dictation would have demanded development of a method for extracting user intent from dictated content. Dictated text is meant to be manually corrected in parallel during dictation, for the SpeechKit system to be trained on the user’s voice. Such correction can not be expected to be conducted in a production system during a patient encounter, which would be a detriment of using the dictation mode.

User specific voice models are used by SpeechKit. These could be loaded when a specific AP logs on to the system, as each AP would need to be identified in order to keep track of who made what observations and for signing the patient record upon completion. One voice model will work with several users – as was exemplified in testing of VISAC where all participants used the same untrained voice model. Also without manual training, the models seem to be trained. The author noticed differences in performance for words frequently used during development, between an unused voice model and the voice model that had been used during development by the author.

For use with standardized clinical terminology, such as SNOMED CT or LOINC, the command-paradigm lends itself well, as specific commands can actuate transmission of specific predefined clinical terms. Standardized clinical terminology codes can also be used for defining the observed values in the FHIR `Observation` that is sent upon selecting an attribute. This can be done by exchanging the attached "valueString" for a "valueCodableConcept".

One drawback with the command paradigm is that the set of available command phrases needs to be defined and generated. No obvious good solution for this has been found during the writing of this thesis. Naive morphological generation has been used, together with manual editing of the results. This has shown to be a time consuming process. In generation of command phrases, search-and-replace was used effectively for different attributes in the same category, indicating that there might exist a set of grammars which can be reused for generation of disparate sets of command phrases. This might allow for less information needed in the specification of voice commands sent to VISAC, if rules for such sets can be stored locally and referenced for the generation of commands instead of transmitting a long list of phrases.

Command phrases should be chosen in a way that mimics the natural speech patterns of AP, while maintaining standardized clinical terminology. In this thesis, redundancy was introduced by adding command phrases of many possible combinations of a set of words, which convey the same message. It became obvious in testing that most combinations were never used, and the AP conformed to similar grammars. Most notably defaulting to "<category> <attribute>", when their first try, or a few tries, did not succeed. Standardized speech is helpful both for data processing purposes as well as for maintaining clarity in communication between AP.

## 5.3 Testing of VISAC

In system testing, it became obvious that the vocabulary was severely lacking for the application to fulfill the use case. Available attributes were not chosen properly and sometimes overlapped, while at other times, needed complementing. This was partly due to a larger final set of terms being expected to be used at time of selecting vocabulary, than what was finally achieved. Lack of time hindered adding a large vocabulary to the application; A feature that was asked for in the first user evaluation with AP in order to gain greater precision in clinical recording. Completeness of clinical terminology was not prioritized towards the end of the project as adding selectable attributes demands time for generation of command phrases as well as some way of displaying the attributes to the user. The final application does not have a scroll function, and can thus not display more options than what can be fitted on screen at one time. The system allows for use of more command phrases, but all commands can then not be displayed in the GUI at the same time.

Possible solutions are adding horizontal scrolling, e.g. using the gyroscope in the HMT-1. Another option is to not show all attributes, and instead just show a subset, or none of the attributes. Selected attributes could be shown, while hiding unselected attributes. A solution like this is less cluttered, and would work well if the vocabulary was complete or nearly complete, so that AP know what attributes are available for reporting. There is likely to exist an Android `View` element which can easily be used for implementing scrollable lists with the layout of the final application. `RecyclerView`, which was used in the first version of the application does have scrolling. In the final version `Table`s were used as these involve less boilerplate code due to a simpler API and thus were quicker to revert to in implementing the final application.

During system testing, many false registrations of commands were noted. The device being constantly listening for commands is a problem, especially if rejection of external speakers can not be achieved. Using a command such as “VISAC start” and “VISAC stop” could make use cumbersome. Stopping VISAC completely from listening to commands would make it impossible to use a voice command to reactivate listening. Not deactivating listening completely, on the other hand, could result in the reactivation command being mistakenly recognized. The reactivation command thus needs to be chosen carefully. The HMT-1 does have a programmable button on the side, which could be used for starting and stopping VISAC listening. Using this would make the device less hands free which might be a problem in some cases. Multiple modes of control might be the safest solution and could also include e.g. using the gyroscope.

During the user tests, VISAC got stuck in a loop for a few iterations, when TTS was active and the device mistakenly recognized what was repeated by the device itself using TTS, as if it was a command phrase. Using an earpiece would remove the problem and simultaneously make it less likely that patient personal information is exposed to an unauthorized bystander. If VISAC provides decision support via an earpiece, this information would not be heard by the partner AP, which might be detrimental to the common task, but also not by bystanders for whom the information is not intended to be disclosed.

White noise was used in the tests. It is unknown how good a model this is for ambulance noise. It is unlikely that ambulance noise would follow the uniform frequency distribution of white noise. This was chosen as no objectively representative noise was available and white noise generation tools are readily available. Adding white noise effectively raises the noise floor and is probably more “frequency dense” than real ambulance noise, which should be unequally concentrated with more energy at certain frequencies of the sound spectrum. Increasing the white noise amplitude should directly decrease the signal to noise-ratio. The test should thus be able to measure the noise rejection capabilities of the HMT-1 microphone array. In this, the positioning of the stereo speakers relative to the user could have an effect on noise rejection as the HMT-1 microphones might show better rejection capabilities in certain directions. The method for producing the white noise, implemented in the application Noice, is unknown to the author and also affects the exact noise profile produced.

A-weighted dB-measurements were used as this is a frequently reported sound level measurement. A-weighting – which models the human ear – might not be optimal for evaluation of speech recognition systems, as these use microphones which are not necessarily A-weighted and might put emphasis on sound content differently than the human ear does.

## 5.4 Test Results

For the speech recognition tests, it should be noted that the speaker voice level was not measured. The voice level was thus not held constant between the different noise level tests. This introduces an important source of inconsistency in measurements. Granted, that humans normally raise their voice when the environment in which they speak is noisy, this voice level difference might not be a problem per se, but it makes comparing the measurements less reliable. Between Figures 4.10 and 4.9, it can be noted that the parameters calculated for using SpeechKit commands deviate differently at 80 dB(A) than for the other noise levels. In other words: among utterances recognized by VISAC as speech, VISAC was better at correctly recognizing the uttered phrase for 80 dB(A) noise than for 93 dB(A) noise. The reasonable conclusion is that the phrases were spoken more loud or more clear, relative to the noise level, at 80 dB(A) than at 93 dB(A). This emphasizes the need for controlling the voice sound level as well as the clarity of speech. A simple solution would be to pre-record the phrases and choose a set sound level when playing this recording during testing.

A word accuracy of between 90.6 and 99.1 % in a SpeechKit dictation-based system during quiet conditions has been reported [29]. These percentages are noted as “accuracy” and it is unclear from the article whether they are acquired from WER calculations or not. WER is the only stated method of measuring speech recognition performance, making it plausible that these percentages correspond to WER-values even as they are reported as “accuracy”. If this is the case, they correspond to WER-values of 0.14 and 0.09 respectively. VISAC performed substantially worse than this with a WER of 0.59 in quiet conditions. The article used SpeechKit for German and French, which might affect the results. The clarity of speech is also a variable that is probable to differ between these two projects. Additionally, what kind of phrases are spoken is certain to affect the outcome. SpeechKit's utilization of LSTM in dictation is likely to result in better recognition performance for sentences longer than a few words.

This marked difference in WER between commands and dictation indicates that using commands is the better choice for the application at hand. This assessment is supported by the WIL and $\text{WIL}_{simpl}$, both showing a 30 % lower loss of information when using commands compared to dictation, in quiet conditions. Studying the $\text{A}_{rec}$ and WIL for commands, we note that even with 42 % of word information lost for quiet conditions, 85 % of uttered commands resulted in the intended action. Adding redundancy by allowing similar phrases for the same action thus seems to have been useful. The $\text{A}_{rec}$-parameters also indicate that for dictation, registered phrases are part of the correct command set only in 27 % of cases for quiet conditions and only 5 % at 80 dB(A). Together with the knowledge of much smaller difference in WER between dictation and commands, this indicates errors in words that carry actionable information, rather than errors in many of the words of a phrase.

With increasing noise levels, the number of phrases that were not registered at all increased, culminating at 43 % for 93 dB(A). This loss of spoken phrases accounts for a large part of recognition problems. When disregarding empty phrases, 98 % of commands resulted in the correct action in quiet conditions and 75 % at 93 dB(A). All in all, recall of command intent was measured to be between 50 % and 75 % at 80 dB(A) – noise levels found inside an ambulance – and about 85 % in quiet conditions. Command phrases with the prepended phrases "patienten har", "det är" and "vi har" were often not correctly recognized or returned empty responses. This is exemplified in several variations in Table 4.3. As the command phrases are long, the risk of introducing errors that make the phrase unintelligible might increase. Further analysis on such properties of different types of commands could be interesting to do. Some command sets contained more phrases than others, which might also be an interesting parameter to explore to see where interference between command sets begins to be a problem.

AP user tests resulted in a SUS mean score of 63.75. This is slightly below that of a "passable" score of 70. Several AP noted the need for a more complete vocabulary for the application to be functional. AP were predominantly skeptical towards using VISAC with TTS, stating that this could be distracting. TTS could also potentially share patient personal information with people that are close by if an earpiece is not used.

# 6

# Conclusions

The field of HMDs for prehospital use is leaving its infancy. There are several manufacturers of HMDs for different use cases and prehospital healthcare has already seen application of HMDs in Norway. This thesis has explored the feasibility of a voice interface for Swedish ambulance care from contemporary commercially available hardware and software. A pilot system was constructed and tested with AP at Örebro ambulance department.

The VISAC system uses the head mounted display HMT-1 (RealWear, Inc., United States), the Dragon Medical SpeechKit (Nuance Communications, Inc., United States) and the text-to-speech engine Acapela TTS (Acapela Group, Europe) to perform a naive ABCDE "primary survey" which is defined in the form of an HL7 FHIR resource. Utilizing a clinical coding standard such as SNOMED CT for coding of clinical terminology permits univocal communication of clinical measurements and prepares for interoperability with other products used in the care chain.

Testing of the finished demonstrator product indicates low usability due to a small implemented vocabulary. A method for extracting the relevant actionable information from the users speech needs refinement for the product to be feasible in a live environment. HMT-1 does function well with SpeechKit and the combination is robust to noise to at least 65 dB(A) but prone to register commands from surrounding speakers. Rejection of external sources of speech and a method of determining when VISAC is spoken to are needed for feasibility in clinical practice.

# Bibliography

[1] G. Reay, J. A. Rankin, L. Smith-MacDonald, and G. C. Lazarenko, “Creative adapting in a fluid environment: an explanatory model of paramedic decision making in the pre-hospital setting,” *BMC Emergency Medicine*, vol. 18, no. 1, p. 42, Nov. 2018, doi:10.1186/s12873-018-0194-1.

[2] B. W. Sundström and K. Dahlberg, “Being Prepared for the Unprepared: A Phenomenology Field Study of Swedish Prehospital Care,” *Journal of Emergency Nursing*, vol. 38, no. 6, pp. 571–577, Nov. 2012, doi:10.1016/j.jen.2011.09.003.

[3] M. K. Corman, “Titrating the Rig: How Paramedics Work in and on Their Ambulance,” *Qualitative Health Research*, vol. 28, no. 1, pp. 47–59, Jan. 2018, publisher: SAGE Publications Inc, doi:10.1177/1049732317739266.

[4] H. Sjölin, V. Lindström, V. Vicente, H. Hult, C. Ringsted, and L. Kurland, “Prehospital emergency nurses’ experiences of care in critical incidents,” *International Emergency Nursing*, vol. 51, p. 100890, Jul. 2020, doi:10.1016/j.ienj.2020.100890.

[5] L. Honeycutt Ed., *PHTLS Prehospital Trauma Life Support.* 6th ed. Canada: Mosby, Inc., 2007, pp. 7-8.

[6] A. Khorram-Manesh, K. Lennquist Montán, A. Hedelin, M. Kihlgren, and P. Örtenwall, “Prehospital triage, discrepancy in priority-setting between emergency medical dispatch centre and ambulance crews,” *European Journal of Trauma and Emergency Surgery*, vol. 37, no. 1, pp. 73–78, Feb. 2011, doi:10.1007/s00068-010-0022-0.

[7] B. L. Bigham, E. Bull, M. Morrison, R. Burgess, J. Maher, S. C. Brooks, and L. J. Morrison, “Patient safety in emergency medical services: executive summary and recommendations from the Niagara Summit,” *Canadian Journal of Emergency Medicine*, vol. 13, no. 1, pp. 13–18, Jan. 2011, publisher: Cambridge University Press, doi:10.2310/8000.2011.100232.

[8] C. Hohenstein, D. Hempel, K. Schultheis, O. Lotter, and T. Fleischmann, “Critical incident reporting in emergency medicine: results of the prehospital reports,” *Emergency Medicine Journal*, vol. 31, no. 5, pp. 415–418, May 2014, doi:10.1136/emermed-2012-201871.

[9] B. L. Bigham, J. Maher, and L. J. Morrison, “Patient safety in emergency medical services: advancing and aligning the culture of patient safety in EMS,” Canadian Patient Safety Institute, Edmonton, Alberta, Canada, oCLC: 1059225230. 2010. [Online]. Available: https://www.deslibris.ca/ID/224678, Accessed on: 2022-03-04.

[10] J. Wihlborg, G. Edgren, A. Johansson, and B. Sivberg, “The desired competence of the Swedish ambulance nurse according to the professionals – A Delphi study,” *International Emergency Nursing*, vol. 22, no. 3, pp. 127–133, Jul. 2014, doi:10.1016/j.ienj.2013.10.004.

[11] A. Jacobsen, A. Bremer, A. Svensson, A. Rantala, A. Strömsöe, A. Hörberg, C. Nilsson, E. Höglund, G. Norberg, H. Andersson, J. Persson, J. Aléx, J. Wihlborg, K. Jepsen, M. Viking, M. Holmberg, S. Hammarbäck, and V. Lindström, “Ambulanssjukvården behöver genomgripande förändringar,” *Dagens Medicin*, Jun. 2021. [Online]. Available: https://www.dagensmedicin.se/opinion/debatt/ambulanssjukvarden-behover-genomgripande-forandringar/, Accessed on: 2022-03-05.

[12] J. Nilsson, S. Johansson, G. Nordström, and B. Wilde-Larsson, “Development and Validation of the Ambulance Nurse Competence Scale,” Jan. 2020, doi:10.1016/j.jen.2019.07.019.

[13] B. L. Bigham, J. E. Buick, S. C. Brooks, M. Morrison, K. G. Shojania, and L. J. Morrison, “Patient safety in emergency medical services: a systematic review of the literature,” *Prehospital emergency care: official journal of the National Association of EMS Physicians and the National Association of State EMS Directors*, vol. 16, no. 1, pp. 20–35, Mar. 2012, doi:10.3109/10903127.2011.621045.

[14] “Behandlingsriktlinjer: Nationella riktlinjer för ambulanssjukvården,” Föreningen för Ledningsansvariga Inom Svensk Ambulanssjukvård, Sweden, Jan. 2018. [Online]. Available: https://www.s112.se/wp-content/uploads/2018/05/SLAS-behandlingsriktlinjer-Vuxen-och-barn-20180102-2.pdf, Accessed on: 2022-03-05.

[15] M. A. Hagiwara, B. A. Sjöqvist, L. Lundberg, B.-O. Suserud, M. Henricson, and A. Jonsson, “Decision support system in prehospital care: a randomized controlled simulation study,” *The American Journal of Emergency Medicine*, vol. 31, no. 1, pp. 145–153, Jan. 2013, doi:10.1016/j.ajem.2012.06.030.

[16] M. A. Hagiwara, B.-O. Suserud, A. Jonsson, and M. Henricson, “Exclusion of context knowledge in the development of prehospital guidelines: results produced by realistic evaluation,” *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine*, vol. 21, no. 1, p. 46, Dec. 2013, doi:10.1186/1757-7241-21-46.

[17] M. Hagiwara, M. Henricson, A. Jonsson, and B.-O. Suserud, “Decision-Support Tool in Prehospital Care: A Systematic Review of Randomized Trials,” *Prehospital and Disaster Medicine*, vol. 26, no. 5, pp. 319–329, Oct. 2011, doi:10.1017/S1049023X11006534.

[18] D. L. Hunt, R. B. Haynes, S. E. Hanna, and K. Smith, “Effects of Computer-Based Clinical Decision Support Systems on Physician Performance and Patient OutcomesA Systematic Review,” *JAMA*, vol. 280, no. 15, pp. 1339–1346, 1998, doi:10.1001/jama.280.15.1339.

[19] M. Andersson Hagiwara, B.-O. Suserud, B. Andersson-Gäre, B.-A. Sjöqvist, M. Henricson, and A. Jonsson, “The effect of a Computerised Decision Support System (CDSS) on compliance with the prehospital assessment process: results of an interrupted time-series study,” *BMC Medical*

*Informatics and Decision Making*, vol. 14, no. 1, p. 70, Aug. 2014, doi:10.1186/1472-6947-14-70.

[20] K. Kawamoto, C. A. Houlihan, E. A. Balas, and D. F. Lobach, “Improving clinical practice using clinical decision support systems: a systematic review of trials to identify features critical to success,” *BMJ*, vol. 330, no. 7494, p. 765, Mar. 2005, publisher: British Medical Journal Publishing Group Section: Information in practice, doi:10.1136/bmj.38398.500764.8F.

[21] A. H. Morris, “Decision support and safety of clinical environments,” *Quality and Safety in Health Care*, vol. 11, no. 1, pp. 69–75, Mar. 2002, doi:10.1136/qhc.11.1.69.

[22] HealthIT.gov, “What is an electronic health record (EHR)? | HealthIT.gov,” 2019. [Online]. Available: https://www.healthit.gov/faq/what-electronic-health-record-ehr (accessed on: 2022-03-05).

[23] C. Sinsky, L. Colligan, L. Li, M. Prgomet, S. Reynolds, L. Goeders, J. Westbrook, M. Tutty, and G. Blike, “Allocation of Physician Time in Ambulatory Practice: A Time and Motion Study in 4 Specialties,” *Annals of Internal Medicine*, vol. 165, no. 11, pp. 753–760, Dec. 2016, publisher: American College of Physicians, doi:10.7326/M16-0961.

[24] T. G. Holzman, “Computer-human interface solutions for emergency medical care,” *Interactions*, vol. 6, no. 3, pp. 13–24, May 1999, doi:10.1145/301153.301160.

[25] Ortivus, “Ortivus vinner ett strategiskt viktigt kontrakt på prehospital patientmonitorering i Region Skåne!” *Ortivus*, 2021. [Online]. Available: https://www.ortivus.com/sv/nyheter/ortivus-vinner-ett-strategiskt-viktigt-kontrakt-pa-prehospital-patientmonitorering-i-region-skane/ (accessed on: 2022-03-05).

[26] L. Ekstedt, “Bäst och sämst med röststyrd diktering,” *Vision*, 2020. [Online]. Available: https://vision.se/tidningenvision/arkiv/2020/nr2/bast-och-samst-med-roststyrd-diktering/, Accessed on: 2022-03-07.

[27] Region Västmanland, “Cosmic integrerad taligenkänning - Region Västmanland,” 2021. [Online]. Available: https://www.regionvastmanland.se/vardgivare/it-stod/it-system/cosmic-integrerad-taligenkanning/ (accessed on: 2022-03-07).

[28] R. Zydowicz, “A speech recognition enabled ambulance call report system for paramedics,” MSc thesis, Ottawa-Carleton Institute for Computer Science, Carleton University, Ottawa, Ontario, Canada, 2011. [Online]. Available: https://curve.carleton.ca/244f546d-0c64-4f11-b309-ef492cbbf50a

[29] L. Meier, J. G. Bauer, and K. Denecke, “Speech-based Documentation in Emergency Medical Services with the Electronic Language Interface for Ambulance Services,” in *2020 IEEE International Conference on Healthcare Informatics (ICHI)*, Nov. 2020, pp. 1–6, iSSN: 2575-2634.

[30] P. Schlosser, “Head-Worn Displays for Emergency Medical Services,” in *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, New York, NY, USA, 2021, pp. 1–5. [Online]. Available: https://doi.org/10.1145/3411763.3443421, Accessed on: 2021-11-05.

[31] W. Vorraber, S. Voessner, G. Stark, D. Neubacher, S. DeMello, and A. Bair, "Medical applications of near-eye display devices: an exploratory study," *International Journal of Surgery (London, England)*, vol. 12, no. 12, pp. 1266–1272, Dec. 2014, doi:10.1016/j.ijsu.2014.09.014.

[32] C. Romare and L. Skär, "Smart Glasses for Caring Situations in Complex Care Environments: Scoping Review," *JMIR mHealth and uHealth*, vol. 8, no. 4, p. e16055, Apr. 2020, company: JMIR mHealth and uHealth Distributor: JMIR mHealth and uHealth Institution: JMIR mHealth and uHealth Label: JMIR mHealth and uHealth Publisher: JMIR Publications Inc., Toronto, Canada, doi:10.2196/16055.

[33] N. J. Wei, B. Dougherty, A. Myers, and S. M. Badawy, "Using Google Glass in Surgical Settings: Systematic Review," *JMIR mHealth and uHealth*, vol. 6, no. 3, p. e54, Mar. 2018, doi:10.2196/mhealth.9409.

[34] T. Mentler, J. Kappel, L. Wünsch, and M. Herczeg, "Photo-enriched Documentation during Surgeries with Google Glass: An Exploratory Usability Study in a Department of Paediatric Surgery," *i-com*, vol. 15, no. 2, pp. 171–178, Aug. 2016, publisher: Oldenbourg Wissenschaftsverlag, doi:10.1515/icom-2016-0017.

[35] T. Mentler, H. Berndt, and M. Herczeg, "Optical Head-Mounted Displays for Medical Professionals: Cognition-supporting Human-Computer Interaction Design," in *Proceedings of the European Conference on Cognitive Ergonomics*, ser. ECCE '16. New York, NY, USA: Association for Computing Machinery, Sep. 2016, pp. 1–8. [Online]. Available: https://doi.org/10.1145/2970930.2970957

[36] A. Widmer and H. Müller, "Using Google Glass to enhance pre-hospital care," in *Swiss e-health summit*, Bern, Switzerland, 2014. [Online]. Available: https://www.researchgate.net/publication/265377411_Using_Google_Glass_to_enhance_pre-hospital_care

[37] R. Galati, M. Simone, G. Barile, R. De Luca, C. Cartanese, and G. Grassi, "Experimental Setup Employed in the Operating Room Based on Virtual and Mixed Reality: Analysis of Pros and Cons in Open Abdomen Surgery," *Journal of Healthcare Engineering*, vol. 2020, p. 8851964, Aug. 2020, doi:10.1155/2020/8851964.

[38] J. Chirapurath, "Mixed Reality Momentum: HoloLens 2 expands globally and new Azure service launches," *Microsoft Azure Blog*, Sep. 2020. [Online]. Available: https://azure.microsoft.com/en-us/blog/mixed-reality-momentum-hololens-2-expands-globally-and-new-azure-service-launches/ (accessed on: 2021-09-14).

[39] "Mixed Reality Intelligence Healthcare Edition," Microsoft and Hypothesis Group, 2020. [Online]. Available: https://info.microsoft.com/rs/157-GQE-382/images/Mixed_Reality_Intelligence_Healthcare_Edition.pdf

[40] Sykehuset Innlandet, "Videoassistert beslutningsstøtte i akuttkjeden," 2022. [Online]. Available: https://sykehuset-innlandet.no/fag-og-forskning/forskning/innovasjon/prehospital-beslutningsstotte (accessed on: 2022-03-08).

[41] S. Price, “Telehealth and emergency care in Norway,” *Health Europa*, 2020. [Online]. Available: https://www.healtheuropa.eu/telehealth-and-emergency-care-in-norway/101277/, Accessed on: 2022-03-08.

[42] K. Denecke, L. Meier, J. G. Bauer, M. Bender, and C. Lueg, “Information Capturing in Pre-Hospital Emergency Medical Settings (EMS),” *Studies in health technology and informatics*, vol. 270, pp. 613–617, Jun. 2020, doi:10.3233/SHTI200233.

[43] N. M. McGeorge, S. Latiff, C. Muller, L. Dong, C. Chewning, D. Friedson-Trujillo, and S. Kane, “Design and Development of a Prototype Heads-Up Display: Supporting Context-Aware, Semi-Automated, Hands-Free Medical Documentation,” *Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care*, vol. 10, no. 1, pp. 18–22, Jun. 2021, publisher: SAGE Publications, doi:10.1177/2327857921101066.

[44] S. V. Blackley, J. Huynh, L. Wang, Z. Korach, and L. Zhou, “Speech recognition for clinical documentation from 1990 to 2018: a systematic review,” *Journal of the American Medical Informatics Association*, vol. 26, no. 4, pp. 324–338, Apr. 2019, doi:10.1093/jamia/ocy179.

[45] F. R. Goss, S. V. Blackley, C. A. Ortega, L. T. Kowalski, A. B. Landman, C.-T. Lin, M. Meteer, S. Bakes, S. C. Gradwohl, D. W. Bates, and L. Zhou, “A clinician survey of using speech recognition for clinical documentation in the electronic health record,” *International Journal of Medical Informatics*, vol. 130, p. 103938, Oct. 2019, doi:10.1016/j.ijmedinf.2019.07.017.

[46] L. Honeycutt Ed., *PHTLS Prehospital Trauma Life Support.* 6th ed. Canada: Mosby, Inc., 2007, pp. xxvii-xxviii.

[47] L. Honeycutt Ed., *PHTLS Prehospital Trauma Life Support.* 6th ed. Canada: Mosby, Inc., 2007, pp. 95-99.

[48] E. Höglund, private communication, Oct. 2021.

[49] M. Cybulski, K. Wiecheć, M. Zieliński, and B. I. Bilski, “How Paramedics Perceive Internal Noise in Ambulance? Sensory Processing Sensitivity (SPS) and Subjective Noise Assessment,” *Noise & Health*, vol. 21, no. 103, pp. 242–247, 2019, doi:10.4103/nah.NAH_10_19.

[50] M. C. T. Hansen, J. H. Schmidt, A. C. Brøchner, J. K. Johansen, S. Zwisler, and S. Mikkelsen, “Noise exposure during prehospital emergency physicians work on Mobile Emergency Care Units and Helicopter Emergency Medical Services,” *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine*, vol. 25, no. 1, p. 119, Dec. 2017, doi:10.1186/s13049-017-0459-9.

[51] J. Dundurs and I. J. Janssen, “Influence of Noise in Ambulance Vehicles on Emergency Service Personnel in North Germany and Latvia,” *Scientific Papers/RSU*, 2017. [Online]. Available: https://www.rsu.lv/en/scientific-papers/influence-noise-ambulance-vehicles-emergency-service-personnel-north-germany-and.html (accessed on: 2022-03-10).

[52] Arbetsmiljöverket, “Buller (AFS 2005:16), föreskrifter,” AFS 2005:16. 2005. [Online]. Available: https://www.av.se/arbetsmiljoarbete-och-inspektioner/publikationer/foreskrifter/buller-afs-200516/, Accessed on: 2022-03-10.

[53] A. N. Rimell, N. J. Mansfield, and G. S. Paddan, “Design of digital filters for frequency weightings (A and C) required for risk assessments of workers exposed to noise,” *Industrial Health*, vol. 53, no. 1, pp. 21–27, Jan. 2015, doi:10.2486/indhealth.2013-0003.

[54] A. B. Nassif, I. Shahin, I. Attili, M. Azzeh, and K. Shaalan, “Speech Recognition Using Deep Neural Networks: A Systematic Review,” *IEEE Access*, vol. 7, pp. 19 143–19 165, 2019, conference Name: IEEE Access, doi:10.1109/ACCESS.2019.2896880.

[55] A. Graves and J. Schmidhuber, “Framewise phoneme classification with bidirectional LSTM and other neural network architectures,” *Neural Networks*, vol. 18, no. 5, pp. 602–610, Jul. 2005, doi:10.1016/j.neunet.2005.06.042.

[56] C. Baber and J. Noyes, “Automatic Speech Recognition in Adverse Environments,” *Human Factors*, vol. 38, no. 1, pp. 142–155, Mar. 1996, publisher: SAGE Publications Inc, doi:10.1518/001872096778940840.

[57] S. Alharbi, M. Alrazgan, A. Alrashed, T. Alnomasi, R. Almojel, R. Alharbi, S. Alharbi, S. Alturki, F. Alshehri, and M. Almojil, “Automatic Speech Recognition: Systematic Literature Review,” *IEEE Access*, vol. 9, pp. 131 858–131 876, 2021, doi:10.1109/ACCESS.2021.3112535.

[58] V. S. Reddy Gade and M. Sumathi, “A Comprehensive Study on Automatic Speaker Recognition by using Deep Learning Techniques,” in *2021 5th International Conference on Trends in Electronics and Informatics (ICOEI)*, Jun. 2021, pp. 1591–1597.

[59] M. Kuhlmann, “Vad är naturligt språk-behandling?” [Online]. Available: https://ai-for-naturligt-sprak.ida.liu.se/content/ETE335-1.pdf

[60] K. V. Ghag and K. Shah, “Comparative analysis of effect of stopwords removal on sentiment classification,” in *2015 International Conference on Computer, Communication and Control (IC4)*, Sep. 2015, pp. 1–6.

[61] S.-G. Malmgren, *Svensk lexikologi.* Lund, Sweden: Studentlitteratur, 1994.

[62] W. B. McGregor, *Linguistics: An Introduction.* 2nd ed. Bloomsbury Academic, Feb. 2015, p. 59.

[63] RealWear, Inc., “Introduction to Designing for the RealWear HMT-1 and HMT-1Z1,” *RealWear*, 2018. [Online]. Available: https://www.realwear.com/blog/designing-realwear-hands-free-wearhf/ (accessed on: 2022-03-11).

[64] Google Developers, “Platform Architecture,” *Android Developers*, 2021. [Online]. Available: https://developer.android.com/guide/platform (accessed on: 2022-03-12).

[65] Google Developers, “Introduction to Activities,” *Android Developers*, 2021. [Online]. Available: https://developer.android.com/guide/components/activities/intro-activities (accessed on: 2022-03-12).

[66] Google Developers, “Layouts,” *Android Developers*, 2021. [Online]. Available: https://developer.android.com/guide/topics/ui/declaring-layout (accessed on: 2022-03-12).

[67] Google Developers, “ViewModel Overview,” *Android Developers*, 2021. [Online]. Available: https://developer.android.com/topic/libraries/architecture/viewmodel (accessed on: 2022-03-12).

[68] P. Charollais, *The JSON Data Interchange Syntax*, ECMA Std. 404, 2017. [Online]. Available: https://www.ecma-international.org/publications-and-standards/standards/ecma-404/, Accessed on: 2022-03-12.

[69] 1177 Vårdguiden, “Läs din journal via nätet,” *1177*, 2021. [Online]. Available: https://www.1177.se/om-1177-vardguiden/e-tjanster-pa-1177-vardguiden/det-har-kan-du-gora-nar-du-loggat-in/las-din-journal/las-din-journal-via-natet/ (accessed on: 2022-03-12).

[70] F. Oemig and R. Snelick, *Healthcare Interoperability Standards Compliance Handbook*. Cham, Switzerland: Springer International Publishing, 2016. [Online]. Available: http://link.springer.com/10.1007/978-3-319-44839-8

[71] E. Åkerblom, “Snomed CT -en översikt v2.1,” Socialstyrelsen, 2022. [Online]. Available: https://www.socialstyrelsen.se/globalassets/sharepoint-dokument/dokument-webb/klassifikationer-och-koder/snomed-ct-oversikt-v2-1.pdf, Accessed on: 2022-03-18.

[72] HL7 International, “HL7 Standards Product Brief - FHIR® R4 (HL7 Fast Healthcare Interoperability Resources, Release 4) | HL7 International.” [Online]. Available: http://www.hl7.org/implement/standards/product_brief.cfm?product_id=491 (accessed on: 2022-03-12).

[73] HL7 International, “Observation - FHIR v4.0.1,” 2019. [Online]. Available: http://hl7.org/fhir/observation-definitions.html (accessed on: 2022-03-12).

[74] HL7 International, “Valueset-observation-status - FHIR v4.0.1,” 2019. [Online]. Available: http://hl7.org/fhir/valueset-observation-status.html (accessed on: 2022-03-12).

[75] J. R. Lewis, “Measuring Perceived Usability: The CSUQ, SUS, and UMUX,” *International Journal of Human-Computer Interaction*, vol. 34, no. 12, pp. 1148–1156, Dec. 2018, publisher: Taylor & Francis Ltd, doi:10.1080/10447318.2017.1418805.

[76] J. Brooke, “SUS: A ’Quick and Dirty’ Usability Scale,” CRC Press, 1996.

[77] A. Bangor, P. Kortum, and J. Miller, “An Empirical Evaluation of the System Usability Scale,” *International Journal of Human-Computer Interaction*, vol. 24, no. 6, pp. 574–594, Aug. 2008, publisher: Taylor & Francis Ltd, doi:10.1080/10447310802205776.

[78] J. Sauro and J. R. Lewis, *Quantifying the User Experience: Practical Statistics for User Research, Second Edition*. Morgan Kaufmann Publishers, 2016.

[79] A. Morris, V. Maier, and P. Green, “From WER and RIL to MER and WIL: improved evaluation measures for connected speech recognition,” in *INTERSPEECH 2004 - ICSLP, 8th International Conference on Spoken Language Processing*, Jeju Island, Korea, 2004. [Online]. Available: https://www.researchgate.net/profile/Phil-Green-4/publication/221478089_From_WER_and_RIL_to_MER_and_WIL_improved_evaluation_measures_for_connected_speech_recognition/links/00b4951f95799284d9000000/From-WER-and-RIL-to-MER-and-WIL-improved-evaluation-measures-for-connected-speech-recognition.pdf, Accessed on: 2022-03-05.

[80] R. Errattahi, A. El Hannani, and H. Ouahmane, “Automatic Speech Recognition Errors Detection and Correction: A Review,” *Procedia Computer*

*Science*, vol. 128, pp. 32–37, Jan. 2018, doi:10.1016/j.procs.2018.03.005.

[81] Y.-Y. Wang, A. Acero, and C. Chelba, "Is word error rate a good indicator for spoken language understanding accuracy," in *2003 IEEE Workshop on Automatic Speech Recognition and Understanding (IEEE Cat. No.03EX721)*, Nov. 2003, pp. 577–582.

[82] G. Navarro, "A guided tour to approximate string matching," *ACM Computing Surveys*, vol. 33, no. 1, pp. 31–88, Mar. 2001, doi:10.1145/375360.375365.

[83] A. C. Morris, "An information theoretic measure of sequence recognition performance," Dalle Molle Institute for Perceptual Artificial Intelligence, Nov. 2003.

[84] Microsoft, "HoloLens 2—Overview, Features, and Specs | Microsoft HoloLens." [Online]. Available: https://www.microsoft.com/en-us/hololens/hardware (accessed on: 2022-03-14).

[85] Microsoft, "Start Developing for Mixed Reality | Microsoft HoloLens." [Online]. Available: https://www.microsoft.com/en-us/hololens/developers (accessed on: 2022-03-14).

[86] RealWear, Inc., "HMT-1 Specifications for Model T1100G." [Online]. Available: http://support.realwear.com/knowledge/hmt-1-specifications-for-model-t1100g (accessed on: 2021-09-14).

[87] Vuzix, "Vuzix M4000 | Smart Glasses with See-Through Display." [Online]. Available: https://www.vuzix.com/products/m4000-smart-glasses (accessed on: 2022-05-30).

[88] Vuzix, "Vuzix M400 | Wearable, Powerful, Smart Glasses." [Online]. Available: https://www.vuzix.com/products/m400-smart-glasses (accessed on: 2022-05-30).

[89] Google, "Tech Specs." [Online]. Available: https://www.google.com/glass/tech-specs/ (accessed on: 2022-05-30).

[90] RealWear, Inc., "REALWEAR HMT-1®," *RealWear*. [Online]. Available: https://www.realwear.com/hmt-1/ (accessed on: 2022-03-17).

[91] RealWear, Inc., "HMT-1 Device Details." [Online]. Available: http://support.realwear.com/knowledge/hmt-1-device-details (accessed on: 2022-03-17).

[92] scooley, evmill, beelia, v-jmathew, joyjaz, v-jodben, and SethPaniagua, "Om hololens 2," *HoloLens 2-maskinvara*, Mar. 2022. [Online]. Available: https://docs.microsoft.com/sv-se/hololens/hololens2-hardware (accessed on: 2022-03-17).

[93] Meticulous Market Research Pvt. Ltd., "Speech and Voice Recognition Market by Size, Share, Forecasts, & Trends Analysis," Oct. 2019. [Online]. Available: https://www.meticulousresearch.com/product/speech-and-voice-recognition-market-5038, Accessed on: 2022-03-17.

[94] "HAPI FHIR," 2022. [Online]. Available: https://github.com/hapifhir/hapi-fhir (accessed on: 2022-03-17).

[95] "Android FHIR SDK," 2022. [Online]. Available: https://github.com/google/android-fhir (accessed on: 2022-03-14).

[96] UralicNLP, Version 1.4.0, [Software]. Mika Hämäläinen, 2022.

[97] Språkteknologigruppen, “Granskas ordböjare,” 2020. [Online]. Available: https://skrutten.nada.kth.se/granskaapi/inflect (accessed on: 2022-03-18).

[98] Noice, Version 1.3.3. Gangwar, Ashutosh, Sep. 2021. [App] Available: https://f-droid.org/en/packages/com.github.ashutoshgngwr.noice/

[99] NIOSH Sound Level Meter App, Version 1.2.5 for iOS. Washington, D.C., USA, National Institute for Occupational Safety and Health, 2021. [App] Available: https://apps.apple.com/us/app/niosh-slm/id1096545820

[100] National Institute for Occupational Safety and Health, “NIOSH Sound Level Meter App | NIOSH | CDC,” 2021. [Online]. Available: https://www.cdc.gov/niosh/topics/noise/app.html (accessed on: 2022-02-23).

[101] E. Crossley, T. Biggs, P. Brown, and T. Singh, “The Accuracy of iPhone Applications to Monitor Environmental Noise Levels,” *The Laryngoscope*, vol. 131, no. 1, pp. E59–E62, 2021, doi:10.1002/lary.28590.

[102] Meld, Version 3.20.4, [Software]. The GNOME Project, 2021.

[103] JiWER, Version 2.3.0, [Software]. GitHub, jitsi, 2021.

[104] Natural Language Toolkit, Version 3.7, [Software]. NLTK Project, 2022.

[105] Snowball stemmer, Version 2.2.0, [Software]. Snowball, 2021.

[106] E. Wickman, “Taluppfattbarhet med strupmikrofon,” 2018. [Online]. Available: http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-353351

# A

# Available Phrases in VISAC

Command phrases available to the user, sorted by selected category or attribute. In total 240 selection phrases were available.

**Category: Luftvägar**
luftväg,
luftvägar,
luftvägen,
luftvägarna,

*Attribute: Fria*
fri,
fria,
vi har fri luftväg,
vi har fria luftvägar,
patienten har fri luftväg,
patienten har fria luftvägar,
det är fri luftväg,
det är fria luftvägar,
luftväg fri,
luftväg är fri,
luftvägen fri,
luftvägen är fri,
luftvägar fria,
luftvägar är fria,
luftvägarna fria,
luftvägarna är fria,
patientens luftväg fri,
patientens luftväg är fri,
patientens luftvägar fria,
patientens luftvägar är fria,
luftväg fri,
luftväg är fri,
luftvägar fria,
luftvägar är fria,
luftvägen fri,
luftvägen är fri,

*Attribute: Ofria*
ofri,
ofria,
inte fri,
inte fria,
ej fri,
ej fria,
vi har ofri luftväg,
vi har ofria luftvägar,
patienten har ofri luftväg,
patienten har ofria luftvägar,
det är ofri luftväg,
det är ofria luftvägar,
luftväg ofri,
luftväg är ofri,
luftvägen ofri,
luftvägen är ofri,
luftvägar ofria,
luftvägar är ofria,
luftvägarna ofria,
luftvägarna är ofria,
patientens luftväg ofri,
patientens luftväg är ofri,
patientens luftvägar ofria,
patientens luftvägar är ofria,
luftväg ofri,
luftväg är ofri,
luftvägar ofria,
luftvägar är ofria,
luftvägen ofri,
luftvägen är ofri,
vi har inte fri luftväg,
vi har inte fria luftvägar,
patienten har inte fri luftväg,
patienten har inte fria luftvägar,
det är inte fri luftväg,
det är inte fria luftvägar,
luftväg inte fri,
luftväg är inte fri,
luftvägen inte fri,
luftvägen är inte fri,
luftvägar inte fria,
luftvägar är inte fria,
luftvägarna inte fria,
luftvägarna är inte fria,
patientens luftväg inte fri,
patientens luftväg är inte fri,
patientens luftvägar inte fria,
patientens luftvägar är inte fria,
luftväg inte fri,
luftväg är inte fri,
luftvägar inte fria,
luftvägar är inte fria,
luftvägen inte fri,
luftvägen är inte fri,
vi har ej fri luftväg,
vi har ej fria luftvägar,
patienten har ej fri luftväg,
patienten har ej fria luftvägar,
det är ej fri luftväg,
det är ej fria luftvägar,
luftväg ej fri,
luftväg är ej fri,
luftvägen ej fri,
luftvägen är ej fri,
luftvägar ej fria,
luftvägar är ej fria,
luftvägarna ej fria,
luftvägarna är ej fria,

patientens luftväg ej fri,
patientens luftväg är ej fri,
patientens luftvägar ej fria,
patientens luftvägar är ej fria,
luftväg ej fri,
luftväg är ej fri,
luftvägar ej fria,
luftvägar är ej fria,
luftvägen ej fri,
luftvägen är ej fri,

*Attribute: Luftvägsstopp*
luftvägsstopp,
vi har luftvägsstopp,
patienten har luftvägsstopp,
det är luftvägsstopp,
vi har stopp i luftväg,
vi har stopp i luftvägen,
vi har stopp i luftvägarna,
patienten har stopp i luftväg,
patienten har stopp i luftvägen,
patienten har stopp i luftvägarna,
det är stopp i luftväg,
det är stopp i luftvägen,
det är stopp i luftvägarna,

**Category: Andning**
andning,
andningen,

*Attribute: Normal*
normal,
vi har normal andning,
patienten har normal andning,
det är normal andning,
patientens andning är normal,
patienten andas normalt,
andas normalt,
andning normal,
andningen normal,
andning är normal,
andningen är normal,
patienten andas normalt,
andas normalt,

*Attribute: Ansträngd*
ansträngd,
vi har ansträngd andning,
patienten har ansträngd andning,
det är ansträngd andning,
patientens andning är ansträngd,
patienten andas ansträngt,
andas ansträngt,
andning ansträngd,
andningen ansträngd,
andning är ansträngd,
andningen är ansträngd,
patienten andas ansträngt,
andas ansträngt,

*Attribute: Nedsatt*
nedsatt,
vi har nedsatt andning,
patienten har nedsatt andning,
det är nedsatt andning,
patientens andning är nedsatt,
patienten andas nedsatt,
andas nedsatt,
andning nedsatt,
andningen nedsatt,
andningär nedsatt,
andningen är nedsatt,
patienten andas nedsatt,
andas nedsatt,

*Attribute: Saknas*
saknas,
andning saknas,
andningen saknas,
ingen andning,
ingen andning hos patienten,
andas inte,
patienten andas inte,
ingen andning,
patienten har ingen andning,
patienten saknar andning,

**Category: Cirkulation**
cirkulation,
cirkulationen,

*Attribute: Normal*
normal,
cirkulation normal,
cirkulationen normal,
vi har normal cirkulation,
patienten har normal cirkulation,
det är normal cirkulation,
normal cirkulation,
cirkulationen är normal,
patientens cirkulation normal,
patientens cirkulation är normal,

*Attribute: Pulslös*
pulslös,
vi har pulslös cirkulation,
patienten har pulslös cirkulation,
det är pulslös cirkulation,
patienten är pulslös,
cirkulation pulslös,
cirkulationen pulslös,
cirkulation är pulslös,
cirkulationen är pulslös,

*Attribute: Trycklös*
trycklös,
vi har trycklös cirkulation,
patienten har trycklös cirkulation,
det är trycklös cirkulation,
patienten är trycklös,
cirkulation trycklös,
cirkulationen trycklös,
cirkulation är trycklös,
cirkulationen är trycklös,

*Attribute: Fyllig*
fyllig,
cirkulation fyllig,
cirkulationen fyllig,
vi har fyllig cirkulation,
patienten har fyllig cirkulation,
det är fyllig cirkulation,

fyllig cirkulation,
cirkulationen är fyllig,
patientens cirkulation fyllig,
patientens cirkulation är fyllig,

*Attribute: Tunn*
tunn,
cirkulation tunn,
cirkulationen tunn,
vi har tunn cirkulation,
patienten har tunn cirkulation,
det är tunn cirkulation,
tunn cirkulation,
cirkulationen är tunn,
patientens cirkulation tunn,
patientens cirkulation är tunn,

**Category: Neurologi**
neuro,
neurologi,
neurologiska symptom,
*Attribute: Fullt vaken*
fullt vaken,
fullt vaken patient,
vi har en fullt vaken patient,
patienten fullt vaken,
patienten är fullt vaken,

*Attribute: Medvetslös*
medvetslös,
medvetslös patient,
vi har en medvetslös patient,
patienten medvetslös,
patienten är medvetslös,

*Attribute: Desorienterad*
desorienterad,
desorienterad patient,
vi har en desorienterad patient,
patienten desorienterad,
patienten är desorienterad

**DEPARTMENT OF ELECTRICAL ENGINEERING**
**CHALMERS UNIVERSITY OF TECHNOLOGY**
Gothenburg, Sweden 2022
www.chalmers.se