Speech Synthesis on the World Wide Web: Evaluation of the Uppsala University Internet Portal

Lisa Arrenius
(limpie@stp.ling.uu.se)

Master's thesis in Computational Linguistics
Språkteknologiprogrammet
(Language Engineering Program)
Uppsala University · Department of Linguistics and Philology

June 4, 2004

Supervisor:
Mats Dahllöf, Uppsala University

On behalf of:
Public Relations Office,
Department of IT strategy
at Uppsala University

Abstract

Accessibility on the World Wide Web is a process of designing web sites, so that the entire community of users may gain access to the information and communication provided by each site, regardless of the user's personal needs or possible disabilities. This process is regulated by official standards and guidelines, foremost the World Wide Web Consortium (W3C); the process is carried out by the web administrators, and in the end affects how the visitors experience the site. In the past five years, there has been an increase in accessibility awareness. Direct steps have been taken to open the web to the general public, by introducing software and alternative sources of information. Following these actions are evaluations, which assess various aspects of the web site: the user's experience during their visit, as well as the administrator's work process.

This thesis evaluates the accessibility and communication of the Uppsala University web portal (http://www.uu.se/), which is equipped with an automated web reading service, Readspeaker. Readspeaker is provided by Phoneticom (http://www.phoneticom.com/) and based on BrightSpeech, a text-to-speech system developed by InfoVox, which synthesizes a textual input into a sound file. The qualitative black-box evaluation includes a questionnaire and follow-up interview questions, posed to a group of six users, who frequently visit the university web portal. An additional assessment was done with web masters at Uppsala University, who shared their opinion of the Readspeaker service from an administrative point of view.

The thesis discusses terms, which are crucial to the assessment of the online text-to-speech service. The terms refer to the discipline of computational linguistics, from which the thesis subject originates, as well as to the accessibility standards Uppsala University's web policy comply with.

The evaluation shows that by providing their visitors with Readspeaker, Uppsala University attempts to adapt the web portal to the recommendations of accessibility, which concern information as well as Design For All. The various categories of assessment show that the users felt a general appreciation for the service, navigation and linguistic abilities. However, the overall request from the user group as well as from the web masters was improved interaction.

Contents

List of figures

Preface

The completion of my master's thesis has been an educational, exciting, and sometimes tiring journey. The subject of computational linguistics has proven to be as diverse and nuanced as I had hoped it to be, when I started my first year in 2000. Working so intensely with one particular topic during the course of a semester is a fascinating ordeal. There have been days when I merged into Readspeaker completely, and then there were those times when I'd rather watch paint dry than go anywhere near that thesis. Now that I have reached the finish line, the memories of tedious literature, slow Internet connections and a Xerox machine on strike become dear. Knowing that I can accomplish such an overwhelming task as putting together a master's thesis, I will be able to go out into the world and fend for myself.

There are so many people along the way that I know wish to thank. Firstly, I would like to extend my greatest appreciation towards my supervisor, Mats Dahllöf, who has known exactly how much of a supervisor I have needed. Thank you for all your answers! Furthermore, I wish to thank the Department of Information and the Department of IT strategy at Uppsala University; especially Mattias Bolkéus Blom, who made sure I had a place at the Administration for my thesis, and Åke Johansson, who has been very supportive in answering my questions. Thanks to Pelle Lindé for helping me with the computer. I am endlessly grateful to each and every one of the helpful and cooperative evaluation participants. Without you, there would not have been a thesis. Thanks to David Öhlin, who helped me with the online questionnaire. Finally, thank you to my loving family, friends and classmates for all your support and encouragement.

Lisa Arrenius

Uppsala, May 2004

1 Introduction

Accessibility is defined as ”providing access” according to the Merriam-Webster’s Online Dictionary (http://www.m-w.com), where the term ”access” is further defined as ”freedom or ability to obtain or make use of”. Since it was launched on the public market, this has been the one of the foremost objectives of Internet and its sub-channels; to provide information in a communicative manner to its users. There are many ways of enforcing this objective, some of which will be brought up in this thesis. Through information technological design, as well as through the developments made within the field of speech technology, the Internet has been made accessible not only to users without any special needs, but to a great part of the disabled community. Where there used to be a limitation imposed on users with little or no means of accessing written text, web sites can to, now a greater extent, be browsed by anyone, despite any special needs the visitor might have. The textual aspect of Internet has now been expanded to include graphics, film and sound clips. The user can log on to another user’s web camera and not only type their conversation but also talk directly to one another, using a microphone and head speakers to convey and receive messages. The extension of the Internet also means that accessibility and the technology to allow accessible sites must be managed and updated on a regular basis. According to Statistics Sweden (Statistiska Centralbyrån), approximately 80 percent of all Swedes between the ages of 16 and 74 have access to a personal computer in their homes. Furthermore, approximately 75 percent of all Swedes between the ages 16 and 74 have access to the Internet in their homes (Statistiska Centralbyrån, 2003.). Not only is the Internet expanding, so are the amount of users accessing its web sites.

One way of allowing the general public to access a web site is to implement an automatic text reader, a so called Text-to-Speech (TTS) system. This system is able to convert a given text input into a sound file, sending it back to the user for instant speech output. The system is particularly convenient on the Web, since many sites base their efficiency on frequent and regular updates. The automatic process is preferred to a pre-recorded voice, due to financial and time-efficiency aspects. One particular instance of automatic web reading systems is

Readspeaker, a TTS service provided by the Uppsala-based company **Phoneticom** (http://www.phoneticom.com/). The company was founded in 1999 by Sound Box Production with the purpose of expanding the field of Internet sound. After an introductory period of development and work shops, the company launched their first product, **Readspeaker One**, a year later. Readspeaker One is a basic automatic text reader, which translates text input into a sound file. Today, the service is used by approximately 80 Swedish and international web sites[1]. Uppsala University signed up with Phoneticom in 2002. This was a part of the action plan to increase accessibility and usability at the university, in accordance with the accessibility resolution the Swedish Government made in 1999 (Näringsdepartementet, 1999).

1.1 Purpose

The purpose of this thesis is to assess the capability and accessibility of the online TTS service, **Readspeaker**, available on Uppsala University's web portal (starting page: http://www.uu.se/). The thesis will evaluate how communicative and accessible the university's web pages are to users browsing the portal with Readspeaker, by means of a questionnaire and follow-up interview questions. The evaluation is performed according to the *qualitative black-box strategy* (see section 3.1.2)[2].

The thesis will examine the Readspeaker service from several perspectives, foremost considering language engineering and the areas therein. The term *language engineering* refers to two fields of discipline, computer science and linguistics, as well as the merge between the two (see Jurafsky & Martin, 2003)

The evaluation is performed on behalf of **the Public Relations Office** and **the Department of IT strategy** at Uppsala University, which base their work on the University's accessibility policy (Larsson, 2003) and Declaration of goals and visions for information technology (Larsson, 2003). The evaluation and the proc-

[1] For a list of company references, see http://www.phoneticom.com/kunder/index.html
[2] A detailed introduction to the qualitative evaluation is found in chapter 3, **Method**. Also see Cole, A. et al, 1995

ess, through which this thesis has reached a conclusion, are, however, carried out independently and the conclusions are my own.

1.2 Material

The TTS service, Readspeaker, is available on the Uppsala University web portal, and is provided on 23 out of the 250 pages in the university web catalogue. The service first translates a given web site's html-code into consecutive written text. The written text is then translated into synthetic speech. The web information is resent to the user as a sound file. There are three types of services available today, Readspeaker One (as mentioned above), **Readspeaker Navigator** and **Readspeaker ReadIt**. The two latter allow the user to interact with the web contents, in a way the first service could not. Navigator is the service most in use on the university web portal, and is therefore the target of this evaluation. As shown in figure 1, it provides the top section of the web window with an added frame of hyperlink icons, which allows the user to choose what type of information they wish to hear: all contents, text or a list of links. The user may stop or abort the service at any time, as well save the sound file to their hard drive.

Figure 1 Readspaker Navigator frame[3]

All that is required of the client is a link to the service server, http://www.-uu.se/lyssna, as well as a standard sound card and speakers in the user's computer. The rest is managed on the provider's end. To simplify the design, the Public Relations Office provides all university webmasters with templates from the web site. The templates are merely recommendations (Johansson, 2004). As opposed to the other services, which only requires a standard web browser, Readspeaker ReadIt also demands JavaScript on the user's computer in order to function. This

[3] From the Uppsala University web portal with **Readspeaker**, http://isi.phoneticom.com/cgi-bin/uursnav

service allows the user to mark any section of the text on the web page and have it read aloud.

All three services are adaptable to the client's lexical and pronunciation needs and requirements, since the lexicon is managed by Phoneticom, the service-provider. They can also be adjusted to read some parts of the text and leave out others. ReadIt can also be adapted to personalize the sound quality and the type of synthetic voice chosen. All three services are based on the BrightSpeech synthesis (see section 2.1.2) produced by Babel InfoVox (Belgium, http://www.babel-tech.com/).

1.3 Outline of the thesis

Chapter 2, **Background**, will introduce the background to speech synthesis, its techniques, development and linguistic challenges, as well as the main outlines of accessibility standardization.

The structure of the evaluation analysis will be based on the ideal synthesis versus the actual performance of the Readspeaker service, estimating the difference between the two poles. This will be discussed in chapter 3, **Method: evaluation** and 4, **Results**. The final chapter (chapter 5, **Conclusions and Future development**) will give an overview of the results as well as a discussion of future solutions. Below is a block diagram of the thesis, containing terms and abbreviations that will be introduced in the upcoming chapters.

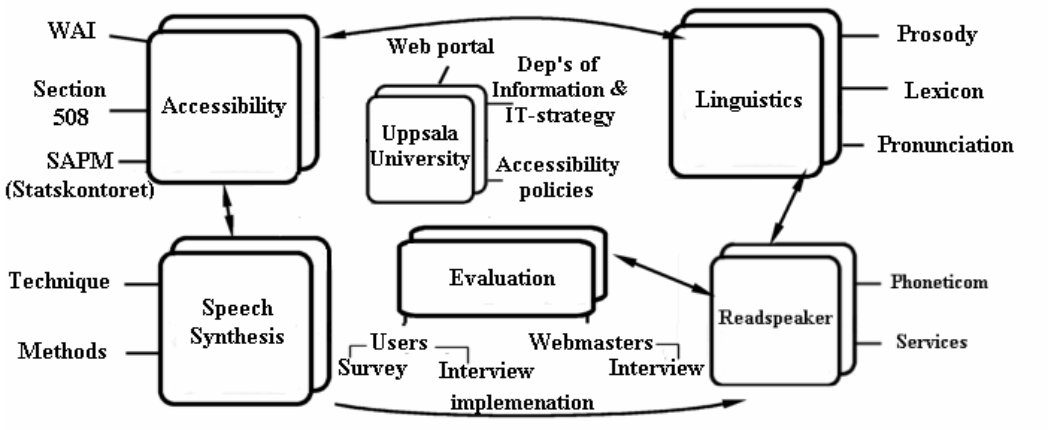

Figure 2 Block diagram of the thesis

2 Background

To provide an understanding for the Readspeaker service on the Uppsala University web portal, as well as its accessibility and communicativity, the introductory portion of this thesis has been to study several areas surrounding the service, its purpose and any regulations that can motivate Uppsala University to invest in and maintain such a service for their Web users. As shown in figure 2 (1.3), the thesis consists of several blocks that, together, contribute to the understanding of the service's presence on the university web.

The areas that are of greatest importance to the evaluation and understanding of the automated web site-reader will be introduced in this chapter. The speech synthesis will be introduced, as well as several crucial linguistic terms. Furthermore, the political as well as social standardization process will be discussed; a process that follows the development of an interactive World Wide Web, and that Uppsala University complies with in attempting to make their web portal accessible to all.

2.1 Speech synthesis

The web-based service Readspeaker, which is evaluated in the thesis, is an automatic translator of a textual input into synthetic speech. There are many steps to the process of translation; each crucial to the actual output quality and the degree of understanding that the user experiences. This process is vital to the comprehension and evaluation of Readspeaker, and will be presented in detail – but first, a brief introduction to the speech synthesis mechanism. According to O'Shaughnessy (2000: 339), a speech synthesis converts "input text (…) into speech waveforms, using algorithms and some form of previously coded speech data," where the input text is entered either by keyboard or optical character recognition (OCR). OCR involves scanning written text into a computer system, resulting in images, and "translating the images into a form that the computer can manipulate…." (Webopedia Computer Dictionary, http://www.webopedia.com/).

The input text can also be collected from a database storage, something that is used especially for a previously synthesized text.

There are several kinds of speech synthesizers that use various methods to process the textual input into a sound output. They are classified by speech unit size, as well as by the parameterization of speech for storage and synthesis (O'Shaughnessy, 2000: 343). Syntheses are most commonly distinguished as being either rule-based or data-driven. **Rule-based syntheses** base their output on an acoustic model of speech production. This method can create a vowel by modelling how the signal is led through a filter, which acts as the human vocal tract, creating a particular resonance (formant) depending on the articulatory part (Fahlstedt, 2004: 10-11). Generalized rules are extracted from the filtered information. The phonetic input of the synthesis is tested on the rules; when a match is found, digital speech is produced by the synthesizer. Rule-based syntheses are also referred to as **formant syntheses**, because the speech information behind the generalized rules is related to formant and anti-formant frequencies and bandwidths (Vinet, 2003: 3). As opposed to data-driven syntheses, this method does not use human speech samples in order to create the output speech. These types of syntheses are considered to have a greater degree of intelligibility[4], as well as be resource-economic, since they do not depend on stored segments of human speech. Since the output is completely artificial, the syntheses of this strategy can lack in naturalness. (Wikipedia, http://en.wikipedia.org/wiki/Main_Page)

Another method for creating artificial speech is merging slices of pre-recorded human speech. This type of synthesis is called **concatenative** or **data-driven**. Two subcategories can be distinguished within this method, **diphone** and **unit-selection**, differing mainly in the size of the units being concatenated. Both methods store the pre-recorded speech units in a database, from which the concatenation originates. Parts of utterances that have not been previously

[4] There are two major assessment terms:
Intelligibility, how much of the spoken language do you understand? How quickly do you become fatigued as a listener?, and **Naturalness**, how much like real speech does the output of the TTS system sound? (Speechworks, n.d. See also section 4.1.1)

processed and stored in the database are “built up from smaller units.” (Simosen et al, 2001: 58).

The diphone data-driven synthesis refers to a storage of all possible diphones in a particular language, which are then merged in accordance with the phonetization of the input text. A diphone is a speech unit consisting of two half-phonemes, or of the phonetic transition in between, e.g.

“Cat”: *silence* + c – c + a – a + t – t + *silence*

The other method, unit selection, collects its speech data from a depository containing units of various lengths, including diphones as well as words and phrases. Each pre-recording is stored in multiple occurrences, pronounced in different prosodic contexts (Skeppstedt, 2002: 8). Hence, this type of synthesis requires an extensive storage facility, and has only recently become a popular method, since memories and performance of computers have increased. According to Hunt & Black (1996),

> the primary motivation for a large database is that with a large number of units available with varied prosodic and spectral characteristics it should be possible to synthesize more natural sounding speech that can be produced with a small set of controlled units.

Since the diphone unit requires digital post-processing in order to incorporate prosodic information, the naturalness of the pre-recordings may be reduced. The unit selection synthesis, however, requires no digital post-processing; it simply concatenates the stored units as they are. Therefore, this type of synthesis usually ”gives the greatest naturalness.” (Wikipedia).

2.1.1 Text-to-Speech Systems

The methods of creating an automated signal, resembling that of human speech (see section 2.1), base their results on a limited input capability. These types of speech synthesizers are created to manage speech units, such as pre-recorded formants or waveforms, concatenating these into human-like speech. They usually have a high quality, but a restricted performance. “The output is limited to combinations of the speech units, usually with their original intonations” (O’Shaughnessy, 2000: 343), which means that the synthesizer has little or no flexibility in dealing with previously unprocessed text. Methods of these types,

which convert text input from formant information, are called **limited-text** or **voice-response systems**.

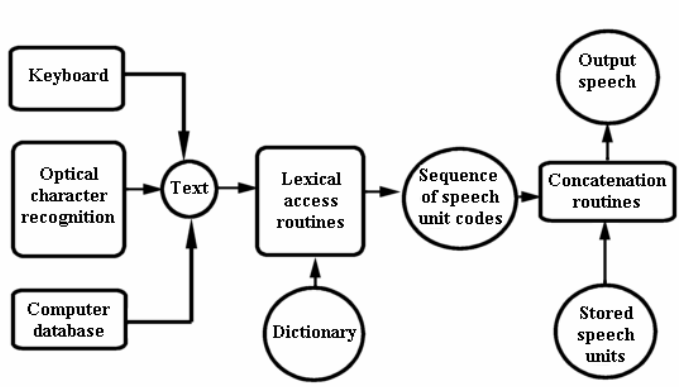

Figure 3 Block diagram of the steps of speech synthesis (based on O'Shaughnessy, 2000: 338)

For a system to be able to convert any textual input into speech, it requires a completely automated analysis of the internal, linguistic structures, as well as conditions that govern the translation from written input to spoken output (see figure 3 above). A system that can manage conversion of any given text, without requiring cost-inefficient manual labor, cannot rely on pre-recorded material alone. In the case of frequently updated texts, like those found to a great extent on the Internet, they would most likely lose their urgency if the process were in any way manual. The system that can accept and process general text is referred to as **unrestricted** or **Text-to-Speech** (TTS)[5]. O'Shaughnessy summarizes this conversion method as

> …processing [that] involves: (1) translating the input sentences into a sequence of linguistic codes to fetch the appropriate stored units, then (2) determining intonation parameters from the text, to vary F0[6] and duration properly. (O'Shaughnessy, 2000: 344)

Dutoit (1997) describes the TTS system as a computer-based system that should be able to read *any* text aloud, regardless of how it has been introduced into the system. Dutoit makes a point of emphasizing that TTS systems, are foremost capable of automatically processing new, i.e. previously unknown information.

[5] The system evaluated in the thesis will be referred to as a TTS.
[6] F0: the fundamental frequency of the vocal fold vibration, see O'Shaughnessy, 2000: 36

This, according to Dutoit, is what separates the TTS system from other talking machines (Dutoit, 1997: 1).

The process of converting text to speech requires a set of rules that translates each grapheme of the text input into phonemes. Each written element of the sequence of text must initially be transcribed into a spoken equivalence. In the definition of TTS systems, which Dutoit uses, this conversion is in the foreground: "…automatic production of speech through a grapheme-to-phoneme transcription of the sentences to utter." (Dutoit, 1997: 2). The system performs the transcription through a two-step process (as shown in figure 4), where, simply put, a **Natural Language Processing module** (NLP) produces a phonetic transcription of the input text, and a **Digital Signal Processing model** (DSP) transforms the transcribed symbols into speech.

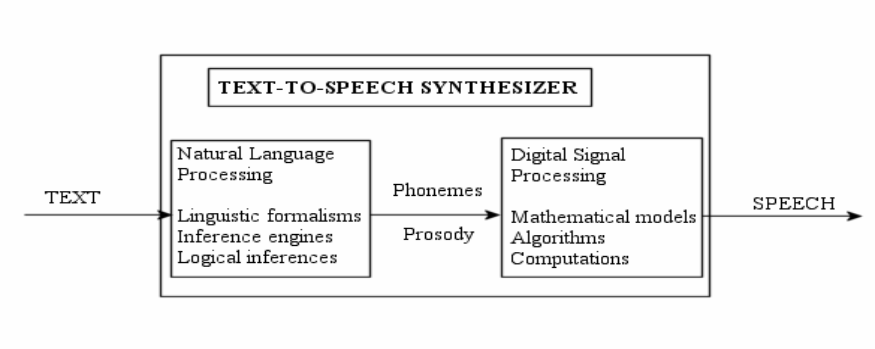

Figure 4 A simple but general functional diagram of a TTS system (based on Dutoit, 1997: 4)

As opposed to a voice-response system, which concatenates isolated words or phrases from a limited vocabulary, the TTS system constructs speech from text "using small stored speech units and extensive language processing" (O'Shaughnessy, 2000: 339). For text expansion and frequent lexical development, the TTS system is therefore far superior to formant synthesis. The key here is language processing, which requires phonetic representation as well as lexicon information through regular expressions.

2.1.2 Readspeaker

This chapter has dealt with the theory behind automated text reading, as well as various methods on which to apply the theory and create speech, independently or

from written input. For efficient conversion of frequently updated text to speech, the common method to use today is a TTS, which is free to process any text in a given language, regardless of the system's familiarity with the given input. A proper solution for a constantly updated information site such as a web portal would therefore be a TTS. Proper here means fast, simple and cost-efficient (Ishishara, 2002: 1).

Readspeaker is a TTS system that synthesizes speech through the InfoVox product, **BrightSpeech**. The synthesizer processes the input text, comparing it to a library of prerecorded human speech units, which, according the product information sheet produces "a clear and natural voice that sounds human and achieves a level of quality that has never been produced until BrightSpeech." (InfoVox, 2003). This will be further elaborated in chapter 4. The BrightSpeech synthesis is available for 18 different languages, including US English and Swedish (Ubult, 2004: 13). The product is adoptable to the specific needs of the client; there is a user dictionary, in addition to the standard word database, which allows the user to include "accurate pronunciations of foreign-specific words, acronyms or proper nouns." (InfoVox, 2003). These customary adjustments are made by regular expressions[7], which explicitly inform the synthesizer how an expression, idiom or abbreviation are pronounced. The regular expressions may also impose an overall change throughout a text, which has a continuous error. Before the synthesizer consults the dictionary, it analyzes the text input morphosyntactically, in order to achieve a prosodically correct output. According to Dutoit, natural prosody relies heavily on syntax (Dutoit, 1997: 5), a contextual piece of information that is required, since "to get a good prosody, the TTS system must 'understand' what it is reading" (Fahlstedt, 2004: 13, my translation).

The process behind the Readspeaker service, provided on the Uppsala University web portal since 2002 (Borgegård, 2002), consists of six steps (Larsson, 2004). First, the service retrieves the selected page from the Phoneticom server. It then revises the page to see if there have been any changes since its last

[7] A **regular expression** (abbreviated as **regexp** or **regex**) is a string that describes a whole set of strings, according to certain syntax rules. (Wikipedia)

translation. This is done by using a so called **checksum**[8], an operation which results in a value. The value is compared to the value, which is stored in the database. If the values are identical, the database will simply send the old translation to the user. Otherwise, the server's html-interpreter reads the page like a browser, excluding any or all html-tags that are not of interest to the end user. The text is then matched to a lexicon of words with their phonetical transcription. The interpreter can be adapted to browse only certain or all parts of the page, depending on the client's needs and the site's structure. Finally, the speech synthesis, BrightSpeech, records the text derived from the html-interpreter, before sending it to the end user as a sound file. See figure 5 below for an overview of the process of Readspeaker.

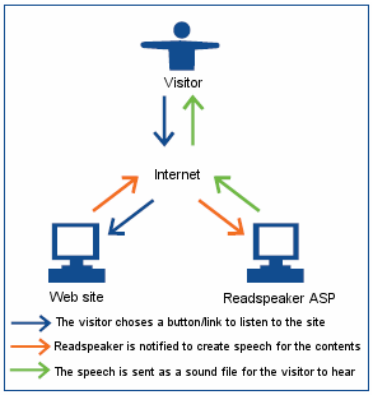

Figure 5 The Readspeaker process (Phoneticom, 2002)

In the process of automatically rendering speech based on a textual input, it is possible to encounter several challenges. Ambiguity, number and abbreviation conversion are a few of the imaginable issues that a TTS may come upon. The Readspeaker (BrightSpeech) synthesizer deals with the latter by pre-processing each occurrence in the input, translating abbreviations into the complete representation, matching dates with the generalized pronunciation rules and so on. Disambiguation is possible in some of the 18 languages available to the synthesis,

[8] A **checksum**: a method of mathematically determining if a series of numbers is correct, a sort of fingerprint for the value. The method can tell if the series is wrong, and in some cases where the error is and what the correct value should be (Susning.nu, http://www.susning.nu , my translation)

and is done by using a part-of-speech tagger. This establishes a linguistic context, from which the TTS performs the disambiguation.

The synthesizer is able to interpret the natural language of an input, choosing the correct synthesis to read the text as naturally as possible. However, this requires a clear, yet simple mark-up in the html-code, as follows:

`<html lang="sv">` or `<!-- isi_language sv -->`[9].

The excerpt indicates that the page contains the Swedish language and should be read by a Swedish synthesis. There are two Swedish syntheses on the Uppsala University web portal: Ingemar, an older version which uses a male vocal setting, and Emma, the new synthesizer with a female voice based on BrightSpeech, as well as an American female synthesis. A comparison of these three syntheses will be presented in chapter 4.

Finally, the company behind Readspeaker, Phoneticom, state in their documentation (as well as in the interview of March 18, 2004) that the main target group for the service of Readspeaker consists of people with reading difficulties, rather than visually disabled people. The service is by no means a competitor with the screen readers and other forms of assistance technology meant for the blind. The service is foremost a complement to web users with dyslexia, decreased eye sight from old age and/or a foreign background (Bergström, 2004).

2.1.3 Swedish prosody

The linguistic term **prosody** is used to describe several aspects of the spoken language. Speech has several dimensions that written language lacks and, when missing or failing in any way, can cause the language output to sound confusing or erroneous. These aspects are **suprasegmental**, i.e. concerns entire morphemes, words or phrases, as opposed to palatalization and velarization, which apply to the single phoneme. Translated into linguistics terms, the aspect of time is **quantity** (or **length**), the aspect of intensity **stress**, and finally, frequency is **intonation**.

[9] Code source of: http://www.uu.se/Webdesign/lyssna/lang.html,

2.1.3.1 Quantity

The quantitative aspect of spoken language depends on the temporal relations in an utterance. A word can have phonemes of various lengths – compare *write-written, get-gotten*. The Swedish language has a so called **V:C** vs. **VC: system** (Lundström-Holmberg & Trampe, 1987: 19). These rules demand that for each short consonant, it must be preceded by a long vowel. Vice versa, each long vowel must be preceded by a short consonant. Furthermore, when the length of a vowel changes, along with the length of the specific word, the pronunciation of it alters as well: e.g. *hat* [hɑ:t] (hate) – *hatt* [hat:] (hat)

2.1.3.2 Stress

There are two types of stress, that both have to do with alternation in volume (Lundström-Holmberg & Trampe, 1987: 19). In human speech, this kind of alternation is regulated in the respiratory muscles, and often indicates that the speaker wishes to emphasize a particular section of the utterance. E.g. *She didn't do* ***it*** as opposed to ***She*** *didn't do it*. These two utterances bear the same truth-conditional meaning, but pragmatically, they tell us two different things. In the first example, the agent is not cleared of all suspicions, only of those particular ones included in 'it'. In the second example, the agent is an innocent bystander, with no involvement at all. The utterances have the same morphosyntactic structure, and the same semantic meaning of the individual words. But when stress is added to a word, the meaning of the entire utterance is altered. Hence, this type of emphasis is called **word stress** (Lundström-Holmberg & Trampe, 1987: 19).

The other type is called **bound stress** (Lundström-Holmberg & Trampe, 1987: 19) and appears in some languages. The purpose of this type of stress is to indicate boundaries between inter-phrasal elements, not necessarily to indicate a change in pragmatic meaning, as does the type mentioned above. Bound stress appears in, among other, Slavic languages, where the stress always lies on the first syllable. Being aware of this fact makes it easy to notice when a word starts, and another ends.

2.1.3.3 Intonation

Variations in tonal level, or pitch, are in linguistic terms called intonation patterns. In human speech, the variations are caused by alternating the frequency of the vocal cord. Varying the intonation patterns in an utterance can indicate a difference between words with otherwise identical phonological representation. This type of intonation is commonly referred to as tonal accent (or simply **tone**). In the Swedish language, there are two intonation variations: **acute** and **grave accent** (Lundström-Holmberg & Trampe, 1987: 19). The acute accent is found in the plural form of 'steg' (step), *stegen*, whereas the grave can be located in, for instance, the definite form of 'stege' (ladder), *stegen*.

Intonation can also cover an entire sentence, indicating whether the utterance is a question (interrogative) or a statement (predicative), as well as if the utterance is expected to be followed by more utterances of the same speaker, or if the speaker is finished. This type of intonation is called **clause intonation** (or simply intonation). A speaker, who is nearly done, will symptomatically use a "**downdrift** (my bold mark), (...) a tendency for high tones to be somewhat lowered when preceded by a low tone, a phenomenon which results in falling intonation." (Katamba, 1989: 241) This phenomenon, according to Katamba, has a physiological explanation, where the speaker gradually runs out of air to support his/her speech.

2.2 Accessibility standards for the Web

Today, there are several guidelines, concerning web page accessibility. Uppsala University complies with two of the foremost standardization guidelines available, the **Web Content Accessibility Guidelines 1.0** (WCAG) (W3C, 1999) and **24-timmarswebben** (The 24 hour Web) (Statskontoret, 2004). To quote the University Policy of Designing Web Contents: "To pursue designing web information, the University should comply with the national and international recommendations, which have been composed, regarding the accessibility of the web sites" (Larsson, 2003: 1, my translation)

Therefore, this part of the background chapter will introduce relevant parts of these two documents and the authorities behind them. There will also be a section

on standardization documents, which are not directly applicable to the University accessibility plan, but that nevertheless are of interest to general accessibility on the web.

2.2.1 WCAG

The guidelines presented in WCAG were put together by the Web Accessibility Initiative (WAI), a committee branch of the **World Wide Web Consortium**[10] (W3C). WAI works towards three specific, long-term goals: 1. Universal Access, 2. Semantic Web, and 3. Web of Trust (W3C, 1994, http://www.w3c.org/Consortium/#mission). The process towards enforcing these goals includes writing and distributing specific guidelines for web authors and users around the world[11].

The primary purpose of the WCAG guidelines is ”to promote accessibility” (W3C, 1999: 1), i.e. one of three main goals of the W3 Consortium, mentioned above. To do so, the guidelines contain more or less exact recommendations of a web site’s design and contents, in order to make the Web transform gracefully, i.e. be accessible to all users, despite any physical or cognitive constraints or disabilities (W3C, 1999: 5).

Although the guidelines aim towards access for disabled users, the main idea for the WCAG and the other areas of the W3C documentations is access for all. If a site is accessible for a disabled user, it will pertain to a user without special needs, as well: ”each accessible design choice generally benefits several disability groups at once and the Web community as a whole.” (W3C, 1999: 1). This idea is usually referred to as **Design for All** (see section 2.3.2).

The WAI web accessibility guidelines are organized into several categories, all concerning accessible design for the Web. Each category consists of one guideline

[10] The Consortium is a global cooperation, founded in 1994, and hosted by three different institutions around the world: the Massachusetts Institute of Technology Laboratory for Computer Science (MIT/CSAIL) in the United States, the European Research Consortium for Informatics and Mathematics (ERCIM) in Sophia-Antipolis in France, as well as the Keio University Shonan Fujisawa Campus in Japan.

[11] The series of guidelines contain, aside from the afore mentioned WCAG, User Agent Accessibility Guidelines, as well as Authoring Tool Accessibility Guidelines. Although important, these two documents will not be introduced in the thesis, simply due to lack of time and space.

and one or more checkpoints. Each checkpoint has a priority level assigned by the Working Group based on the checkpoint's impact on accessibility from one down to three (my bold marks, to show where the priorities differ[12]):

> [Priority 1]
>
> A web content developer **must** satisfy this checkpoint. Otherwise, one or more groups will find it **impossible** to access information in the document. Satisfying this checkpoint **is a basic requirement for some groups to be able to use** Web documents.
>
> [Priority 2]
>
> A web content developer **should** satisfy this checkpoint. Otherwise, one or more groups will find it **difficult** to access information in the document. Satisfying this checkpoint **will remove significant barriers to accessing** Web documents.
>
> [Priority 3]
>
> A web content developer **may** satisfy this checkpoint. Otherwise, one or more groups will find it **somewhat difficult** to access information in the document. Satisfying this checkpoint **will improve access to** Web documents. (W3C, 1999: 7)

These guidelines make it easy for the reader to take into account precisely what is valid for their web site design. The recommendations are clear and well-formulated, and referrals to other sections, as well as other W3C guidelines, are present throughout the documentation. The guidelines concern all parts of the web design and browsing stages.

Guideline 1 (W3C, 1999: 8-10) discusses the issue of auditory and visual contents on a web page. Checkpoint 1.1 states that each non-text content (such as pictures, videos, and pre-recorded audio) must be provided as a text equivalent. The text equivalent serves an important function for visually impaired users, who rely on supportive technology, such as screen readers and speech synthesizers to browse the Internet. However, there are users who benefit from non-text contents, such as illiterate or dyslectic, which is why this guideline emphasizes the presence and frequent update of both text and non-text contents. This issue is also emphasized in Guideline 3, which discusses the use of markup and style sheets, i.e. layout specifications for a particular web site.

[12] In presenting the guidelines, the words **must**, **should** and **may** will show which priority the guideline has.

Guideline 2 (W3C, 1999: 10) stresses the importance of color-independent web pages. Color alone must not be used to convey information, since neither colorblind users nor devices with non-color displays will receive the information.

The issue of natural language usage is discussed in Guideline 4 (W3C, 1999: 12-13). A language markup clarifies the pronunciation and interpretation of foreign text. Checkpoint 4.1 states that a document's text and any text equivalent must clearly identify changes in the natural language through markup. Assistive technology, such as speech synthesizers and braille devices, can automatically switch to the new language, if it is clarified in the source code. Moreover, a natural language markup allows search engines to find key words and identify documents in a requested language.

Guidelines 5 (W3C, 1999: 13-14) and 6 (W3C, 1999: 14-15) further state the importance of graceful transformation, concerning web site tables and pages featuring new technologies. According to checkpoint 6.3 (W3C, 1999: 15), a page must be usable although scripts or applets are turned off or not supported. If it is not, an alternative page with equivalent information must be provided.

User control is discussed in guidelines 7 (W3C, 1999: 15-17) and 8 (W3C, 1999: 17), concerning automatic changes in contents as well as embedded objects. The configurations of automatic changes to a layout must be accessed by the user; if it is not, the guideline recommends avoiding these automatic alterations.

Guideline 9 (W3C, 1999: 17-18) recommends that a page is accessible through various input devices, e.g. mouse, keyboard, voice, head wand, or other. The guideline stresses the importance of text equivalents to non-text contents, which allow users to interact without a pointing device (see Guideline 1 above).

The issue of frequently updated technology, related to the Web, is discussed in guidelines 10 (W3C, 1999: 18-20) and 11 (W3C, 1999: 20-21). These guidelines recommend using W3C technologies, i.e. standard, accessible formats endorsed by the consortium, such as HTML and CSS. If non-WC3 formats, e.g. PDF, Shockwave, etc, are used, an alternative version of the content should be provided, according to checkpoint 11.4 (W3C, 1999: 21). The alternative version must be updated as regularly as the primary page.

Guideline 12 (W3C, 1999: 21-22) states that a page must have context and orientation information. This means that each frame of the page must be titled, according to checkpoint 12.1 (W3C, 1999: 21-22). Furthermore, guideline 13 (W3C, 1999: 22-23) recommends that the navigation mechanisms are provided clearly. Each link on a page should explicitly state its target, according to checkpoint 13.1, and related links may be grouped and identified.

Finally, guideline 14 (W3C, 1999: 23) recommends that web documents are clear and simple, which means that page layout is consistent, graphics are recoqnizable, and the language is easy to understand. The page must be written in a clear and simple language, to promote effective communication (checkpoint 14.1). Checkpoint 14.3 stresses that the style of presentation may be consistent across pages.

2.2.2 The 24-hour web (Sweden)

The Swedish Agency for Public Management (SAPM, http://www.stats-kontoret.se) is associated with the Swedish Government. The Agency's main purpose is to promote the development of a just, democratic, and efficient public sector. It acts in five different areas, which all apply to the general purpose:

1, Investigations and evaluations

2, Modernization of state administrations

3, Procurement coordination

4, Electronic civic information

5, Administration and disposal of premises (Statskontoret, 2004)

The areas of interest for this thesis are numbers two and four, modernization of state administrations and electronic civic information. These two created the other standardization recommendations, which Uppsala University comply with in their web accessibility policy: 24-timmarswebben (the 24-hour web).

The purpose of this branch of the SAPM is first and foremost to develop web sites that are adapted to its purpose, efficient and simple to manage for as many users as possible. In 1997, the 24-hour authorities network was first named after the idea of a constantly open, accessible online government agency that works in

the open and efficiently, which makes it easy for any citizen or business to access. Today, the name is used by the Government, not only for authorities but also for county councils and municipalities.

The 24-hour authorities document (Statskontoret, 2002) discusses issues concerning accessible web pages (Statskontoret, 2002: 51-54), for instance the term **Design for All** as well as recommendations to comply with the W3C's guidelines for content accessibility (see section 2.3.1). Furthermore, the document stresses the importance of communicating in a clear and concise language. A text should not only be easy to understand, but it should also be to the point, avoiding irrelevant contents (Statskontoret, 2002: 55).

The concept Design for All came about in 2002 by the European Design for All e-accessibility Network. The term is referred to throughout the 24-hour authorities documentation, and can be summarized as follows:

> Design for All requires that the developer of a product or a service considers the human variation, when it comes to qualities, capabilities, and preferences. Design for All opposes the idea of an average person. (Statskontoret, 2002: 24, my translation)

It is important to remember that designing web sites for a particular group, excluding others, requires a greater long-term effort, since the greater the general accessibility is, the more the site reaches out to a larger community.

2.2.3 Section 508 (USA)

Up until now, this section of the chapter has introduced the two standardization guidelines, which Uppsala University endorses in its efforts to develop accessibility on the university web portal. There are several other interesting documents available, which play an important role in the development of Web access for all. The Unites States of America passed a law in 1998, claiming Information Technology accessibility for the general public. The law, **Section 508**, is an amendment of the 1973 Rehabilitation Act, an act whose purpose was "to empower individuals with disabilities to maximize employment, economic self-sufficiency, independence, and inclusion and integration into society" (U.S. Rehabilitation act, 1973). The Section 508 standards merely apply to the Federal sector; they do not concern the private sector. However, if a member of the private sector is

contracted to design a web site for a member of the Federal sector, the new web site must comply with the Standards.

The amendment applies to all institutions that fall under the Federal sector, including those in the areas of communication, computing, and presentation. Furthermore, it covers the full range of **electronic** and **information technologies** in this particular sector. Electronic and information technology are defined in the Section 508 standards as: ”any equipment or interconnected system or subsystem of equipment, that is used in the creation, conversion, or duplication of data or information.” (U.S. Rehabilitation act, 1973)

There is one particular section of this document that is of great interest to this thesis, paragraph 1194.22: *Web-based Intranet and Internet Information and Applications* (Section 508, n.d). This paragraph states that the Section 508 standards comply with the W3C’s Web Accessibility Guidelines, by presenting a paraphrased list of 12 WCAG checkpoints (see section 2.3.1). Among others, a text equivalent must be provided for each non-text element, as well as for any multimedia presentation (see WCAG Checkpoint 1.1 and 1.2, W3C, 1999). Moreover, the alternative pages that provide these text equivalents shall be updated whenever the primary page changes (see WCAG Checkpoint 11.4, W3C, 1999). The list also contains five additional points to the WCAG, stating for instance the importance of providing the users with a method of skipping repetitive navigation links. Furthermore, paragraph 1194.2 states that agencies shall provide individuals with disabilites with the information and data involved by an alternative means of access.

2.2.4 *e*Europe 2002

In the Uppsala University Policy of Web Accessibility, the referral to WCAG is done through its main endorser, the European Union *e*Europe 2002 Action Plan. *e*Europe was formed at the initiative of the European Commission in 1999. The objective of this formation was to bring Europe online (Council of the EU, 2000: 1). The Action Plan was adopted at the Feira European Council in Portugal, in June of 2000, with the purpose of ensuring that the targets set by the Lisbon

European Council are reached by defining the necessary measures (Council of the EU, 2000: 1). In order to achieve this purpose, the European Council set a deadline for the Action Plan, since ”tight deadlines in critical areas for the new economy is one of the key driving forces of the *e*Europe initiative.” (Council of the EU, 2000: 3). The deadline was set to 2002, with reservation of long-term projects within the Action plan needing more time.

Among the objectives found in the Action Plan, there are three that are of special interest to this thesis. Objective 1 b, *Faster Internet for researchers and students,* (EU, 2000: 8), which states that high-speed networks will open new possibilities for collaborative learning and researching in and between public and private sector. Furthermore, objective 2 c, *Participation for all in the knowledge-based economy* discusses the urgency of “special attention (…) given to disabled people and the fight against ‘info-exclusion’” (EU, 2000: 17). This is to be managed through a possible legislation, which will ensure the conformity and accessibility of the information society (EU, 2000: 17). Finally, objective 3 b of the *e*Europe Action plan, titled *Government online: electronic access to public services* (EU, 2000: 22), calls for new technologies to make information as accessible as possible. The aim is to ensure easy access to essential public data for all users. More on how these objectives apply to the Readspeaker service on the Uppsala University web portal will be discussed in chapter 4.

3 Method: evaluation

There are numerous theories that discuss the appropriate means of reaching as fair an evaluation as possible. The ideal method would include the participation of a representative[13], impartial group of individuals, whose replie to the questions posed by the researcher, would be unaffected by any form of bias (see section 3.2.2) or subjectivity. This type of situation is difficult to establish, but it is up to the researcher to find a situation that resembles it to the greatest possible extent. This chapter will introduce and motivate the methodology behind the thesis' evaluation, its conditions and shortcomings. Furthermore, the composition of the questionnaire and interview will be discussed. Finally, the participants will be introduced, as well as how they were asked to partake in the evaluation.

3.1 Evaluation strategy

There are four major evaluation strategies, which are applicable to different kinds of systems and situations, discussed in Cole et al (1995). The **adequacy** strategy evaluates the suitability of a system as regards a specific user group. The **diagnostic** strategy is most commonly used by system developers in order to evaluate the "taxonimization of the space of possible inputs" (Cole et al, 1995: 476), i.e. the system's capacity to store and sort incoming data. A combination of the two methods mentioned above is the **formative** strategy, which evaluates a system's performance with respect to its intended users and use. Finally, Cole et al refers to the **performance** strategy, which is similar to the diagnostic strategy mentioned above, and applies to the performance value of the entire system.

Each strategy can illustrate a particular usability perspective of the system, depending on the evaluation's goal users. Cole et al distinguish between two usability methods: the **black box** and the **glass box** evaluation (Cole et al, 1995: 476). For every system there are a number of individuals involved, representing a specific target group: the system developer, the system administrator, or the end

[13] Representative results are results that to the greatest extent give a correct representation of the conditions in the group from which the participants have been selected. (Holme & Solvang, 1997)

user[14]. The latter is defined above, since this is the target group of greatest interest to the Readspeaker evaluation. *The strategy used in this thesis is the* ***qualitative*** (see section 3.1.2) *adequacy black box.*

3.1.1 Black box vs. glass box

An end user (see definition above) has little or no insight in the actual process of a particular system. What goes on after the user has entered the input and until the output presents itself is merely of some interest to the user, for instance how long the process takes. A black box strategy (see figure 6 below), according to Cole et al (1995), evaluates the system as a whole, as opposed to glass box evaluations, which assess the separate components and how they together form a system entity (Cole et al, 1995: 476). Therefore, this evaluation, which focuses on the end user and their opinions and comments regarding the Readspeaker service, will use the black box strategy.

Figure 6 Simple diagram of the black box adequacy evaluation

A glass box strategy would be applicable if the system's developers and administrators were to participate in the evaluation. But since the purpose of this thesis is to assess the accessibility of the Uppsala University Web portal, the most direct path to take is to ask the end users themselves. However, an additional part of this thesis' evaluation process has been to ask the opinions of the webmasters at the University. This part has not been structured according to the evaluation methodology, due to its less significant role toward meeting the purpose. It is, nevertheless, interesting to hear how the administrators feel towards the service and implementation thereof, in contrast or conformity with the opinions of the end users (see chapter 4).

[14] An end user is defined as "a natural person who makes use of resources for application purposes", (Brown et al, 2002)

3.1.2 Qualitative vs. quantitative

Holme & Solvang (1997) distinguishes between two kinds of evaluation methods that concern the composition of the evaluation group. In **quantitative evaluation**, the focus lies on the statistical representativity. Conclusions made from data originating from a group of 500 participants are statistically more reliable than data originating from a group of five participants. If the evaluation, however, wishes to capture the in-depth experiences and opinions of participants, chosing a quantitative strategy may be unwise. The time and effort that goes into interviewing that large a group, as well as analyzing the information, would not necessarily lead to more substantial results. The size of the data would most likely be too extensive to handle within a limited time frame.

The foremost objective of a **qualitative evaluation** is to come to terms with certain factors, and how they influence the evaluation. This method focuses on the underlying conditions behind opinions, rather than the opinions themselves (Holme & Solvang, 1997: 92). In the qualitative evaluation, there are fewer issues concerning the validation of data, since the researcher and the participant have a closer connection. The researcher has a better understanding of the conditions that shape the participants' answers and comments (Holme & Solvang, 1997: 93). In a situation of familiarity between the researcher and the participant, issues may arise that need special attention and analysis. There may be cases of bias, as mentioned earlier (see section 3.3), and there may be issues of prejudice[15] and expectations that obstruct, rather than support, the evaluation. Holme & Solvang (1997) mention that problems may arise when the participants act and respond as they believe the researcher *expects* them to, rather than in a completely truthful manner. However, the method of qualitative, to a greater extent than quantitative, evaluation has the ability of identifying and understanding these issues. As the researcher processes and reprocesses the answers provided by the participants, there is an increased awareness, which may reduce evaluation problems, such as the researcher's prejudice, misconceptions between the interviewer and subjects,

[15] Prejudice: Socially founded, subjective opinions of the phenomenon subject to the evaluation (Holme & Solvang, 1997: 95)

and effects of bias between the researcher and the evaluation participant. The qualitative research process, described in e.g. Holme & Solvang (1997), is based on an interaction between the researcher and evaluation participants. The information extracted from a first questioning with the participants can lead to further questions and new conclusions, which in turn may raise new questions. This cycle of evaluation may lead to a more valid outcome. However, it may also cause the participant to adapt their answers, as a pattern of questions/answers becomes obvious over time.

In sum, the strategy of quality – despite certain negative aspects – is suitable in a situation of few participants, and few participants are equally suitable when dealing with qualitative issues, such as an in-depth understanding for underlying factors, which influence a certain individual's opinion.

3.1.3 Evaluation method: questionnaire and interview

The process of collecting the information from the participants demanded that the following matters were taken into consideration. Most importantly, the evaluation had to be shaped and performed on the basis of the participants' terms. Not only would this make the participants more comfortable, but it would also lead to a greater and more analysis-worthy collection of data. The focal point of the qualitative interview, according to Holme & Solvang (1997: 99), is the interviewee. Hence, it is vital that the questions are as adaptable to the participant's disposition and to the interview situation as possible. However, since the data must be analyzable, there has to be some type of structure to the questions as well. Ejlertsson (1996: 81) stresses the importance of posing questions that are categorical and more or less similar throughout each interview.

Thus, the more flexible a question is the more qualitative the evaluation. However, the more structured the question, the easier it is for the researcher to analyze its answers (Bell, 1995: 74). It seems efficient to find a middle path, between the genuinely qualitative and the more structured evaluation method, which is why the method of evaluating Readspeaker is divided into two parts for

each participant: a questionnaire and a follow-up interview, exploring the opinions that the questionnaire revealed.

3.1.3.1 Questionnaire and interview

The evaluation took place during a period of two weeks in April 2004, at the Public Relations Office, St Olofsgatan 10 B in Uppsala – with the exception of the very last session, which was held at the participant's office in the Student Union house. Each evaluation session took between 40 minutes and 1 hour, and was held on an individual basis, i.e. with one participant and myself present. The participant was initially asked to browse the portal for as long as they wished, in order to become as comfortable with Readspeaker as possible, before being presented with a 10 question written questionnaire on paper (see Appendix A). Two of the participants answered the questionnaire online, at a later stage, in order to use their screen readers. As the participant browsed the Internet, I sat as quietly as possible in the background, taking notes on problems or special situations the participant encountered (see section 4.1). Since the users may experience my presence as stressful or influential in any way, they were asked to acquaint themselves with Readspeaker before coming to the evaluation. By doing so, they could shorten the evaluation session, as well. The questionnaire was followed by a longer interview, consisting of six categories of discussion, with several attendant questions per category (see Appendix C). The interview was recorded on tape, and then transcribed for the upcoming analysis.

There were several categories of interest to the evaluation of Readspeaker, which formed the structure of the questionnaire and interview. Since the purpose of the thesis was to assess the usability as well as the accessibility of the Readspeaker service, the questions of **Navigation** and speech naturalness (**Language**) were obvious candidates. Since the service platform is the Web, it must be able to integrate with the elements and plug-ins found online. Therefore, another category that was included in the evaluation was the functionality of the service on **the Web**. Finally, the qualitative evaluation should ask questions about the participant's individual needs and behaviors. Categories that tended to these

issues were **Personal needs –** how the participants responded to written and spoken text – and the participant's **Future use** of Readspeaker.

There are seven major categories of questions in a questionnaire, each with a particular type of answer (Bell, 1995: 74-75):

- Open questions
- Multiple answers
- Categories
- Ranking
- Scale
- Quantity
- Table

The main part of the questionnaire included in the evaluation of Readspeaker consists of **multiple, graded answers**, in order to account for the participants' opinions, and simultaneously prepare for a manageable analysis of the answers. Questions 1–4 establish the participant's experience of the Internet and of Readspeaker; questions 5–7 reveal their opinions of the service, itself. The latter part is explored further in the interview. In questions 8-10 (the background section), the participant is asked to state their gender, age and occupation. These questions are of type **category**, where only one answer, linked to one of several distinctions, is possible.

3.1.4 Two perspectives: end user and webmaster

Since the purpose of the thesis is to assess how well Uppsala University's web portal addresses its users and allows them to interact with the information provided on the web page, it is natural that the participants of the two-part evaluation are frequent visitors, who are acquainted with the specifications of this particular web portal, including terms, titles and information that applies to a particular community (see section 4.1.1). This thesis also discusses the opinions of the people behind the web portal, the webmasters, as a comparative aspect to the user evaluation. The communication is a symmetrical process between the service provider and the representative of Uppsala University, web manager Åke

Johansson of the Public Relations Office, who in turn provides the departments with information concerning the Web (Johansson, 2004). The end users have little influence in the administrative or developmental stage, which is why it is necessary to include the webmasters in the evaluation. This was done by means of an e-mail correspondence between randomly chosen webmasters and myself (see section 4.1.2).

3.2 Evaluation issues

In the process of preparing for the evaluation, as well as during the actual performance, several questions have been raised that are related to the selections made and possible difficulties that may arise, as the work progressed. This section will discuss some of the more evident issues, which the evaluation has raised. It is vital to acknowledge these issues, as they may have a great impact on the results of the evaluation. Lifting them to the surface, as well as keeping them in mind while interpreting the results, could lead to a more just end result.

3.2.1 Ethical considerations

This thesis' ethical considerations are based on the guidelines (2002: passim) presented by the Council of Science (http://www.codex.uu.se/index.html). These guidelines list four main requirements for the protection of the individual throughout the research process: the requirement of **information**, **consent**, **confidentiality** and **use**.

The information requirement stipulates that the researcher must inform the individuals involved of the project's purpose and contents (p 7). Participation in any form is always voluntary, and the people involved may at any time terminate their participation. This is established further by the consent requirement, which states that the participants decide for how long, to what extent and on what terms they are involved in the evaluation (p 10). According to this requirement, there are three types of evaluations, **active, passive** or **a combination of the two**; each type requiring a certain degree of openness between the researcher and participants. In the case of this thesis, the evaluation was active, i.e. the

participants were observed through a first-hand study, which requires up-front information at the time of the evaluation (p 7). Each participant was therefore informed and approved of the questionnaire and interview questions at the time of the evaluation, as well as the fact that they would be included in a published report, searchable on the Internet.

The requirement of confidentiality is vital to the project, since the participants were guaranteed full anonymity in order to take part – a guarantee formulated to ensure "difficulty for outsiders to identify individuals or groups of individuals" (p 13, my translation). Furthermore, protecting the participants from having their identity revealed in any way, in the course of the evaluation or in the following report, is far more important than the effect, which the specific piece of information may have on the conclusions (p 13).

The final requirement states that "data concerning individuals, collected for the purpose of research, may not be used or lent for commercial purposes…" (p 14). Since the evaluation is ordered by Uppsala University, who in turn is a client of the commercial company, Phoneticom, this requirement may be an issue. However, the material presented to the employers will merely be the final, confidential report. No personal data will be revealed to the commercial interests.

Finally, it is important that the participants are allowed access to the findings and final conclusions of the evaluation, in advance. According to one of two recommendations stated in the Principals (p 15), the researcher should provide the people involved with ethically sensitive sections, controversial interpretations in the report before it is published. Furthermore, it is recommended that the participants are asked if they wish to read the final report. Both recommendations were followed in the Readspeaker evaluation, when an opportunity arose or if a participant specifically asked.

3.2.2 Research methodology

An important issue concerning research methodology is that of **research bias**, which is discussed in Bell (1995: 94). Bias may appear in interview situations, undermining the validity of the resulting data. Bell claims that there are several

factors that cause the interviewer to unconsciously influence the participant. Most commonly, these stem from the social connection between the two, such as the participant's will to appeal to the interviewer or an underlying hostility between the interviewer and subject; these phenomena are commonly referred to as response or interview effects (Borg, 1981: 87). Holme & Solvang (1996: 106-107) discusses the various theatrical elements that can influence the outcome of the evaluation, including the **role of the interviewer/-ee** and **the actors' ability to participate**. The role has to do with the expectations that the interviewer and partaker may have of each other, positive or negative, that are based on a personal relationship outside the interview situation or through social status differences.

Another important factor that may cause bias in a research situation is the actors, i.e. the interviewers, ability to acknowledge signals and moods, which the participant generates. The dilemma of ensuring that the information the interviewee gives is voluntary, while retaining one's own desire for information is important to acknowledge, and to the greatest extent avoid.

3.3 Evaluation participants

Although a qualitative evaluation does not focus on the statistical generalization that may be emphasized otherwise, it is still vital that the participants selected to take part are relevant to the target of the evaluation (Holme & Solvang, 1996: 101). This section of the method chapter will discuss the process of selecting and contacting participants for the evaluation of Readspeaker; who they are and why these particular individuals were chosen, as opposed to any others. The section will also discuss the issues behind selecting and establishing a connection to the participants, what problems arose and why.

3.3.1 Participant representation

The basis for forming a group of participants was twofold: the methodology of the evaluation, as well as the target group for Readspeaker (see section 2.1.2). For a qualitative evaluation of an automatic reading tool on a domain-specific forum, such as the Uppsala University web portal, a group of no more than 10 students

are required. The particpants should have some type of communicative deficiency, which could be remedied by Readspeaker. After having determined the number of people, with regards to the guidelines of a qualitative evaluation, and their personal abilities, the next step was to contact prospective participants. This was done in several ways. Birgitta Hydén, head manager for students with disabilities, included an advertisement for participants, in an EU-funded project that went out to all students with dyslexia. I also had an advertisement in the student newspaper, Ergo, on April 2, 2004:

> Användare sökes!
>
> Är du student vid Uppsala universitet? Använder du Internet dagligen eller ofta? Har du erfarenhet av datoriserad uppläsning av webben eller skulle du vilja pröva denna funktion? 5-8 personer sökes till ett studentprojekt med syftet att utvärdera
>
> Lyssna-funktionen på Uppsala universitets webbportal. Intresserad av att delta? Kontakta Lisa Arrenius, på lisal@passagen.se eller 073-621 47 97 (Ergo, 2004: 22)
>
> (Users wanted!
>
> Are you a student at Uppsala University? Do you use Internet daily or often? Do you have any experience of computerized reading of the web or would like to try this service? 5-8 people are wanted for a student project with the purpose of evaluating the Readspeaker service on the Uppsala University web portal. Are you interested in participating? Contact Lisa Arrenius on lisal@passagen.se or 073-621 47 97)

The above notice was also sent out as a mass mail to all computational linguistics students. Furthermore, there were personal inquiries made to friends and acquaintances. Finally, a group was put together, composed of seven students and one student union administrator. In the course of preparation, two of the students decided to leave the evaluation group. The group now consisted of Carin, dyslectic, Barbara, non-native Swedish, Erik and Fredrik, both blind, and Ulrika and Anna, without any special needs[16] - a small group, which represents a variety of communicative deficiencies.

The process of establishing a group of partakers was not initiated until the beginning of April, i.e. mid-term. The time prior to this was used to collect

[16] The participants are presented under assumed names, due to confidentiality.

background information, do research and talk to experts on the various subject matters. The plan of inviting participants when most of the background research was done was well planned and laid a good foundation for the understanding of the results to come. However, people are never predictable, and faced with the requirements of the research ethics (see section 3.2.1), it was sometimes difficult to coordinate the evaluation process. The two people who initially were intrested, but decided to leave, had to be replaced on short notice. Fortunately, the evaluation had not commenced, so losing data was not an issue. All in all, the participants were most accomodating, and the evaluation went smoothly.

4 Results of the evaluation

In this chapter, the results of the evaluation will be discussed, on the basis of several factors. The focus will be on the participants' personal opinions, comments and reflections regarding the Readspaker service on the Uppsala University web portal. These were revealed during individual evaluation sessions with the participants, consisting of a questionnaire and a follow-up interview (see section 3.1.3). Since the purpose of this thesis is to assess the accessibility and usability of Readspeaker, the opinions and comments presented by the participants will be discussed in light of the accessibility standards, which the university policies of IT and web design comply with, and which are presented in chapter 2. The evaluation analysis will also be discussed in the light of several guidelines on how to assess a TTS product (see section 4.3).

4.1 Users

Each evaluation session was divided into three segments: browsing (by launching Readspeaker Navigator on the Uppsala University starting page), a questionnaire and a follow-up interview (see section 3.1.3.1). Every segment contained similar themes, but the means of addressing these themes altered. Although the general evaluation situation was artificial, the initial segment of the evaluation was less structured, more on the terms of the participants than the latter. The three segments together support the analysis of the Readspeaker service in the eyes of its end users. Here is a summary of the evaluation interviews (figure 7):

	Anna	**Barbara**	**Carin**	**Diana**	**Erik**	**Fredrik**
Personal Comments *Any particularly good/bad elements?*	...it feels good (...) a innovative complement to the usual web sites	It sounds good but it needs improvement. It reads just everything, and that's not necessary.	(...) if it doesn't improve it is rather unnecessary ...it's great that it exists, (...) good that it has various functions	...better than I expected. ...I'm fascinated that it works... Ingemar stuttered a bit	...it gives many a chance that for various reasons have difficulties reading web pages	It's not made for the blind, that's obvious

	Anna	**Barbara**	**Carin**	**Diana**	**Erik**	**Fredrik**
Language *How well did you understand the voice? How was the pronunciation?*	... easy to listen to with a correct stress on syllables... ...Emma sounded softer and (...) more natural	Pretty well, but I had problems with the stress. No period at the end of a row, (...) difficult to under-stand. ...it can read long texts so well	The American page works fine but the Swedish is just chaotic. ...you understand the voice but you don't remember what it said ...correct pronuncation would be good	I thought the voice was easily understood, both the male and female ...strange when someone with a Swedish accent read (...) a German or English page	...the speech is generally good ...very well, I'm an experienced speech synthesis user I prefer Emma to Ingemar, she is more natural	Good speech, easy to hear Is it possible to adapt the speed? For e-mail addresses, is there a possibility of reading out each letter?
The Web *What worked/did not work?*	Not where I surfed at least	It can handle everything.	...some pages looked weird, otherwise nothing	..the scroll list, I don't think it read that part	You can't go to forms	It seems to read everything Can you do a search?
Navigation *What was good/not so good?*	...the spontaneous reaction was to click the link number ..oh the box, I could just as well click the link with the mouse	...pretty slow ...it works but it's slow when you try to listen to the text	After three links you have to count on the screen. ...if she could talk slower ...you can type the link	It works fairly well, but I think you have to get used to it	...skip the standard links... "Skip to main contents"	...good shortkey commands (...)unnecessary steps (...) I probably want the page read
Personal needs *Do you prefer written or spoken text?*	Both, actually. CD's along with course literature is a terrific complement	For a large materal I prefer written... ...it helps if I have a voice to associate with the text	Combined... ...listening and reading at the same time. That works the best.	...if I found a long text (...) it's not healthy to read at the computer	I use a lot of Braille... ...I only use speech as a support	A fast speech synthesis is faster than a fast Braille reader I use them parallel...

	Anna	Barbara	Carin	Diana	Erik	Fredrik
Future use *Will you use Readspeaker in the future? Why/why not?*	I don't think I will use it. It's quite time-consuming to hear a long list of links (...) Possibly (...) read text, its best feature.	I think I will use it. ...if I need to listen to Swedish, I'll log on to Uppsala University's web page and listen	Perhaps a bit, actually, more than before. ...there should be some kind of instruction	Since I've tested it, if I found a longer text (...) I'd be frustrated surfing with it.	Probably not that much since I have my equipment	I don't think I will ...perhaps if I'm somewhere and forgot my equipped computer

Figure 7 Table of quotations from the interviews

The questionnaire answers (see appendix B) show that the six participants were all experienced web users, visiting the Internet at least daily. It was important to establish this, since the responses they made to the following questions could be viewed on the basis of their Internet experience. Furthermore, the participants, all but one being students at Uppsala University, visit the university web portal at least a few times a month, or as much as several times a day. They are all rather familiar with the structure of the web portal's pages, structure and contents. However, none of the participants had used the Readspeaker service on a regular basis. Four of the participants had used it on occasion, and two had never used it prior to the evaluation session. It was therefore very interesting to observe as each participant browsed the university web portal with the TTS switched on, since any previous experience may have influenced the results of the evaluation.

4.1.1 Personal comments

So what did the participants think of the Readspeaker service? According to follow-up interview each partaker underwent, the most prominent of spontaneous reactions was "…it was better than I had expected" (Diana)[17] and "I was actually very surprised how good it sounded" (Anna). Some of the users were surprised

[17] All quotations are my translation, since the evaluation sessions were all held in Swedish.

that a computer could sound so natural, especially Erik, Anna and Diana. I asked Erik, who is blind, how well he understood the synthetic voice, and he replied "Very well." (Erik) Anna was asked if there was anything special she appreciated, whereupon she replied "…the voice was very good. It was nice to listen to and really correct stress on syllables and sentence structure…" (Anna). Carin was not equally positive in her personal comment. She even went so far as to call the service "unnecessary, if it is not improved it is pretty unnecessary." (Carin) Furthermore, she did not appreciate the Swedish synthesis' pronunciation; it was simply awkward and did not work at all, according to her. Barbara agrees in her personal comment that the service needs improvement, and that she hardly understood any of the Swedish synthesis if she did not follow along in the text.

4.1.2 Natural language mark-up

The participants found that international pages in the university web portal may be a problem for the Readspeaker service. Barbara wanted to read an English text about Nobel laureates at the university, but was forced to read along with the service in order to understand to whom Readspeaker was referring. Since the text consisted of English text and Swedish personal names, Barbara found the outcome confusing: "If there are any non-English or non-American words, it is difficult to understand what it is saying (…)" (Barbara). Diana agrees that this is a problem. She visited the German version of the starting page, but could not understand any of the contents, since the Swedish synthesis was launched. "(…) it became weird when someone with a Swedish accent that read when we chose some English-speaking or German-speaking page." (Diana) According to Erik, however, it is a matter of getting used to the language mixture. Being an experienced speech synthesis user, he claims he does not notice the oddities of pronunciation. "(…) I even listen to an English synthesis reading Swedish texts (…)" (Erik). The issue of the reader's multilingualism, introduced in section 2.1.2, is crucial to a web site that expects to have international visitors. Yet, the solution to this problem is quite simple: a mark-up code in the web page's source tells the synthesis which language synthesizer to use (Bergström, 2004).

4.1.3 Navigation

Navigation is another issue, which the participants have various opinions on. If the *Read all* or *Read links* is chosen in the Readspeaker Navigator panel (see section 1.2), the service announces the number of links. The links are then numbered as they are read aloud. The user is able to go from page to page by entering the link number in the Readspeaker frame navigation box. Once on the new page, the user is required to relaunch the service, in order to hear the contents read aloud. The evaluation of the navigation must be analyzed in the light of temporary factors, such as a slow connection or a glitch in the system. The participants generally agreed that this means of navigating was functional, although rather difficult to learn. Anna comments this by saying "…first, I hardly understood how I should go further when they were counting out all those numbers. The first spontaneous reaction was to click the number when they said it." (Anna) It was not obvious that the link numbers and the box in the upper frame were in any way connected. She goes on by saying that "I could just as well click the link with the mouse." (Anna). Fredrik agrees that the process of navigating Readspeaker contained unnecessary steps. "If I go to a new page, I have to hit the key to get there, then start the reader. I probably want the page read aloud, if I go there." (Fredrik) Since Readspeaker Navigator allows the user to decide what part of the web contents they wish to hear, the relaunching for every page is understandable, according to Fredrik, who still requests some kind of initial default reader. "…you can stop it if you don't agree with what it is reading." (Fredrik)

The figures in the questionnaire (see appendix B) reveal that the group of participants agreed that the navigation was neither excellent nor poor. Three of them found the navigation to be fine, three of them thought Readspeaker did not navigate that well. The ability to interpret the link numbers was highly appreciated by the dyslectic representative, Carin, who may encounter difficulties if the service merely required exact wording.

There were several features the participants requested. Aside from the possibility to launch the service automatically, as well as navigating independently of the box in the Readspeaker frame, speed adaptation of the voice is something sev-

eral of the participant mention is lacking of the service. ”You can’t remember what [the synthesis] said. Then you forget the links. If three links have passed, you still have to count on the screen (...) if she talked slower (...).” (Carin). The two blind participants, who claim to be experienced speech synthesis users, insist on the feature of speed adaptation. They both feel that this is the most apparent difference between functionality and user-unfriendliness. Diana experienced that several pages were not included in the Readspeaker service. She stated in her questionnaire that she used the university web portal to search for literature on the university library site. This was not possible with Readspeaker, since the library pages were not included in the service range. An error message told her that she had chosen a site that is not supported by the Readspeaker Navigator Service, and she was given the choice to visit the site without Readspeaker. This message appears again when she tries to do a search from the starting page. Fredrik requested the possibility of spelling out certain crucial information, such as e-mail addresses.

4.1.4 Interaction

Interaction is a key feature of the Web, a feature that Readspeaker has not quite incorporated. As Erik points out during his browsing session, ”How am I going to get to the input box?” (Erik), a question that summarizes the difficulty Readspeaker has to handle interactivity between the user and system. The problem of lacking interaction is evident throughout the different evaluation sessions, but with varying degrees of consequences for the participants. Diana simply returns to the starting page of the portal with Readspeaker turned off, to perform her search; the other visually abled participants solve this issue in the same way. Erik, on the other hand, has no alternative means of interacting with the page, and is forced to manually browse the pages in search of the particular piece of information he is interested in. Hence, interaction is the divide between the screen reader, the tool of visually impaired users, and the Readspeaker service. It is through this feature that the proclaimed target group of Readspeaker is established. The target group can handle the flaws of the TTS service; visually impaired cannot.

4.1.5 Personal needs

The function of **being read to** as opposed to reading along with the synthesis is an interesting point of discussion in the evaluation sessions. For a user with no or little need for a web reader, like Diana and Anna, the purpose of Readspeaker diminishes when the pronunciation fails. They claim that it is just as easy, or even easier, to "click directly with the mouse on the link" (Anna), browsing the web the way they are used to. Most of the participants would prefer the Readspeaker, if it were a supplement in the case of having one long, consecutive text to read, as opposed to a navigational tool for a multi-article page with numerous hyperlinks. In her interview, Diana mentions ergonomics as a reason to use Readspeaker instead of reading on a screen, something she feels is not healthy. She associates to riding in a car, when she would prefer a book on tape, as opposed to reading the book herself. The two blind participants agree that Readspeaker assists them foremost when they are away from their personal computer. In this case, they would not have access to their own screen reader, an aid they are accustomed to, which guides them from the computer start-up and onwards. Fredrik explains that the screen reader is linked to a Braille-keyboard, which cooperates with the screen reader, and informs him of the contents located directly under the keyboard marker. He is able to receive instant feedback from the screen reader of his whereabouts on the page, when navigating around on the web site.

4.1.6 Future use

Will the participants start using the service regularly? There was a difference in opinion here. Barbara uses Readspeaker to improve her Swedish pronunciation. She appreciates the help she gets from Readspeaker in this area, and will continue to use it. Erik and Fredrik may use it if they are at another computer, than their own, which are provided with a personal screen reader. Carin says she will most likely use it more than previously, and that it suits her reading needs. She is pleased that the service exists. However, she requests better instructions from the university, as to where the Readspeaker service can be found and how it works. She tells me that as a new, dyslectic student, she was presented with extensive

information regarding her education – where to find and order talking books etc, but no information about the website. Anna and Diana, however, are not so sure they will use Readspeaker in the future. They compare the service with their usual means of browsing; Anna and Diana agree that the service is too slow for them. However, they both find that the service is functional when it comes to reading consecutive texts, a situation they could imagine would benefit from Readspeaker.

4.2 Webmasters

After the final evaluation session, it was time to investigate the other aspect of the Readspeaker service: the administration. The webmasters are responsible for managing the pages, which the participants had browsed and evaluated with the use of Readspeaker Navigator. Since the foremost purpose of questioning the web administrators was to establish an understanding for the effort behind the evaluated interface, there was no intricate, theoretical process in contacting and formulating questions for these participants. Nevertheless, their comments are just as interesting for the evaluation, as a whole. The participating webmasters were chosen randomly from the Uppsala University web catalogue and contacted via e-mail with a few questions, concerning the service from an administrative point of view. Six administrators were contacted, three replied. The extent of the replies varied, but they were still comprehensive enough for this part of the evaluation. They all agree that adding the service to their particular site of the portal was an easy task; all that is required of them is a few lines of html-coding. This is an important condition in spreading accessibility. Another crucial aspect is that the service does not require bandwidth of the client, since the information is uploaded to Phoneticom's server for processing. However, there are a few drawbacks. The administrators feel they have little or no control over what content is read aloud, as well as how the page is navigated; a feature that is requested by the web administrators. They wish to be able to denote what parts of the contents are to be read, and what parts should be left out. This issue came up in the main evaluation, as well; the student participants complained that the service read everything. There was no possibility of hearing the latest update or links specific for the

particular page. Nor could the service leave out certain areas of text, e.g. copyright information.

4.3 Discussion

Several of the comments, which arose in the evaluation, can be discussed in the light of the standardization documents introduced in chapter 2. Regulations, which are signed and ratified by numerous nations, validate the opinions brought forth by the participants.

The ethical guidelines presented by the Council of Science (2002), discussed in chapter 3, apply to the Readspeaker evaluation in several ways. Foremost, the urgency of the consent requirement (see section 3.2.1) arose when establishing a connection with presumptive participants who later decided not to become involved. Although it caused the project a minor temporal set-back, it was important to realize that the participants' interest went before that of the thesis. Since the participants were active, in the respect that they were observed and interviewed first-hand (see section 3.2.1), each partaker was informed of the structure and contents of the evaluation as they were contacted, as well as before their respective evaluation session. Confidentiality was granted each participant, through assumed names and a well thought-through presentation of their interviews, so that no piece of information would reveal their identity.

According to SpeechWorks (2003), the responsibility of developing a functional TTS lies not only with the service provider, but also with the clients. The authors and administrators that provide the system with text to read aloud must be aware of that fact that ”the output problems are caused by the input text.” The White paper *Writing text to improve synthetic speech* discusses several solutions to problems related to text-to-speech systems, such as Readspeaker. There are two main elements that must be considered when writing for a TTS: **context** and **text formatting**. In a contextual environment, a symbol or word with various meanings may be interpreted erroneously, due to ambiguity. Usually, ambiguity is handled in the pre-processing stage of the synthesis (see section 2.1.2). However, when this is not possible, it is up to the web site author to formulate the text in an

unambiguous way. It is important to bare in mind that text-to-speech systems ”read precisely what you enter.” (SpeechWorks, 2003: 4) This makes basic linguistic issues, such as punctuation and spelling, all the more consequential. As Barbara pointed out, a text with no puncuation at the end of the row was difficult to understand.

Furthermore, Statskontoret (2002: 55) states that it is recommended the language on the web page be complaisant, clear and to the point. It is therefore the responsibility of the administrator and author of the web page, not only to make the visitor understand the contents, but also the TTS. The evaluation showed that browsing the web with Readspeaker can be tedious, since the service reads everything on the page, regardless of its news value. Statskontoret proposes a solution to this, by stating that ”the text should (...) be disposed to state the most important first, allowing the reader to read as much as they are interested in knowing.” (Statskontoret, 2002: 55)

Carin said in her interview that she was pleased the Readspeaker service existed, an opinion held by several of the participants, as well as by the webmasters. According to the *e*Europe Action Plan of 2002, the presence of Readspeaker on the Uppsala University web portal is necessary, since ”Public sector web sites (...) must be designed to be accessible (...)” (Council of European Union, 2000: 18). In accordance with objective 2 a of this action plan (2000: 17-18), the Readspeaker service acts towards **participation for all**, a notion discussed in several accessibility documents, some of which are introduced in this thesis (see section 2.3). Although only approximately 9 per cent of the pages included in the web portal are linked to the service (see section 1.2), it is still possible to access a majority of the pages through an initial launch of Readspeaker on the starting page.

The action plan succeeding *e*Europe 2002, *e*Europe 2005 (Council of European Union, 2002), proposes **Virtual campuses for all students**, a feature that ensures that ”all universities offer on-line access for students and researches to maximize the quality and efficiency of learning processes and activities.” (Council of European Union, 2002: 12), and which should be enforced by the year 2005. This proposal entails that the entire Uppsala University web portal shall be accessible

to all its students, and that an aid like Readspeaker must be as functional and wide-spread as possible. The students participating in the evaluation were able to browse a great portion of the university web portal, and since numerous departments rely on their web pages for the spreading of information, schedules etc., it is important that Readspeaker functions throughout the portal. There were several difficulties, as shown by the interview replies. Diana could not visit the online library, and Erik was not able to use the search function.

In the evaluation, there were no remarks regarding the service's ability to handle images. According to WCAG (Web Content Accessibility Guidelines) 1.0, each non-text element must be provided with a text equivalent[18], a regulation, which is also mentioned implicitly in the Uppsala University Policy for Web design: "The University should strive towards allowing all users within a target group to receive the same information. (…)" (Larsson, 2003) Since the web templates supplied by the university's Public Relations Office implement the web policy, the administrators are informed of the importance of the text equivalents. Subsequently, the portal's pages are well-marked, for the benefit of Readspeaker as well as for its users.

[18] Equivalent, according to WCAG: Content is "equivalent" to other content when both fulfill essentially the same function or purpose upon presentation to the user. (W3C, 1999: 27)

5 Conclusions and future development

5.1 Overall summary

The purpose of this thesis was to assess the online TTS service, Readspeaker, (provided by Phoneticom) on the Uppsala University web portal. The service evaluation was done mainly from the perspective of the end user, researching how communicative and accessible the university's web pages are to users browsing the portal with Readspeaker. The evaluation was performed according to the black-box strategy, which considers aspects that are of interest to the end user, i.e. what the system requires to process certain information and how it responds to these requirements. The composition of the participants was done with respect to the qualitative method of evaluation. This method does not require a large group of participants, since the main focus is the individual's detailed opinions, rather than the statistical validity of a large collection of homogeneous data. The formation was also made on the basis of Readspeaker's target group, as explained by its providing company, Phoneticom: users with reading disabilities, rather than those who are visually impaired. An additional, smaller evaluation was done from an administrative point of view, where webmasters at the university were asked their opinion of Readspeaker.

The user evaluation (see section 4.1), which consisted of a questionnaire and follow-up interview answers from six participants, revealed Readspeaker as being a relatively unknown, infrequently used service with several advantages. The general consensus of the group was that the service was favorable when dealing with long, consecutive texts, and that the naturalness of the speech was superior to their expectations. The overall weakness of the speech synthesis is the lack of interaction; it does not support any type of fill-in forms, such as search boxes and scroll menus. There were, however, several positive reactions to the naturalness of the artificial voice, even though very few pages could handle multilingualism. Although the navigation lacked somewhat in efficiency, the group experienced it as being functional, especially since it allows both letter and numerical inquiries.

The Readspeaker service's functionality from the webmasters' perspective lies foremost in the simplicity of its implementation (see section 4.2). There is no excess work required of the web site administrator in order to provide the pages' visitors with the speech synthesis; nor must the client host an extensive storage facility, since the pre-processing, lexicon database and synthesis are all located on the provider's server. The webmasters seem to agree with the users that the disadvantage of the service is the lack of interactivity; they have little possibility of adapting the service to suit their own needs.

In conclusion, by providing their visitors with the Readspeaker service, the Uppsala University web portal meets the applicable parts of the accessibility guidelines, with which the official policies of the university comply. However, it is crucial that the webmasters take their end of the responsibility of improving and developing the online text-to-speech service.

5.2 Future development

The Readspeaker service has been in development since Phoneticom was established in 2000, and provided at the Uppsala University web portal since 2002. The requests made by the evaluation participants are all reasonable points of elaboration, in order to increase the accessibility of the service. Interaction is a key word, which represents the basis for the Internet, and which needs to be present in Readspeaker. The ability to search pages, fill in forms as well as control the contents to be read aloud are essential elements that both users and administrators insist on.

Furthermore, the users should be able to understand the text being read, which means improving the pronunciation and textual rhythm, as well as expanding the pre-processing of the specific client-based lexicon, including abbreviations and personal names; a task for the providers. It also means an increased responsibility of the online administrators and web authors: marking up the natural language of the specific page, writing with the TTS' behavior and limitations in mind (see section 4.3), as well as specifying the title or headline of all pages. An additional task for the webmasters is to update their respective sites, by linking to the

Readspeaker service. As of today, only 9 per cent of the pages found in the university web catalogue are equipped with such a link.

The evaluation of Readspeaker's functionality, which was presented in this thesis, contributes merely in a small way to the overall assessment of web accessibility. It is crucial that the evaluation proceeds along with the web and service development. The awareness of the people behind the codes, graphics, and tables must be present, through continuous user evaluations during the process of development. The standardized documents introduced in the thesis (see section 2.3), which the university web policy refer to, discuss numerous aspects of accessibility that were not presented here, but with equal relevance to the general progress of web accessibility.

References

Babel Info-Vox, 2003: *BrightSpeech – a human-sounding Text-to-Speech application*, Mons

Bell, Judith, 1995: *Introduktion till forskningsmetodik*, 2 ed, Open university press, UK

Borg, W R., 1981: *Applying educational research: A practical guide for teachers*, New York

Borgegård, Johannes, 2002: *Universitetets webb får talförmåga* (06.06 2003), Uppsala http://info.uu.se/press.nsf/pm/universitetets.webb.id06.html

Brown, A & Haas, H (eds.), 2003: *Web Services Glossary* W3C Working Draft (08.08.2003) (http://www.w3.org/TR/2003/WD-ws-gloss-20030808/)

Chisholm et al (W3C), 1999: *Web Content Accessibility Guidelines 1.0*, Wisconsin http://www.w3.org/TR/WAI-WEBCONTENT/

Cole, Ronald A (ed. in chief), 1995: *Survey of the State of the Art in Human Language Technology*, Cambridge

Council of the European Union, 2000: *eEurope 2002 Action Plan*, Brussels (http://europa.eu.int/information_society/eeurope/2002/index_en.htm)

Dutoit, Thierry, 1997: *An Introduction to Text-To-Speech Synthesis*, Kluwer Academic Publishers, Dordrecht

Ejlertsson, Göran, 1996: *Enkäten i praktiken. En handbok i enkätmetodik*, Studentlitteratur, Lund

Fahlstedt, Ellinor, 2004: *SVT Talande textremsa. En användarutvärdering av två talsynteser i en ny tillämpning*, Royal Institute of Technology (KTH), Stockholm

Holme, I M & Solvang, B K, 1996: *Forskningsmetodik. Om kvalitativa och kvantitativa metoder* Studentlitteratur Lund

Hunt, A J & Black, A W, 1996: *Unit selection in a concatenative speech synthesis system using a large speech database*, Japan

Ishishara, Roberta Ph.D, 2002: *TTS Overview A text-to-speech overview*, TTS Update White Paper, http://www.tmaa.com/tts/TTS%20overview.pdf

Jurafsky, Daniel & Martin, J H, 2000: *Speech and language processing: an introduction to natural language processing, computational linguistics and speech recognition*, New Jersey

Katamba, F, 1989: *An Introduction to Phonology*, Longman Group, Essex

Larsson, Lars-Elwe (Uppsala University's Department of IT infrastructure and telephony), 2003: *IT policy*, Dnr UFV 2002/1475, Uppsala

Larsson, Lars-Elwe (Uppsala University's Department of IT-strategy), 2003: *Utformning av webbinnehåll ur ett tillgänglighetsperspektiv*, Dnr UFV 2003/587, Uppsala

Lundström-Holmberg, Eva & af Trampe, Peter, 1987: *Elementär Fonetik*, Studentlitteratur Lund

Ministry of Industry (The Swedish Government), 1999: *Ett informationssamhälle för alla,* prop. 1999/2000:86, Stockholm

O'Shaughnessy, Douglas, 2000: *Speech Communications, Human and machine* 2nd Ed, New York

Section 508 – US Law for accessibility (15.04.2004) http://www.section508.gov/index.cfm?FuseAction=Content&ID=11

Simosen B et al, 2001: *Synthesised speech with unit selection,* Eindhoven

Skeppstedt, Maria, 2002: *Swedish text-to-speech in Festival Speech Synthesis System,* Stockholm

SpeechWorks White Paper, n.d: *Assessing Text-to-Speech System Quality* (30.03.2004) http://www.tmaa.com/tts/white_papers.htm.

SpeechWorks White Paper, n.d: *Writing Text to Improve Synthetic Speech,* (30.03.2004) http://www.tmaa.com/tts/white_papers.htm

Statistiska centralbyrån, 2004: *Privatpersoners användning av datorer och Internet* Stockholm

Statskontoret, 2002: *24-timmarswebben – rekommendationer och råd för 24-timmarsmyndighetens webbplats* Stockholm

Ubult, Mats (Hjälpmedelsinstitutet), 2004: *IT som talstöd – Dokumentation av ett seminarium* Stockholm

Vetenskapsrådet, 2002: *Forskningsetiska principer inom humanistisk-samhällsvetenskaplig forskning* Stockholm

Vinet, Roman, 2003: *Enhancing rule-based synthesizers using concatenative synthesis*, Royal Institute of Technology (KTH), Stockholm

Evaluation data

1. Anna, 2004; personal interview, April 15, 2004
2. Barbara, 2004; personal interview, April 20, 2004
3. Carin, 2004; personal interview, April 20, 2004
4. Diana, 2004, personal interview, April 20, 2004
5. Erik, 2004; personal interview, April 21, 2004
6. Fredrik, 2004; personal interview, April 23, 2004

Interviews

1. Bergström, Niclas, Phoneticom AB, 2004; personal interview, March 18, 2004
2. Johansson, Åke, Uppsala University Department of Information, 2004; personal interview, March 30, 2004
3. Larsson, Fredrik, Phoneticom AB, 2004: personal interview, March 18, 2004

Appendix A

Enkät: utvärdering av Lyssna-funktionen på Uppsala universitets webbportal
(Survey: evaluation of the Listen-service on the Uppsala University web portal)
2004-04-15

1. Hur ofta använder du Internet? (How often do you use the Internet?)
Flera gånger om dagen (Several times a day) ☐
Varje dag (Once a day) ☐
Flera gånger i veckan (Several times a month) ☐
Några gånger i månaden (A few times a month) ☐
Nästan aldrig (Rarely) ☐

2. Hur ofta surfar du Uppsala universitets sidor?
(How often do you surf the Uppsala University site?)
Flera gånger om dagen (Several times a day) ☐
Varje dag (Once a day) ☐
Flera gånger i veckan (Several times a month) ☐
Några gånger i månaden (A few times a month) ☐
Nästan aldrig (Rarely) ☐

3. Vad gör du i huvudsak på Uppsala universitets sidor? (Flera svar kan anges)
(What do you mainly do on the Uppsala University site? (Multiple answers possible)**)**
Besöker min institutions sida ☐
(Visit my department's page)
Besöker andra institutioner än min egen ☐
(Visit other departments than my own)
Besöker min sida på Studentportalen ☐
(Log on to the Student Portal)
Söker / reserverar litteratur ☐
(Search /make reservations for literature)
Tar reda på senaste nytt på universitetet ☐
(Find out the latest news at the University)
Läser Ergo (Read Ergo) ☐
Annat (Other) ☐

_____
(specificera) (specify)

4. Har du surfat på Internet med Lyssna-funktionen tidigare?
(Have you surfed the Internet with Readspeaker before?)
Ja, ofta (Yes, frequently) ☐
Ja, någon enstaka gång (Yes, occassionally) ☐
Nej (No) ☐

5. Hur upplevde du att navigeringen fungerade med Lyssna?
(How did Readspeaker perform navigation?)
Mycket bra (Very well) ☐

Bra (Fine) ☐
Mindre bra (Not very well) ☐
Dåligt (Poorly) ☐
Kommentar (Comment): _____

_____

_____

6. Vilken röst föredrog du att lyssna på?
(Which synthetic voice did you prefer to listen to?)
Ingemar (manlig) (male) ☐
Emma (kvinnlig) (female) ☐
Amerikansk kvinnlig (American female) ☐

7. Hur upplevde du din favoritröst?
(How did you like your preferred voice?)
Mycket bra (Very goodl) ☐
Bra (Fine) ☐
Mindre bra (Not very good) ☐
Dåligt (Poor) ☐
Kommentar (Comment): _____

_____

_____

Bakgrundsinformation: (Background Information)

Kön: (Gender)
Man (Male) ☐
Kvinna (Female) ☐

Ålder: (Age)
15-25 ☐
26-35 ☐
36-50 ☐
51- ☐

Sysselsättning: (Occupation)
Student, humaniora (Arts) ☐
Student, naturvetenskap (Science) ☐
Student, teknik (Technology) ☐
Student, medicin (Medicine) ☐
Annat (Other) ☐

_____ (specify/specificera)

Appendix B

Enkät: utvärdering av Lyssna-funktionen på Uppsala universitets webbportal
(Survey: evaluation of the Listen-service on the Uppsala University web portal)
2004-04-15

Six replies in total.
Five students, one student administrator.
Four female, two male.
Four in the ages 26 to 35, two in 15 to 25.
Target group representatives: one dyslectic, one non-Swedish speaker.
Two blind users.

Hur ofta använder du Internet? (How often do you use the Internet?)

Flera gånger om dagen (Several times a day)	5
Varje dag (Once a day)	1
Flera gånger i veckan (Several times a month)	-
Några gånger i månaden (A few times a month)	-
Nästan aldrig (Rarely)	-

Hur ofta surfar du Uppsala universitets sidor?
(How often do you surf the Uppsala University site?)

Flera gånger om dagen (Several times a day)	1
Varje dag (Once a day)	1
Flera gånger i veckan (Several times a week)	3
Några gånger i månaden (A few times a month)	1
Nästan aldrig (Rarely)	-

Vad gör du i huvudsak på Uppsala universitets sidor? (Flera svar kan anges)
(What do you mainly do on the Uppsala University site? (Multiple answers possible)**)**

Besöker min institutions sida (Visit my department's page)	5
Besöker andra institutioner än min egen (Visit other departments than my own)	-
Besöker min sida på Studentportalen (Visit the Student Portal)	4
Söker / reserverar litteratur (Search /make reservations for literature)	1
Tar reda på senaste nytt på universitetet (Find out the latest news at the University)	2
Läser Ergo (Read Ergo)	1
Annat (Other)	1

Search people via the electronic telephone list

_____

(specificera) (specify)

Har du surfat på Internet med Lyssna-funktionen tidigare?
(Have you surfed the Internet with Readspeaker before?)

Ja, ofta (Yes, frequently)	-
Ja, någon enstaka gång (Yes, occassionally)	4
Nej (No)	2

Hur upplevde du att navigeringen fungerade med Lyssna?
(How did the Listen-service perform navigation?)

Mycket bra (Very well)	-
Bra (Fine)	3
Mindre bra (Not very well)	3
Dåligt (Poorly)	-

Kommentar (Comment):

- Vill ha snabbval med siffror utan att behöva skriva in på sidan. (Wants shortcuts with numbers without having to type it on the page (in the box))
- Det fungerar bra med texter (särskilt långa texter, men inte speciellt bra med länkar). (It works fine with texts (especially longer texts, but not as well with links))
- Krångligt att hänga med vad som sas, komma ihåg alla olika länkar som rabblades upp för snabbt. (Difficult to keep up with what is said, remembering all the different links that are listed to quickly)
- Bättre än väntat. Frustrerande långsamt. (Better than expected. Frustratingly slow)
- Inskrivningsfält/forms omöjligt att hitta (Input fields/forms impossible to find)
- För långsamt, man har svårt att välja ut just den info man vill ha, avsevärt mindre anpassad för blinda än ordinarie skärmläsare. (Too slow, difficult to choose the particular information you want, considerably less adapted to blind users than ordinary screen readers)

Vilken röst föredrog du att lyssna på?
(Which synthetic voice did you prefer to listen to?)

Ingemar (manlig) (male)	1
Emma (kvinnlig) (female)	3
Amerikansk (American)	2

Hur upplevde du din favoritröst?
(How did you like your preferred voice?)

Mycket bra (Very well)	3
Bra (Fine)	2
Mindre bra (Not very well)	1
Dåligt (Poorly)	-

Kommentar (Comment):

- E: Bra betoning! Låter mycket naturligt! (Good intonation! Sounds very authentic!)
- E: Mer flytande uttal (More fluent pronunciation)

- A: Uttalet var klarare och talet gick inte så fort (The pronunciation was clearer, and the speech was not so fast)
- I: Hakade upp sig och läste samma ord flera gånger men orden blev tydligare (Got caught and read the same word over again, but the words were clearer)
- A: Van talsyntesanvändare, men Emma och amerikanska är de mest människoliknande jag har hört (used to speech syntheses, but Emma and the American are the most human-like I have heard)
- E: Tydlig. (Clear)

Bakgrundsinformation: (Background Information)

Kön: (Gender)

Man (Male)	2
Kvinna (Female)	4

Ålder: (Age)

15-25	2
26-35	4
36-50	-
51-	-

Sysselsättning: (Occupation)

Student, humaniora (Arts)	1
Student, naturvetenskap (Science)	3
Student, teknik (Technology)	
Student, medicin (Medicine)	
Annat (Other)	2

- Student, social science
- Administrator, Uppsala student union

Appendix C

Intervjufrågor (Interview questions):

Allmän utvärdering – Egna kommentarer (General evaluation – Personal comments)
Fanns det några särskilt bra/dåliga inslag? Vilka? (Are there any specifically good/bad elements? Which?)
Vad fattas/är överflödigt? (What did the service lack? What was superfluous?)

Språk: (Language)
Hur väl förstod du rösten? Vad var bra/dåligt i uttalet? (How well did you understand the voice? What in the pronunciation was good/bad?)

Webben: (The Web)
Vad fungerar/fungerar inte? (What works/does not work?)

Navigering: (Navigation)
Vad var bra/mindre bra? (What was good/not so good?)

Behov: (Personal needs)
Föredrar du talad framför skriven text? Om ja, varför? (Do you prefer spoken text to written? If yes, why?)

Framtid: (The Future)
kommer du att använda Lyssna-funktionen framgent? Varför/varför inte?
Om nej, vad fattas för att du ska vilja använda funktionen regelbundet?
(Will you use the Readspeaker-service in the future? Why/why not? If no, what is missing for you to use it frequently?)