Evaluation of a Computer Assisted Language Learning System for Swedish Language Learners

HELENA BERGSTRÖM

KTH Computer Science
and Communication

Master of Science Thesis
Stockholm, Sweden 2007

Evaluation of a Computer Assisted Language Learning System for Swedish Language Learners

H E L E N A   B E R G S T R Ö M

Evaluation of a Computer Assisted Language Learning System for Swedish language learners

Abstract

This master thesis is a part of a project conducted by Centre for Speech Technology and the Language Department at KTH. The focus of the master thesis has been on the implementation and evaluation of Ville, a virtual language teacher. Ville is a program for pronunciation training in the Swedish language for beginners. The program has been used in combination with SWELL, a web course in Swedish for beginners. The subjects of this research project were eight foreign master students and one foreign doctoral student from the department.

When the project started the subjects had the same prerequisites. None of the students had any social relations with Swedish native speakers and they all had a minor knowledge of Swedish. The project started off with a Swedish language test. The test consisted of a number of Swedish words for the subjects to pronounce and record. The test was done in the same system environment as Ville and with the help of flash-cards. There was a qualitative interview done in relation to the test, to be able to determine the subjects' level of knowledge in Swedish. After the test the subjects were sent home to practise with Ville and SWELL during a month. In the end of the month another test was done, identical with the first test. The results from the two tests were evaluated and analyzed by 16 independent "non-experts" in the field.

Utvärdering av ett språkinlärningsverktyg

Sammanfattning

Detta examensarbete har varit en del av ett projekt drivet av Centrum för Talteknologi och Språkinstitutionen på KTH. Examensarbetets fokus har varit på att implementera och utvärdera Ville, en virtuell språklärare, ett program för uttalsträning i det svenska språket för nybörjare i svenska. Åtta stycken utländska master studenter och en utländsk doktorand på institutionen har under en månads tid använt Ville i kombination med en webbaserad kurs i svenska. Kursen i svenska heter SWELL och är utvecklad av Språkinstitutionen på KTH.

Förutsättningarna för försökspersonerna var lika för alla när projektet startade. Ingen av studenterna hade något socialt umgänge med svenskar och deras kunskaper i svenska var mycket begränsade. Projektet initierades med ett test i svenska. Testet gick ut på att de nio försökspersonerna fick uttala och spela in ett antal svenska ord. Testet gjordes i samma programmiljö som Ville och med hjälp av "flashcards". Vid testtillfället gjordes en kvalitativ intervju för att få en tydligare bild av personens kunskaper i svenska. Därefter skickades försökspersonerna hem med uppgiften att träna med Ville och SWELL under en månad. Efter månaden var slut gjordes ett nytt test identiskt med det första testet. Resultaten från dessa två test bedömdes sedan av sexton stycken oberoende "icke-experter" på området.

Acknowledgements

Table of Contents

List of Abbreviations

ASR	Automatic speech recognition
CALL	Computer-assisted language learning
CAPT	Computer-assisted pronunciation teaching
L1	First language
L2	Second language
SLA	Second language acquisition
ECA	Embodied conversational agent
VLT	Virtual language teacher

1 Introduction

One of the most difficult things when learning a new language is the adaptation of a new accent. The degree of difficulty probably depends to some extent on the difference between your mother tongue and the new language that you are learning. Other important parameters are language skills and interest in learning new languages. Also you will need a lot of assistance and attention when you are a beginner. This can be quite a challenge for a teacher with 30-35 students, giving each and everyone of them enough attention and the feedback needed to improve.

Ville is a research project and a program developed at the Centre for Speech Technology at KTH, with the aim to teach pronunciation to people learning the Swedish language. The program teaches pronunciation with two different approaches, perception and production. Ville only teaches pronunciation and glossary so it is necessary to fill in the gaps with the help of SWELL, a web-based distance course in Swedish. SWELL teaches for example grammar and will hopefully give the subjects an understanding of the structure of the language.

The purpose of this thesis is to investigate whether it is possible for a pronunciation tutoring program, a virtual language teacher, to give similar or almost similar support to the student as a teacher does? Is it a good complement to classroom training? In this master thesis I have made an attempt to evaluate Ville and SWELL to see whether the student has improved their Swedish or not by using the two programs.

1.1 Presentation of the subjects of study

In the selection of subjects for this research project we used the archive of the Language Department. The subjects were listed in the archive since they had applied for a Swedish classroom course during the year, but not been accepted. A qualitative interview was done with the subjects when they came to the department for the first time. The interview revealed a lot of useful information about the subject's linguistic background. Even though their language skills varied some, they all appeared to be at the same level in Swedish. None of the subjects had attended a Swedish course. They had all been in Sweden since August 2005 and socialized with very few Swedish people. At KTH all their lectures were in English and only a couple of them listened to Swedish music, radio or watched Swedish television.

The subjects in this research project are briefly and anonymously presented in chapter 4. From now on I will refer to them as S1, S2 etc. This presentation together with the results will hopefully make it easier to follow the personal development of each subject.

1.2 Purpose and method

This master thesis presents the experiences made from the evaluation of Ville, the virtual language teacher in combination with SWELL. The goal of this study is to investigate the efficiency of Ville in particular and CALL- CAPT programs in general.

An experimental study was performed with the nine subjects presented earlier. The study took place during the month of June 2006 and was initiated with a test. The test was composed of 97 different words and was identical to a second test made at the end of the month. The two tests were based upon words from an earlier study made by Anne-Marie Öster (1999) to cover typical difficulties that foreign people might have in the Swedish language.

When the subject came to the department for the first test, they were given a CD with Ville and were instructed on how to use the program. The subjects were advised to use the program as much as possible, but at least 20 min per day. At the same meeting there was a questionnaire made with the subject to determine his/her knowledge in Swedish. They had earlier been introduced to SWELL by Cecilia Melin-Weissenborn from the Language Department.

A post-questionnaire was made with each of the subjects to get an understanding of what their experience was in using the two programs. This was made when the subjects came to the department for the second test. This is presented in (4.2). Additionally we analyzed the outcome of the study with the help of 16 non-experts in the area. For this we used a program called Judge in the Spruce package that I will describe more in detail later. The results are analyzed and presented in general in chapter 4 and individual results are presented in (4.1).

Some data were collected during the study to get extra material for future research. The subjects were asked to send in the sound files they produced when using the production part in Ville. The production component of Ville will be described more in detail in (2.2).

The report includes theoretical background which considers CALL, CAPT in general (2.1) and Ville (2.2) and SWELL (2.3) in particular. Chapter three is a description of the method, followed by a presentation of the results (4) and chapter five sums up with a discussion.

2 Theoretical background

CALL (Computer Assisted Language Learning) is a computer based tool for foreign language learning or for special training in your mother tongue. CALL touches four different areas of research; psychology, computer science, linguistics and pedagogy. CALL comes from the discipline language learning and applied linguistics. CALL has a lot in common with the discipline computer linguistics but has always been separated according to (Borin, 2002) because of a misunderstanding about each others disciplines, different culture and different language learning ideologies. Borin points out that we would benefit from bringing these two areas together. Borin thinks that the development of CALL would benefit from computer linguistics in the following ways:

- By together with a language teacher design good CALL applications.
- Evaluation of different CALL applications.
- By creation and improvements of tools to evaluate and design such applications.
- Through doing a realistic assessment of whether a certain function is possible to achieve and in that case what it would cost to achieve it. Is it relevant to use a CALL system in this case?

CAPT is an abbreviation for Computer Assisted Pronunciation Training, which is a variant of a CALL system. Ville which will be described more in detail later is an example of a CAPT system.

2.1 Evaluation of language learning

One of the main reasons why one would like to evaluate language learning is to see how much can be learnt with a certain teaching method. The evaluation of language learning is often very subjective. One of the important tools, for a teacher to evaluate the student's language skills, is a language test. Three common language tests are the written, oral and listening tests. With the results the teacher can get an understanding about the student's level of language.

2.1.1 Evaluation CALL

Just as evaluating language learning is hard, it is difficult to evaluate CALL systems. Here with the help of (Chapelle, 2001) I will try to structure some principles for evaluating CALL.

Table 1, Principles and Implications

Principle	Implication
Evaluation of CALL is a situation-specific argument.	CALL developers need to be familiar with criteria for evaluation which should be applied relative to a particular context.
CALL should be evaluated through two perspectives: judgemental analysis of software and planned tasks, and empirical analysis of learners performance	Methodologies for both types of analyses are needed.
Criteria for CALL task quality should come from theory and research on instructed SLA (Second Language Acquisition).	CALL evaluators need to keep up with and make links to research on instructed SLA.
Criteria should be applied in view of the purpose of the task.	CALL asks should have a clearly articulated purpose.
Language learning potential should be the central criterion in evaluation of CALL.	Language learning should be one aspect of the purpose of CALL tasks.

Since the evaluation is a complex issue it needs to engage not only the researchers but every one that uses the CALL system such as the learners and teachers. The table below, (Chapelle, 2001) represents different levels of analysis for CALL evaluation.

Table 2, Different levels of analysis for CALL evaluation

Level of analysis	Object of evaluation	Example question	Method of evaluation
1	CALL software	Does the software provide learners the opportunity for interactional modifications to negotiate meaning?	Judgemental
2	Teacher-planned CALL activities	Does the CALL activity designed by the teacher provide learners the opportunity to modify interaction for negotiation of meaning?	Judgemental
3	Learners' performance during CALL activities	Do learners actually interact and negotiate meaning while they are working in a chat room?	Empirical

The first level of analysis refers to the CALL software. This is the usual target for evaluation of CALL. It might be considered the easiest component to evaluate but one must not forget the relevancy of the users. This is considered in level 2 and 3 of the analysis.
In the second level the teacher's performance and involvement is taken into account. The teacher is very much involved in how the CALL is used in class, as well as how it is introduced and the structures around it. As pointed out, "It's not so much the program, as what you do with it" by (Jones, 1986) in (Chapelle, 2001). The third level focuses on the learner's performance. It highlights how the empirical data reflects the way the learner uses CALL.

2.2 CALL, CAPT systems generally

This chapter will describe a few different kinds of CALL systems comparable to Ville. This is to see how other CALL systems of the same type as Ville operate. These systems have been developed both in the regime of University Departments of Speech Technology but also by companies in the same area.

2.2.1 Fluency

Fluency is a CALL system developed at the Language Technologies Institute at Carnegie Mellon University. Fluency uses state-of-the-art speech recognition technology (SPHINX II) also developed at CMU (see figure 1). This interactive software helps users perfect their accents in a foreign language. The system is said to detect pronunciation errors, such as duration mistakes and incorrect phones, and offers visual and aural suggestions as to how to correct them. According to (Sobkowiak, 2003) however, the process of detecting pronunciation errors is a hit-and misses procedure. Even if Fluency evaluates and corrects the pronunciation of single segments, it is difficult for the user to understand the correction criterias. The user can also listen to himself and to a native speaker.
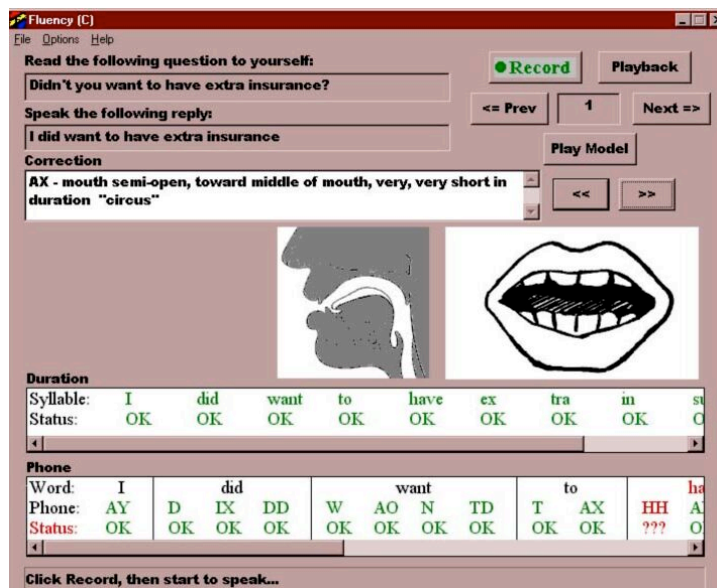
Figure 1, Fluency

2.2.2 Pronunciation Power

The Pronunciation Power was developed by Blackstone Multimedia Corporation and is, according to their web site, used and recommended by over 4000 universities, colleges etc. The Pronunciation Power 2 uses game-like exercises such as listen-record-compare with single words, minimal pairs and full sentences (see figure 2). Other functionality built into Pronunciation Power includes a vocal-tract cross-section, lip animations, and STAIR exercises (stress, timing, articulation, intonation, rhythm).

According to (Sobkowiak, 2003) Pronunciation Power is flexible which makes is excellent for learning under different conditions and settings. The input is relatively little since the system is using Automatic Speech Recognition, although the system doesn't automatically evaluate the users' pronunciation which is criticized by Sobkowiak.
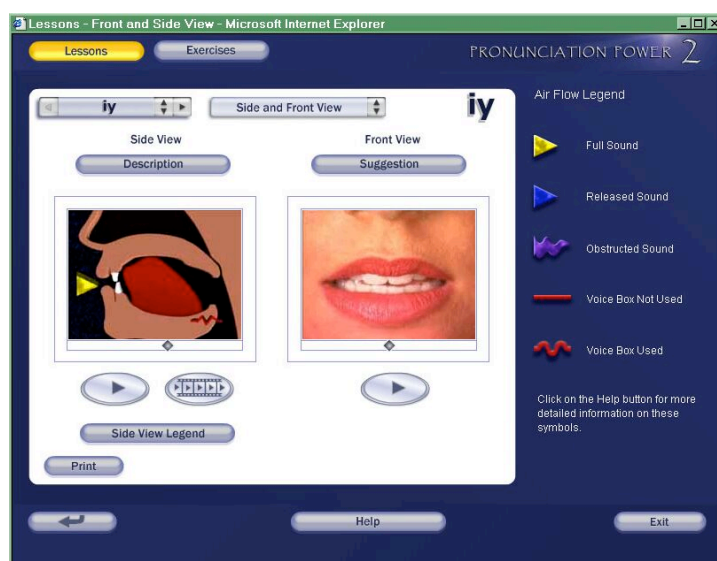
Figure.2, Pronunciation Power

2.2.3 Better Accent Tutor

The Better Accent Tutor was developed by the company Better Accent 1997-2007.
It is a tool for improving you American English pronunciation. One of the three factors that is most important for the comprehensibility according to Better Accent are; intonation, stress and rhythm. The Better Accent Tutor analyses the users' utterance on the basis of these three factors and points out the errors in comprehensive way to the user (see figure 3).
The learners' pitch and intensity graphs are displayed by ASR for inspection and comparison.
The disadvantage of the Better Accent Tutor is according to (Sobkowiak, 2003) that the flexibility of the application is compromised due to the narrow focus and modest content.
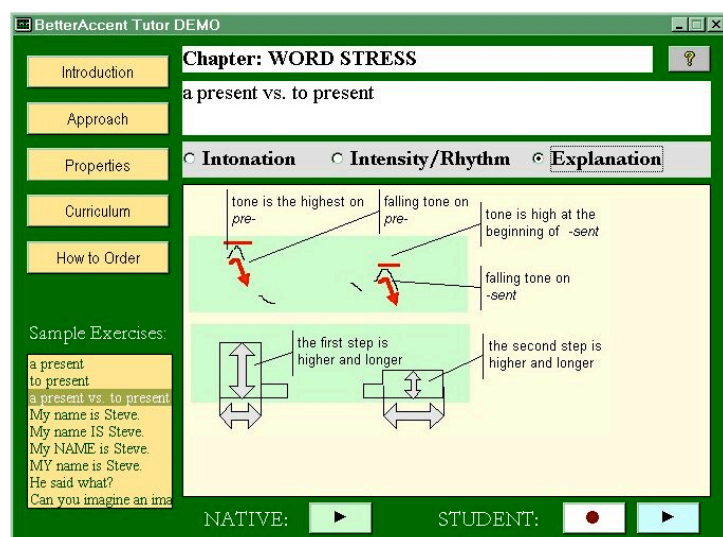
Figure3, Better Accent Tutor

2.3 Ville- The Virtual Language Tutor

2.3.1 Ville- an introduction

Here is an introduction to Ville, the program used within this research project. Every component and step will be described and some of them with pictures to visualize the program.

Ville (Virtual Language Learning Entity) is an embodied conversational agent (ECA) and a virtual language teacher. His purpose is to give feedback to a language learner and encourage her to do better. The user interacts with Ville using a headset. The student records utterances which are analysed by the program. All utterances as well as results of exercises and tests are saved in a folder with username, and the date and time as a name tag. This will be used for future research and may also be used to determine whether the student has improved or not.

There are four parties who have been involved in the research project for new types of applications for CALL, and Ville is a result of this project. The parties are:

- Centre for Speech Technology (CTT), KTH Stockholm
- The Unit for Language and Communication, KTH, Stockholm, Sweden
- The Natural Interactive Systems Laboratory (NISLab), University of Southern Denmark
- Department of Linguistics and Scandinavian Studies (ILN), University of Oslo, Norway

Ville is a project built around speech and language technology research and development. The system provides a user interface with the agent "Ville" and a platform with a universal architecture that can be extended modularly. The components in Ville are described according to (Wik, 2004):

Talking head - The talking head is used for the embodying of the tutor. The talking head has the features lip-sync to synthetic and natural speech and also the signs frowning, nodding and eyebrow movement among more.

Pronunciation Analyser - This is the most important part of the system. This is the part that detects pronunciation errors and gives correction suggestions. A person learning a new language is most likely to make pronunciation errors if a distinction that carries a meaning in Swedish does not carry any meaning in the users' native language.
Pronunciation errors can be divided into phonetic errors and prosodic/rhythmic errors. Examples of phonetic errors are the substitution of phonemes, embracing insertions and omissions. A typical prosodic error is stress at the wrong place or the wrong vowel length. Depending on the target language and the learner's linguistic background the learner makes certain errors and not all phonetic and prosodic/rhythmic errors are covered by the Pronunciation Analyser at present.
The Pronunciation Analyser is detecting the pronunciation errors with the help of forced alignment. Given a text and a sound file, forced alignment analysis can find the time segments of every phoneme and compare it with a reference.

2.3.2 Ville- a presentation of features

Ville has been designed in a way so that hopefully in the future can cover the most common difficulties that the foreign people learning Swedish might have. Detectors are used in Ville to detect pronunciation errors. The first detector to be implemented in Ville was the phoneme duration detector and the process of extending Ville to detect other pronunciation errors is continuous. The four different pronunciation errors that Ville can detect at the moment are as follows: duration, lexical stress, insertion and reduction. The exercises in the production part are made so they will practise these kinds of problems.

There are two types of pronunciation training that you can practise with Ville, perception and production training. Below I will give some examples, visualized with figures, to each of these two areas, all according to (Wik, 2006).

Perception
The perception part contains four different kinds of exercises. These are; Lexical stress, minimal pairs duration, minimal pairs vowels, and Animals (see figure 4).
You can choose whether you would like to do these as an exercise or a test, and you can select the time you want to spend on the exercise (see figure 5). The difference between exercise and test is that you will get feedback during the exercise whether you pressed the right button or not. In the test case you will only get a score after the test is finished which sums up the errors and right answers that you have made.
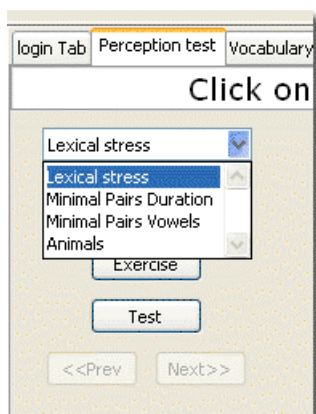
Figure 4, Perception exercises
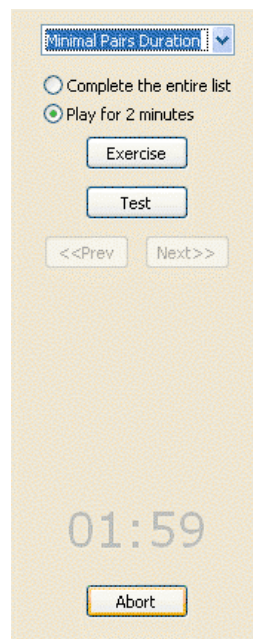
Figure 5, Test or exercise

Lexical stress

In the Swedish language the lexical stress can be on any syllable. That's why this exercise is very good and important for a French person learning Swedish since in French the stress is always on the last syllable. As an explanation to Figure 6 the u means unstressed syllable and the__ means a stressed syllable. The squares are clickable and when you do you will hear Ville say a word that matches the lexical stress pattern (see figure 6).
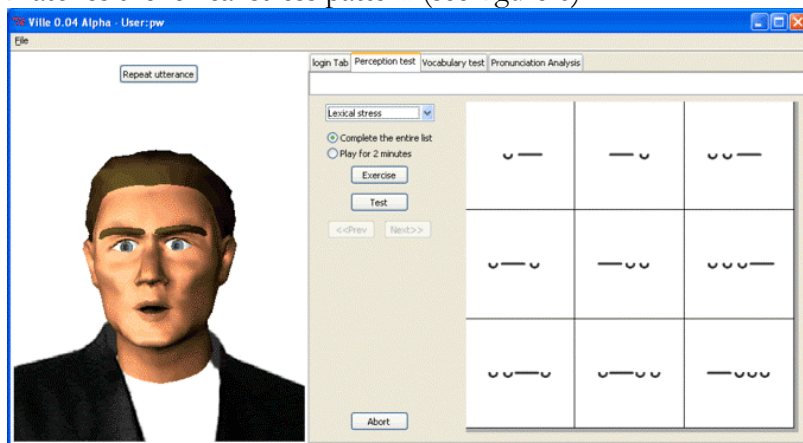
Figure 6, Lexical stress exercise

Minimal pairs

If two words differ by one phone only and have different meanings they are said to be minimal pairs. An example of minimal pair duration is "kina-kinna" (see figure 7), and an example of minimal pair vowel is "lukta-lykta". These minimal pairs can be tricky for some foreign people to differentiate, which is why it is important to practise.
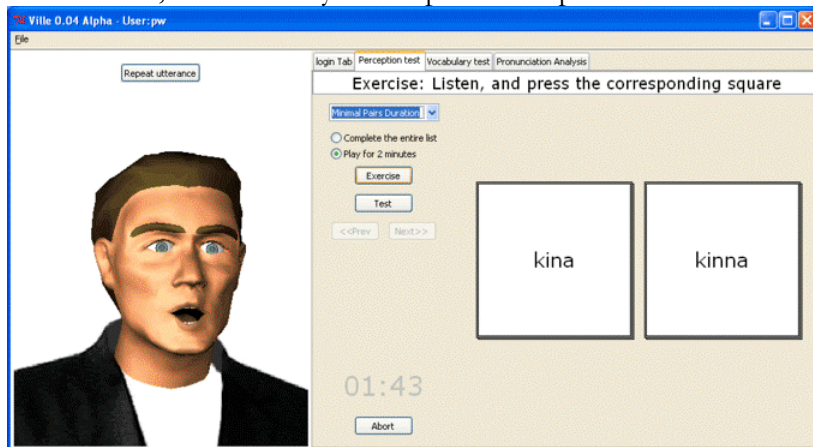
Figure 7, Minimal pairs exercise

Animals

One of the exercises contains flash cards of different animals. This is an example of a good vocabulary exercise to begin with since the animals are easily recognized (see figure 8).
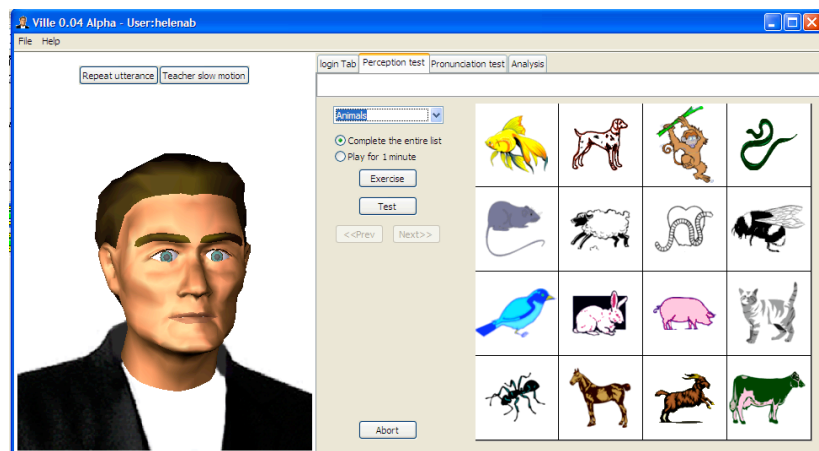
Figure 8, Animal vocabulary exercise

Production

Ville also has a production feature, where your utterances are analysed and you can get help with the pronunciation. The production tab is composed of several exercises. The exercises make use of a deck of flashcards which content can be selected from a list (see figure 9).
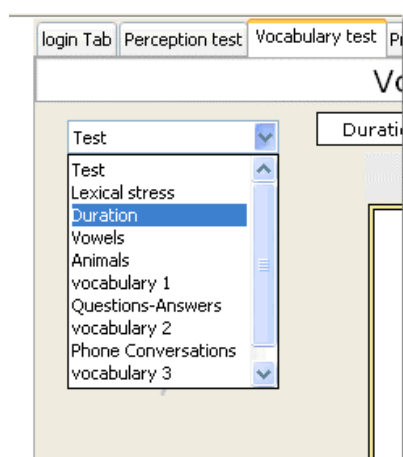
Figure 9, Vocabulary test

The first four exercises are the same as in the production part. The others are sentences and words taken from the first chapters of SWELL. Each exercise in the production tab has a set of ionic buttons to give feedback to the user. Green light means correct pronunciation and red light means incorrect pronunciation (see figure 10).

Figure 10, Feedback on pronunciation

Pronunciation error analyzer

Some of the pronunciation errors such as duration, lexical stress, insertion and reduction faults are detectable by the analyzer. These errors are detected by the detectors mentioned earlier. When you are practising vowel/consonant duration problems the duration detector is activated. The lexical stress detector detects the stressed/unstressed syllables.

The reduction and insertion analyzers take care of the problems that foreign people might have with inserting an extra vowel in the word or reducing the word with one consonant. A good example is the Swedish word Stockholm which will be pronounced "Estocolmo" by a Hispanic person.

Further pronunciation error analysis

When a user is practicing duration and they get a red light, they can click on the red light to get an analysis of the error. The users' utterance is compared in a spectrogram with the teachers' utterance (see figure 11). The spectrogram help to understand where in the utterance the pronunciation error was made.
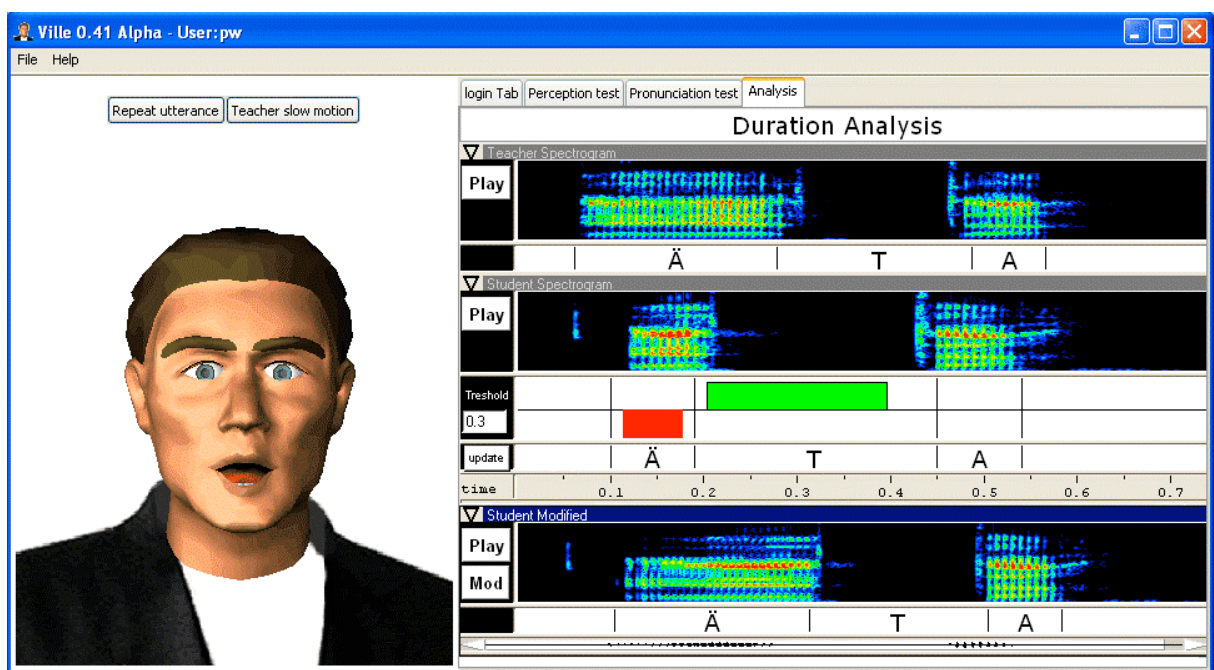
Figure 11, Pronunciation error analyzer

2.4 SWELL

This is a brief introduction to SWELL since Ville is the main target in this master thesis. SWELL is a Swedish distance course designed as an alternative to the classroom course given by the language unit at KTH. SWELL covers Swedish grammar, introduction to new vocabulary, and some listening exercises, but does not have an interactive pronunciation part Ville was therefore considered a good complement to SWELL. SWELL consists of 12 chapters. The subjects were told to go through chapter one in SWELL before the first test, since the chapter includes a brief introduction to Swedish as well as the phonetic alphabet (see figure 12). It was important to give the subjects a hint how Swedish should be pronounced in order to do the first test.

Figure 12, The Swedish alphabet

SWELL teaches grammar and how to build a sentence in Swedish. Also an important part of SWELL is to read and listen to the texts, see the example below (figure 13). You click on the icon below the text and follow the text while listen to it.

Figure 13, Listening exercise

2.5 Standard deviation - Theory

Standard deviation is a common statistical measure, used to measure how widely spread the values in a data set is. If the data points are all close to the mean, then the standard deviation is close to zero. If many data points are far from the mean, then the standard deviation is far from zero. If all the data values are equal, then the standard deviation is zero.

$$s_N = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (x_i - \overline{x})^2}$$

The exact standard deviation definition is:
(Wolfram, Mathworld, 1999)

2.6 Statistical significance - Theory

Statistical significance means that the results are reliable. The t-test is a method often used to decide whether the results are statistically significant or not. A variant of the method, the single-sample t-test, is applied in this study. The single sample t-test is intended to use the data from a single sample to test hypotheses about a single population. In this study we only have a single-sample since we don't have a reference group and no results to compare with the subjects results. This implies the use of a null-hypothesis, which means that the subjects made no improvement. If the result from the t-test < 5 % the null-hypothesis is rejected (there is a statistical significance), if the t-test result > 5 % the null-hypothesis is kept (there is no statistical significance).

Single-sample T-test: $t = \dfrac{\hat{x} - \mu}{S_{\overline{x}}}$

$\mu = 500$, which means that there is no difference between the first and second test.
$\hat{x}$: The sample mean. One sample is the analysis of a subject made by one judge.

Standard error: $S_{\overline{x}} = \sqrt{\dfrac{S^2}{n}}$

S^2 : Standard deviation
n : Number of samples, in this case 16, the number of judges.

2.7 Cronbach alpha – Theory

The Cronbach alpha is not a statistical method it is a coefficient of reliability. Cronbach alpha describes how well a set of items measures a single unidimensional latent construct. In this case how well the judge's judgements agree.

$$\alpha = \frac{N \cdot \overline{r}}{(1 + (N - 1) \cdot \overline{r})}$$

N is equal to the number of items and r-bar is the average inter-item correlation among the items. So when increasing the number of items you will also increase the Cronbach alpha. The cut off value of   for being acceptable is 0.70. Values > 0.70 indicate that the set of items agrees. (Journal of Extension, 1999)

3 Historical background

3.1 CALL, CAPT and foreign language teaching in the 50's to 70's

In the 50's and early 60's empiricist theory was predominant in language teaching, pedagogically audiolingualism as it was described by Stern in (Levy, 1997). The approach emphasized the use of the target language in a spoken form and students were expected to learn the language through practice. The teacher presented new vocabulary and structures through dialogues which students learned through imitation and repetition. Levy tells about another important influence during the 60's called programmed instruction. One of the leading researchers in this area was B. F Skinner. He suggested that the use of teaching machines could improve the more individualized language training for the student. With the help of instructional steps or "frames" the student would get instant feedback on his/her responses.
The systematic and routine character of the exercises presented in the audio-lingual approach was picked up by the software developers of the time, who realized that these exercises were actually programmable on the computer.

The PLATO (Programmed Logic for Automatic Teaching Operations) project is considered to have initiated CALL. The project initialized in 1960 and was a sort of email system which enabled the users to "talk" to each other by notes files. The notes files were of two kinds, one for communicating teacher to student and vice versa and one for general announcements to all users. The student records were saved to ease the student's improvement, for the teacher's information and further research. The PLATO project didn't cover all the language learners' needs (no speech production and understanding for example) but it was a good tool for practising vocabulary and grammar.

3.2 CALL, CAPT and foreign language teaching in the late 70's to 80's

In the late 70's the focus in language training came to be more on the complexity of language teaching and learning and the needs of the individual learner. Notable was the humanistic methods which engaged the whole person, their emotions and feelings, all according to Levy.

In the 80's the interest in CALL grew dramatically thanks to the microcomputer. It was invented already in 1973 but introduced to society in the 80's. The Storyboard program was one of the big software for microcomputer introduced at this time. Storyboard, written by John Higgins is a text-reconstruction program where the aim is to reconstruct a text word by word using textual clues such as title, introductory material etc. Storyboard is an authoring program which enables the teacher or the students to easier write their own texts and these are then saved in the program for future use. According to Levy Storyboard is a good example of how CALL software evolve as the concept and technology develops.

3.3 CALL, CAPT and foreign language teaching in the 90's

Few would argue that one of the most dramatic and important technology developments in the 1990's was the Internet. This meant a greater access to material, people and learning

environments. The CALL project described below is a good example of how the Internet was used in the early 90's.

The International Email Tandem Network begun in 1993 and was initiated by Helmut Brammers. The project is described by (Levy, 1997) as language learning by computer mediated communication using the internet. This project enabled university students from around the world to be linked together and learn languages in tandem. The Tandem Network consisted of subnets which had their own bilingual forum where students could discuss with each other. The Tandem Network also included a database which students could add new information to, or access to get learning materials from.

In this way the language learning goes beyond the capacity and offerings of a language department and lacks in a way the need of language teachers. The interactions between the students are based on how they want to learn and teach a language and are more individualized. In this project the email work accompanied a formal language course but there is no reason why it couldn't be a stand alone tool for language learning. This gives the user more options to access language learning in a more non-traditional way.

4 Method

The major part of the participants in this study was registered in the archive of the language unit, since they had applied for a Swedish classroom course earlier during the year but not been accepted. Only the exchange students at KTH are guaranteed a seat on the classroom course. The remaining subjects were friends (also master students) of these master students and finally there was one PhD student from the department of Speech Technology. The subjects were equally represented over gender and age, four women and five men from the age of 23 to 42. The variation of mother tongue and origin was also satisfying. The participants' specialization of study varied, from civil engineering to computer science. Very few of the subjects had any former experience with using CALL. Only a couple of them had used a CALL program for English learning. None of them had taken a Swedish course before and their knowledge in Swedish was minor. As an appreciation for the participation every student was guaranteed an advantage in the selection of students for the autumn Swedish classroom course.

4.1 Subject instructions

When the subjects came to the department for the first time, we had an introduction meeting with them and they were shown how to use SWELL and Ville. The subjects were told to finish chapter one before they came back to the department for the pre test. During the period of the research project the subjects were instructed to finish chapter two to four before the post test. This was only the minimum effort, but we thought that if we pushed them to hard they wouldn't sign up for the project.

4.2 Pre-Post test

The first test was carried out at the end of May and was initiated with home studies in the Swedish distance course. One month passed by between the pre and post tests. The tests were constructed of 97 different words. These words had been selected from the vocabulary of a test that was created by Cecilia Melin-Weissenborn and Anne-Marie Öster. These words were supposed to represent the different difficulties that foreign people experience when they are learning Swedish. The subjects got help with the adjustment of the microphone, and were also instructed on the speech level they should use to do a clear recording.

4.3 Pre-Post questionnaire

The pre questionnaire (appendix A) was done in order to collect information about the linguistic background of the subjects as well as their knowledge of Swedish. The questionnaire also gave an indication on the probability for improvement owing to Swedish friends, watching Swedish television and listening to Swedish music.
After finishing up the research project with a second test a post questionnaire was done. This was done to collect the thoughts and suggestions of the nine subjects for improvement of the system. The questionnaire was both a qualitative (4.2) and a quantitative (appendix B) interview. The questions were taken from (Öster, 2006) and were questions used for evaluating a hearing improvement program for speech therapists. Since the questions were general, a few extra questions were added to make the questionnaire more suited for this study.

4.4 Equipment

4.4.1 Test environment

The pre-post tests were designed in the same program environment as Ville with the talking head and flashcards (figure 14). Every word was recorded by pushing the space button, which could be done as many times as the user needed to make a satisfying recording. The subjects could then move between the words with the next and previous buttons.
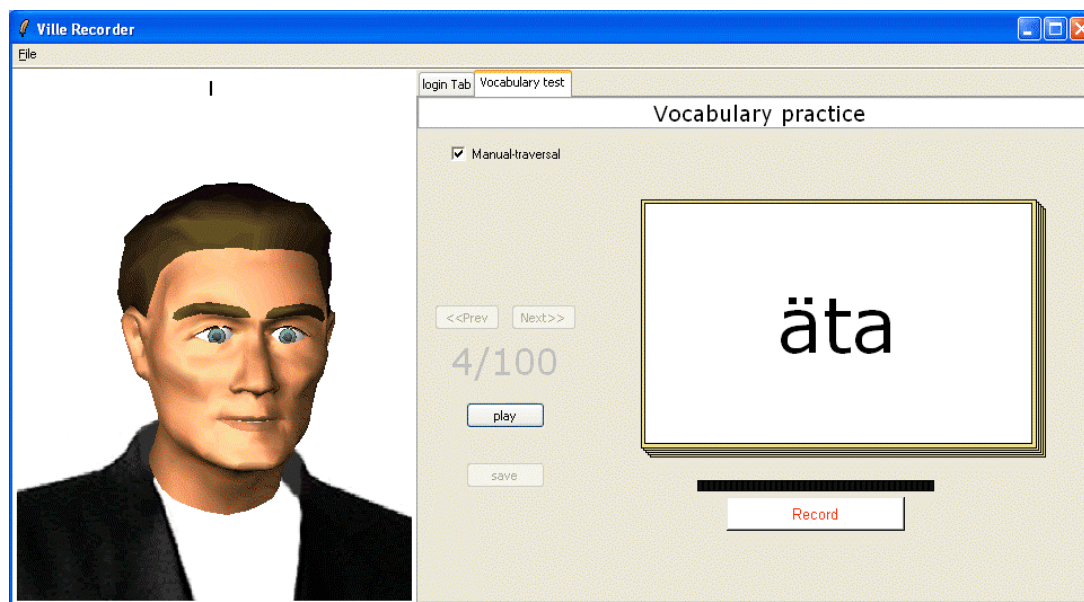
Figure 14, Pre- post test design

4.4.2 Hardware and environment

The hardware used was an IBM ThinkPad T42 and a headset with a microphone attached to it (Creative HS-300). The test was carried out in the same room every time. It was a small non sound isolated office at the Department of Speech Technology.

4.5 Data evaluation

4.5.1 How the data was evaluated

To be able to judge the results of the two tests we engaged a number of "objective" judges. They were all non-experts in the area of CALL, CAPT and had no former knowledge of language learning evaluation. The judges that participated were mostly friends and family to me. One must take into consideration that maybe the judges answers and judgements could somewhat be colored by the fact that they know me and wanted the results of the study to be successful. To overcome this effect the two tests were mixed up in "Judge" which is the program used for the evaluation. A code was added to each word-pair, this to ease the recognition for me. Judge is described more in detail below. The judges also used an IBM ThinkPad and the same type of headset that the subjects used.

4.5.2 What data was evaluated

Not all 97 words were presented to the judges. They were given a set of 10 word pairs from all nine subjects, so 90 word pairs in total. Each word pair consisted of one utterance from each test. Ex. Word pair = utterance test1 + utterance test2.

My though was to not give them too much material since they might get tired at the end of the exercise and not put enough effort into the judgements.

Table 3, The words presented to the judges

bil
bok
etta
herr
här
sju
syr
väg
vägg
äta

4.6 Spruce

In order to compare the two sound files some sort of program or method was needed. The choice fell on an already existing tool, Spruce (Granqvist, 1996-98). Spruce contains of three different tools: Glue, Judge and Visor. The first two components are described in more detail below.

4.6.1 Glue

Spruce contains a program called Glue. Glue is used to "glue" two files together. In this case it is sound files that are "glued" together as a pair and saved as one file. Then in the next step you can use Glue again to glue together all these pairs of sound files into one big file. This file contains all the pairs and can be used either in Judge. Judge makes it possible to compare these pairs of sound files connected to each of the two language tests. In this way, for example, you can pick out the word "herr" from test one and compare it with the same word in test two.

4.6.2 Judge

Below is a session in Judge (see figure 15). When you have started a session the word you hear is displayed below the buttons. This is visualized in the white field.

Figure 15, Judge

You can also choose parameters on the scroll bar (see figure 16). In this research project I used the parameter "Första bäst" (The first one is the best) for the left side of the scroll bar, "Lika bra" (Equal) for the middle, and "Sista bäst" (The last one is the best). These can be adjusted in the Judge settings where you can also for example choose to show the .GLU labels which are the name of the word pair file that was glued together in Glue. An example of a .GLU label is herr1x.
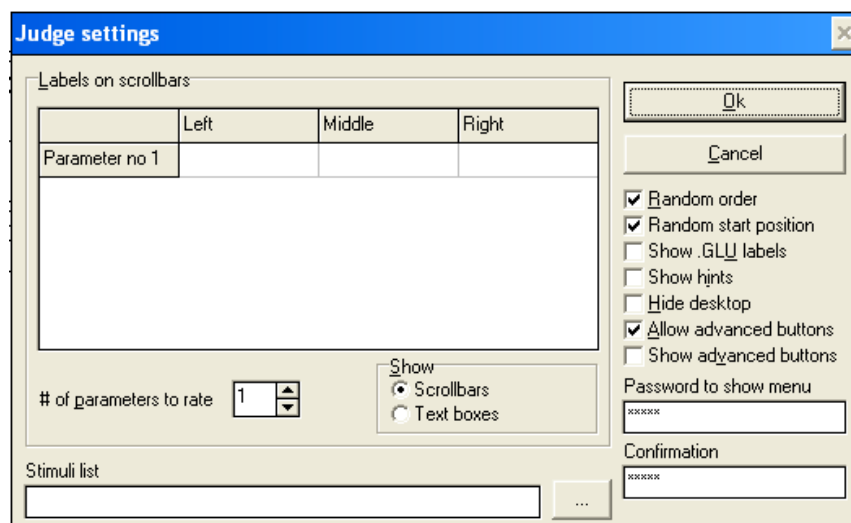
Figure 16, Judge settings

The words are shown with the file name which consists of three parts: the word, a number from one to nine (defining the subject it belongs to) and a cryptic x or y at the end. The x and y stands for the order of the sound files. An x is displayed if the sound file from the first test comes before the sound file from the last test. A y is displayed if the sound file from the second test comes before the one from the first test. In the earlier example "herr", which was two sound files glued together, was shown in Judge as herr1x. This represents, subject one and that the sound file from the first test came first.

5 Results

Below is a short presentation of the nine subjects in this study.

Tables 4- 12, Short presentation of the subjects

S1	
Gender	Female
Age	30
Country	Spain
Mother tongue	Spanish
Second language	English

S2	
Gender	Female
Age	29
Country	Pakistan
Mother tongue	Urdu
Second language	English, Punjabi

S3	
Gender	Male
Age	35
Country	Peru
Mother tongue	Spanish
Second language	English

S4	
Gender	Male
Age	42
Country	Kenya
Mother tongue	English
Second language	-

S5	
Gender	Male
Age	27
Country	Greece
Mother tongue	Greek
Second language	English

S6	
Gender	Male
Age	24
Country	India
Mother tongue	Tamil
Second language	English
Other languages	French, Hindi

S7	
Gender	Female
Age	27
Country	Russia
Mother tongue	Russian
Second language	English
Other languages	German

S8	
Gender	Male
Age	25
Country	Pakistan
Mother tongue	Urdu
Second language	English
Other languages	Punjabi

S9	
Gender	Female
Age	23
Country	China
Mother tongue	Chinese
Second language	English

5.1 Individual Results in numbers

Table 13, Individual results

	Time spent (per week)		Result		
Subject	Ville	SWELL	Average	Median	Standard deviation
S1	-	>60 min	501	581	342
S2	-	>60 min	508	506	291
S3	0-30 min	0-30 min	488	498	192
S4	-	30-60 min	574	557	260
S5	>60 min	>60 min	540	494	312
S6	30-60 min	>60 min	597	639	317
S7	30-60 min	30-60 min	526	495	291
S8	>60 min	>60 min	562	507	259
S9	30-60 min	0-30 min	544	583	257

Table explanations

How much time spent (per week), with the two programs, is according to the subjects themselves. The data was taken from the post questionnaire.

The results are presented as if the pre-test was presented to the judges as the first stimuli. If the two tests were switched so the second test came first the result were presented as 1000-y. Only because the scale was from 0 to 1000 and to avoid a negative result (<0) the solution was to take the largest value minus the judges value.

Theory for standard deviation see chapter 2.5.

5.2 Individual Results, graphs and subject facts

The utterances from subject one to nine has been judged and compared by the 16 judges, presented in more detail below. The outcome of these judgments has been put into the following graphs. These graphs are meant to serve as a complement to the overall picture of the students´ improvement and will show the differences in the judges' opinion.

The graph displays one student at a time and the average judgment of that student from each and every one of the sixteen judges. As a reference the average number of languages spoken among the subjects was 2.44.

5.2.1 Subject 1
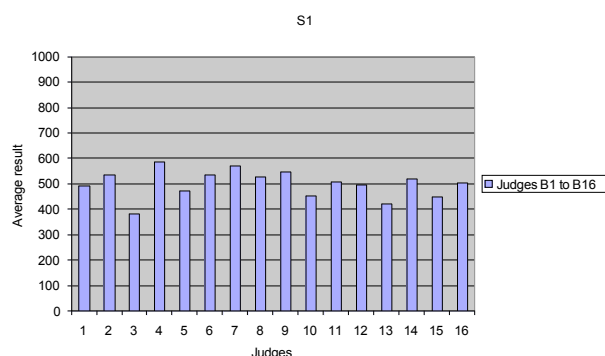
Diagram 1, Subject 1

Why do you want to learn Swedish, what are your goals?
I think it's useful since I live and make my PhD here.
On a scale from 1-5 (5 = very important, 1 = not important), how important is it for you to achieve a good Swedish accent? 5
Time spent on Ville/How much time is required to make progress? (According to the subject.) I didn't use the program.
Time spent on SWELL/How much time is required to make improvement?
(According to the subject.) > 60 min per week/ At least 60 min per day.
The final result of subject 1 was 501, which is an average result compared to the other subjects.

5.2.2 Subject 2
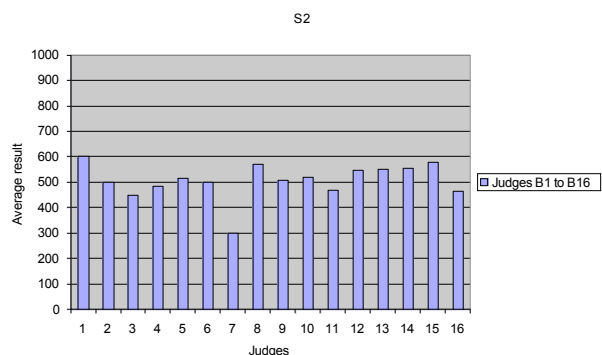
Diagram 2, Subject 2

Why do you want to learn Swedish, what are your goals?
I like learning new languages, it helps me to know the country and communicate with the people.
On a scale from 1-5 (5 = very important, 1 = not important), how important is it for you to achieve a good Swedish accent? 4
Time spent on Ville/How much time is required to make progress? (According to the subject.) I didn't use the program enough to be able to judge.
Time spent on SWELL/How much time is required to make improvement?
(According to the subject.) > 60 min per week/ About 2 hrs per day.
The final result of subject 2 was 508, which is an average result compared to the other subjects.

5.2.3 Subject 3
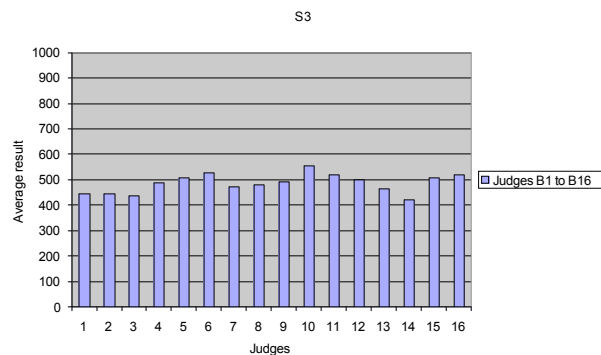
Diagram 3, Subject 3

Why do you want to learn Swedish, what are your goals?
For practical use and to communicate and to be able to communicate.
On a scale from 1-5 (5 = very important, 1 = not important), how important is it for you to achieve a good Swedish accent? 3
Time spent on Ville/How much time is required to make progress? (According to the subject.) 0-30 min per week/ 60 min per week
Time spent on SWELL/How much time is required to make improvement? (According to the subject.) 0-30 min per week/ > 120 min per week
The final result of subject 3 was 488, which is a result slightly under average compared to the other subjects.

5.2.4 Subject 4
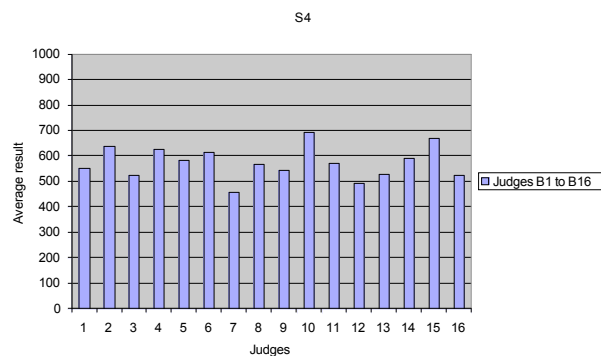
Diagram 4, Subject 4

Why do you want to learn Swedish, what are your goals?
Since I live in the country I want to learn the language. It sounds romantic.
On a scale from 1-5 (5 = very important, 1 = not important), how important is it for you to achieve a good Swedish accent? 5
Time spent on Ville/How much time is required to make progress? (According to the subject.) I didn't use the program.
Time spent on SWELL/How much time is required to make improvement? (According to the subject.) 30-60 min per week/60 min per week
The final result of subject 4 was 574, which is a result slightly above average compared to the other subjects.

5.2.5 Subject 5
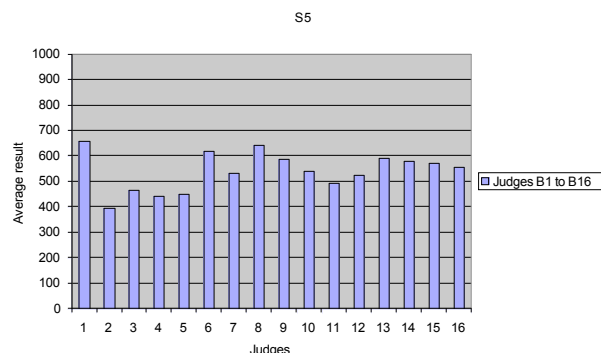
Diagram 5, Subject 5

Why do you want to learn Swedish, what are your goals?
To get a better understanding of the culture and it's required for my PhD studies.
On a scale from 1-5 (5 = very important, 1 = not important), how important is it for you to achieve a good Swedish accent? 4
Time spent on Ville/How much time is required to make progress? (According to the subject.) > 60 min per week/ 30-45 min per day
Time spent on SWELL/How much time is required to make improvement? (According to the subject.) > 60 min per week/ > 2 hrs per day
The final result of subject 5 was 540, which is a result slightly above average compared to the other subjects.

5.2.6 Subject 6
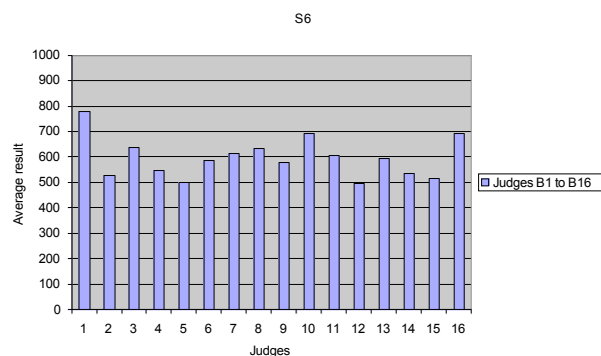
Diagram 6, Subject 6

Why do you want to learn Swedish, what are your goals?
To be able to interact with people.
On a scale from 1-5 (5 = very important, 1 = not important), how important is it for you to achieve a good Swedish accent? 4
Time spent on Ville/How much time is required to make progress? (According to the subject.) 30-60 min per week/ > 60 min per week
Time spent on SWELL/How much time is required to make improvement? (According to the subject.) > 60 min per week/ > 60 min per week
The final result of subject 6 was 597, which is a result slightly above average compared to the other subjects.

5.2.7 Subject 7

S7
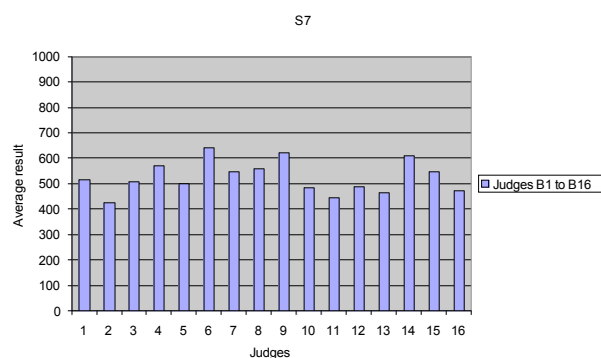
Diagram 7, Subject 7

Why do you want to learn Swedish, what are your goals?
I like languages and it's an interesting experience to learn a new language. It was my favorite subject in school.
On a scale from 1-5 (5 = very important, 1 = not important), how important is it for you to achieve a good Swedish accent? 3
Time spent on Ville/How much time is required to make progress? (According to the subject.) 30-60 min per week/ 15 min per day
Time spent on SWELL/How much time is required to make improvement? (According to the subject.) 30-60 min per week/ 15-30 min day
The final result of subject 7 was 526, which is a result slightly above average compared to the other subjects.

5.2.8 Subject 8

S8
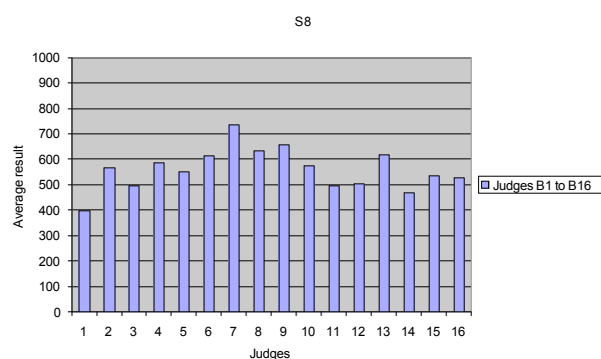
Diagram 8, Subject 8

Why do you want to learn Swedish, what are your goals?
I like to learn new languages. I think it's important to know Swedish since I'm working in the country.
On a scale from 1-5 (5 = very important, 1 = not important), how important is it for you to achieve a good Swedish accent? 3-4
Time spent on Ville/How much time is required to make progress? (According to the subject.) > 60 min per week/ > 60 min per week
Time spent on SWELL/How much time is required to make improvement? (According to the subject.) > 60 min per week/ 1-2 hrs per day
The final result of subject 8 was 562, which is a result above average compared to the other subjects.

5.2.9 Subject 9
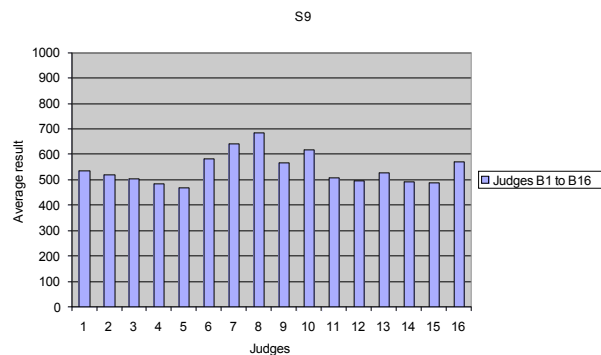
Diagram 9, Subject 9

Why do you want to learn Swedish, what are your goals?
I want to be able to read the newspaper, to get the information and to communicate.
On a scale from 1-5 (5 = very important, 1 = not important), how important is it for you to achieve a good Swedish accent? 3
Time spent on Ville/How much time is required to make progress? (According to the subject.) 30-60 min per week/ > 60 min per week
Time spent on SWELL/How much time is required to make improvement? (According to the subject.) 0-30 min per week/ 30-60 min per week
The final result of subject 9 was 544, which is a result slightly above average compared to the other subjects.

5.3 Statistics and measures

5.3.1 Statistical Significance – Result

A t-test was done on the results from each subject and also for the group of subjects.

Table 14, T-test result for each subject

Test value = 500, 95% Confidence Interval of the Difference			
Judgement Subject n	**T-test value**	**Degrees of Freedom**	**Significance**
1	0.049	159	0.961 Not Significant
2	0.366	159	0.715 Not Significant
3	-0.817	159	0.415 Not Significant
4	3.622	159	0.000 Significant
5	1.635	159	0.104 Not Significant
6	3.862	159	0.000 Significant
7	1.138	159	0.257 Not Significant
8	3.000	159	0.003 Significant
9	2.165	159	0.032 Significant

Table 15, T-test result for the group of subjects

Test value = 500, 95% Confidence Interval of the Difference				
Judgement	**T-test value**	**Degrees of freedom**	**Significance**	**Mean difference**
1-9	5.048	1440	0.000 Significant	537,85- 500 = **37,85**

This implies, according to the theory of statistical significance, that the results are statistically significant. Thanks to the greater number of degrees of freedom the result of the t-test for the group of subjects is significant.

The mean difference describes how much the judgement differs from 500, which was very little in this case.
For theory of t-test and statistical significance see chapter 2.6.

5.3.3 Cronbach alpha – Result

The result of the computation of Cronbach alpha
    = 0.9391

This result implies that the judges' opinion agrees, since values > 0.70 indicate that the set of items agrees.
For the theory of Cronbach alpha se chapter 2.7.

The standard deviation, t-test values and Cronbach alpha value are taken from a computation in SPSS made upon my measured values by Mattias Heldner at the Department of Speech, Music and Hearing, Stockholm.

5.4 The Judges- a presentation

The judges were selected from my family and friends and were all non-experts in the research area. They were of varying sex and age, from 23 to 58, and had Swedish as their mother tongue. The overall attitude was positive and everyone seemed to make an effort with the assignment. They were all handed the same information before starting the "judgement" in Judge. The only instructions they were given was the interpretation of the parameters on the scrollbar. "Första bäst" meant that the first word sounded more Swedish (e.g. had a more Swedish accent and was more understandable) then the second word. "Andra bäst" meant the other way around, that the second word sounded more Swedish. They were also told that they could play every two-word pair as many times as they wanted to, to make sure that they were satisfied with their judgment. The judgments were stored as results in a text-file and these were valued as described before in the table explanations.

5.5 Post-Questionnaire

The answers to the questionnaire

The following questions were posed to the subjects when they came to the department for the post test after finishing the test period. These answers will hopefully serve as a channel for the subjects in the survey to speak their mind and give suggestions for improvements, complain about things they disliked and give some input on what to keep. It was also useful for us in this project to get more information about how much the users had worked with the two programs, how big their motivation were and how useful they considered the programs to be. This facilitated the work on evaluating Ville and SWELL.

Here are the answers to the qualitative interview. It is not a summary but a representation of the subjects' answers. The quantitative part of this interview is placed in the appendix.

Ville

Do you think that any previous knowledge about computers is necessary to work with the system?
Comments: "Of course the basic level to be able to start a CD for example. "The program itself is not complicated. With the assumption that it would come with a manual. An inexperienced user won't know how to open a profile."

How did the system meet with your expectations?
Comments: "I had some unexpected problems with it. The recording was a problem for me. The warning message you get should be in English, otherwise it's hard to understand."
"I liked that the pronunciation was very clear." "I didn't have any expectations. It met my expectations, I'm quite satisfied."

Was the system easy to work with?
Comments: "The system was easy, but the problems that arose were not so easy to solve."
"I didn't have any difficulties figuring out the exercises. But there were some bugs." "As a study program it's very easy to work with."

Did you consider the training meaningful?
Comments: "Yes but I didn't have the time to work with it enough. With a lot of other stuff to do I didn't have the time to focus on the training with the programs." "For a first approach with the language it's good. It could be a supplement for a normal course but it could not substitute a teacher. Good to work at home with accent spelling etc."

What is your opinion on the talking head?
Comments: "I don't take so much notice of the movements of the face. I listened more to the sounds. It would be better if you could see inside the mouth." "It's very useful to see the mouth expression. The particular model needs improvements. Movements of the mouth could be more natural, not so detailed. It gives a feeling of someone helping you."

How was the system from a pedagogical point of view?
Comments: "In the way of pronunciation and listening it's pedagogical. But the lacking of meaning of the words is less pedagogical." "As a substitute it was pedagogical. No sense of competition for the user. No qualitative feed-back. Not the same feed-back as a teacher. "Modify the interface placing the cards under the face." "Pedagogical but doesn't force you to study. But I lived in Sweden for 10 months and I didn't study Swedish. With this program I got the possibility to train Swedish, and the accent. I would never have studied Swedish if it wasnt for the program."

Was the system reliable?
Comments: **"**No**,** because it wasnt working as it should." "I didn't have any problems with it." "If the bugs were fixed it would be very reliable. Some of the exercises had the wrong answers; I have checked it with my friend who is studying to become a Swedish language teacher."

Did you want more instructions about the program before starting with it?
Comments: **"**Some more instructions would have been good, for example in English on the CD or on a paper." "At least a two hour guide practice would be very good." "Didn't know which exercises which did or didn't work. I figured it out by myself." "I didnt have any problems with the program. The instructions were enough for me, and the program was easy to install."

Did you miss the possibility to train something? If yes, which one?
Comments: "I don't have so much academically knowledge so it's hard to say. I would like to be able to interact, conversate with Ville. It could be a girl instead of a guy because it is easier to relate to the female voice teacher, like a mother, unconditional love."

How much do you think one should train with the system per week on average to make improvements?
Comments: **"**60 min per week. It is important to be self motivated." "30-45 min per day." "More than 60 min per week." "15 min per week is a minimum. The more the better."

Do you think you would have improved more if you had worked more with the system?
Comments: "Yes up to a certain threshold, point. But then you would need a teacher. With a bigger verity it would be better. It could be a categorized level of difficulty. Manual for the user would also be good, how to do the exercises." "Some, because I just used the system. I didnt speak with native speakers so much."

<u>SWELL</u>
Do you think that any previous knowledge about computers is necessary to work with the system?
Comments: "I think its need for some knowledge of computers to enter the program. If you not, you can use 20 minutes or so to get into the program." "Familiar with Internet, know how to save files." "The presentation and everything was easy to understand. Everyone can work with it."

How did the system meet with your expectations?
Comments: **"**Of course it can be improved more. Some connections are missing between the basic English knowledge and the basic Swedish knowledge. The tempus were not all very detailed, it could have been more instructions "It was helpful, but lacked the possibility to get feed-back. I have a friend who is studying to become a teacher in the Swedish language and after chapter three she helped me. It then got really hard. The grammar and theory is progressing fast. The theory was very hard and the exercises were too easy, but it gives you confidence."

Was the system easy to work with?
Comments:." "Everything was very well explained how to use it." "The scroll with the chapters was good." "There is lots of room for improvement. Relocate the links for the sound files. Add an icon to "lyssna". Sound file should be before the text." "I used a similar pronunciation program in English, it was very useful."

Did you consider the training meaningful?
Comments: **"**Yes, because I was totally ignorant of the language in the beginning." **"**Yes but I didnt have the time to work with it enough. With a lot of other stuff to do I didnt have the time to focus on the training with the programs." "It gives some material to start with. I will have an advance when starting the class room course." "It's more meaningful if you have a teacher, but this is also meaningful."

How was the system from a pedagogical point of view?
Comments: "It could be improved; more examples could be added as well as exercises and vocabulary." "It goes fast for a beginner. I would have wanted a teacher, some physical contact. To be able to ask questions. Good for highly educated, but not so good for less educated people."

Was the system reliable?
Comments: "One of the five listening parts was not working as it should." "I didn't have any problem." "The system is too fast."

Did you want more instructions about the program before starting with it?
Comments: "I was showed how to access it and I didnt need any extra help." "It would be useful with an introduction in English in one color, and the Swedish in another color." "Its easy, the simplest program."

Did you miss the possibility to train something? If yes, which one?
Comments: "I would have wanted more pronunciation training. They who read the texts are too fast." "Material helps you to build up a theoretical background. In a combination with Ville it's good, since they have different purposes."

How much do you think one should train with the system per week on average to make improvements?
Comments: "At least one hour per day. If you really want to learn a language you need to spend a lot of time learning it." "More than 120 minutes per week. Swell took more time than Ville." "At least two hours every day."

Do you think you would have improved more if you had worked more with the system?
Comments: "The more you work the more you learn." "It's a good system for self studying." "You can improve up to a certain point, same as with Ville. If you would have gone through the twelve chapters you would have improved more."

6 Discussion

Some individual improvements in the Swedish language have been done among the subjects. This improvement could be a result of the subjects training with Ville and SWELL but could also be explained by the subjects' new interest of learning the new language. Since the individual plays a big role in language learning it is hard to find the truth about what has most impact on the learning. It is also tricky to decide on how, how much and what to teach.

One month is too short for learning a new language. However this kind of experiment could give a good indication on the efficiency of the programs used. Unfortunately we were not able to test Ville in combination with, for example, ordinary classroom teaching or other equivalent pronunciation training programs. Neither did we have a reference group to compare with. This might have helped to analyze the results and to give a more correct picture of the effectiveness of Ville. Furthermore we had too few subjects, with individual differences, to be able to judge upon the outcome of this project.

6.1 Lessons learnt

More material to base the results upon would have been preferable. We didn't get hold of the files recorded during the study. Every time the student performed one of the exercises in Ville a file was recorded and saved at the hard drive of their computer. The students were told to send these files in to us for evaluation, but very few of them did so.

This is another important lesson to learn. One should not give the subjects too much freedom and give them clear and direct instructions to follow during the research study. If not you will risk to not getting as much information as you need.

From the beginning the students were more or less at the same level and the effort and time put in was very varying. The students were more or less at the same level even after the project was finished, that was somewhat surprising. A good example is subject number 5, who had a fairly good Swedish accent already at the beginning and had, according to the judges, only a slightly better accent at the end. But if a person with Swedish as their mother tongue would have taken the two tests she would probably not have improved since she's fluent from the beginning. Then the result could have been ~500 = equally good. This kind of evaluation (in Judge) only gives the information whether the subject has improved or not. An additional method, to evaluate the improvement of the subjects knowledge in Swedish, could have been a perception test. This wouldn't have required the judges and would have measured automatically and with more reliability than a test with the judges.

With the outcome of this study known, one month is too short to get any reliable results, especially when the subjects only did the research project on the side of their regular studies. From the beginning we under-estimated the time and effort needed on behalf of the subjects. Also we thought that no one would join the research project if we would set the demand too high. Hence we did not push them hard and we set the minimum of work effort, with both Ville and SWELL, to 20 min per day.

Another thing that complicated the study and the outcome was problems using Ville. One of the subjects did not have his own computer. Another subject had problems with bugs in the program and unfortunately did not inform us about it , which resulted in her not using the program at all. A third subject simply did not use Ville.

6.2 Future work

6.2.1 Suggestions for improvement and development of Ville

- Develop more exercises in existing areas, for example, conversations and vocabulary.
- Include a vocal-tract cross-section to see how the sound is produced in the mouth.
- Include a translation in English with the Swedish word, for example, on the vocabulary flashcards and conversations.
- Build new detectors for pronunciation error detection.
- Expand Ville to other languages.

6.2.2 Suggestions for improvement in similar future CALL evaluation projects

- Give the subjects a good incentive to join and fulfill the research project.
- Overestimate the time needed for the study, do not rush it/the subjects.
- Be sure to collect all the information you might think you need and more in case you will need it.
- Have the dates of the tests and all other deadlines set already from the beginning.
- Set the rules and demands on the subjects straight from the beginning.
- Make sure every one of the subjects have the equipment required and all the material and programs in order.
- Have clear deadlines when the subjects are to send in material etc.
- Check with the subjects every week that they are on track.
- Use several, different methods to evaluate the improvement to reassure that you get the correct picture.

References

BORIN, L. (2002) *What have you done for me lately? The fickle alignment of NLP and CALL.* Paper presented at the EuroCALL 2002, Jyväskylä, Finland.

CHAPELLE, C. (2001) *Computer Applications in Second Language Acquisition (Foundations for teaching, testing and research).* Cambridge University Press.

GRANQVIST S. (1996-98) *Spruce signal workstation add-on package*, Stockholm, Sweden.
See the web page of AB Nyvalla DSP at http://www.nyvalla-dsp.se/ for more information.

HINCKS, R. (2005) *Computer Support for Learners of Spoken English.* Doctoral Thesis, KTH School of Computer Science and Communication, Stockholm.

LEVY, M. (1997) *Computer-Assisted Language Learning. Context and Conceptualization.* Oxford University Press.

ÖSTER, A-M. (1999) *Strategies and results from spoken L2 teaching with audio-visual feedback.* KTH Department of Speech, Music and Hearing, Stockholm.

ÖSTER, A-M. (2006) *Computer-Based Speech Therapy Using Visual Feedback with Focus on Children with Profound Hearing Impairments.* Doctoral Thesis, KTH School of Computer Science and Communication, Stockholm.

SOBKOWIAK, W. (2003) *Pronunciation in EFL CALL.* Adam Mickiewicz University, Poznan, Poland

WIK, P. (2004) *Fonetik.* Designing a virtual language tutor. KTH Department of Speech, Music and Hearing, Stockholm.

WIK, P. (2006) *Manual for Ville.* http://www.speech.kth.se/ville/swell2.html, Stockholm.

WOLFRAM, MATHWORLD (1999) http://mathworld.wolfram.com/StandardDeviation.html

BETTER ACCENT TUTOR. (1997-2007) http://www.betteraccent.com/

JOURNAL OF EXTENSION. (1999) http://www.joe.org/joe/1999april/tt3.html

Appendices

A. Pre-Questionnaire for the subjects of research

-When did you arrive in Sweden?

_____

-Is it your first stay here?

_____

-Do you have relatives living in Sweden, Swedish husband/wife?

_____

-Do have Swedish friends?

_____

-If so how much do you see them, and do you speak Swedish to them on occasion?

_____

-Do you listen to Swedish radio, if so how much time per week?

_____

-Do you regularly watch Swedish television, if so what do you watch and how much?

_____

-Do you go to the cinema to watch Swedish films or foreign films with Swedish subtitles, if so how often?

_____

-Do you listen to Swedish music, if so how much per week?

_____

-Why do you want to learn Swedish, what are your goals with the new language?

_____

-Is it important for you to achieve a good Swedish accent, on a scale from 1-5 how important is it?

_____

-How tall are you?

_____

-Do you have lectures in Swedish?

_____

B. Post-Questionnaire (quantitative)

The first part of the questionnaire considers Ville and the last part is about SWELL the Swedish distance course.

Ville

	Do you think that any previous knowledge about computers is necessary to work with the system?		
	A lot	Some	None
S1			
S2		X	
S3		X	
S4			
S5		X	
S6		X	
S7			X
S8			X
S9			X

	How did the system meet with your expectations?			
	Very well	Well	Not so well	Badly
S1				
S2				
S3	X			
S4				
S5	X			
S6	X			
S7		X		
S8		X		
S9		X		

	Was the system easy to work with?			
	Very easy	Easy	Not so easy	Difficult
S1				
S2		X		
S3	X			
S4				
S5	X			
S6		X		
S7	X			
S8		X		
S9	X			

	Did you consider the training meaningful?			
	Very Meaningful	Meaningful	Not so Meaningful	Meaningless
S1				
S2		X		
S3		X		
S4				
S5	X			
S6		X		
S7		X		
S8	X			
S9		X		

	What is your opinion on the talking head?			
	Very useful	Useful	Not so useful	Pointless
S1				
S2				
S3			X	
S4				
S5	X			
S6		X		
S7		X		
S8		X		
S9	X			

	How was the system from a pedagogical point of view?			
	Very pedagogical	Pedagogical	Not so pedagogical	Unpedagogical
S1				
S2				
S3	X			
S4				
S5		X		
S6		X		
S7		X		
S8		X		
S9	X			

	Was the system reliable?			
	Very reliable	Reliable	Not so reliable	Unreliable
S1				
S2				X
S3	X			
S4				
S5		X		
S6	X			
S7		X		
S8	X			
S9	X			

	Did you want more instructions about the program before starting with it?		
	A lot	Some	None
S1			
S2		X	
S3		X	
S4			
S5			X
S6	X		
S7			X
S8		X	
S9		X	

	Did you miss the possibility to train something? If yes, which one?		
	A lot	Some	None
S1			
S2			
S3		X	
S4			
S5		X	
S6			X
S7			X
S8			X
S9		X	

	How much did you train with the system per week on average?		
	0-30 min	30-60 min	>60 min
S1			
S2			
S3	X		
S4			
S5			X
S6		X	
S7		X	
S8			X
S9		X	

	Do you think you would have improved more if you had worked more with the system?		
	A lot	Some	None
S1			
S2	X		
S3	X		
S4			
S5		X	
S6	X		
S7		X	
S8	X		
S9	X		

SWELL

	Do you think that any previous knowledge about computers is necessary to work with the system?		
	A lot	Some	None
S1			X
S2		X	
S3		X	
S4		X	
S5		X	
S6			X
S7			X
S8		X	
S9			X

	How did the system meet with your expectations?			
	Very well	Well	Not so well	Badly
S1		X		
S2	X			
S3	X			
S4		X		
S5		X		
S6	X			
S7		X		
S8		X		
S9	X			

	Was the system easy to work with?			
	Very easy	Easy	Not so easy	Difficult
S1			X	
S2	X			
S3	X			
S4			X	
S5		X		
S6	X			
S7	X			
S8		X		
S9		X		

	Did you consider the training meaningful?			
	Very Meaningful	Meaningful	Not so Meaningful	Meaningless
S1		X		
S2	X			
S3		X		
S4	X			
S5	X			
S6	X			
S7		X		
S8	X			
S9		X		

	How was the system from a pedagogical point of view?			
	Very pedagogical	Pedagogical	Not so pedagogical	Unpedagogical
S1		X		
S2		X		
S3	X			
S4		X		
S5		X		
S6		X		
S7		X		
S8		X		
S9	X			

	Was the system reliable?			
	Very reliable	Reliable	Not so reliable	Unreliable
S1	X			
S2				
S3	X			
S4		X		
S5	X			
S6	X			
S7		X		
S8	X			
S9			X	

	Did you want more instructions about the program before starting with it?		
	A lot	Some	None
S1			X
S2			X
S3			X
S4		X	
S5			X
S6		X	
S7			X
S8			X
S9			X

	Did you miss the possibility to train something? If yes, which one?		
	A lot	Some	None
S1			X
S2			X
S3			X
S4		X	
S5		X	
S6			X
S7			X
S8			X
S9			X

	How many chapters did you do?
S1	1
S2	4
S3	4
S4	2,5
S5	4
S6	3
S7	2
S8	8
S9	2

	How much did you train with the system per week on average?		
	0-30 min	30-60 min	>60 min
S1			X
S2			X
S3	X		
S4		X	
S5			X
S6			X
S7		X	
S8			X
S9	X		

	Do you think you would have improved more if you had worked more with the system?		
	A lot	Some	None
S1	X		
S2	X		
S3	X		
S4	X		
S5		X	
S6	X		
S7		X	
S8	X		
S9		X	

C. Subject summaries

Summary subject 1

Subject 1 had used Swell less 60 min per week, and not Ville at all. The final result of subject 1 was 501, which is an average result compared to the other subjects. (The scale was from 0-1000 where 0 indicated that the utterance from the first test was the best. 1000 indicated that the utterance from the last test was the best, and 500 imply that there was no difference between the utterances.) This subject knew 2 languages before start learning Swedish which is under average.

Summary subject 2

Subject 2 had used Swell less 60 min per week, and not Ville at all. The final result of subject 2 was 508, which is an average result compared to the other subjects. (The scale was from 0-1000 where 0 indicated that the utterance from the first test was the best. 1000 indicated that the utterance from the last test was the best, and 500 imply that there was no difference between the utterances.) This subject knew 3 languages before start learning Swedish which is above average. (Average number of languages spoken among the subjects was 2.44.)

Summary subject 3

Subject 3 had used Ville and Swell respectively 0-30 min per week. The final result of subject 3 was 488, which is a result slightly under average compared to the other subjects. (The scale was from 0-1000 where 0 indicated that the utterance from the first test was the best. 1000 indicated that the utterance from the last test was the best, and 500 imply that there was no difference between the utterances.) This subject knew 2 languages before start learning Swedish which is under average. (Average number of languages spoken among the subjects was 2.44.)

Summary subject 4

Subject 4 had used Swell 30-60 min per week, and Ville not at all. The final result of subject 4 was 574, which is a result slightly above average compared to the other subjects. (The scale was from 0-1000 where 0 indicated that the utterance from the first test was the best. 1000 indicated that the utterance from the last test was the best, and 500 imply that there was no difference between the utterances.) This subject knew 1 language before start learning Swedish which is under average. (Average number of languages spoken among the subjects was 2.44.)

Summary subject 5

Subject 5 had used Ville and Swell respectively more than 60 min per week. The final result of subject 5 was 540, which is a result slightly above average compared to the other subjects. (The scale was from 0-1000 where 0 indicated that the utterance from the first test was the best. 1000 indicated that the utterance from the last test was the best, and 500 imply that there was no difference between the utterances.) This subject knew 2 languages before starting to learn Swedish, which is under average. (Average number of languages spoken among the subjects was 2.44.)

Summary subject 6

Subject 6 had used Ville 30-60 min/week and Swell > 60 min per week. The final result of subject 6 was 597, which is a result slightly above average compared to the other subjects. (The scale was from 0-1000 where 0 indicated that the utterance from the first test was the best. 1000 indicated that the utterance from the last test was the best, and 500 imply that there was no difference between the utterances.) This subject knew 4 languages before start learning Swedish which is above average. (Average number of languages spoken among the subjects was 2.44.)

Summary subject 7

Subject 7 had used Ville and Swell respectively 30-60 min per week. The final result of subject 7 was 526, which is a result slightly above average compared to the other subjects. (The scale was from 0-1000 where 0 indicated that the utterance from the first test was the best. 1000 indicated that the utterance from the last test was the best, and 500 imply that there was no difference between the utterances.) This subject knew 3 languages before start learning Swedish which is above average. (Average number of languages spoken among the subjects was 2.44.)

Summary subject 8

Subject 8 had used Ville and Swell respectively more than 60 min per week. The final result of subject 8 was 562, which is a result above average compared to the other subjects. (The scale was from 0-1000 where 0 indicated that the utterance from the first test was the best. 1000 indicated that the utterance from the last test was the best, and 500 imply that there was no difference between the utterances.) This subject knew 3 languages before start learning Swedish which is above average. (Average number of languages spoken among the subjects was 2.44.)

Summary subject 9

Subject 9 had used Ville and Swell respectively more than 60 min per week. The final result of subject 9 was 544, which is a result slightly above average compared to the other subjects. (The scale was from 0-1000 where 0 indicated that the utterance from the first test was the best. 1000 indicated that the utterance from the last test was the best, and 500 imply that there was no difference between the utterances.) This subject knew 2 languages before start learning Swedish which is under average. (Average number of languages spoken among the subjects was 2.44.)